



REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE



ECOLE NATIONALE POLYTECHNIQUE

Département de l'hydraulique

LABORATOIRE DE RECHERCHE DES SCIENCES DE L'EAU LRS-EAU

Thèse présentée par

BOUKHAROUBA Khadidja
Ingénieur d'état et Magister en hydraulique

Pour l'obtention du diplôme de Doctorat d'Etat en Hydraulique

Thème

**ELABORATION D'UN MODELE STOCHASTIQUE
DE TYPE BOX-JENKINS POUR LA PREDICTION SIMPLE ET
FILTRE DE KALMAN POUR LA PREDICTION MULTI-SITE DES
APPORTS DES OUEDS AU NORD D'ALGERIE**

Soutenue publiquement le 15 Décembre 2010 devant le jury :

Présidente	BELMOKHTAR	Oum Hani	Professeur	ENP
Directeur de thèse	KETTAB	Ahmed	Professeur	ENP
Examineur	YOUCEF ETTOUMI	Fatiha	Professeur	USTHB
Examineur	BOUKHATALA	Djamel	Professeur	ENP
Examineur	BENMAMAR	Saâdia	MC-(A)	ENP
Examineur	BESSENASSE	Mohammed	MC-(A)	U.BLIDA
Examineur	DOUAOUI	Abdelkader	MC-(A)	U.CHLEF
Invité	TAIBI	Rachid	Directeur général	ANRH/MRE

2010

REMERCIEMENTS

Je tiens à remercier en premier lieu, mon encadreur le professeur Ahmed KETTAB pour ses multiples éclairs de génie ! Car si en définitive cette thèse a pu être soutenue c'est grâce à l'aide et aux conseils qu'il m'a prodigués, qu'il trouve ici l'expression de ma reconnaissance entière. Je remercie également les membres de jury qui m'ont fait l'honneur d'accepter de présider, d'examiner, ou d'assister en qualité d'invité à la soutenance de cette thèse.

Je tiens également à remercier l'ensemble de personnes qui ont contribué à la réalisation de cette thèse. Parmi celles-ci, je remercie le professeur Zekai SEN d'Istanbul Technical University, pour son inestimable aide. Ces interventions au sujet du filtrage linéaire optimal, son suivi de l'évolution de mes travaux, ses critiques pertinentes et ses conseils m'ont guidée tout au long de la réalisation de ce travail. En plus ses encouragements et sa confiance sans cesse renouvelée m'ont été d'un grand apport, puisse-t-il trouver ici l'expression de toute ma gratitude.

Par ailleurs, je manifeste une gratitude particulière au professeur Mamdouh SHAHIN de la Hollande. Les différentes discussions qu'il m'a accordées depuis le début de cette thèse jusqu'à sa finalisation m'ont été d'un apport certain. Ses critiques constructives, ses remarques et ses suggestions ont enrichi le travail et m'ont permis de pousser plus loin la réflexion sur certains aspects de cette thèse. Son expérience longue dans le domaine de l'hydrologie a été mise entièrement à mon profit, et je n'aurais certainement pas bénéficié d'un meilleur service linguistique pour mes publications internationales en langue anglaise.

Cette thèse a également bénéficié des enrichissements du professeur Olivier TALAGRAND du Laboratoire de Météorologie Dynamique (LMD) de Paris, les correspondances et les discussions qu'il m'a accordées ont beaucoup éclairé ma lanterne et contribué à l'amélioration de cette thèse. J'y associe les professeurs Hervé le TREUT du LMD de Paris pour sa collaboration et le professeur André TURGEON, Chaire industrielle CRSNG/ hydro-Québec de l'Ecole Nationale Polytechnique de Montréal, pour les multiples discussions enrichissantes. J'accorde une mention particulière à Karim DAHAK, Office-line d'Alger pour son aide financière, Hamid ALI et Mme Leila SRIDI de l'ANRH d'Alger pour leur aide au sujet des données et au professeur Bertrand COTE de l'université de Sherbrook pour ses précieux conseils.

Enfin, j'ai une reconnaissance spéciale à Mounira LALLALI qui était pour moi une formidable source d'énergie. Son assistance et ses encouragements continus, l'aide inépuisable qu'elle n'a pas cessé de me procurer ces années durant, son soutien opérationnel sans relâche et ses sacrifices ces dernières années en particulier, ont tous contribué à la réalisation de ce travail dans un contexte conviviale, je ne saurais jamais comment lui rendre hommage, qu'elle trouve ici le témoignage de toute ma considération et mon profond respect.

En dépit de tous ces soutiens, quelques erreurs ou insuffisances d'analyse restent de mon entière responsabilité.

LISTE DES ABBREVIATIONS ET SYMBOLES

t	le temps.
Z_t	série des observations (ou mesures) au temps t .
ε_t	résidus au temps t .
B	opérateur de retard tel que : $B^k Z_t = Z_{t-k}$ pour $k=1,2,\dots$
Δ	opérateur de différence tel que : $\Delta X_t = X_t - X_{t-1}$, $\Delta = 1 - B$
Δ^d	la $d^{i\text{eme}}$ différence de X_t , telle que : $\Delta^d X_t = (1 - B)^d X_t$
Δ_s	opérateur de différence saisonnière tel que : $\Delta_s X_t = (1 - B^s) X_t = X_t - X_{t-s}$
Δ_s^d	la $d^{i\text{eme}}$ différence saisonnière de X_t , telle que : $\Delta_s^d X_t = (1 - B^s)^d X_t$
$\psi(B)$	fonction de transfert reliant Z_t à ε_t tel que $Z_t = \mu + \psi(B)\varepsilon_t$.
$\bar{X}, E(X), \mu$	moyenne
$\phi(B)$	moyenne, espérance mathématique de X . polynôme caractéristique d'ordre p tel que : $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$.
ϕ_p	paramètre autorégressif d'ordre p .
AR(p)	modèle autorégressif d'ordre p .
BJ	Box et Jenkins
MA(q)	modèle moyenne mobile d'ordre q .
ARMA(p, q)	modèle Autorégressif d'ordre p et de moyenne mobile d'ordre q .
ARIMA(p, d, q)	modèle Autorégressif d'ordre p et de moyenne mobile d'ordre q d'ordre de différenciation d .
SARIMA(p, d, q)x(P, D, Q) _s ,	modèle ARIMA(p, d, q) saisonnier d'ordre autorégressif saisonnier P et de moyenne mobile saisonnière Q , d'ordre de différenciation saisonnière D .
$\hat{X}_{k/k}$	estimation de l'état X au temps k en étant au temps k .
$\hat{X}_{k/k-1}$	prédiction de l'état X au temps k en étant au temps $k-1$.
Z_k	mesure de l'état X au temps k .
H_k	matrice de mesure au temps k .
K_k	matrice du gain du filtre de Kalman.
$P_{k/k}$	matrice de covariance d'erreur de prédiction au temps k .
I	matrice unité.
FK	Filtre de Kalman
$P_{k/k-1}$	prédiction de la covariance d'erreur au temps k en étant au temps $k-1$.
$\phi_{k+1/k}$	matrice de transition de l'état X du temps k au temps $k+1$.
$P_{k+1/k}$	prédiction de la covariance d'erreur d'estimation de l'état au temps $k+1$ sachant qu'on est au temps k .
Q_k	matrice de covariance d'erreur du système.

R_k	matrice de covariance d'erreur de mesure.
TEM	test d'égalité de la moyenne.
TEV	test d'égalité de la variance.
$\sigma^2, \sigma_X^2, Var(X)$	variance, variance de X
σ_ε^2	variance des résidus.
EQM	erreur quadratique moyenne.
EAM	erreur absolue moyenne.
EM	erreur moyenne.
det	déterminant.
Cov_x	covariance de X .
V^T	transposée de V .
$\rho(k), \rho_k, Corr(X_t, X_{t+k})$	fonction d'auto-corrélation au pas k .
$\gamma_k, \gamma(k)$	fonction d'auto-covariance au pas k .
k, h, i	indices de temps.
ACF	fonction d'auto-corrélation
PACF	fonction d'auto-corrélation partielle
AIC	critère d'information de Akaike
BIC	critère d'information bayésien
SC	critère de Schwarz
IC	intervalle de confiance
$.. \sim \mathcal{N}(0,1)$.. suit une loi Normale de paramètres (0,1)
$.. \sim BB$ (ou WN)	..est un bruit blanc (ou de l'anglais : white noise)

ملخص

ان تحليل وفهم الظواهر الهيدرولوجية يتطلب نمودجات رياضية تاخذ بعين الاعتبار الطبيعة الستوكاستيكية للمتغيرات الهيدرولوجية وتغيراتها الزمنية. هذا العمل يندرج في نفس السياق، حيث يهتم بالبحث عن نمودج ستوكاستيكي للتنبؤ بالسيلان في الوديان الجزائرية. في هذا المضمراخذت عينة من عشرة محطات هيدروغرافية وجاءت نتائج استخدام عملية بوكس و جنكينز بنمودجات مختلفة من نوع ARIMA ذات النتائج المقبولة. ومن اجل التخلص من شرط الثبوت الزمني، قادننا استعمال طرق التنبؤ الديناميكي الى مرشح كالمان الذي يتميز بخاصية النتائج المثلى و بالطبيعة الديناميكية للنمودج الرياضي التي تسمح باخذ بعين الاعتبار التغير الزمني للظواهر الهيدرولوجية. النتيجة هي الحصول على عملية انية للتنبؤ متعدد المواقع بحيث لا توجد عليها قيود لا على مستوى الزمن ولا على مستوى المكان بالاضافة الى كونها تتاقلم بشكل تلقائي مع التغيرات في الشروط المناخية للجهاز الهيدرولوجي المعني بالدراسة. هذا النمودج الاخير يمكن من تحسين اخطاء التنبؤ المتحصل عليها بواسطة نمودجات ARIMA بشكل ملحوظ كما ان لديه ميزة كبيرة تتمثل في الافادة باخطاء التباير للتنبؤ وذلك بشكل دقيق.

الكلمات المفاتيح: ستوكاستيك, ARIMA, مرشح كالمان, تنبؤ متعدد المكان, سيلان الاودية.

Résumé

L'analyse et a compréhension des phénomènes hydrologiques requiert des modèles mathématiques qui prennent en charge la nature stochastique des variables hydrologiques ainsi que leur variabilité temporelle. Dans ce contexte, l'objectif du présent travail est la recherche d'un modèle stochastique pour la prévision des apports liquides des rivières en Algérie. A cet effet, une dizaine de stations hydrométriques ont été considérées et le recours aux processus de Box-Jenkins a conduit à différents modèles paramétriques stationnaires de type ARIMA dont les résultats sont acceptables. Dans le but de se libérer de la contrainte de stationnarité, le recours aux méthodes de prédiction dynamique nous a orientés vers le filtre de Kalman qui est caractérisé par l'optimalité de ses résultats et par la nature dynamique du modèle mathématique qui permet de rendre compte de la variabilité temporelle du processus hydrologique étudié. Le résultat est une opération de prédiction multi-site en ligne, où l'opérateur n'est fixé ni par le temps ni par l'espace, en revanche il s'adapte automatiquement aux changements dans les conditions météorologiques du système hydrologique en question. Ce dernier modèle a permis l'amélioration considérable des erreurs de prédiction obtenues par les modèles ARIMA, et présente l'avantage majeur de procurer avec exactitude la covariance des erreurs de prédiction.

Mots clés : stochastique, ARIMA, filtre de Kalman, prédiction multi-site, apport des rivières.

Abstract

The analysis and understanding of hydrological phenomenon requires mathematical models which carry out the stochastic nature of the hydrological variables and their temporal variability. In this context, the aim of the present work is to find a stochastic model for the prediction of streamflows in Algeria. Hence, ten hydrometric stations have been considered and the utilization of Box-Jenkins processes lead to different stationary ARIMA parametric models whose results are acceptable. In order to be free of the stationarity condition, the utilization of dynamic prediction methods conducts us to Kalman filter which is characterized by its results optimality and by the dynamic nature of the mathematical model allowing to hold the studied hydrological process temporal variability. The result is an online multi-site prediction operation where the operator is not bound to time or space, but rather adapts automatically itself to the hydrological system meteorological conditions changements. The latest model allows for a considerable assesment in the prediction errors obtained by ARIMA models and presents a great advantage by providing with the exact prediction covariance error.

Key words: stochastic, ARIMA, Kalman filter, multi-site prediction, streamflows.

LISTE DES FIGURES

Numéro	Titre	Page
Fig. 0.1	Situation géographique des 10 stations hydrométriques au Nord d'Algérie.	24
Fig.I.2.1	Apports annuels observés à la station hydrométrique BENI BAHDEL entre 1925 et 1993.	43
Fig.I.2.2	Apports mensuels observés à la station hydrométrique AIN BERDA entre 1963 et 1993.	43
Fig. II.2.1	Apports annuels originaux observés à AIN BERDA (1963-1993)	92
Fig. II.2.2	Ajustement à la loi Normale des apports annuels observés à AIN BERDA (1963-1993)	93
Fig. II.2.3	ACF des apports annuels observés à AIN BERDA (1963-1993)	94
Fig. II.2. 4	PACF des apports annuels observés à AIN BERDA (1963-1993)	94
Fig. II.2.5	Apports annuels originaux observés à BENI BAHDEL (1924-1992)	95
Fig. II.2.6	Différences premières des apports annuels observés à BENI BAHDEL (1924-1992)	97
Fig. II.2.7	Ajustement Normal des différences premières des apports annuels observés à BENI BAHDEL (1924-1992)	97
Fig. II.2.8	ACF des apports annuels différenciés observés à BENI BAHDEL (1924-1992)	98
Fig. II.2.9	PACF des apports annuels différenciés observés à BENI BAHDEL (1924-1992)	98
Fig. II.2.10	ACF des résidus du modèle ARIMA(0,1,2)	102
Fig. II.2.11	PACF des résidus du modèle ARIMA(0,1,2)	102
Fig. II.2.12	Ajustement des résidus du modèle ARIMA(0,1,2) à La loi de Gauss	104
Fig. II.2.13	Prédictions des apports annuels à BENI BAHDEL obtenues par le modèle ARIMA(0,1,2)	106
Fig. II.2.14	Apports annuels originaux observés à BOUCHEGOUF (1948-1993)	108
Fig. II.2.15	Ajustement à la loi Normale des apports annuels observés à BOUCHEGOUF (1948-1993)	109
Fig. II.2.16	ACF des apports annuels observés à BOUCHEGOUF (1948-1993)	110
Fig. II.2.17	PACF des apports annuels observés à BOUCHEGOUF (1948-1993)	110
Fig. II.2.18	Apports annuels originaux observés à BOUHNIFIA (1934-1992)	112
Fig. II.2.19	Ajustement Normal des apports annuels originaux observés à BOUHNIFIA (1934-1992)	113
Fig. II.2.20	des apports annuels observés à BOUHNIFIA (1934-1992)	113
Fig. II.2.21	PACF des apports annuels observés à BOUHNIFIA (1934-1992)	114
Fig. II.2.22	ACF des résidus du modèle ARIMA(1,0,3)	117
Fig. II.2.23	PACF des résidus du modèle ARIMA(1,0,3)	117
Fig. II.2.24	Ajustement des résidus du modèle ARIMA(1,0,3) à La loi de Gauss	119
Fig. II.2.25	Prédictions des apports annuels à BOUHNIFIA obtenues par le modèle ARIMA(1,0,3)	120
Fig. II.2.26	Apports annuels originaux observés à CHEFFIA (1966-1992)	122
Fig. II.2.27	Ajustement à la loi Normale des apports annuels observés à CHEFFIA (1966-1992)	123
Fig. II.2.28	ACF des apports annuels observés à CHEFFIA (1966-1992)	124
Fig. II.2.29	PACF des apports annuels observés à CHEFFIA (1966-1992)	124
Fig. II.2.30	Apports annuels originaux observés à KSOB (1966-1992)	126
Fig. II.2.31	Ajustement Normal des apports annuels observés à KSOB (1966-1992)	127
Fig. II.2.32	ACF des apports annuels observés à KSOB (1966-1992)	127
Fig. II.2.33	PACF des apports annuels observés à KSOB (1966-1992)	128
Fig. II.2.34	ACF des résidus du modèle ARIMA(2,0,0)	131
Fig. II.2.35	PACF des résidus du modèle ARIMA(2,0,0)	131
Fig. II.2.36	Ajustement des résidus du modèle ARIMA(2,0,0) à La loi de Gauss	133
Fig. II.2.37	Prédictions des apports annuels à KSOB obtenues par le modèle ARIMA(2,0,0)	135
Fig. II.2.38	Apports annuels originaux observés à MEFROUCHE (1943-1992)	136
Fig. II.2.39	Ajustement Normal des apports annuels observés à MEFROUCHE (1943-1992)	137

Fig. II.2.40	ACF des apports annuels observés à MEFROUCHE (1943-1992)	138
Fig. II.2.41	PACF des apports annuels observés à MEFROUCHE (1943-1992)	138
Fig. II.2.42	ACF des résidus du modèle ARIMA(2,0,3)	141
Fig. II.2.43	PACF des résidus du modèle ARIMA(2,0,3)	141
Fig. II.2.44	Ajustement des résidus du modèle ARIMA(2,0,3) à La loi de Gauss	143
Fig. II.2.45	Prédictions des apports annuels à MEFROUCHE obtenues par le modèle ARIMA(2,0,3)	144
Fig. II.2.46	Apports annuels originaux observés à MIREBECK (1968-1993)	146
Fig. II.2.47	Ajustement à la loi Normale des apports annuels observés à MIREBECK (1968-1993)	147
Fig. II.2.48	ACF des apports annuels observés à MIREBECK (1968-1993)	148
Fig. II.2.49	PACF des apports annuels observés à MIREBECK (1968-1993)	148
Fig. II.2.50	Apports annuels originaux observés à PIERRE DU CHAT (1952-1993)	150
Fig. II.2.51	Premières différences régulières des apports annuels observés à PIERRE DU CHAT (1952-1993)	151
Fig. II.2.52	Ajustement Normal des premières différences régulières des apports annuels observés à PIERRE DU CHAT (1952-1993)	151
Fig. II.2.53	ACF des premières différences régulières des apports annuels observés à PIERRE DU CHAT (1952-1993)	152
Fig. II.2.54	PACF des premières différences régulières des apports annuels observés à PIERRE DU CHAT (1952-1993)	152
Fig. II.2.55	Apports annuels originaux observés à REMCHI (1948-1993)	154
Fig. II.2.56	Ajustement Normal des apports annuels observés à REMCHI (1948-1993)	155
Fig. II.2.57	Ajustement Normal de la racine carrée des apports annuels observés à REMCHI (1948-1993)	155
Fig. II.2.58	ACF de la racine carrée des apports annuels observés à REMCHI (1952-1993)	156
Fig. II.2.59	PACF de la racine carrée des apports annuels observés à REMCHI (1952-1993)	156
Fig. II.2.60	ACF des résidus du modèle ARIMA(1,0,3)	159
Fig. II.2.61	PACF des résidus du modèle ARIMA(1,0,3)	160
Fig. II.2.62	Ajustement des résidus du modèle ARIMA(1,0,3) à La loi de Gauss	162
Fig. II.2.63	Prédictions des apports annuels à REMCHI obtenues par le modèle ARIMA(1,0,3)	163
Fig. II.3.1	Apports mensuels observés à AIN BERDA	169
Fig. II.3.2	ACF des apports mensuels	171
Fig. II.3.3	PACF des apports mensuels	171
Fig. II.3.4	ACF des résidus du modèle SARIMA(1,0,0)(0,1,1) ₁₂	175
Fig. II.3.5	PACF des résidus du modèle SARIMA(1,0,0)(0,1,1) ₁₂	175
Fig. II.3.6	Résidus du modèle SARIMA(1,0,0)(0,1,1) ₁₂	177
Fig. II.3.7	Prédictions des apports mensuels à AIN BERDA obtenues par le modèle SARIMA(1,0,0)(0,1,1) ₁₂	178
Fig. II.3.8	Apports mensuels observés à BENI BAHDEL	180
Fig. II.3.9	ACF des apports mensuels	182
Fig. II.3.10	PACF des apports mensuels	182
Fig. II.3.11	ACF des résidus du modèle SARIMA(1,1,1)(2,1,1) ₁₂	185
Fig. II.3.12	PACF des résidus du modèle SARIMA(1,1,1)(2,1,1) ₁₂	186
Fig. II.3.13	Résidus du modèle SARIMA(1,1,1)(2,1,1) ₁₂	188
Fig. II.3.14	Prédictions des apports mensuels à BENI BAHDEL obtenues par le modèle SARIMA(1,1,1)(2,1,1) ₁₂	189
Fig. II.3.15	Apports mensuels observés à BOUCHEGOUF	190
Fig. II.3.16	ACF des apports mensuels	192
Fig. II.3.17	PACF des apports mensuels	192
Fig. II.3.18	ACF des résidus du modèle SARIMA(1,0,0)(0,1,1) ₁₂	195
Fig. II.3.19	PACF des résidus du modèle SARIMA(1,0,0)(0,1,1) ₁₂	196
Fig. II.3.20	Résidus du modèle SARIMA(1,0,0)(0,1,1) ₁₂	198
Fig. II.3.21	Prédictions des apports mensuels à BOUCHEGOUF obtenues par le modèle SARIMA(1,0,0)(0,1,1) ₁₂	199

Fig. II.3.22	Apports mensuels observés à BOUHNIFIA	200
Fig. II.3.23	ACF des apports mensuels	202
Fig. II.3.24	PACF des apports mensuels	202
Fig. II.3.25	ACF des résidus du modèle SARIMA(1,1,1)(0,1,1) ₁₂	205
Fig. II.3.26	PACF des résidus du modèle SARIMA(1,1,1)(0,1,1) ₁₂	206
Fig. II.3.27	Résidus du modèle SARIMA(1,1,1)(0,1,1) ₁₂	208
Fig. II.3.28	Prédictions des apports mensuels à BOUHNIFIA obtenues par le modèle SARIMA(1,1,1)(0,1,1) ₁₂	209
Fig. II.3.29	Apports mensuels observés à CHEFFIA	210
Fig. II.3.30	ACF des apports mensuels	212
Fig. II.3.31	PACF des apports mensuels	212
Fig. II.3.32	ACF des résidus du modèle SARIMA(1,1,1)(0,1,1) ₁₂	215
Fig. II.3.33	PACF des résidus du modèle SARIMA(1,1,1)(0,1,1) ₁₂	216
Fig. II.3.34	Résidus du modèle SARIMA(1,1,1)(0,1,1) ₁₂	218
Fig. II.3.35	Prédictions des apports mensuels à CHEFFIA obtenues par le modèle SARIMA(1,1,1)(0,1,1) ₁₂	219
Fig. II.3.36	Apports mensuels observés à KSOB	221
Fig. II.3.37	ACF des apports mensuels	223
Fig. II.3.38	PACF des apports mensuels	223
Fig. II.3.39	ACF des résidus du modèle SARIMA(0,0,3)(0,1,1) ₁₂	226
Fig. II.3.40	PACF des résidus du modèle SARIMA(0,0,3)(0,1,1) ₁₂	227
Fig. II.3.41	Résidus du modèle SARIMA(0,0,3)(0,1,1) ₁₂	229
Fig. II.3.42	Prédictions des apports mensuels à KSOB obtenues par le modèle SARIMA(0,0,3)(0,1,1) ₁₂	230
Fig. II.3.43	Apports mensuels observés à MEFROUCHE	231
Fig. II.3.44	ACF des apports mensuels	233
Fig. II.3.45	PACF des apports mensuels	233
Fig. II.3.46	ACF des résidus du modèle SARIMA(1,0,0)(0,1,1) ₁₂	237
Fig. II.3.47	PACF des résidus du modèle SARIMA(1,0,0)(0,1,1) ₁₂	237
Fig. II.3.48	Résidus du modèle SARIMA(1,0,0)(0,1,1) ₁₂	239
Fig. II.3.49	Prédictions des apports mensuels à MEFROUCHE obtenues par le modèle SARIMA(1,0,0)(0,1,1) ₁₂	241
Fig. II.3.50	Apports mensuels observés à MIREBECK	242
Fig. II.3.51	ACF des apports mensuels	244
Fig. II.3.52	PACF des apports mensuels	244
Fig. II.3.53	ACF des résidus du modèle SARIMA(2,0,0)(0,1,1) ₁₂	248
Fig. II.3.54	PACF des résidus du modèle SARIMA(2,0,0)(0,1,1) ₁₂	248
Fig. II.3.55	Résidus du modèle SARIMA(2,0,0)(0,1,1) ₁₂	250
Fig. II.3.56	Prédictions des apports mensuels à MIREBECK obtenues par le modèle SARIMA(2,0,0)(0,1,1) ₁₂	251
Fig. II.3.57	Apports mensuels observés à PIERRE DU CHAT.	253
Fig. II.3.58	ACF des apports mensuels	255
Fig. II.3.59	PACF des apports mensuels	255
Fig. II.3.60	ACF des résidus du modèle SARIMA(1,0,0)(1,1,1) ₁₂	258
Fig. II.3.61	PACF des résidus du modèle SARIMA(1,0,0)(1,1,1) ₁₂	259
Fig. II.3.62	Résidus du modèle SARIMA(1,0,0)(1,1,1) ₁₂	261
Fig. II.3.63	Prédictions des apports mensuels à PIERRE DU CHAT obtenues par le modèle SARIMA(1,0,0)(1,1,1) ₁₂	262
Fig. II.3.64	Apports mensuels observés à REMCHI	263
Fig. II.3.65	ACF des apports mensuels	265
Fig. II.3.66	PACF des apports mensuels	265
Fig. II.3.67	ACF des résidus du modèle SARIMA(0,0,1)(0,1,1) ₁₂	268
Fig. II.3.68	PACF des résidus du modèle SARIMA(0,0,1)(0,1,1) ₁₂	269
Fig. II.3.69	Résidus du modèle SARIMA(0,0,1)(0,1,1) ₁₂	271
Fig. II.3.70	Prédictions des apports mensuels à REMCHI obtenues par le modèle	272

SARIMA(0,0,1)(0,1,1)₁₂

Fig. III.2.1	Prédictions obtenues à la station hydrométrique AINBERDA	309
Fig. III.2.2	Prédictions obtenues à la station hydrométrique BENI BAHDEL	310
Fig. III.2.3	Prédictions obtenues à la station hydrométrique BOUCHEGOUF	310
Fig. III.2.4	Prédictions obtenues à la station hydrométrique BOUHNIFIA	311
Fig. III.2.5	Prédictions obtenues à la station hydrométrique CHEFFIA	311
Fig. III.2.6	Prédictions obtenues à la station hydrométrique KSOB	312
Fig. III.2.7	Prédictions obtenues à la station hydrométrique MEFROUCH	312
Fig. III.2.8	Prédictions obtenues à la station hydrométrique MIREBECK	313
Fig. III.2.9	Prédictions obtenues à la station hydrométrique PIERRE DU CHAT	313
Fig. II.2.10	Prédictions obtenues à la station hydrométrique REMCHI	314
Fig. III.2.11	Prédictions des apports annuels aux dix stations hydrométriques(1968)	315
Fig. III.2.12	Prédictions des apports annuels aux dix stations hydrométriques(1969)	315
Fig. III.2.13	Prédictions des apports annuels aux dix stations hydrométriques(1970)	316
Fig. III.2.14	Prédictions des apports annuels aux dix stations hydrométriques(1971)	316
Fig. III.2.15	Prédictions des apports annuels aux dix stations hydrométriques(1972)	317
Fig. III.2.16	Prédictions des apports annuels aux dix stations hydrométriques(1973)	317
Fig. III.2.17	Prédictions des apports annuels aux dix stations hydrométriques(1974)	318
Fig. III.2.18	Prédictions des apports annuels aux dix stations hydrométriques(1975)	318
Fig. III.2.19	Prédictions des apports annuels aux dix stations hydrométriques(1976)	319
Fig. III.2.20	Prédictions des apports annuels aux dix stations hydrométriques(1977)	319
Fig. III.2.21	Prédictions des apports annuels aux dix stations hydrométriques(1978)	320
Fig. III.2.22	Prédictions des apports annuels aux dix stations hydrométriques(1979)	320
Fig. III.2.23	Prédictions des apports annuels aux dix stations hydrométriques(1980)	321
Fig. III.2.24	Prédictions des apports annuels aux dix stations hydrométriques(1981)	321
Fig. III.2.25	Prédictions des apports annuels aux dix stations hydrométriques(1982)	322
Fig. III.2.26	Prédictions des apports annuels aux dix stations hydrométriques(1983)	322
Fig. III.2.27	Prédictions des apports annuels aux dix stations hydrométriques(1984)	323
Fig. III.2.28	Prédictions des apports annuels aux dix stations hydrométriques(1985)	323
Fig. III.2.29	Prédictions des apports annuels aux dix stations hydrométriques(1986)	324
Fig. III.2.30	Prédictions des apports annuels aux dix stations hydrométriques(1987)	324
Fig. III.2.31	Prédictions des apports annuels aux dix stations hydrométriques(1988)	325
Fig. III.2.32	Prédictions des apports annuels aux dix stations hydrométriques(1989)	325
Fig. III.2.33	Prédictions des apports annuels aux dix stations hydrométriques(1990)	326
Fig. III.2.34	Prédictions des apports annuels aux dix stations hydrométriques(1991)	326
Fig. III.2.35	Prédictions des apports annuels aux dix stations hydrométriques(1992)	327
Fig. II.2.36	Gain du filtre de Kalman	330
Fig. II.2.37	Covariance d'erreur de prédiction	331
Fig. III.2.38	Innovations du FK à AINBERDA (1968-92)	332
Fig. III.2.39	Innovations du FK à BENIBAHDEL (1968-92)	332
Fig. III.2.40	Innovations du FK à BOUCHEGOUF (1968-92)	333
Fig. III.2.41	Innovations du FK à BOUHNIFIA (1968-92)	333
Fig. III.2.42	Innovations du FK à CHEFFIA (1968-92)	333
Fig. III.2.43	Innovations du FK à KSOB (1968-92)	334
Fig. III.2.44	Innovations du FK à MEFROUCH (1968-92)	334
Fig. III.2.45	Innovations du FK à MIREBECK (1968-92)	334
Fig. III.2.46	Innovations du FK à PIERRE DU CHAT (1968-92)	335
Fig. III.2.47	Innovations du FK à REMCHI (1968-92)	335
Fig. III.2.48	Fonction d'auto-corrélation des innovations à AINBERDA	336
Fig. III.2.49	Fonction d'auto-corrélation des innovations à BENIBAHDEL	336
Fig. III.2.50	Fonction d'auto-corrélation des innovations à BOUCHEGOUF	336
Fig. III.2.51	Fonction d'auto-corrélation des innovations à BOUHNIFIA	337
Fig. III.2.52	Fonction d'auto-corrélation des innovations à CHEFFIA	337
Fig. III.2.53	Fonction d'auto-corrélation des innovations à KSOB	337
Fig. III.2.54	Fonction d'auto-corrélation des innovations à MEFROUCHE	338
Fig. III.2.55	Fonction d'auto-corrélation des innovations à MIREBECK	338
Fig. III.2.56	Fonction d'auto-corrélation des innovations à PIERRE DU CHAT	338

Fig.III.2.57	Fonction d'auto-corrélation des innovations à PIERRE DU CHAT	339
Fig.III.2.58	Ajustement Normal aux innovations de AINBERDA	342
Fig.III.2.59	Ajustement Normal aux innovations de BENIBAHDEL	342
Fig.III.2.60	Ajustement Normal aux innovations de BOUCHEGOUF	342
Fig.III.2.61	Ajustement Normal aux innovations de BOUHNIFIA	343
Fig.III.2.62	Ajustement Normal aux innovations de CHEFFIA	343
Fig.III.2.63	Ajustement Normal aux innovations de KSOB	343
Fig.III.2.64	Ajustement Normal aux innovations de MEFROUCH	344
Fig.III.2.65	Ajustement Normal aux innovations de MIREBECK	344
Fig.III.2.66	Ajustement Normal aux innovations de PIERRE DU CHAT	344
Fig.III.2.67	Ajustement Normal aux innovations de REMCHI	345
Fig.II.2.68	Ajustement Normal aux innovations de AINBERDA tronquées des 5 premières valeurs	346
Fig.III.2.69	Ajustement Normal aux innovations de REMCHI tronquées des 5 premières valeurs	346
Fig.III.2.70	Comparaison de la moyenne des erreurs de prédiction entre le Filtre de Kalman (FK) et les modèles ARMA	350
Fig.III.2.71	Comparaison de la variance des erreurs de prédiction entre le Filtre de Kalman (FK) et les modèles ARMA	350
Fig.III.2.72	Comparaison de la déviation standard des erreurs de prédiction entre le Filtre de Kalman (FK) et les modèles ARMA	350
Fig.III.2.73	Comparaison du minimum des erreurs de prédiction entre le Filtre de Kalman (FK) et les modèles ARMA	351
Fig.III.2.74	Comparaison du maximum des erreurs de prédiction entre le Filtre de Kalman (FK) et les modèles ARMA	351

LISTE DES TABLEAUX

Numéro	Titre	Page
Tableau 0.1	Nom et coordonnées des 10 stations hydrométriques au Nord d'Algérie.	24
Tableau 0.2	Destination des barrages correspondant aux cinq stations hydrométriques considérées.	25
Tableau II.1.1	comportement des corrélogrammes et corrélogrammes partiels des processus ARMA	81
Tableau II.2.1	Statistiques des apports annuels observés à AIN BERDA (1963-1993)	92
Tableau II.2.2	Calcul des séries différenciées (Données des 5 premières années)	92
Tableau II.2.3	Statistiques descriptives des séries différenciées (1924-1992)	93
Tableau II.2.4	ACF des apports annuels observés à AIN BERDA (1963-1993)	94
Tableau II.2.5	PACF des apports annuels observés à AIN BERDA (1963-1993)	95
Tableau II.2.6	Statistiques des apports annuels observés à BENI BAHDEL (1924-1992)	95
Tableau II.2.7	Calcul des séries différenciées (données des 5 premières années)	96
Tableau II.2.8	Statistiques descriptives des séries différenciées (1924-1992)	96
Tableau II.2.9	ACF des apports annuels stationnaires	99
Tableau II.2.10	PACF des apports annuels stationnaires	99
Tableau II.2.11	Comparaison des modèles ajustés.	100
Tableau II.2.12	Estimation des paramètres du modèle retenu ARIMA(0,1,2) avec constante	101
Tableau II.2.13	Estimation des paramètres du modèle retenu ARIMA(0,1,2) sans constante	101
Tableau II.2.14	Statistiques des erreurs du modèle ARIMA(0,1,2)	101
Tableau II.2.15	ACF des résidus du modèle ARIMA(0,1,2)	103
Tableau II.2.16	PACF des résidus du modèle ARIMA(0,1,2)	103
Tableau II.2.17	Prédictions des apports annuels à BENI BAHDEL pour les 12 ans futurs	105
Tableau II.2.18	Prédictions avec ARIMA(0,1,2)	107
Tableau II.2.19	Statistiques des apports annuels observés à BOUCHEGOUF (1948-1993)	108
Tableau II.2.20	Calcul des séries différenciées (Données des 5 premières années)	109
Tableau II.2.21	Statistiques descriptives des séries différenciées (1948-1993)	109
Tableau II.2.22	ACF des apports annuels observés à BOUCHEGOUF (1948-1993)	110
Tableau II.2.23	PACF des apports annuels observés à BOUCHEGOUF (1948-1993)	111
Tableau II.2.24	Statistiques des apports annuels observés à BOUNNIFIA (1934-1992)	111
Tableau II.2.25	Calcul des séries différenciées (Données des 5 premières années)	112
Tableau II.2.26	Statistiques descriptives des séries différenciées (1934-1992)	112
Tableau II.2.27	ACF des apports annuels stationnaires	114
Tableau II.2.28	PACF des apports annuels stationnaires	114
Tableau II.2.29	Comparaison des modèles ajustés	115
Tableau II.2.30	Estimation des paramètres du modèle ARIMA(1,0,3)	116
Tableau II.2.31	Statistiques des erreurs du modèle ARIMA(1,0,3)	116
Tableau II.2.32	ACF des résidus du modèle ARIMA(1,0,3)	117
Tableau II.2.33	PACF des résidus du modèle ARIMA(1,0,3)	118
Tableau II.2.34	Prédictions des apports annuels à BOUHNIFIA pour les 12 ans futurs	120
Tableau II.2.35	Prédictions avec ARIMA(1,0,3)	121
Tableau II.2.36	Statistiques des apports annuels observés à CHEFFIA (1966-1992)	122
Tableau II.2.37	Calcul des séries différenciées (Données des 5 premières années)	123
Tableau II.2.38	Statistiques descriptives des séries différenciées (1966-1992)	123
Tableau II.2.39	ACF des apports annuels observés à CHEFFIA (1966-1992)	124
Tableau II.2.40	PACF des apports annuels observés à CHEFFIA (1966-1992)	125
Tableau II.2.41	Statistiques des apports annuels observés à KSOB (1966-1992)	125
Tableau II.2.42	Calcul des séries différenciées (Données des 5 premières années)	126
Tableau II.2.43	Statistiques descriptives des séries différenciées (1966-1992)	126
Tableau II.2.44	ACF des apports annuels stationnaires	128
Tableau II.2.45	PACF des apports annuels stationnaires	128
Tableau II.2.46	Comparaison des modèles ajustés	129

Tableau II.2.47	Estimation des paramètres du modèle retenu ARIMA(2,0,0) avec constante	129
Tableau II.2.48	Statistiques des erreurs du modèle ARIMA(2,0,0)	130
Tableau II.2.49	ACF des résidus du modèle ARIMA(2,0,0)	131
Tableau II.2.50	PACF des résidus du modèle ARIMA(2,0,0)	132
Tableau II.2.51	Prédictions des apports annuels à KSOB par le modèle ARIMA(2,0,0)	134
Tableau II.2.52	Prédictions des apports annuels à KSOB pour les 12 ans futurs	135
Tableau II.2.53	Statistiques des apports annuels observés à MEFROUCHE (1943-1992)	136
Tableau II.2.54	Calcul des séries différenciées (Données des 5 premières années)	137
Tableau II.2.55	Statistiques descriptives des séries différenciées (1943-1992)	137
Tableau II.2.56	ACF des apports annuels stationnaires	138
Tableau II.2.57	PACF des apports annuels stationnaires	139
Tableau II.2.58	Comparaison des modèles ajustés	140
Tableau II.2.59	Estimation des paramètres du modèle retenu ARIMA(2,0,3) avec constante	140
Tableau II.2.60	Statistiques des erreurs du modèle ARIMA(2,0,3)	141
Tableau II.2.61	ACF des résidus du modèle ARIMA(2,0,3)	142
Tableau II.2.62	PACF des résidus du modèle ARIMA(2,0,3)	142
Tableau II.2.63	Prédictions avec ARIMA(2,0,3)	145
Tableau II.2.64	Prédictions des apports annuels à MEFROUCHE pour les 12 ans futurs	145
Tableau II.2.65	Statistiques des apports annuels observés à MIREBECK (1968-1993)	146
Tableau II.2.66	Calcul des séries différenciées (Données des 5 premières années)	147
Tableau II.2.67	Statistiques descriptives des séries différenciées (1968-1993)	147
Tableau II.2.68	ACF des apports annuels observés à MIREBECK (1968-1993)	148
Tableau II.2.69	PACF des apports annuels observés à MIREBECK (1968-1993)	149
Tableau II.2.70	Statistiques des apports annuels observés à PIERRE DU CHAT (1952-1993)	149
Tableau II.2.71	Calcul des séries différenciées (Données des 5 premières années)	150
Tableau II.2.72	Statistiques descriptives des séries différenciées (1952-1993)	150
Tableau II.2.73	ACF des apports annuels stationnaires	152
Tableau II.2.74	PACF des apports annuels stationnaires	153
Tableau II.2.75	Statistiques des apports annuels observés à REMCHI (1948-1993)	153
Tableau II.2.76	Calcul des séries différenciées (Données des 5 premières années)	154
Tableau II.2.77	Statistiques descriptives des séries différenciées (1948-1993)	154
Tableau II.2.78	ACF de la racine carrée des apports annuels	156
Tableau II.2.79	PACF de la racine carrée des apports annuels	157
Tableau II.2.80	Comparaison des modèles ajustés	158
Tableau II.2.81	Estimation des paramètres du modèle retenu ARIMA(1,0,3) avec constante	158
Tableau II.2.82	Statistiques des erreurs du modèle ARIMA(1,0,3)	159
Tableau II.2.83	ACF des résidus du modèle ARIMA(1,0,3)	160
Tableau II.2.84	PACF des résidus du modèle ARIMA(1,0,3)	160
Tableau II.2.85	Prédictions des apports annuels à REMCHI pour les 12 ans futurs	163
Tableau II.2.86	Prédictions avec ARIMA(1,0,3)	164
Tableau II.2.87	Récapitulatif des principaux résultats obtenus à l'échelle annuelle.	165
Tableau II.2.88	Statistiques des erreurs des modèles ajustés à l'échelle annuelle.	165
Tableau II.3.1	Statistiques des apports mensuel observés à AIN BERDA (Sep1963 à Aout 1993)	169
Tableau II.3.2	Calcul des séries différenciées (données des 20 premiers mois)	170
Tableau II.3.3	Statistiques descriptives des séries différenciées pour les 372 mois.	170
Tableau II.3.4	ACF des apports mensuels stationnaires	172
Tableau II.3.5	PACF des apports mensuels stationnaires	172
Tableau II.3.6	Comparaison des modèles ajustés	173
Tableau II.3.7	Estimation des paramètres du modèle retenu SARIMA(1,0,0)(0,1,1) ₁₂	174
Tableau II.3.8	Statistiques des erreurs du modèle SARIMA(1,0,0)(0,1,1) ₁₂	174
Tableau II.3.9	ACF des résidus du modèle SARIMA(1,0,0)(0,1,1) ₁₂	176
Tableau II.3.10	PACF des résidus du modèle SARIMA(1,0,0)(0,1,1) ₁₂	176
Tableau II.3.11	Prévisions des apports mensuels pour les 24 mois futurs à AIN BERDA	179
Tableau II.3.12	Statistiques des apports mensuels observés à BENI BAHDEL (Sep1924 à Aout 1993)	180
Tableau II.3.13	Calcul des séries différenciées (données des 20 premiers mois)	181
Tableau II.3.14	Statistiques descriptives des séries différenciées pour les 828 mois.	181

Tableau II.3.15	ACF des apports mensuels stationnaires	182
Tableau II.3.16	PACF des apports mensuels stationnaires	183
Tableau II.3.17	Comparaison des modèles ajustés	184
Tableau II.3.18	Estimation des paramètres du modèle retenu SARIMA(1,1,1)(2,1,1) ₁₂	184
Tableau II.3.19	Statistiques des erreurs du modèle SARIMA(1,1,1)(2,1,1) ₁₂	184
Tableau II.3.20	ACF des résidus du modèle SARIMA(1,1,1)(2,1,1) ₁₂	186
Tableau II.3.21	PACF des résidus du modèle SARIMA(1,1,1)(2,1,1) ₁₂	187
Tableau II.3.22	Prévisions des apports mensuels à BENI BAHDEL pour les 24 mois futurs	189
Tableau II.3.23	Statistiques des apports mensuels observés à BOUCHEGOUF (Sep1948 à Aout 1993)	190
Tableau II.3.24	Calcul des séries différenciées (données des 20 premiers mois)	191
Tableau II.3.25	Statistiques descriptives des séries différenciées pour les 408 mois.	191
Tableau II.3.26	ACF des apports mensuels stationnaires	192
Tableau II.3.27	PACF des apports mensuels stationnaires	193
Tableau II.3.28	Comparaison des modèles ajustés	194
Tableau II.3.29	Estimation des paramètres du modèle retenu SARIMA(1,0,0)(0,1,1) ₁₂	194
Tableau II.3.30	Statistiques des erreurs du modèle SARIMA(1,0,0)(0,1,1) ₁₂	194
Tableau II.3.31	ACF des résidus du modèle SARIMA(1,0,0)(0,1,1) ₁₂	196
Tableau II.3.32	PACF des résidus du modèle SARIMA(1,0,0)(0,1,1) ₁₂	197
Tableau II.3.33	Prévisions des apports mensuels à BOUCHEGOUF pour les 24 mois futurs	199
Tableau II.3.34	Statistiques des apports mensuel observés à BOUHNIFIA (Sep1934 à Aout 1993)	200
Tableau II.3.35	Calcul des séries différenciées (données des 20 premiers mois)	201
Tableau II.3.36	Statistiques descriptives des séries différenciées pour les 708 mois.	201
Tableau II.3.37	ACF des apports mensuels stationnaires	202
Tableau II.3.38	PACF des apports mensuels stationnaires	203
Tableau II.3.39	Comparaison des modèles ajustés	204
Tableau II.3.40	Estimation des paramètres du modèle retenu SARIMA(1,1,1)(0,1,1) ₁₂	204
Tableau II.3.41	Statistiques des erreurs du modèle SARIMA(1,1,1)(0,1,1) ₁₂	204
Tableau II.3.42	ACF des résidus du modèle SARIMA(1,1,1)(0,1,1) ₁₂	206
Tableau II.3.43	PACF des résidus du modèle SARIMA(1,1,1)(0,1,1) ₁₂	207
Tableau II.3.44	Prévisions des apports mensuels à BOUHNIFIA pour les 24 mois futurs	209
Tableau II.3.45	Statistiques des apports mensuels observés à CHEFFIA (Sep1966 à Aout 1992)	210
Tableau II.3.46	Calcul des séries différenciées (données des 20 premiers mois)	211
Tableau II.3.47	Statistiques descriptives des séries différenciées pour les 324 mois.	211
Tableau II.3.48	ACF des apports mensuels stationnaires	212
Tableau II.3.49	PACF des apports mensuels stationnaires	213
Tableau II.3.50	Comparaison des modèles ajustés	214
Tableau II.3.51	Estimation des paramètres du modèle retenu SARIMA(1,1,1)(0,1,1) ₁₂	214
Tableau II.3.52	Statistiques des erreurs du modèle SARIMA(1,1,1)(0,1,1) ₁₂	214
Tableau II.3.53	ACF des résidus du modèle SARIMA(1,1,1)(0,1,1) ₁₂	216
Tableau II.3.54	PACF des résidus du modèle SARIMA(1,1,1)(0,1,1) ₁₂	217
Tableau II.3.55	Prévisions des apports mensuels à CHEFFIA pour les 24 mois futurs	220
Tableau II.3.56	Statistiques des apports mensuels observés à KSOB (Sep1966 à Aout 1992)	221
Tableau II.3.57	Calcul des séries différenciées (données des 20 premiers mois)	222
Tableau II.3.58	Statistiques descriptives des séries différenciées pour les 324 mois.	222
Tableau II.3.59	ACF des apports mensuels stationnaires	223
Tableau II.3.60	PACF des apports mensuels stationnaires	224
Tableau II.3.61	Comparaison des modèles ajustés	225
Tableau II.3.62	Estimation des paramètres du modèle retenu SARIMA(0,0,3)(0,1,1) ₁₂	225
Tableau II.3.63	Statistiques des erreurs du modèle SARIMA(0,0,3)(0,1,1) ₁₂	225
Tableau II.3.64	ACF des résidus du modèle SARIMA(0,0,3)(0,1,1) ₁₂	227
Tableau II.3.65	PACF des résidus du modèle SARIMA(0,0,3)(0,1,1) ₁₂	228
Tableau II.3.66	Prévisions des apports mensuels à KSOB pour les 24 mois futurs	230
Tableau II.3.67	Statistiques des apports mensuels observés à MEFROUCHE (Sep1943 à Aout 1992)	231
Tableau II.3.68	Calcul des séries différenciées (données des 20 premiers mois)	232
Tableau II.3.69	Statistiques descriptives des séries différenciées pour les 600 mois.	232

Tableau II.3.70	ACF des apports mensuels stationnaires	234
Tableau II.3.71	PACF des apports mensuels stationnaires	234
Tableau II.3.72	Comparaison des modèles ajustés	235
Tableau II.3.73	Estimation des paramètres du modèle retenu SARIMA(1,0,0)(0,1,1) ₁₂	235
Tableau II.3.74	Statistiques des erreurs du modèle SARIMA(1,0,0)(0,1,1) ₁₂	236
Tableau II.3.75	ACF des résidus du modèle SARIMA(1,0,0)(0,1,1) ₁₂	238
Tableau II.3.76	PACF des résidus du modèle SARIMA(1,0,0)(0,1,1) ₁₂	238
Tableau II.3.77	Prévisions des apports mensuels à MEFROUCHE pour les 24 mois futurs	240
Tableau II.3.78	Statistiques des apports mensuels observés à MEFROUCHE (Sep1968 à Aout 1993)	242
Tableau II.3.79	Calcul des séries différenciées (données des 20 premiers mois)	243
Tableau II.3.80	Statistiques descriptives des séries différenciées pour les 324 mois.	243
Tableau II.3.81	ACF des apports mensuels stationnaires	245
Tableau II.3.82	PACF des apports mensuels stationnaires	245
Tableau II.3.83	Comparaison des modèles ajustés	
Tableau II.3.84	Estimation des paramètres du modèle retenu SARIMA(2,0,0)(0,1,1) ₁₂	246
Tableau II.3.85	Statistiques des erreurs du modèle SARIMA(2,0,0)(0,1,1) ₁₂	247
Tableau II.3.86	ACF des résidus du modèle SARIMA(2,0,0)(0,1,1) ₁₂	249
Tableau II.3.87	PACF des résidus du modèle SARIMA(2,0,0)(0,1,1) ₁₂	249
Tableau II.3.88	Prévisions des apports mensuels à MIREBECK pour les 24 mois futurs	252
Tableau II.3.89	Statistiques des apports mensuels observés à PIERRE DU CHAT (Sep1952 à Aout 1993)	253
Tableau II.3.90	Calcul des séries différenciées (données des 20 premiers mois)	254
Tableau II.3.91	Statistiques descriptives des séries différenciées pour les 504 mois	254
Tableau II.3.92	ACF des apports mensuels stationnaires	255
Tableau II.3.93	PACF des apports mensuels stationnaires	256
Tableau II.3.94	Comparaison des modèles ajustés	257
Tableau II.3.95	Estimation des paramètres du modèle retenu SARIMA(1,0,0)(1,1,1) ₁₂	257
Tableau II.3.96	Statistiques des erreurs du modèle SARIMA(1,0,0)(1,1,1) ₁₂	257
Tableau II.3.97	ACF des résidus du modèle SARIMA(1,0,0)(1,1,1) ₁₂	259
Tableau II.3.98	PACF des résidus du modèle SARIMA(1,0,0)(1,1,1) ₁₂	260
Tableau II.3.99	Prévisions des apports mensuels à PIERRE DU CHAT pour les 24 mois futurs	262
Tableau II.3.100	Statistiques des apports mensuels observés à REMCHI (Sep1948 à Aout 1993)	263
Tableau II.3.101	Calcul des séries différenciées (données des 20 premiers mois)	264
Tableau II.3.102	Statistiques descriptives des séries différenciées pour les 516 mois.	264
Tableau II.3.103	ACF des apports mensuels stationnaires	265
Tableau II.3.104	PACF des apports mensuels stationnaires	266
Tableau II.3.105	Comparaison des modèles ajustés	267
Tableau II.3.106	Estimation des paramètres du modèle retenu SARIMA(0,0,1)(0,1,1) ₁₂	267
Tableau II.3.107	Statistiques des erreurs du modèle SARIMA(0,0,1)(0,1,1) ₁₂	267
Tableau II.3.108	ACF des résidus du modèle SARIMA(0,0,1)(0,1,1) ₁₂	269
Tableau II.3.109	PACF des résidus du modèle SARIMA(0,0,1)(0,1,1) ₁₂	270
Tableau II.3.110	Prévisions des apports mensuels à REMCHI pour les 24 mois futurs	273
Tableau II.3.111	Récapitulatif des principaux résultats obtenus à l'échelle mensuelle.	274
Tableau II.3.112	Statistiques des erreurs des modèles ajustés à l'échelle mensuelle.	274
Tableau III.2.1	Prédictions multi-sites des apports liquides annuels (1968-1992) aux dix stations hydrométriques.	328
Tableau III.2.2	Innovations du filtre de Kalman (1968-1992).	328
Tableau III.2.3	Erreurs relatives en % des prédictions multi-sites des apports annuels aux dix stations hydrométriques.	329
Tableau III.2.4	Auto-corrélations des innovations à AINBERDA	339
Tableau III.2.5	Auto-corrélations des innovations à BENIBAHDEL	339
Tableau III.2.6	Auto-corrélations des innovations à BOUCHEGOUF	339
Tableau III.2.7	Auto-corrélations des innovations à BOUHNIFIA	340
Tableau III.2.8	Auto-corrélations des innovations à CHEFFIA	340
Tableau III.2.9	Auto-corrélations des innovations à KSOB	340

Tableau III.2.10	Auto-corrélations des innovations à MEFROUCHE	340
Tableau III.2.11	Auto-corrélations des innovations à MIREBECK	340
Tableau III.2.12	Auto-corrélations des innovations à PIERRE DU CHAT	340
Tableau III.2.13	Auto-corrélations des innovations à REMCHI	341
Tableau III.2.14	Test de Box-Pierce pour les innovations du filtre.	341
Tableau III.2.15	Test d'adéquation du Chi-deux de l'ajustement normal à 95% de confiance des innovations du filtre.	345
Tableau III.2.16	Moyenne, déviation standard et erreur relative des valeurs observés et prédites des apports liquides annuels (1968-1992).	347
Tableau III.2.17	Moyenne et déviation standard des valeurs observées et prédites des apports liquides annuels aux 10 stations hydrométriques (1968-1992).	347
Tableau III.2.18	Comparaison des erreurs de prédiction entre Filtre de Kalman (FK) et modèles ARMA	349

TABLE DES MATIERES

INTRODUCTION GENERALE.....	22
----------------------------	----

PARTIE I

PREAMBULE I.....	27
------------------	----

Chapitre I.1

NOTIONS DE BASE POUR ABORDER LES MODELES ARIMA ET LE FILTRE DE KALMAN

I.1.1 INTRODUCTION.....	28
I.1.2 PROBABILITES ET VARIABLES ALEATOIRES.....	28
I.1.2.1 Probabilités.....	28
I.1.2.2 L'espace de probabilité	29
I.1.2.3 Variable aléatoire réelle (v.a.)	29
I.1.3 CARACTERISATION D'UNE VARIABLE ALEATOIRE.....	30
SCALAIRE	
I.1.3.1 Fonction de répartition et fonction de densité de probabilité.....	30
I.1.3.2 Moyenne et Variance.....	31
I.1.3.3 Distribution Normale ou de Gauss.....	32
I.1.4 CARACTERISATION D'UNE VARIABLE ALEATOIRE.....	33
VECTORIELLE	
I.1.4.1 Fonction de répartition et fonction de densité de probabilité.....	33
I.1.4.2 Moyenne et Covariance.....	33
I.1.4.3 Vecteur aléatoire de Gauss.....	34
I.1.5 INDEPENDANCE ET PROBABILITE CONDITIONNELLE.....	34
I.1.6 PROCESSUS STOCHASTIQUE ET BRUIT BLANC.....	35
I.1.6.1 Processus stochastique	35
I.1.6.2 Bruit Blanc	35
I.1.7 ESTIMATION STOCHASTIQUE LINEAIRE.....	36
I.1.7.1 Modèles Espace-Etat.....	36
I.1.7.2 Modèles Boite Noire.....	37
I.1.7.3 Caractérisation Stochastique.....	38
I.1.8 LE FILTRAGE LINEAIRE.....	39
I.1.9 CONCLUSION.....	40

Chapitre I.2

LES SERIES CHRONOLOGIQUES ET LES PROCESSUS STOCHASTIQUES EN HYDROLOGIE

I.2.1	INTRODUCTION.....	41
I.2.2	LES SERIES CHRONOLOGIQUES.....	42
I.2.2.1	Définition.....	42
I.2.2.2	Types importants de séries chronologiques.....	42
I.2.2.3	Exemples de séries chronologiques.....	42
I.2.2.4	Composantes principales des séries chronologiques.....	43
I.2.2.5	Quelques types de décomposition.....	44
I.2.2.6	Objectifs de l'étude des séries chronologiques.....	45
I.2.2.7	Types de modèles ajustés aux séries chronologiques.....	45
I.2.3	LES PROCESSUS STOCHASTIQUES (PS).....	46
I.2.3.1	Définition.....	46
I.2.3.2	Auto-corrélations totales et auto-corrélations partielles d'un processus stochastique.....	47
I.2.3.3	Processus stochastique stationnaire.....	49
I.2.3.4	Processus stochastique ergodique.....	50
I.2.3.5	Innovations d'un processus stochastique.....	51
I.2.4	CONCLUSION.....	52

Chapitre I.3

LES MODELES STOCHASTIQUES LINEAIRES DES SERIES CHRONOLOGIQUES EN HYDROLOGIE

I.3.1	LES PROCESSUS LINEAIRES.....	53
I.3.1.1	Opérateur de retard.....	53
I.3.1.2	Opérateur de différence.....	54
I.3.1.3	Filtre linéaire.....	54
I.3.1.4	Processus linéaire.....	55
I.3.2	REPRESENTATION DE WOLD.....	55
I.3.3	LES PROCESSUS STOCHASTIQUES LINEAIRES ARIMA.....	56
I.3.3.1	Modèle moyenne mobile d'ordre infini MA(∞).....	56

I.3.3.2 Processus moyenne mobile d'ordre fini MA(q).....	57
I.3.3.3 Processus autorégressif d'ordre fini AR(p)	59
I.3.3.4 Les processus mixtes ARMA(p,q).....	64
I.3.3.5 Les processus mixtes intégrés ARIMA(p,d,q).....	71
I.3.3.6 Les processus mixtes saisonniers intégrés SARIMA(p,d,D,q).....	72
I.3.4. CONCLUSION.....	73

CONCLUSION I.....	74
-------------------	----

PARTIE II

PREAMBULE II.....	75
-------------------	----

Chapitre II.1

LA METHODE DE BOX – JENKINS ET LA MODELISATION STOCHASTIQUE LINEAIRE DES SERIES CHRONOLOGIQUES

II.1.1 INTRODUCTION	76
II.1.2 STATIONARISATION ET DESAISONALISATION.....	77
II.1.2.1 Transformation et différenciation.....	77
II.1.2.2 Détermination de l'ordre de différenciation	78
II.1.3 DETERMINATION DU TYPE ET DE L'ORDRE DU MODELE.....	78
II.1.3.1 Paramètres AR.....	79
II.1.3.2 Paramètres MA.....	79
II.1.3.3 Paramètres ARIMA	80
II.1.3.4 Paramètres SARIMA(p,d,q)(P,D,Q) _s	80
II.1.4 ESTIMATION DES PARAMETRES DU MODELE.....	81
II.1.4.1 Rappels : l'estimateur du maximum de vraisemblance.....	81
II.1.4.2 Application aux modèles ARMA(p,q).....	82
II.1.5 VALIDATION	83
II.1.5.1 Tests de signification des paramètres.....	83
II.1.5.2 Tests de Bruit Blanc.....	83
II.1.5.3 Critères de comparaison de modèles	88
II.1.6 PREDICTION.....	88
II.1.6.1 Transformation de la série	88
II.1.6.2 Prédicteur pour un processus ARMA.....	88
II.1.7 CONCLUSION	90

Chapitre II.2

MODELISATION ARIMA ET PREDICTION DES APPORTS ANNUELS

II.2.1	INTRODUCTION.....	91
II.2.2	APPLICATION ET DISCUSSION DES RESULTATS	91
II.2.2.1	AIN BERDA.....	91
II.2.2.1.1	Stationnarisation	92
II.2.2.1.2	Identification.....	93
II.2.2.2	BENI BAHDEL	95
II.2.2.2.1	Stationnarisation.....	96
II.2.2.2.2	Identification	97
II.2.2.2.3	Estimation.....	100
II.2.2.1.4	Validation.....	102
II.2.2.2.5	Prédiction.....	105
II.2.2.3	BOUCHEGOUF.....	108
II.2.2.3.1	Stationnarisation	108
II.2.2.3.2	Identification	109
II.2.2.4	BOUHNIFIA	111
II.2.2.4.1	Stationnarisation.....	112
II.2.2.4.2	Identification	113
II.2.2.4.3	Estimation.....	115
II.2.2.4.4	Validation	116
II.2.2.4.5	Prédiction.....	119
II.2.2.5	CHEFFIA.....	122
II.2.2.5.1	Stationnarisation.....	122
II.2.2.5.2	Identification	123
II.2.2.6	KSOB	125
II.2.2.6.1	Stationnarisation.....	125
II.2.2.6.2	Identification	127
II.2.2.6.3	Estimation.....	129
II.2.2.6.4	Validation	130
II.2.2.6.5	Prédiction.....	133
II.2.2.7	MEFROUCHE	136
II.2.2.7.1	Stationnarisation.....	136
II.2.2.7.2	Identification	137
II.2.2.7.3	Estimation.....	139
II.2.2.7.4	Validation	141
II.2.2.7.5	Prédiction	144
II.2.2.8	MIREBECK.....	146
II.2.2.8.1	Stationnarisation	146
II.2.2.8.2	Identification	147
II.2.2.9	PIERRE DU CHAT	149
II.2.2.9.1	Stationnarisation.....	150
II.2.2.9.2	Identification.....	151

II.2.2.10	REMCHI.....	153
II.2.2.10.1	Stationnarisation.....	154
II.2.2.10.2	Identification.....	155
II.2.2.10.3	Estimation.....	157
II.2.2.10.4	Validation.....	159
II.2.2.10.5	Prédiction.....	162
II.2.3	RECAPITULATIF ET CONCLUSIONS.....	164

Chapitre II.3

MODELISATION ARIMA ET PREDICTION DES APPORTS MENSUELS

II.3.1	INTRODUCTION.....	168
II.3.2	APPLICATION ET DISCUSSION DES RESULTATS.....	168
II.3.2.1	AIN BERDA.....	169
II.3.2.1.1	Stationnarisation.....	169
II.3.2.1.2	Identification.....	170
II.3.2.1.3	Estimation des modèles ajustés.....	173
II.3.2.1.4	Validation du modèle SARIMA(1,0,0)(0,1,1) ₁₂	175
II.3.2.1.5	Prédiction.....	178
II.3.2.2	BENI BAHDEL.....	180
II.3.2.2.1	Stationnarisation.....	180
II.3.2.2.2	Identification.....	181
II.3.2.2.4	Estimation des modèles ajustés.....	183
II.3.2.2.3	Validation du modèle SARIMA(1,1,1)(2,1,1) ₁₂	185
II.3.2.2.5	Prédiction.....	188
II.3.2.3	BOUCHEGOUF.....	190
II.3.2.3.1	Stationnarisation.....	190
II.3.2.3.2	Identification.....	191
II.3.2.3.3	Estimation des modèles ajustés.....	193
II.3.2.3.4	Validation du modèle.....	195
II.3.2.3.5	Prédiction.....	198
II.3.2.4	BOUHNIFIA.....	200
II.3.2.4.1	Stationnarisation.....	200
II.3.2.4.2	Identification.....	201
II.3.2.4.3	Estimation des modèles ajustés.....	203
II.3.2.4.4	Validation du modèle SARIMA(1,1,1)(0,1,1) ₁₂	205
II.3.2.4.5	Prédiction.....	208
II.3.2.5	CHEFFIA.....	210
II.3.2.5.1	Stationnarisation.....	210
II.3.2.5.2	Identification.....	211
II.3.2.5.3	Estimation des modèles ajustés.....	213
II.3.2.5.4	Validation du modèle SARIMA(1,1,1)(0,1,1) ₁₂	215

II.3.2.5.5 Prédiction	218
II.3.2.6 KSOB.....	221
II.3.2.6.1 Stationnarisation.....	221
II.3.2.6.2 Identification	222
II.3.2.6.3 Estimation des modèles ajustés	224
II.3.2.6.4 Validation du modèle SARIMA(0,0,3)(0,1,1) ₁₂	226
II.3.2.6.5 Prédiction	229
II.3.2.7 MEFROUCHE.....	231
II.3.2.7.1 Stationnarisation.....	231
II.3.2.7.2 Identification	232
II.3.2.7.3 Estimation des modèles ajustés	235
II.3.2.7.4 Validation du modèle SARIMA(1,0,0)(0,1,1) ₁₂	236
II.3.2.7.5 Prédiction	240
II.3.2.8 MIREBECK.....	242
II.3.2.8.1 Stationnarisation.....	242
II.3.2.8.2 Identification	243
II.3.2.8.3 Estimation des modèles ajustés	246
II.3.2.8.4 Validation du modèle SARIMA(2,0,0)(0,1,1) ₁₂	247
II.3.2.8.5 Prédiction	250
II.3.2.9 PIERRE DU CHAT.....	253
II.3.2.9.1 Stationnarisation.....	253
II.3.2.9.2 Identification	254
II.3.2.9.3 Estimation des modèles ajustés	256
II.3.2.9.4 Validation du modèle SARIMA(1,0,0)(1,1,1) ₁₂	258
II.3.2.9.5 Prédiction	261
II.3.2.10 REMCHI.....	263
II.3.2.10.1 Stationnarisation.....	263
II.3.2.10.2 Identification	264
II.3.2.10.3 Estimation des modèles ajustés	266
II.3.2.10.4 Validation du modèle SARIMA(0,0,1)(0,1,1) ₁₂	268
II.3.2.10.5 Prédiction	271
II.2.3 RECAPITULATIF ET CONCLUSIONS.....	273
CONCLUSION II.....	277

PARTIE III

PREAMBULE III	279
Chapitre III.1	

LA METHODE DU FILTRE DE KALMAN POUR LA PREDICTION EN HYDROLOGIE

III.1.1 INTRODUCTION.....	283
III.1.2 LES SYSTEMES LINEAIRES.....	285
III.1.2.1 Variables et équations d'état.....	285
III.1.2.2 Forme matricielle des équations d'état.....	285
III.1.2.3 Résolution des équations d'état (matrice de transition).....	286
III.1.3 CONDITIONS D'APPLICATION DU FILTRE DE KALMAN.....	288
III.1.3.1 Modèle mathématique adopté.....	288
III.1.3.2 Digitalisation du modèle mathématique.....	289
III.1.3.3 Les entrées et sorties du filtre de Kalman.....	290
III.1.4 MECANISME D'ESTIMATION DU FILTRE DE KALMAN.....	292
III.1.4.1 Prédiction d'un signal analytique.....	292
III.1.4.2 Prévision d'un signal aléatoire.....	292
III.1.4.3 Estimation d'un signal mixte.....	292
III.1.4.4 Développement des formules récurrentes du FK.....	294
III.1.5 CRITERES D'OPTIMALISATION.....	297
III.1.6 STABILITE DU FILTRE VIS-A-VIS DE LA DIVERGENCE.....	298
III.1.7 CONCLUSION.....	299

Chapitre III.2

APPLICATION DU FILTRE DE KALMAN A LA PREDICTION MULTI-SITE DES APPORTS ANNUELS

III.2.1 INTRODUCTION.....	300
III.2.2 DONNEES DE LA PREDICTION MULTI-SITE.....	301
III.2.3 HYPOTHESES.....	301
III.2.4 FORMULATION ESPACE-ETAT DU MODELE ADOPTE.....	301
III.2.5 CONDITIONS INITIALES.....	303
III.2.5.1 Vecteur d'état initial et matrice de covariance d'erreur associée.....	304
III.2.5.2 Matrice de covariance des bruits du système et celle des bruits des mesures.....	304
III.2.5.3 Matrice de transition d'état et matrice de mesure.....	305
III.2.6 LES EQUATIONS RECURRENTES DU FILTRE DE KALMAN.....	305
III.2.6.1 Matrice gain du filtre.....	305
III.2.6.2 Utilisation de la mesure pour actualiser l'estimation.....	306
III.2.6.3 Matrice de covariance d'erreur associée à l'estimation (actualisée) corrigée par la mesure.....	306
III.2.6.4 Prédiction un pas en avant de l'estimation (actualisée) corrigée par la mesure...	307
III.2.6.5 Matrice de covariance de l'erreur de prédiction un pas en avant.....	307
III.2.7 RESULTATS ET INTERPRETATIONS.....	308
III.2.7.1 Prédictions multi-sites.....	308
III.2.7.2 Optimalité des résultats.....	327

III.2.7.3 Analyse des innovations du filtre (résidus).....	332
III.2.7.4 Moyenne et déviation standard des prédictions multi-sites	347
III.2.8 COMPARAISON DES RESULTATS DE PREDICTION AVEC CEUX DES MODELES ARMA	348
III.2.9 CONCLUSION.....	352
CONCLUSION III.....	354
CONCLUSION GENERALE ET RECOMMANDATIONS FUTURES.....	356
REFERENCES BIBLIOGRAPHIQUES.....	361

INTRODUCTION GENERALE

Entre les besoins en eau toujours croissants d'une société en plein développement socio-économique et la nature semi-aride de son climat, l'Algérie est un pays qui souffre des pénuries d'eau qui deviennent de plus en plus spectaculaires. En effet, sur les 100 milliards de mètres cubes reçus annuellement sur sa partie Nord, sous forme de précipitations, seulement 4,8 milliards de mètres cubes sont actuellement mobilisables dans environ 130 barrages, actuellement en exploitation (**Kettab et al , 2008**). Si en plus, l'on retranche les pertes sous leurs différentes formes, on se retrouve avec un chiffre beaucoup plus faible pour faire face aux besoins socio-économiques, qui ne cessent de perturber la balance de l'offre et de la demande de cette ressource. Pour un développement durable d'un tel pays, tout le monde s'accorde sur un point : le passage par la mise en place d'un système de gestion rigoureuse de cette ressource rendant compte de toutes les composantes (sociales, économiques et environnementales) est incontournable (**Boukharouba et Kettab, 2003b**).

Dans cette perspective, la modélisation et la prédiction des processus hydrologiques tels que procurés par la nature, jouent un rôle primordial dans la planification et la gestion des ressources hydriques, en plus elles permettent d'apporter des outils d'aide à la décision (**Koutsoyiannis et al., 2008**). A cet effet, il est important de mettre en place des modèles dynamiques qui prennent en charge la nature complexe, la non linéarité et les variations temporelles des processus hydrologiques (**Lettenmaier et Wood, 1993**), ainsi que leurs inter-corrélations qui sont liées essentiellement aux variations climatologiques et physiographiques des bassins en questions.

En effet, l'avènement des variables hydrologiques dans la nature est le résultat d'un certain nombre d'éléments physiques interconnectés très complexes, parmi lesquels les facteurs climatiques et physiographiques jouent un rôle prédominant. De telles variables, comme les précipitations, les débits de rivières qui en résultent ou encore les apports liquides, sont le produit de phénomènes complexes qui varient continuellement dans le temps et qui peuvent être mesurés par un nombre fini d'observations. Ces observations indiquent que les variables hydrologiques sont stochastiques dans la nature (**Yevjevich, 1971**) et non linéaires (**Amorcho & Orlob, 1961**), ceci est une conséquence de (1) la variabilité temporelle des processus d'érosion et, de dépôt (2) des changements climatiques (3) l'incertitude de l'état dans le temps et (4) le transfert d'énergie du cycle hydrologique dont le caractère est non linéaire.

Dans la pratique, les entrées et les sorties des systèmes hydrologiques sont considérées comme des variables aléatoires mesurables à des points fixes du temps, formant ainsi des séries chronologiques au pas de temps journalier, hebdomadaire, mensuel ou annuel. Cependant, les données obtenues incluent une composante bruitée dont les paramètres statistiques peuvent être estimés à partir des mêmes données (**Von Storch et al., 2001**).

D'une façon générale, les variables aléatoires en hydrologie constituent des contraintes pouvant affecter soit directement ou indirectement la performance finale de n'importe quel plan d'action dans le domaine de l'eau, et limitent par leur nature les interventions des décideurs. En effet, ces variables aléatoires peuvent entraîner un risque important lorsqu'il s'agit d'élaborer une

stratégie dont la réalisation des objectifs dépend du comportement de ces dernières. D'où l'idée d'étudier et de comprendre le processus stochastique sous-jacent. En d'autres termes, l'étude du comportement passé moyennant des enregistrements chronologiques de ces variables pourrait nous renseigner sur le comportement futur de ces dernières. Par conséquent, une aide considérable est mise à la disposition des décideurs du secteur afin qu'ils puissent élaborer leurs stratégies. Ceci est fait en leur offrant une possibilité de prévoir le risque en partie et par conséquent la réduction de ses effets.

La recherche de ces mécanismes revient à déterminer dans quelle mesure ces variables aléatoires sont strictement aléatoires. C'est-à-dire dans quel sens sont-elles non prévisibles. En d'autres termes, existe-t-il des parties structurelles reconnaissables au sein de ces variables?

Une étude détaillée des phénomènes hydrologiques requiert donc des modèles mathématiques qui sauraient prendre en ligne de compte la nature stochastique des variables hydrologiques ainsi que leur variabilité temporelle. Le planificateur pourrait ainsi procéder soit à la simulation du mécanisme générateur en question soit à sa prédiction.

Dans ce contexte les objectifs du présent travail peuvent être résumés en quelques points essentiels comme suit:

1. étudier dans quelle mesure les variables hydrologiques, considérés souvent comme variables purement aléatoires, ne contiennent-elles pas des parties structurées pouvant être prévisibles.
2. mettre les séries chronologiques de ces variables en équation en vue de faire leur prévision.
3. étudier la possibilité d'application du Filtre de Kalman aux données hydrologiques de la rive sud de la méditerranée, notamment le Nord d'Algérie.
4. procéder à la prédiction multi-site de ces données.

Pour mettre en application ces idées et concepts, un choix a été fait sur la variable hydrologique des apports liquides des rivières en raison, d'une part de l'intérêt particulier qu'elle présente en étant un facteur lié directement ou indirectement au développement du secteur socio-économique du pays et d'autre part de la disponibilité relative de l'information en temps et en espace. A cet effet, le tri et la sélection de la multitude de données disponibles au niveau de l'Agence Nationale des Ressources Hydrauliques a été fait selon des critères qui répondent principalement aux objectifs déjà cités du présent travail. Le choix est alors fixé sur une dizaine de stations hydrométriques qui sont réparties entre les hydro-systèmes Est Ouest et centre du Nord d'Algérie tel qu'illustré par la Fig. 0.1 suivante, dont cinq d'entre elles se trouvent sur des barrages déjà en exploitation.



(source Agence Nationale des Barrages et Transferts)

Fig. 0.1 Situation géographique des 10 stations hydrométriques au Nord d'Algérie.

Le Tableau 0.1 regroupe les informations relatives aux coordonnées de ces stations, selon l'Agence Nationale des Ressources Hydrauliques. Le Tableau 0.2 résume les informations concernant les barrages sur lesquels se trouvent cinq d'entre les stations hydrométriques considérées.

Tableau 0.1 Nom et coordonnées des 10 stations hydrométriques au Nord d'Algérie.

No	Station	Nom	Latitude (N)	Longitude (E)	Élévation(m)
01	140501	Boucheghouf	40 50 60	05 97 80	95
02	140601	Mirebeck	40 82 20	06 04 00	10
03	140602	Ain Berda	40 75 00	05 85 30	55
04	160703	Remchi	39 00 35	04 20 00	85
05	160801	Pierre du chat	39 04 62	04 20 62	52
06	160403	Beni-Bahdel	38 56 60	04 26 00	645
07	111504	Bouhanifia	39 24 80	02 65 10	295
08	31501	Cheffia	40 67 75	06 34 00	121
09	050902	Ksob	39 80 30	02 48 12	570
10	160701	Mefrouch	38 72 80	04 03 25	1110

(Source : Agence Nationale des Ressources Hydrauliques)

Tableau 0.2 Destination des barrages correspondant aux cinq stations hydrométriques considérées.

No.	Barrage	Wilaya	Capacité estimée (Hm3)	Volume régularisé (Hm3/an)	Destination
01	Bénibahdel	Tlemcen	55.5	48	AEP-Oran et IRR. périmètre de Maghnia
02	Bouhanifia	Mascara	48.4	Transfert	Transfert Fergoug et AEP-Hacine
03	Chéffia	El Tarf	167.1	95	AEP-Annaba, AEI-El Hadjar et IRR. périmètre de Bounamoussa
04	Ksob	M'sila	15.6	20	IRR. périmètre du Ksob
05	Mefrouch	Tlemcen	14.5	17	AEP- Tlemcen

(Source : Agence Nationale des Barrages et Transferts)

Concernant le type de modèles utilisés pour la mise en équation des séries chronologiques concernées en vue de la prédiction, sujet de la deuxième partie de cette thèse, le choix a porté sur les processus stochastiques de Box et Jenkins. Ce choix est à notre humble avis justifié par deux raisons : la première est que ces derniers se prêtent bien à la modélisation des variables hydrologiques telles que les apports des rivières qui sont généralement caractérisés par une forte dépendance temporelle entre les valeurs successives. Ceci est validé par les résultats de la multitude d'applications faites dans le domaine de l'eau, depuis l'introduction de cette méthode dans les années soixante dix du siècle dernier jusqu'à ce jour, et ce dans toutes les régions du monde, il suffit de se référer aux références bibliographiques pour le constater (**Brillinger, 1976 ; Doornik & Ooms, 2003 ; Droesbeke et al., 1989 ; Geweke & Porter-Hudak, 1983 ; Hipel & McLeod, 1994 ; Koutsoyiannis, 2008**) entre autres. La deuxième raison est que la théorie de ces modèles qui revient à une trentaine d'années déjà n'est pas assez vulgarisée en Algérie pour être bien exploitée dans le domaine de la modélisation hydrologique. Cela dit, l'absence de cette théorie des programmes pédagogiques nationaux de formation en hydraulique que se soit au niveau de la graduation ou de la post-graduation, n'a fort heureusement pas empêché quelques applications ambitieuses de cette théorie de voir le jour, malheureusement ceci reste toujours restreint et demande la multiplication des contributions pour une meilleure exploitation de la méthode.

A cet effet, la méthode de Box et Jenkins pour la modélisation stochastique linéaire des séries chronologiques a été exposée en détail dans le premier chapitre de la deuxième partie de cette thèse. L'accent a été mis sur les outils de base de la méthode ainsi que sur son algorithme, tout en faisant ressortir son caractère itératif. Le second chapitre est consacré à l'application de la méthode aux données annuelles observées dans les dix stations hydrométriques. C'est un travail de modélisation stochastique uni-variée de type ARIMA, dont le résultat escompté est la mise en équation des différentes séries chronologiques ainsi que la prédiction de leurs valeurs futures. Dans le troisième chapitre, le pas de temps des séries chronologiques a été réduit et l'application de la méthode de Box et Jenkins est réalisée sur les données mensuelles des dix stations précédentes. La saisonnalité introduite par les données mensuelles fera appel à l'usage des modèles non stationnaires et saisonniers SARIMA multiplicatifs. Le résultat serait une série de 10 modèles de prédiction saisonniers représentant chacun la structure de dépendance interne d'une série chronologique.

Les notions de base ainsi que les concepts et outils mathématiques nécessaires à la compréhension des fondements théoriques de cette méthode ainsi que celle du filtre de Kalman, abordé dans la troisième partie de ce travail sont présentés dans la première partie de la thèse.

Dans le but de se libérer de la contrainte de stationnarité qui est à la base de la méthode de Box et Jenkins et dont le résultat est un modèle paramétrique dont les paramètres ne changent pas dans le temps, le recours aux méthodes de prédiction dynamique nous a orientés vers le filtre de Kalman. Ce choix est justifié par deux raisons : la première est l'optimalité de ses résultats, la deuxième est la nature dynamique du modèle mathématique qui permet de rendre compte de la variabilité temporelle du processus hydrologique étudié.

En effet, le Filtre de Kalman discret linéaire est l'un des outils les plus puissants et les plus adaptés aux situations complexes où tout change dans le temps (**Boone, 2000**). Il se base sur le concept des moindres carrés et offre une méthode de filtrage qui jouit d'une propriété très importante, à savoir l'optimisation séquentielle. Cette dernière est une conséquence directe de la structure de feed-back que possède le filtre de Kalman, et qui fait que le modèle est recalé au fur et à mesure que les nouvelles mesures arrivent (**Brown & Hwang, 1997**). Ceci, est un grand avantage puisque ça permet de prendre en considération les variations dans le temps des paramètres du modèle, ce qui est d'une importance primordiale en hydrologie où l'on a souvent affaire à des modèles non parfaitement stationnaires.

La fonction de base du filtre de Kalman est la procuration des estimations de l'état courant du système, mais il peut également servir à la prévision des valeurs futures de la variable décrite ou à l'amélioration des estimations antérieures de la dite variable (**Sen & Latif, 2002**).

Dans le cadre du présent travail, nous allons investir et développer une approche par le FK pour la modélisation stochastique et la prédiction multi-site des apports liquides annuels, en plus de l'amélioration des erreurs associées. Le résultat serait une opération de prédiction multi-site en ligne, où l'opérateur ne sera pas fixé par le temps encore moins par l'espace, en revanche il s'adaptera automatiquement aux changements dans les conditions météorologiques du système hydrologique en question. C'est l'objet de la troisième partie de cette thèse.

PARTIE I

Introduction à la modélisation stochastique linéaire des variables hydrologiques

PREAMBULE I

Une série chronologique est définie comme une suite d'observations correspondant à la même variable, indexée par le temps. L'ordre dans lequel ces observations sont échantillonnées importe peu en statistiques, car il s'agit d'apporter des jugements sur une population à partir d'un échantillon de la même population. Pour l'analyse des séries chronologiques, particulièrement en hydrologie, les choses sont différentes et l'ordre des observations revêt une importance primordiale. La périodicité de la série n'importe en revanche pas : il peut s'agir de mesures quotidiennes, mensuelles, trimestrielles, annuelles... voire même sans périodicité.

Lorsqu'une série est stable autour de sa moyenne, on parle de série stationnaire, inversement lorsqu'une série n'est pas stable autour de sa moyenne ou dont la moyenne n'est pas constante, on parle de série non stationnaire : par exemple si la moyenne croît sur l'ensemble de l'échantillon on parle de tendance. Enfin lorsqu'on observe des phénomènes qui se reproduisent à des périodes régulières, on parle de phénomène saisonnier.

Tout comme en probabilités/statistiques, quand on parle de processus stochastique/série chronologique, il faut bien comprendre la différence entre le processus théorique qui génère des données, sa réalisation telle qu'on l'observe sur les échantillons historiques à notre disposition qu'on appelle séries chronologiques, les futures réalisations et le modèle qu'on construit afin de tâcher de le représenter.

Pour approcher ce problème, on est bien souvent amené à faire quelques hypothèses nous permettant de recourir sans gêne à des outils mathématiques particuliers. En pratique, on considère que c'est impossible de connaître la distribution d'une série chronologique $\{X_t\}, t \geq 0$, on s'intéresse par conséquent à la modélisation de la distribution conditionnelle (à priori constante dans le temps) de $\{X_t\}$ via sa densité : $f(X_t | \mathbf{X}_{t-1})$ conditionnée sur l'historique du processus : $\mathbf{X}_{t-1} = (X_{t-1}, X_{t-2}, \dots, X_0)$. Il s'agit donc d'exprimer X_t en fonction de son passé. Le résultat de cette approche conditionnelle fournit une décomposition en termes de prévision-erreur, selon laquelle : $X_t = E[X_t | \mathbf{X}_{t-1}] + \varepsilon_t$. Où $E[X_t | \mathbf{X}_{t-1}]$ est la composante de X_t qui peut donner lieu à une prévision, quand l'historique du processus, \mathbf{X}_{t-1} est connu et où ε_t représente les informations imprévisibles. Dans ce cas, le processus stochastique est non paramétrique se définissant à partir de la distribution jointe des observations ou de ses premiers moments. On peut de la même façon définir un processus stochastique paramétrique se définissant au contraire à partir d'un mécanisme générateur qui est indexé par des paramètres c'est le cas des processus ARMA.

L'objectif de cette première partie est l'introduction des concepts théoriques de base qui sont indispensables à la compréhension des fondements théoriques de cette démarche. A cet effet, et dans le but de pouvoir bien appréhender les séries temporelles en hydrologie et les processus stochastiques linéaires pouvant leur être associés, cette première partie débutera par un premier chapitre qui introduira les notions et concepts mathématiques indispensables à la compréhension de la partie deux et trois de ce travail. Le deuxième chapitre introduira le concept de processus stochastique/ série chronologique en mettant l'accent sur le lien et la différence entre les deux notions et en insistant sur la relation d'antériorité qui relie l'ensemble de toutes les observations. Enfin, le troisième chapitre présentera comment une série chronologique peut être générée par un processus stochastique linéaire, l'accent sera particulièrement mis sur les fondements et les caractérisations théoriques des différents modèles ARMA qui seront préconisés, dans la deuxième partie de ce travail, pour la modélisation des séries chronologiques.

Chapitre I.1

NOTIONS DE BASE POUR ABORDER LES MODELES ARIMA ET LE FILTRE DE KALMAN

I.1.1 INTRODUCTION

D'une manière évidente, le recours à la méthode de Box- Jenkins ou à celle du Filtre de Kalman en vue d'une application à la prédiction hydrologique, comme dans n'importe quel domaine des sciences de l'ingénieur, requiert la connaissance de quelques notions de bases sur lesquelles reposent les fondements théoriques de ces dernières. Dans ce sens un rappel au sujet de quelques notions élémentaires de probabilités et des variables aléatoires s'avère indispensable et le passage par les processus stochastiques et les systèmes linéaires introduisant des variables d'état est incontournable.

Ce chapitre qui regroupe ces notions et ces concepts de base peut donc être considéré comme une introduction d'une part à la méthode de Box-Jenkins dont l'objectif est la modélisation paramétrique linéaire des séries chronologiques dans un contexte stochastique, et d'autre part à la méthode du filtre de Kalman discret dont l'objectif est l'estimation optimale de l'état d'un système linéaire, tout en restant dans un contexte stochastique.

A cet égard, nous allons d'abord rappeler les notions de probabilités et variables aléatoires. Ceci sera fait d'une manière assez simple pour en saisir les concepts sans pour autant nous alourdir avec des formulations mathématiques qui sont parfois indigestes. Ensuite, nous allons rappeler comment on caractérise mathématiquement (ou stochastiquement) les signaux aléatoires (ou stochastiques). D'autres concepts tels que les bruits blancs gaussiens, les modèles espace-état, les modèles boîte noire, et le filtrage qui seront rencontrés plus tard sont également inclus.

I.1.2 PROBABILITES ET VARIABLES ALEATOIRES

Les définitions sont extraites des références (**Papoulis, 1991 ; Beran, 1994 et 1994a**).

I.1.2.1 Probabilités

La plupart d'entre nous ont une notion sur ce qu'on appelle un évènement aléatoire, ou sur la probabilité de réalisation d'un évènement dans un espace de probabilités. D'une manière très

simple, la probabilité qu'un évènement discret, tel le jet d'une pièce de monnaie par exemple, favorisant un évènement particulier A est défini par :

$$P(A) = \frac{\text{Nombre de cas favorables à } A}{\text{Nombre de cas possibles}} \quad (\text{I. 1.1})$$

La probabilité de réalisation de A ou B est :

$$P(A \cup B) = P(A) + P(B) \quad (\text{I. 1.2})$$

Dans le cas où les évènements A et B sont indépendants, la réalisation de l'un n'influe pas sur la réalisation de l'autre et la probabilité de A et B

$$P(A \cap B) = P(A) \cdot P(B) \quad (\text{I. 1.3})$$

La probabilité de réalisation de A , sachant que B est déjà réalisé est appelée probabilité conditionnelle :

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \quad (\text{I. 1.4})$$

Cette définition triviale de la notion de probabilité fait appel à d'autres notions telles que l'espace de probabilité et les variables aléatoires, que nous allons reprendre d'une manière assez exhaustive.

I.1.2.2 L'espace de probabilité

Un espace de probabilités est un triplet (Ω, \mathcal{A}, P) où

- Ω est l'ensemble des résultats possibles d'une expérience.
- \mathcal{A} est une classe de sous-ensembles de Ω (appelés évènements) formant σ -algèbre, i.e.
 - i. $\Omega \in \mathcal{A}$
 - ii. $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
 - iii. $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$, pour toute suite $\{A_1, A_2, \dots\} \subseteq \mathcal{A}$
- $P : \mathcal{A} \mapsto [0, 1]$ est une fonction qui assigne à chaque évènement $A \in \mathcal{A}$, un nombre $P(A) \in [0, 1]$ appelé la probabilité de A , telle que
 - i. $P(\Omega) = 1$
 - ii. Si $\{A_i\}_{i=1}^{\infty}$ est une suite d'évènements disjoints, alors $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

I.1.2.3 Variable aléatoire réelle (v.a.)

- **Définition (heuristique)**

Une variable aléatoire réelle X est une variable à valeurs réelles dont le comportement peut être décrit par une loi de probabilité. En général cette loi de probabilité est décrite par une fonction de distribution de probabilité.

$$F(x) = P[X \leq x] \quad (\text{I. 1.5})$$

- **Définition (formelle)**

Une variable aléatoire X réelle est une fonction X :

$X: \Omega \rightarrow \mathcal{R}$, telle que

$X^{-1}([-\infty, x]) \equiv \{\omega \in \Omega: X(\omega) \leq x\} \in \mathcal{A}, \forall x \in \mathcal{R}$ (fonction mesurable). La loi de distribution de X est définie par :

$$F(x) = P[X^{-1}([-\infty, x])] \quad (\text{I.1.6})$$

Une variable aléatoire peut être continue ou discrète comme elle peut être scalaire ou vectorielle.

I.1.3 CARACTERISATION D'UNE VARIABLE ALEATOIRE SCALAIRE

I.1.3.1 Fonction de répartition et fonction de densité de probabilité

Soit X une v.a scalaire et continue, la probabilité des évènements se calcule sur des intervalles. Une fonction très connue pour représenter la probabilité des v.a est définie par la fonction de distribution cumulative ou fonction de répartition $F(x)$ (Chatefield, 2001). $F(x)$ associe à tout réel, la probabilité de l'évènement ($X < x$) et on note

$$F(x) = P(X < x) \quad (\text{I.1.7})$$

Les propriétés de cette fonction sont les suivantes :

- $\lim_{x \rightarrow +\infty} F(x) = 1$ et $\lim_{x \rightarrow -\infty} F(x) = 0$,
- $\forall x_1, x_2, P(x_1 \leq X < x_2) = F(x_2) - F(x_1)$,
- $F(x)$ est monotone, non décroissante, discrète ou continue selon que la v.a X prenne des valeurs discètes ou continues.

Si $F(x)$ est dérivable, alors sa dérivée est appelée fonction de densité de probabilité notée $f(x)$ tel que :

$$f(x) = \frac{dF(x)}{dx} \quad (\text{I.1.8})$$

Soit

$$f(x)dx = P(x \leq X < x + dx)$$

Elle possède les propriétés suivantes :

- $f(x)$ est une fonction non négative
- $\int_{-\infty}^{+\infty} f(x) dx = 1$

Il faut noter selon la 2eme propriété de $F(x)$ que la probabilité sur n'importe quel intervalle $[a, b]$ est définie par

$$P(a \leq X < b) = \int_a^b f(x) dx$$

Ce qui fait qu'au lieu de faire la somme des probabilités, tel qu'en équation (I.1.1) pour des événements discrets, on procède à l'intégration de $f(x)$ sur l'intervalle qui nous intéresse.

Cela étant, il faut savoir que pour caractériser et manipuler mathématiquement une v.a donnée, on utilise également les moments de cette v.a. Ce qui nous intéressera dans le cadre de ce travail c'est les deux premiers moments à savoir le moment d'ordre un qui est connu par la moyenne et le moment centré d'ordre deux qui exprime la variance.

I.1.3.2 Moyenne et Variance

La notion de moyenne est familière à nous tous. Par exemple pour une v.a X qui prend ses valeurs dans un ensemble discret de N essais, $x_i, i = 1:N$, la moyenne de l'échantillon de N valeurs est notée

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_N}{N} \quad (\text{I.1.9})$$

Sachant que la probabilité de x_i , $P(x_i) = 1/N = P_i$.

$$\bar{X} = P_1 x_1 + P_2 x_2 + \dots + P_N x_N$$

$$\bar{X} = \sum_{i=1}^N P_i x_i \quad (\text{I.1.10})$$

Si on refait l'expérience un nombre infini de fois ($N \rightarrow \infty$), on espère d'avoir $(P_i N)$ fois l'évènement x_i . Cette notion d'essai un nombre infini de fois conduit à la définition conventionnelle de la valeur espérée ou « Espérance mathématique ».

Ainsi, dans le cas discret l'espérance mathématique d'une v.a X , notée $E(X)$ coïncide avec sa moyenne \bar{X} .

$$E(X) = \sum_{i=1}^N P_i x_i \quad (\text{I.1.11})$$

Avec

- $x_i : (i = 1:N) N$ réalisations possibles de,
- P_i : leurs probabilités correspondantes. C'est la probabilité que $X = x_i$, notée $P(X = x_i) = P_i$

D'une manière similaire, pour les v.a continues

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx \quad (\text{I.1.12})$$

On note que (15) et (16) peuvent être appliquées à des fonctions de v.a, tel que

$$E(g(X)) = \sum_{i=1}^N P_i g(x_i) \quad (\text{I.1.13})$$

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x) f(x) dx \quad (\text{I. 1.14})$$

Comme il a déjà été mentionné, l'espérance d'une v.a est connue par le « moment d'ordre un ». En général le moment d'ordre k d'un v.a est obtenu en choisissant la fonction $g(x) = x^k$ dans les équations (I.1.13) et (I.1.14) selon le cas discret ou continu tel que

$$E(X^k) = \sum_{i=1}^N P_i (x_i^k) \quad (\text{I. 1.15})$$

$$E(X^k) = \int_{-\infty}^{+\infty} x^k f(x) dx \quad (\text{I. 1.16})$$

Le « moment d'ordre deux » est obtenu en prenant $k = 2$, dans (I.1.15) et (I.1.16) on obtient

$$E(X^2) = \sum_{i=1}^N P_i (x_i^2) \quad (\text{I. 1.17})$$

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x) dx \quad (\text{I. 1.18})$$

Un moment d'un intérêt particulier pour nous est le « moment centré d'ordre deux » ou variance d'une v.a. Il est obtenu en prenant $g(x) = x - E(x)$ et en appliquant (I.1.17) ou (I.1.18) selon le cas discret ou continu tel que

$$\begin{aligned} \text{Variance de } X &= E \left[(X - E(X))^2 \right] \\ \text{Var}(X) &= E(X^2) - E(X)^2 \end{aligned} \quad (\text{I. 1.19})$$

La variance est une propriété statistique très utile pour l'étude des événements aléatoires car elle exprime la dispersion autour de leurs moyennes ou en d'autres termes à quel point ils sont « bruités ». La racine carrée de la variance est appelée « déviation standard » son avantage c'est qu'elle donne cette dispersion dans la même unité de mesure de la v.a en question (**Hosking, 1996**).

I.1.3.3 Distribution Normale ou de Gauss

C'est une distribution de probabilité spéciale (**Hirsch et al., 1993 ; Hendry, 1993**). Elle a été historiquement populaire dans la modélisation des événements aléatoires pour une variété de raisons : d'une part, la plupart des v.a naturelles sont Normalement distribuées ou presque, d'autre part sous certaines conditions modérées, il peut être prouvé que la somme de v.a avec n'importe quelle distribution de probabilité, tend vers une distribution Normale. Le théorème central limite établit formellement cette propriété.

L'intérêt (mathématique) des v.a Gaussiennes c'est qu'elles sont entièrement caractérisées par leurs moments d'ordre un et deux.

Etant donnée une v.a continue X qui est Normalement distribuée, avec une moyenne μ et une variance σ^2 , notée $X \rightarrow N(\mu, \sigma^2)$. La fonction de densité de probabilité de X est donnée par :

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2}\right) \quad x \in]-\infty, +\infty[\quad (\text{I. 1.20})$$

- Le premier moment : $E(x) = \mu$
- Le deuxième moment centré: $E[(x - \mu)^2] = \sigma^2$

Il est important à ce niveau de noter que n'importe quelle fonction linéaire d'une v.a Normale est également une v.a Normale. Si $X \rightarrow N(\mu, \sigma^2)$ et $Y = aX + b$ alors $Y \rightarrow N(a\mu, a^2\sigma^2)$. La fonction de densité de probabilité de Y est la suivante :

$$f(y) = \frac{1}{\sqrt{2\pi a^2 \sigma^2}} \exp\left(-\frac{1}{2} \frac{(y - (a\mu + b))^2}{a^2 \sigma^2}\right) \quad (\text{I. 1.21})$$

Enfin si X_1 et X_2 sont deux v.a indépendantes et Normales avec des moyennes μ_1 et μ_2 et des variances σ_1^2 et σ_2^2 respectivement, alors leur somme est complètement caractérisée.

$$Z = (X_1 + X_2) \rightarrow N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

Graphiquement la fonction de densité de probabilité est une cloche.

I.1.4 CARACTERISATION D'UNE VARIABLE ALEATOIRE VECTORIELLE

I.1.4.1 Fonction de répartition et fonction de densité de probabilité

Soit une v.a vectorielle à q dimensions prenant ses valeurs dans \mathcal{R}^q

- Sa fonction de répartition est

$$F(x_1, x_2, \dots, x_q) = P(X_1 < x_1, X_2 < x_2, \dots, X_q < x_q) \quad (\text{I. 1.22})$$

- Sa fonction de densité de probabilité est

$$f(x_1, x_2, \dots, x_q) = \frac{\partial^q F(x_1, x_2, \dots, x_q)}{\partial x_1 \dots \partial x_q} \quad (\text{I. 1.23})$$

I.1.4.2 Moyenne et Covariance

Pour les moments, ce qui nous intéresse c'est uniquement le vecteur des moments d'ordre un (vecteur moyen) et la matrice des moments centrés d'ordre deux (matrice de covariance)

- Moyenne :

$$E(X) = E(E(x_1), E(x_2), \dots, E(x_q))^T \quad (\text{I. 1.24})$$

– Covariance :

$$Cov_x = E[(x - E(x))(x - E(x))^T] \quad (I.1.25)$$

L'élément $Cov_x(i, j)$ de la ligne i et la colonne j de la matrice de covariance Cov_x , vérifie

$$Cov_x(i, j) = \int_{\mathcal{R}^2} (x_i - E(x_i))(x_j - E(x_j)) dF(x_i, x_j) \quad (I.1.26)$$

La matrice de covariance est définie, positive et symétrique.

I.1.4.3 Vecteur aléatoire de Gauss

Pour un vecteur aléatoire Gaussien X de dimension, de moyenne m et de covariance C ,

$$f(x) = \frac{1}{(2\pi)^{q/2} \sqrt{\det C}} \exp\left(-\frac{1}{2} (x - m)^T C^{-1} (x - m)\right) \quad (I.1.27)$$

Avec :

- C^{-1} : matrice inverse de C ,
- T : indice transposé,
- det : déterminant.

On rappelle qu'un tel vecteur peut être généré à partir d'un vecteur Gaussien Normalisé (c'est-à-dire de moyenne nulle et de covariance unité) de la façon suivante :

$$X = m + VN$$

V est choisie telle que : $VV^T = C$

Dans ce qui suit on parle de variable (resp. signal) aléatoire qu'il s'agisse de variable (resp. signal) scalaire ($q = 1$) ou vectorielle à q composantes.

I.1.5 INDEPENDANCE ET PROBABILITE CONDITIONNELLE

A l'instar des v.a discrètes et les équations (I.1.1) et (I.1.2), l'indépendance et la probabilité conditionnelle peuvent être définies pour les v.a continues.

Deux v.a continues sont dites statistiquement indépendantes si leur probabilité conjointe est égale au produit de leurs probabilités individuelles :

$$f(x, y) = f(x) \cdot f(y) \quad (I.1.28)$$

En plus la règle de Bayes découle de (I.1.8) offrant une manière de spécifier la fonction de densité de probabilité de la v.a x conditionnée par la réalisation à priori de la v.a y . La règle de Bayes est la suivante :

$$f(x/y) = \frac{f(x/y) \cdot f(y)}{f(x)} \quad (I.1.29)$$

Dans les paragraphes précédents, notre attention était focalisée sur les caractéristiques spatiales

des v. a., comme il a déjà été mentionné, la magnitude de la variance d'une v.a nous donne une idée sur sa dispersion autour de sa moyenne ou en d'autres termes : à quel point cette v.a est bruitée. Cependant la variance ne nous apprend rien sur le taux ou la propagation de ce bruit au fil du temps. Un rappel succinct des caractéristiques temporelles et par conséquent spectrales s'impose. Les définitions sont tirées de (**Yevjevich, 1972**).

I.1.6 PROCESSUS STOCHASTIQUE ET BRUIT BLANC

Les processus stochastiques seront étudiés avec les détails nécessaires dans la deuxième partie de cette thèse, pour l'instant on se contente de leur définition qui est indispensable pour la compréhension des prochains paragraphes.

I.1.6.1 Processus stochastique

On appelle processus stochastique une suite $\{X_t\}$ de variables aléatoires indicées par le temps (t). Il faut bien noter que

- les X_t ne sont pas forcément indépendants les uns des autres,
- la loi de X_t dépend de t

La notion de processus généralise celle de variable aléatoire : la réalisation d'un processus est une série chronologique (trajectoire).

I.1.6.2 Bruit Blanc

Un processus X_t est un bruit blanc s'il constitue un échantillon d'espérance nulle, non corrélé et de même variance.

On écrit :

$$X_t \sim BB(0, \sigma^2) \text{ ou}$$

$$X_t \sim WN(0, \sigma^2) \quad (\text{de l'anglais } white \text{ noise})$$

Un tel processus est l'exemple d'un processus stationnaire, il n'a ni tendance ni mémoire : la connaissance de la valeur du processus à une date donnée n'apporte aucune information pour la prédiction de sa valeur à une date ultérieure.

La fonction d'auto-corrélation d'un bruit blanc est donc nulle partout sauf quand ($k = 0$). En d'autres termes :

$$\rho(k) = \begin{cases} \text{Constante } A & \text{si } k = 0 \\ 0 & \text{Ailleurs} \end{cases} \quad (\text{I. 1.30})$$

Cette propriété temporelle qui fait d'un bruit blanc complètement non corrélé avec lui-même sauf dans le temps présent ($k = 0$) conduit à sa qualification d'indépendance. C'est une propriété qui est très importante dans la mesure où elle fournit une première indication sur l'indépendance d'un processus.

I.1.7 ESTIMATION STOCHASTIQUE LINEAIRE

En pratique, il existe plusieurs approches pour estimer l'état inconnu d'un système linéaire à partir d'un groupe de mesures. Plusieurs de ces méthodes ne prennent pas en compte la nature « bruitée » des mesures, due à des origines diverses.

Sous l'hypothèse de linéarité et d'un bruit Gaussien w , l'état x et la sortie y sont également des signaux (vectoriels) Gaussiens, qui sont donc complètement caractérisés par leurs premiers et deuxièmes moments. Avant de procéder au calcul de ces caractéristiques stochastiques proprement dit, ce qui fera l'objet du (FK) étudié dans la Partie III de cette thèse, un rappel concernant les modèles espace état et boîte noire s'impose. L'introduction aux systèmes linéaires faisant intervenir des variables d'état sera faite dans le chapitre III.2 concernant la méthode du filtre de Kalman.

I.1.7.1 Modèles Espace-Etat

L'étude de systèmes physiques, émettant au cours du temps des signaux déterminés par des états internes non observés, a conduit à développer en traitement du signal (une branche des sciences de l'ingénieur) les modèles dits espace-état.

Les modèles espace-état intègrent la distinction entre les variables observées (le signal) et les variables cachées (l'état interne). Ils sont constitués :

- d'une ou plusieurs équation(s) d'état décrivant la manière dont les variables cachées sont générées à partir de leur retard et d'innovations.
- d'une ou plusieurs équation(s) de mesure décrivant la manière dont les variables observées sont générées par les variables cachées et les résidus.

Dans leur version élémentaire, les modèles espace-état reposent sur un certain nombre d'hypothèses principales : les équations de mesure et d'état sont linéaires ; les bruits d'observation et d'innovation sont des bruits blancs ; les variables cachées suivent à un instant initial donné une loi gaussienne. À ces dernières, se sont ajoutées des hypothèses secondaires permettant de simplifier les procédures d'estimation telles que: l'indépendance entre les bruits d'observation et d'innovation (condition d'inversibilité) et l'indépendance entre la variable cachée initiale et ces bruits (condition de causalité) (**Brown & Hwang, 1997**).

Ces modèles sont une convenance notationnelle conçue essentiellement pour des problèmes d'estimation et de contrôle en traitement de signal. Les applications pratiques dans d'autres domaines, tel que l'hydrologie, ont en fait un outil incontournable pour l'obtention de bons résultats d'estimation.

Les modèles espace-état peuvent s'écrire sous une forme dite développée qui implique, sous certaines hypothèses, que la variable de mesure s'écrit sous la forme d'une moyenne mobile.

Considérons un système dynamique décrit par une équation de différence (ou différentielle dans le cas continu) d'ordre n ayant la forme suivante :

$$y_{i+1} = a_{0,i} y_i + \dots + a_{n-1,i} y_{i-n+1} + u_i \quad \text{avec } i \geq 0 \quad (\text{I. 1.31})$$

Où $\{u_i\}$ est un processus de bruit aléatoire, de moyenne nulle et de fonction d'auto-corrélation telle que :

$$E(u_i, u_j) = \Phi_{uu}(\tau) = \rho(\tau) \quad (\text{I. 1.32})$$

Avec des valeurs initiales $\{y_0, y_{-1}, \dots, y_{-n+1}\}$ qui sont de v. a centrées (moyenne nulle) ayant une matrice de covariance de dimension $n \times n$ telle que :

$$P_0 = E(y_{-j}, y_{-k}) \quad j, k \in \{0, n-1\} \quad (\text{I. 1.33})$$

Supposons en plus que les bruits u_i sont indépendants du processus à estimer, cette équation peut être réécrite dans la forme Espace –Etat comme suit :

$$\begin{bmatrix} y_{i+1} \\ y_i \\ y_{i-1} \\ \cdot \\ \cdot \\ y_{i-n+2} \end{bmatrix} = \begin{bmatrix} a_0 & a_1 & a_{n-2} & a_{n-1} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y_i \\ y_{i-1} \\ y_{i-2} \\ \cdot \\ \cdot \\ y_{i-n+1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{bmatrix} u_i \quad (\text{I. 1.34})$$

qui conduit à un modèle Espace-état

$$\begin{cases} x_{i+1} = Ax_i + Gu_i \\ y_i = [1 \ 0 \ 0 \ \dots \ 0]x_i \end{cases}$$

ou à sa forme générale

$$\begin{cases} x_{i+1} = Ax_i + Gu_i & (\text{I. 1.35}) \\ y_i = Hx_i & (\text{I. 1.36}) \end{cases}$$

L'équation (I.1.35) représente l'état du système. Elle montre comment le nouvel état x_{i+1} est modélisé comme combinaison linéaire de l'état précédent x_i et du bruit u_i . L'équation (I.1.36) est l'équation de mesure. Elle montre comment les mesures ou les observations y_i dérivent de l'état interne du système. Ces deux équations sont souvent référées comme modèle du système et modèle de mesure, elles servent de base pour toutes les méthodes d'estimation linéaire, et en particulier le filtre de Kalman qui fait l'objet de toute cette démarche (**Goodwin & Sign, 1985**).

I.1.7.2 Modèles Boite Noire

Le problème essentiel, dans ce cas, est d'estimer les états internes d'un système linéaire, en ayant accès uniquement aux sorties du système qui sont représentées par les mesures (les entrées sont également accessibles mais cet aspect sera omis pour nous car on travaillera sur des séries chronologiques). C'est ce qu'on appelle un problème « Boite noire » on accède à ses sorties par voie de mesure mais on ne peut pas observer ce qu'il ya à l'intérieur (**Chiu, 1978**).

Les différentes approches à ce problème sont typiquement basées sur les modèles espace état. Un modèle de système type est celui qui modélise la transformation de l'état de ce système par une équation de différence linéaire et stochastique similaire à (I.1.35) telle que :

$$x_k = Ax_{k-1} + Bu_k + w_k \quad (\text{I. 1.37})$$

en plus, il ya une forme similaire au modèle de mesure (I.1.36) qui modélise la relation entre l'état du système et la mesure. Elle peut être représentée par une expression linéaire de l'équation (I.1.36) tel que :

$$z_k = Hx_k + v_k \quad (\text{I.1.38})$$

Les termes w_k et v_k sont des v.a représentant respectivement les bruits de l'état et de la mesure.

Note

Il est tout à fait rationnel d'insister à ce stade sur le fait que les mesures ne sont pas exprimées uniquement en fonction de l'état mais peuvent être exprimées par n'importe quelle combinaison linéaire de ces états.

I.1.7.3 Caractérisation Stochastique

Sous l'hypothèse d'un système linéaire et d'un bruit ω Gaussien, l'état x et la sortie y sont également des processus (vectoriels) Gaussiens qui sont donc entièrement caractérisés par leurs moments d'ordre un et deux. Le calcul de ces caractéristiques stochastiques sont données par le théorème suivant pour les deux cas continu et discret :

- **Dans le cas continu**

Théorème :

Soit le système linéaire continu :

$$\dot{x}(t) = Ax(t) + M\omega(t)$$

$\omega(t)$ est un bruit blanc Gaussien stationnaire centré de densité spectrale de puissance W . On note $m(t_0)$ et $P(t_0)$ la moyenne et la covariance de l'état initial $x(t_0)$ (lui aussi aléatoire). Alors $x(t)$ est un signal aléatoire gaussien de moyenne $m(t)$ et de covariance $P(t)$ tel que :

- $m(t) = E[x(t)] = \exp[A(t - t_0)] m(t_0)$
- $P(t) = E[(x(t) - m(t))(x(t) - m(t))^T]$ vérifiant l'équation différentielle

$$\dot{P}(t) = AP(t) + P(t)A^T + MWM^T$$

Si le système est stable (toutes les valeurs propres de A sont à partie réelle négative) on tend vers un régime permanent (stationnaire) :

$\dot{P} = 0$ et $P(t) = P$ vérifie alors l'équation de LYAPUNOV continue

$$AP + PA^T + MWM^T = 0$$

Si l'on considère l'équation de sortie : $y(t) = Hx(t)$ alors la matrice de covariance $P_y(t)$ de $y(t)$ vérifie :

$$P_y(t) = HP(t)H^T$$

(Si l'on considère un bruit blanc de mesure alors cette covariance est infinie).

- **Dans le cas discret**

Théorème :

Soit un système linéaire discret :

$$x(k+1) = Ax(k) + M\omega(k)$$

$\omega(k)$ est un bruit pseudo-blanc gaussien centré de densité spectrale W ($E[(\omega(k)\omega(k+j))^T] = W\delta(j)$). Soit $m(0)$ et $P(0)$ la moyenne et la covariance de l'état initial $x(k_0) = x(0)$ (lui aussi aléatoire gaussien et indépendant de $\omega(k)$). On montre que $x(k)$ est un signal aléatoire gaussien de moyenne $m(k)$ et de covariance $P(k)$, tel que :

$$- m(k) = E[x(k)]$$

$$- P(k) = E[(x(k) - m(k))(x(k) - m(k))^T] \quad \text{vérifiant l'équation récurrente de LYAPUNOV}$$

$$P(k+1) = AP(k)A^T + MWM^T$$

Si le système est stable (toutes les valeurs propres de A sont de module inférieur à 1) on tend vers un régime permanent: $P(k+1) = P(k) = P$ vérifie alors l'équation de LYAPUNOV discrète :

$$P = APA^T + MWM^T$$

I.1.8 LE FILTRAGE LINEAIRE

D'une manière générale, le problème du filtrage consiste à déterminer des estimateurs de variables du système lorsque l'environnement présente des perturbations aléatoires. Le but est de déterminer un système (filtre) optimal au sens des moindres carrés (**Brown & Hwang, 1997**).

Si l'on dispose d'une série d'observations Z_t qui sont les mesures d'une variable X_t , entachées d'erreurs de mesure e_t d'origines diverses:

$$Z_t = X_t + e_t \quad (\text{I.1.39})$$

Si la variable X_t est décrite par un modèle d'évolution, le problème du filtrage est celui d'obtenir des estimations optimales au sens de la minimisation de la variance d'erreur entre la variable réelle et son estimation notées \hat{X}_t à partir des observations Z_1, Z_2, \dots, Z_t . En d'autres termes, essayer de supprimer en quelque sorte la présence du bruit e_t d'où le nom : filtrage.

Si on considère un système dont on dispose de mesures jusqu'au temps t , On peut chercher à estimer la valeur de l'état X à un instant donné τ sachant les valeurs des mesures précédentes, noté $(\hat{X}_\tau / Z_1, Z_2, \dots, Z_t)$. Suivant la valeur de τ trois sortes de problèmes de l'estimation de l'état du système X_τ au temps τ sont intéressantes:

- si $\tau < t$ il s'agit d'un problème de lissage ;
- si $\tau = t$ il s'agit d'un problème de filtrage ;
- si $\tau > t$ il s'agit d'un problème de prédiction.

Le problème du filtrage optimal, au sens de Kalman, est le même, il s'agit de trouver, pour le système dynamique en question, la meilleure estimation $\hat{X}_\tau / Z_1, Z_2, \dots, Z_t$ de l'état au sens du critère de la variance conditionnelle minimum.

Le filtre de Kalman permet d'appréhender directement le cas d'un système non stationnaire multivariable.

I.1.9 CONCLUSION

Ce premier chapitre a été une introduction aux modèles ARIMA et au filtrage de Kalman. Il nous a permis d'introduire les concepts de base et de présenter les outils mathématiques pour analyser les processus stochastiques (continus et discrets) et leur transmission dans les systèmes dynamiques linéaires. Il nous a permis d'introduire la notion de bruit blanc qui présente l'une des hypothèses de base pour la méthode de Box-Jenkins et celle de bruit blanc gaussien, qui a l'avantage d'être entièrement caractérisé par sa fonction d'auto-corrélation et qui présente l'une des hypothèses de base du Filtre de Kalman. Il nous a permis également, d'introduire les notions des modèles espace-état, qui forment la base de la formulation du filtre de Kalman ainsi que l'introduction de la notion du filtrage de Kalman.

Chapitre I.2

LES SERIES CHRONOLOGIQUES ET LES PROCESSUS STOCHASTIQUES EN HYDROLOGIE

I.2.1 INTRODUCTION

Les phénomènes hydrologiques sont des processus qui évoluent dans le temps et dans l'espace tels qu'ils soient en partie prédictibles (déterministiques) et en partie aléatoires (**Chow, 1964**). Quelques fois, la variabilité aléatoire l'emporte sur la variabilité déterministique, ce qui justifie l'hydrologie dans le traitement de tels processus comme étant purement aléatoires (**Hirsch *et al.*, 1993**).

Généralement, une variable X_t pouvant être quantifiée, est dite aléatoire quand sa valeur dans le temps ou dans l'espace est incertaine (**Bras & Rodriguez-Iturbe, 1993**). L'ensemble de toutes les variables aléatoires $\{X_t\}$, $t = 1, \dots, T$ où T est l'espace du temps, décrit le comportement du processus stochastique étudié au temps t , et la valeur qu'elle prend et qu'on peut mesurer est appelée « observation ». Une série temporelle mesure donc la variabilité d'un processus continu, mesuré à des intervalles de temps discrets. Elle peut être considérée comme un échantillon d'une population générée par un mécanisme stochastique qui évolue selon certaines lois probabilistes tout à fait inconnues (**Bras & Rodriguez-Iturbe, 1985 ; Brockwell & Davis, 1990**).

Faire référence à un processus stochastique lorsque l'on discute d'une variable hydrologique signifie d'une part qu'il existe un mécanisme générateur de la variable en question (processus), et que d'autre part cette variable est aléatoire (stochastique) (**Papoulis, 1991**). En pratique, on dispose d'un ensemble de réalisations observées sur un espace de temps discrétisé noté $\{X_t\}$ $t = 1, \dots, T$ et il importe de comprendre que l'on ne connaît alors qu'une réalisation particulière du processus sur la période de temps considérée (une trajectoire).

Ainsi pour pouvoir caractériser un (PS) par l'étude d'une série chronologique on est souvent amenés à faire des hypothèses simplificatrices telles que la stationnarité, ou l'ergodicité (**Davidson , 1994**). Mais il se trouve, en observant les graphiques des séries temporelles, que celles ci ne sont pas toujours stationnaires, c-à-d qu'elles sont dépendantes du temps. Dans ce cas, on est poussés à les rendre stationnaires par des artifices mathématiques connus, afin de respecter les hypothèses de travail.

I.2.2 LES SERIES CHRONOLOGIQUES

En hydrology, les données constituent souvent des séries d'observations sur une ou plusieurs variables faites à différents pas de temps réguliers, et présentent généralement des liaisons internes entre elles (Fortin et al., 1997 ; Robinson, 1994a). L'objectif principal de l'étude des séries chronologiques est justement l'étude de telles liaisons pour des fins d'applications différentes (Koutsoyiannis & Georgakakos, 2006 ; Koutsoyiannis & Montanari, 2007 ; See et al., 2007 ; Stedinger et al., 1993 ; Vaté, 1993).

I.2.2.1 Définition

Une série chronologique, appelée également « série temporelle » est une suite d'observations répétées d'une variable X_t (ou de plusieurs variables, c.à.d., un vecteur (X_t^1, X_t^2, \dots)), $\{X_t, t \in T\}$ où T est un intervalle de temps (Bloomfield, 1976).

I.2.2.2 Types importants de séries chronologiques

Il existe deux types de séries chronologiques selon que l'intervalle de temps T est discret ou continu.

- **Série discrète** : Dans le cas où l'intervalle de temps T est discret, c.à.d., $t = 1, 2, \dots, T$ où T est le nombre ($T \subseteq N$: nombres entiers) d'observations équidistantes, répétées avec une certaine fréquence, par exemple, une fois par an, par mois, par jour, etc., la série chronologique est dite discrète.
- **Série continue** : Dans le cas où l'intervalle de temps T est continu, ($T \subseteq \mathcal{R}, \mathcal{R}^+$: nombres réels, réels positifs), comme c'est le cas du domaine de la physique par exemple, où la variable peut être observée de façon continue, la série chronologique est dite continue.

I.2.2.3 Exemples de séries chronologiques

- Apports annuels en (Hm^3) observés à la station hydrométrique BENI BAHDEL sur une période de 68 ans (1925-1993).

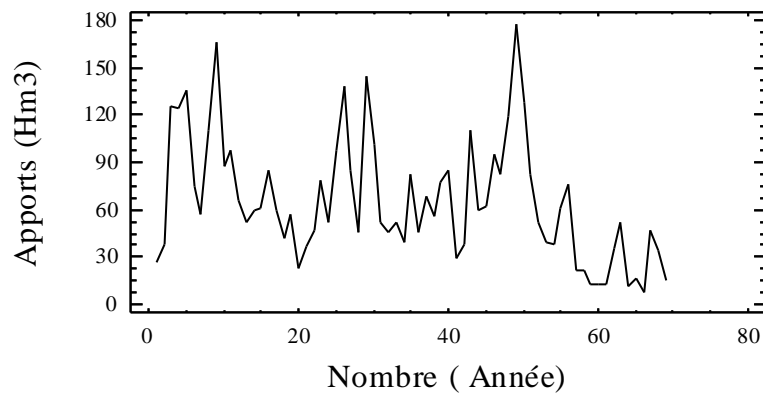


Fig. I.2.1 Apports annuels observés à la station hydrométrique BENI BAHDEL entre 1925 et 1993.

- Apports mensuels observés à la station hydrométrique AIN BERDA entre Septembre 1963 et Aout 1993.

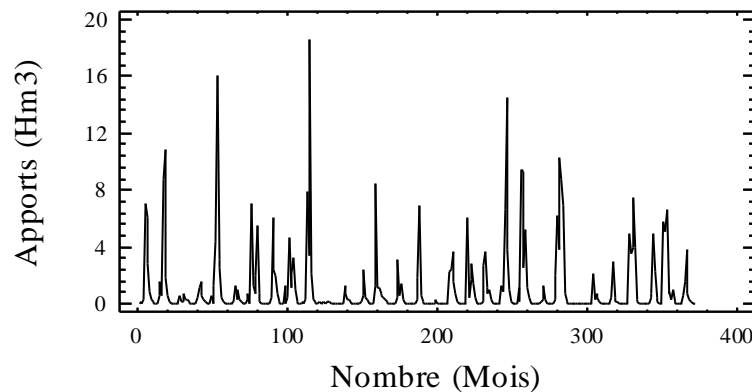


Fig. I.2.2 Apports mensuels observés à la station hydrométrique AIN BERDA entre 1963 et 1993.

I.2.2.4 Comportements des séries chronologiques

Les séries chronologiques sont souvent caractérisées par des comportements qui traduisent l'existence de relations particulières au sein d'elles qui sont décelables par simple inspection visuelle (**Salas, 1993**). Puisque l'objectif est de décrire et d'analyser ce genre de chroniques, il devient impératif donc de proposer des modèles qui intègrent ces différentes caractéristiques (**Cohn & Lins, 2005 ; Cox, 1981 ; Muluye & Coulibaly, 2007**).

Ces caractéristiques forment les composantes d'une série chronologique qu'il faudra analyser en les dissociant les unes des autres, c.à.d. en considérant une série comme issue de la combinaison des différentes composantes, tel que chacune d'elles ait une évolution simple.

- **La tendance ($T_i, 1 \leq n$)** : Elle représente l'évolution à long terme de la grandeur étudiée, et traduit l'aspect général de la série. C'est une fonction monotone, souvent polynomiale.

- **Les variations saisonnières ($S_i, 1 \leq n$)** : Sont liées au rythme imposé par les saisons météorologiques, ou encore par des activités sociales ou économiques.

- **Les cycles ($c_i, 1 \leq n$)** : Regroupent des variations à période moins précise autour de la tendance, par exemple les phases météorologiques de sécheresse et d'humidité. Ces phases durent généralement plusieurs années, mais n'ont pas de durée fixe. Sans informations spécifiques, il est généralement très difficile de dissocier la tendance du cycle, c'est pour cette raison que la composante de tendance regroupe pour la plupart du temps aussi les cycles.

- **Les fluctuations irrégulières (ou bruit) ($e_i, 1 \leq n$)** : Sont des variations de faible intensité et de courte durée, et de nature aléatoire, ce qui signifie qu'elles ne sont pas complètement explicables et donc impossible à prévoir. A cause de leur faible intensité, par rapport aux autres composantes, elles ne sont pas clairement apercevables dans les graphiques. Elles sont apercevables après enlèvement du signal et la question qui s'impose alors est de savoir si elles contiennent encore du signal ou elles sont vraiment un bruit.

- **Les variations accidentelles ou aberrantes** : Sont des valeurs isolées anormalement élevées ou faibles de courte durée. Ces observations brusques de la série sont généralement explicables (inondations dues à une crue, erreur de transcription,...). La plupart de temps, ces accidents sont intégrés dans la série de bruits (fluctuations irrégulières).

- **Les points de changement** : Ce sont des points où la série change complètement d'allure. Ils sont normalement explicables et imposent une analyse séparée de la série, par tronçons.

En résumé, nous considérons une série chronologique comme le résultat de trois composantes principales :

T_i la tendance (intégrant éventuellement un cycle)

S_i les variations saisonnières

e_i les variations aléatoires

I.2.2.5 Quelques types de décomposition

Après avoir détecté graphiquement les composantes présentes, il faut proposer un modèle parmi les suivants, sachant que toutes les autres combinaisons sont également possibles:

- **le modèle additif**

$$X_i = T_i + S_i + e_i, 1 \leq i \leq n \quad (\text{I.2.1})$$

- **le modèle multiplicatif**

$$X_i = T_i \times S_i \times e_i, 1 \leq i \leq n \quad (\text{I.2.2})$$

- **le modèle mixte** : Il s'agit là des modèles où addition et multiplication sont utilisées. On peut supposer, par exemple, que la composante saisonnière agit de façon multiplicative, alors que les fluctuations irrégulières sont additives :

$$X_i = f_i(1 + S_i) + e_i, 1 \leq i \leq n \quad (\text{I.2.3})$$

Enfin, la modélisation stochastique des séries temporelles commence en observant leur graphique et en cherchant une décomposition additive ou multiplicative.

I.2.2.6 Objectifs de l'étude des séries chronologiques

Les séries chronologiques sont le plus simple exemple d'une thématique plus large : l'estimation et la prédiction des processus stochastiques, c.à.d. des familles des variables aléatoires X_t (**Yevjevich, 1972**). Leur particularité, par rapport aux autres types de données statistiques, tient à la présence d'une relation d'antériorité qui ordonne l'ensemble des observations.

On se propose d'estimer la valeur de la variable X_t en un point t quelconque connaissant les valeurs X_{t_i} aux points t_i , pour $i = 1, \dots, N$. Le but principal est le choix d'un modèle raisonnable, qui permettra à partir des valeurs connues la prédiction des valeurs inobservables (comme les valeurs futures des séries chronologiques, ou moins accessibles physiquement, couteuses, etc.) avec une élimination du bruit d'observation éventuel (**Makridakis, 1998**).

- **Description et modélisation** : Le but est ici de déterminer les différentes composantes d'une série chronologique X_t , en particulier, obtenir la série corrigée des variations saisonnières (dessaïsonalisation). Pour les séries stationnaires, on peut aussi chercher à modéliser la série à l'aide d'un modèle ARMA, par exemple dans le but de faire de la prédiction ou de la simulation.

- **Prédiction** : Sur la base des observations X_1, \dots, X_T le but est de faire une prédiction, à la date T , de la réalisation $T + h$, notée $\hat{X}(T + h)$, de manière à ce que la somme des carrés des erreurs de prédiction soit minimale. Dans le cas des modèles ARMA, de nombreuses relations existent à fin de faire de la prédiction avec un intervalle de confiance.

- **Filtrage** : C'est un lissage qui consiste à transformer une série chronologique de façon à détecter (pour éliminer ou au contraire conserver) certaines caractéristiques (composante saisonnière, observations aberrantes, ... etc.). Cette méthode permet aussi de détecter les ruptures au sein d'une série chronologique.

I.2.2.7 Types de modèles ajustés aux séries chronologiques

Il existe deux catégories de modèles pour rendre compte d'une série temporelle. Les premiers considèrent que les données sont une fonction du temps ($y = f(t)$). Cette catégorie de modèle peut être ajustée par la méthode des moindres carrés, ou d'autres méthodes itératives. L'analyse des modèles par transformée de Fourier est une version sophistiquée de ce type de modèle (**Bloomfield, 1976 ; Brillinger & Krishnaiah, 1983**).

Une seconde catégorie de modèles cherche à déterminer chaque valeur de la série en fonction des valeurs qui la précèdent ($X_t = f(X_{t-1}, X_{t-2}, \dots)$). C'est le cas des modèles ARIMA (*Auto-*

Regressive Integrated Moving Average). Cette catégorie de modèles a été popularisée et formalisée par **Box & Jenkins (1970)**.

A noter que le choix de l'un ou l'autre type de modèle est surtout théorique: est-il raisonnable de penser que dans un phénomène donné, les points sont fondamentalement fonction des points précédents et de leurs erreurs, plutôt qu'un signal, périodique ou non, entaché de bruit. On peut noter cependant que souvent, on a recours à l'analyse de variance pour traiter les séries temporelles. Or une des assumptions majeures de l'ANOVA est que les résidus des différentes mesures ne sont pas auto-corrélés. Ce n'est évidemment pas le cas si la performance à l'essai t est liée à la performance réalisée à l'essai $t-1$.

- **Modèles déterministes** : C'est des modèles où la théorie des probabilités n'intervient pas. La série chronologique peut être considérée comme
 - une fonction déterministe du temps : $X_t = f(t)$,
 - ou une fonction de récurrence : $X_t = f(t, X_{t-1}, X_{t-2}, \dots)$
 Pourvu que la fonction $f(\cdot)$ et si nécessaire, les valeurs passées de X_t soient connues, il faut noter qu'un modèle déterministe permet une prédiction parfaite.

- **Modèles stochastiques** : C'est des modèles où les éléments X_t de la série chronologique sont considérés comme des variables aléatoires. Lorsqu'on considère une série $(X_t, t \in T)$ de variables aléatoires, on parle de processus stochastique (ou fonction aléatoire). La théorie des processus stochastiques est la base théorique de l'étude des modèles stochastiques. Il faut noter que c'est cette classe de modèles qui va être préconisée pour la modélisation des séries chronologiques abordées dans le contexte du présent travail.

I.2.3 LES PROCESSUS STOCHASTIQUES (PS)

I.2.3.1 Définition : Soit T un ensemble non vide. Un processus stochastique (PS) sur T est une collection de v.a $X_t: \Omega \rightarrow \mathcal{R}$ telle qu'à chaque élément $t \in T$ est associée une v.a X_t . Le processus s'écrit alors :

$$\{X_t: t \in T\}$$

Si $T = \mathcal{R}$ le processus est continu

Si $T = \mathbb{Z}$ le processus est discret.

On appellera ici processus stochastique (à temps discret) une suite $\{X_t\}$ de variables aléatoires indicée par le temps. Il faut bien noter que

- les X_t ne sont pas forcément indépendants les uns des autres,
- la loi de X_t dépend de t

En bref, la notion de processus généralise celle de variable aléatoire : la réalisation d'un processus est une série chronologique appelée également une trajectoire (**Brockwell & Davis, 1996**).

I.2.3.2 Auto-corrélations totales et auto-corrélations partielles d'un processus stochastique

Il s'agit d'apprécier les dépendances existant entre les observations constituant une trajectoire du processus. L'emploi du coefficient de corrélation paraît alors naturellement indiqué, d'où l'intérêt de la fonction d'auto-corrélation. Cependant, la corrélation entre X_t et X_{t+k} résulte en partie de celles existant éventuellement entre X_t et X_{t+1} , X_t et X_{t+2} , ..., X_t et X_{t+k-1} , d'où l'intérêt de la fonction d'auto-corrélation partielle.

- **Définition 1 :** On étudie la « mémoire » d'un (PS) en calculant son auto-corrélation au pas k , notée ρ_k :

$$\rho_k = \text{Corr}(X_t, X_{t+k}) = \frac{\text{Cov}(X_t, X_{t+k})}{\sqrt{\text{Var}(X_t)\text{Var}(X_{t+k})}} \quad (\text{I. 2.4})$$

qui mesure le lien entre les valeurs du processus à deux dates distantes de k .

Pour un processus stationnaire, $\rho(k)$ prend une forme plus simple :

$$\rho_k = \frac{\text{Cov}(X_t, X_{t+k})}{\text{Var}(X_t)} = \frac{\gamma_k}{\gamma_0} \quad (\text{I. 2.5})$$

On note déjà que la symétrie de la fonction d'auto-covariance entraîne celle de la fonction d'auto-corrélation : $\rho_k = \rho_{-k}$

Ainsi pour une série chronologique de N éléments, la suite de toutes les auto-corrélations contient toutes les informations sur la mémoire de cette série (**Koutsoyiannis, 2002**). On l'estime au moyen de l'auto-covariance comme suit:

$$\hat{\gamma}_k = \frac{1}{N} \sum_{t=k+1}^N (X_t - \bar{X})(X_{t-k} - \bar{X}) \quad (\text{I. 2.6})$$

$$\bar{X} = \frac{1}{N} \sum_{t=1}^N X_t \quad (\text{I. 2.7})$$

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\sum_{t=k+1}^N (X_t - \bar{X})(X_{t-k} - \bar{X})}{\sum_{t=1}^N (X_t - \bar{X})^2} \quad (\text{I. 2.8})$$

Il faut noter que l'on utilise N observations pour calculer la moyenne et la variance, alors que l'on en utilise que $N - k$ pour calculer γ_k . Donc quand $k \rightarrow N$, l'estimateur de γ_k tend vers zéro si le processus est stationnaire en covariance. Si le processus des X est un bruit blanc, alors toutes les observations sont des réalisations indépendantes d'un même processus stochastique et les auto-corrélation estimées $\hat{\rho}_k$ sont asymptotiquement normales.

$$\sqrt{N} \hat{\rho}_k \rightarrow N(0, 1) \quad (\text{I. 2.9})$$

si bien que $[-1.96\sqrt{N}, 1.96\sqrt{N}]$ forme un intervalle de confiance asymptotique à 5% des ρ_k autour de zéro. Le graphique des auto-corrélations (auto-corrélogramme) accompagné de son intervalle de confiance permet donc de juger si celles-ci sont nulles ou non. Maintenant on sait que si un processus est non-stationnaire, l'estimateur $\hat{\rho}_k$ tend vers 1 quand $N \rightarrow \infty$.

Dans la pratique, c-à-d, pour N fini, les autocorrélations d'une série non-stationnaire vont décroître lentement. Ce sera donc une indication pour différencier la série de manière à la rendre stationnaire.

Les auto-corrélations partielles permettent également de caractériser un processus stochastique. Par partielle il faut entendre conditionnelle à une partie du passé.

- **Définition 2 :** On définit l'auto-corrélation partielle de retard k comme la corrélation entre $(X_t - X_t^*)$ et $(X_{t-k} - X_{t-k}^*)$ où X_t^* désigne la régression de X_t sur les $k - 1$ valeurs $\{X_{t-1}, X_{t-2}, \dots, X_{t-k+1}\}$:

$$\begin{aligned} \phi_k &= \text{Corr}(X_t - X_t^*, X_{t-k} - X_{t-k}^*) \\ \phi_k &= \frac{\text{Cov}(X_t - X_t^*, X_{t-k} - X_{t-k}^*)}{\sqrt{\text{Var}(X_t - X_t^*)\text{Var}(X_{t-k} - X_{t-k}^*)}} \end{aligned} \quad (\text{I. 2.10})$$

Avec

$$X_t^* = \sum_{j=1}^{k-1} \alpha_j X_{t-j} \quad \text{et} \quad X_{t-k}^* = \sum_{j=1}^{k-1} \beta_j X_{t-j+1}$$

Où les α_j et β_j sont les coefficients des régressions.

Cette quantité rend compte de l'intensité de la liaison entre X_t et X_{t-k} en supprimant les liaisons induites par des variables intermédiaires $\{X_{t-1}, X_{t-2}, \dots, X_{t-k+1}\}$. On peut ainsi remarquer que pour tout processus, $\phi_1 = \rho_1$ puisque qu'il n'y a aucune variable intermédiaire entre X_t et X_{t-1} .

Les auto-corrélations partielles sont généralement estimées par récurrence car il n'existe pas de forme simple des estimateurs des ϕ_k . Ainsi, on sait que ϕ_2 vaut :

$$\phi_2 = \frac{\text{Corr}(X_t, X_{t-2}) - \text{Corr}(X_t, X_{t-1})\text{Corr}(X_{t-2}, X_{t-1})}{\sqrt{(1 - \text{Corr}(X_t, X_{t-1})^2)(1 - \text{Corr}(X_{t-2}, X_{t-1})^2)}}$$

et peut donc être estimé, en utilisant la stationnarité, par :

$$\hat{\phi}_2 = \frac{\hat{\rho}_2 - \hat{\rho}_1^2}{1 - \hat{\rho}_1^2} \quad (\text{I. 2.11})$$

Note

Pour un processus stochastique Gaussien de moyenne nulle ($E(X) = 0$), la fonction d'auto-corrélation est d'un intérêt particulier car elle le caractérise complètement (**Harvey, 1981**).

I.2.3.3 Processus stochastique stationnaire

Lorsque chacune des variables X_t vérifie $E(X_t) < \infty$ la loi du processus est partiellement résumée par l'espérance des différentes variables et par leurs covariances. Ces moments dépendent en général du temps, ce qui est gênant quand on veut tirer de l'information sur la loi sous-jacente de ce processus à partir de l'observation de réalisations de ce dernier. Pour pouvoir obtenir une accumulation d'information on est amenés à considérer des processus dits stationnaires (**Nerlove, 1979 ; Yevjevich, 1972**).

- **Définition :** *Un processus est stationnaire au second ordre si :*
 - $\forall t \in Z, E(X_t) = m(t) = m$ indépendant de t .
 - $\forall t \in Z, E(X_t^2) < \infty$
 - $\forall t, k \in Z, \text{Cov}(X_t, X_{t+k}) = \gamma(k)$

Ce type de stationnarité est aussi une propriété d'invariance des deux premiers moments par translation dans le temps. Mais on ne dit rien sur les moments d'ordre supérieur.

Il existe un autre type de stationnarité appelée stationnarité stricte (**Johnston, 1984**) et qui consiste à supposer que la distribution jointe de X_t et X_{t+k} ne dépend pas de t , mais seulement de k . Ici, c'est la distribution jointe qui est invariante par translation. Cette propriété est plus forte que la précédente car un processus stationnaire au second ordre peut posséder des moments d'ordre supérieur qui ne sont pas invariants par translation. La notion de stationnarité stricte et de stationnarité au second ordre se confondent par contre pour les processus Gaussiens car ceux-ci sont entièrement résumés par leurs deux premiers moments.

En fin, il convient de rappeler que la plupart des séries ne sont pas stationnaires, mais on peut essayer quand-même de se ramener à ce cas par des transformations (logarithmes, Box-Cox, différenciations, etc.) (**Gourieroux & Monfort, 1990**).

Note

La variance d'un (PS) centré et stationnaire est la valeur de sa fonction d'auto-corrélation à l'origine. Si $k = 0$ alors $\text{Cov}(X_t, X_t) = \text{var}(X_t) = \sigma_X^2$

- **Quel est l'intérêt pratique de la stationnarité ?**

Interpréter une série hydrologique de précipitations, d'apports ou de températures en termes de processus stochastique sur les années 1960-1990 par exemple, signifie que si l'on pouvait remonter le temps jusqu'en 1960 et rétablir les conditions initiales existant alors, nous obtiendrions une série de réalisations jusqu'en 1990 (et au-delà) différentes de celles que l'on trouve aujourd'hui dans les annuaires de l'Agence Nationale des Ressources Hydrauliques mais toutes aussi « raisonnables ». Cette maîtrise du temps et de l'environnement nous échappant encore, en plus l'expérimentation étant bien souvent impossible en hydrologie, ce qui rend la caractérisation complète d'un processus stochastique particulier, a priori une tâche impossible. Pour preuve, si l'on admet que X_t est la réalisation de la variable aléatoire X_t et si nous nous limitons, par exemple aux deux seuls premiers moments d'un processus que l'on suppose de plus gaussien, il nous faudrait donc estimer :

- d'une part : $E(X_1), E(X_2), \dots, E(X_N)$ soit N termes.
- D'autre part :

$$\begin{bmatrix} V(X_1) & \cdots & Cov(X_1, X_N) \\ \vdots & \ddots & \vdots \\ Cov(X_N, X_1) & \cdots & V(X_N) \end{bmatrix} \text{ soit } N(N+1)/2$$

Il est alors évident que cette tâche est impossible si l'on se souvient que l'on ne dispose que d'une seule trajectoire et donc de seulement N réalisations $\{X_1, X_2, \dots, X_N\}$, d'où l'importance du concept de stationnarité, qui justifie l'estimation des modèles statistiques observables une seule fois, ce qui est le cas souvent dans les séries temporelles hydrologiques. Dans ces conditions il est nécessaire d'imposer des restrictions sur les propriétés des processus si l'on veut avancer.

L'objet de l'une de ses restrictions est justement l'hypothèse de stationnarité dans laquelle nous allons travailler pour mener à bien l'analyse des séries chronologiques, il faut rappeler qu'elle nous permet de prendre n'importe quelle origine de temps pour observer N réalisations successives du processus, sans affecter leur distribution associée.

L'hypothèse d'ergodicité présente l'objet d'une 2^{ème} restriction sur les propriétés des (PS).

I.2.3.4 Processus stochastique ergodique

Le problème central dans la théorie des processus stochastiques est l'estimation de leurs diverses statistiques. L'ergodicité est un moyen qui mène à ce but, elle consiste à admettre que la loi des grands nombres est vérifiée même sur un ensemble de réalisations non indépendantes (**Priestley, 1981**).

- **Définition :** On appelle ergodique, un (PS) stationnaire X_t de \mathcal{R}^Z dont les lois jointes peuvent arbitrairement être déterminées, avec précision, à partir d'un échantillon $\{X_1, X_2, \dots, X_N\}$ suffisamment long. D'une manière formelle, X_t est dit ergodique si pour toute suite X de \mathcal{R}^Z et toute fonction g intégrable sur la loi des trajectoires de X , la limite suivante existe.

$$\lim_{n \rightarrow \infty} \left(\frac{1}{N} \right) [g(X) + gov(X) + \dots + gov^N(X)] = \int_{\mathcal{R}} g dq \quad (\text{I. 2.12})$$

Avec

$v: \mathcal{R}^Z \rightarrow \mathcal{R}^Z$, fonction de translation dans le temps définie par:

$$v^i(X_t) = X_{t+i} \quad (\text{I. 2.13})$$

q : la probabilité associée à X sur \mathcal{R}^Z (loi de la trajectoire de).

- **Exemple**

Si l'on veut déterminer le premier moment de X_t , en prenant $g = f(X_1, X_2, \dots, X_k)$, dans l'hypothèse que l'espérance mathématique de $f(X_1, X_2, \dots, X_k)$ existe, alors:

$$\lim_{n \rightarrow \infty} \left(\frac{1}{N} \right) [f(X_1, X_2, \dots, X_k) + f(X_2, X_3, \dots, X_{k+1}) + \dots + f(X_{n+1}, \dots, X_{n+k})] = Ef(X_1, X_2, \dots, X_k)$$

le premier moment est ainsi déterminé. De même si l'on veut déterminer la loi jointe de X_1, X_2, \dots, X_k , il suffit de prendre $f = I_A$ (indicatrice de A, A: partie mesurable de \mathcal{R}^Z), dans ce cas on aura:

$$\lim_{n \rightarrow \infty} \left(\frac{1}{N} \right) [I_A(X_1, X_2, \dots, X_k) + I_A(X_2, X_3, \dots, X_{k+1}) + \dots + I_A(X_{n+1}, \dots, X_{n+k})] = P(X_1, X_2, \dots, X_k)$$

la loi jointe de X_1, X_2, \dots, X_k est ainsi complètement déterminée, quand $N \rightarrow \infty$ par X_1, X_2, \dots, X_k .

On peut ainsi prendre n'importe quelle origine de temps pour observer k réalisations successives du processus sans affecter la distribution associée à ces k réalisations.

I.2.3.5 Innovations d'un processus stochastique

Dans la littérature l'appellation innovation correspond à la partie non prévisible du processus, elle est proche de la notion de résidus (**Vanmarcke, 1983**).

- **Définition :** X_t étant un processus stationnaire au second ordre, on appelle innovation du processus à la date t la variable définie par:

$$X_t - E(X_t | I_{t-1}) \quad (\text{I. 2.14})$$

Où I_{t-1} est l'ensemble d'information sur le processus disponible au temps t .

L'ensemble d'information peut comprendre tout le passé du processus. L'espérance est prise ici au sens d'espérance conditionnelle. Il s'agit de la meilleure prévision de X_t conditionnellement à un ensemble d'information.

I.2.4 CONCLUSION

Dans ce chapitre nous avons projeté la lumière sur les séries chronologiques et les mécanismes sous-jacents qui les génèrent. Nous avons montré l'intérêt et la particularité des séries chronologiques par rapport aux autres types de données statistiques et nous avons vu que cela tient à la présence d'une relation d'antériorité qui ordonne l'ensemble des observations. Nous avons vu également que cette mémoire qui est contenue dans les séries chronologiques fait appel aux mécanismes générateurs sous-jacents qui sont de nature stochastique, d'où leur appellation « processus stochastiques ». La relation et la différence entre les séries chronologiques et processus stochastiques a été mise en exergue par la même occasion. Nous avons également introduit la notion de stationnarité en montrant son intérêt dans la caractérisation de ces processus générateurs et présenté les artifices mathématiques pour l'obtenir, dans le cas d'une série chronologique non stationnaire. La notion d'ergodicité qui est à la base de la caractérisation stochastique a également été abordée. L'introduction de ces notions forme l'arrière plan théorique indispensable à la compréhension des parties suivantes.

Chapitre I.3

LES MODELES STOCHASTIQUES LINEAIRES DES SERIES CHRONOLOGIQUES EN HYDROLOGIE

I.3.1 LES PROCESSUS LINEAIRES

On appelle processus linéaire toute série temporelle qui puisse être représentée par un modèle linéaire après transformation, par exemple $\log(X_t) = a + bt + \varepsilon_t$. En réalité, un modèle linéaire est une approximation de modèles nettement plus complexes et ils ont la particularité d'être très flexibles et estimables avec un faible nombre d'observations.

I.3.1.1 Opérateur de retard

On aura souvent à considérer une variable en fonction de son passé. Il est donc commode de définir un opérateur qui transforme une variable X_t en sa valeur passée. C'est l'opérateur retard désigné par la lettre B (de l'anglais *backward shift operator*) et tel que pour tout $k \in N$:

$$B^k X_t = X_{t-k} \quad (\text{I. 3.1})$$

et pour α scalaire, $B\alpha = \alpha$. On peut utiliser l'opérateur B comme un chiffre, il peut multiplier et diviser tel que : si $BX_t = X_{t-1}$, alors : $X_t = B^{-1}X_{t-1}$ ce qu'on note parfois : $X_t = FX_{t-1}$, où F est l'opérateur avancé (de l'anglais *forward shift*).

L'opérateur B sera utilisé particulièrement à l'intérieur de polynômes notés par exemple:

$$\psi(B) = \psi_0 + \psi_1 B + \psi_2 B^2 + \dots + \psi_k B^k \quad (\text{I. 3.2})$$

Alors:

$$\psi(B)X_t = \psi_0 X_t + \psi_1 X_{t-1} + \psi_2 X_{t-2} + \dots + \psi_k X_{t-k} \quad (\text{I. 3.3})$$

Les opérations usuelles telles que l'addition, multiplication, division et inverse sont possibles sur l'ensemble des polynômes de retard avec les mêmes propriétés que sur les séries

entières. On retiendra en premier deux valeurs particulières des polynômes de retard $\psi(0)$ qui donne la valeur du premier coefficient du polynôme, son terme constant a_0 et $\psi(1)$ qui fournit la somme des coefficients de ce même polynôme. Enfin, l'opérateur $(1 - B)$ joue un rôle spécial dans la mesure où il permet de prendre la différence première d'une série chronologique :

$$(1 - B)X_t = X_t - X_{t-1} \quad (I.3.4)$$

I.3.1.2 Opérateur de différence

L'opérateur de différence Δ est défini par :

$$\Delta X_t = X_t - X_{t-1} \quad (I.3.5)$$

En introduisant l'opérateur de retard B , on peut obtenir :

$$\Delta X_t = X_t - BX_t = (1 - B)X_t$$

Ce qui fait que :

$$\Delta = 1 - B \quad (I.3.6)$$

Ce mode d'écriture sous forme de polynôme en B est en fait très pratique mais totalement formel. Il ne faut garder en esprit que, quand on écrit $(1 - B)X_t$, on définit à partir d'une série X_t , une nouvelle série qui à l'instant t , fait correspondre la différence entre la valeur de la série observée à l'instant t et celle observée à l'instant $t - 1$. Il est généralement utilisé pour éliminer la tendance linéaire d'une série. Si cette tendance n'est pas linéaire, il peut être répété d fois jusqu'à disparition complète de la tendance, dans ce cas d est appelé « ordre de différenciation » et la $d^{\text{ième}}$ différence de X_t peut être définie comme suit :

$$\Delta^d X_t = (1 - B)^d X_t \quad \forall d \in \mathbb{N} \quad (I.3.7)$$

D'une manière similaire la différence saisonnière de la série est définie par :

$$\Delta_s X_t = (1 - B^s)X_t = X_t - X_{t-s} \quad (I.3.8)$$

elle permet d'éliminer une saisonnalité de période s et peut également être répétée d fois, d est alors appelé « ordre de désaisonnalisation ». On utilise fréquemment les opérateurs Δ_{12} pour les données mensuelles, Δ_7 pour les données journalières, Δ_{24} pour les données horaires, etc.

$$\Delta_s^d X_t = (1 - B^s)^d X_t \quad (I.3.9)$$

Il faut bien noter que l'opérateur Δ^d raccourcit la série des X_t de d valeurs puisque $\Delta^d X_t$ n'est défini que pour $t > d$. La même chose pour l'opérateur Δ_s qui raccourcit la série de s valeurs, les s premières servant de références pour la saisonnalité.

I.3.1.3 Filtre linéaire

On appelle filtre linéaire une transformation d'une série entrée $\{X_t\}$, en série sortie $\{Y_t\}$ par application du polynôme retard $\psi(B)$:

$$Y_t = \psi(B)X_t = \left(\sum_{j=-n}^m \psi_j B^j \right) X_t = \sum_{j=-n}^m \psi_j X_{t-j} \quad (\text{I. 3.10})$$

I.3.1.4 Processus linéaire

Un processus Y_t est dit linéaire en ε_t , s'il existe une série entrée $\{\varepsilon_t\}_{t \in R}$ telle qu'il puisse se mettre sous la forme suivante :

$$Y_t = \psi(B)\varepsilon_t = \left(\sum_{j=-\infty}^{+\infty} \psi_j B^j \right) \varepsilon_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j} \quad (\text{I. 3.11})$$

Où ε_t est un bruit blanc de moyenne nulle et de variance connue σ^2 et où $\sum_{-\infty}^{\infty} \psi_j^2 < \infty$.

Il faut remarquer que beaucoup de processus stochastiques avec dépendance sont obtenus par des transformations linéaires de bruits blancs comme nous allons voir dans les paragraphes suivants.

I.3.2 REPRESENTATION DE WOLD

Un processus stochastique non paramétrique se définit à partir de la distribution jointe des observations ou de ses premiers moments. Un processus stochastique paramétrique se définit au contraire à partir d'un mécanisme de génération qui est indexé par des paramètres. Il est possible de caractériser ce mécanisme de manière très générale au moyen du théorème de Wold. Ce théorème montre que tout processus stationnaire peut être représenté de manière unique par la somme de deux composantes indépendantes, une composante régulière parfaitement prévisible parfois appelée déterministe et une composante stochastique:

Théorème : tout processus stationnaire au sens faible et de moyenne zéro, $\{X_t\}$, admet la représentation suivante :

$$X_t = \sum_{j=0}^{+\infty} \psi_j \varepsilon_{t-j} + d_t \quad (\text{I. 3.12})$$

où $\sum_{j=0}^{+\infty} \psi_j \varepsilon_{t-j}$ est la composante linéaire stochastique, d_t est la composante déterministe de X_t , et où $\psi_0 = 1$ et $\sum_{j=0}^{\infty} \psi_j^2 < \infty$.

Le terme ε_t est un bruit blanc qui représente l'erreur faite en prévoyant X_t à partir d'une fonction linéaire de son historique $\mathbf{X}_{t-1} = \{X_{t-j}\}_{j=1}^{\infty}$:

$$\varepsilon_t = X_t - E(X_t | \mathbf{X}_{t-1}) \quad (\text{I. 3.13})$$

la variable d_t est non-corrélée aux ε_{t-j} , pour tout $j \in \mathbb{N}$, bien que d_t puisse être prévue arbitrairement bien à partir d'une fonction linéaire de \mathbf{X}_{t-1} :

$$d_t = E(d_t | \mathbf{X}_{t-1}) \quad (\text{I. 3.14})$$

Ce théorème est à la base de la modélisation des séries temporelles stationnaires. La composante stochastique est exprimée sous la forme de ce que l'on appelle un processus moyenne mobile infinie $MA(\infty)$ et la série temporelle à modéliser est centrée c-à-d, de moyenne zéro. Un des buts de la modélisation consiste à approximer cette moyenne mobile infinie par un processus ayant un nombre fini de paramètres. C'est ce que l'on va voir dans la prochaine section en étudiant les processus AR, MA et ARMA. Les références sont nombreuses nous citons entre autres (**Granger, 1977** ; **Harvey, 1981** ; **Pourahmadi, 2001** ; **Papoulis, 1991** ; **Hendry & Clements, 1998** ; **Yevjevich, 1971**)

I.3.3 LES PROCESSUS STOCHASTIQUES LINEAIRES ARIMA

I.3.3.1 Modèle moyenne mobile d'ordre infini $MA(\infty)$

- **Définition :** On appelle processus $MA(\infty)$ un processus linéaire $X_t, t \in \mathbb{Z}$ vérifiant une relation :

$$X_t = \sum_{i=-\infty}^{\infty} \psi_i \varepsilon_{t-i}, \forall t \in \mathbb{Z} \quad (\text{I. 3.15})$$

où ε_t est un bruit blanc de variance σ_ε^2 .

On peut montrer qu'à la limite pour $\sum_{i=-\infty}^{\infty} \psi_i^2 < \infty$ le processus X_t constitué de la somme infinie de chocs aléatoires converge.

Remarquons qu'en définissant : $Y_t = \mu + X_t$ nous obtenons le processus linéaire

$$Y_t = \mu + \sum_{i=-\infty}^{\infty} \psi_i \varepsilon_{t-i}, \forall t \in \mathbb{Z} \quad (\text{I. 3.16})$$

qui est également convergent pour $\sum_{i=-\infty}^{\infty} \psi_i^2 < \infty$, μ étant une constante.

En introduisant l'opérateur de retard B dans l'équation précédente on obtient :

$$Y_t = \mu + \psi(B)\varepsilon_t = \mu + \left(\sum_{i=-\infty}^{\infty} \psi_i B^i \right) \varepsilon_t \quad (\text{I. 3.17})$$

- **Espérance et variance**

$$E(X_t) = 0 \quad (\text{I. 3.18})$$

$$Var(X_t) = E(X_t^2) = \sigma_\varepsilon^2 \sum_{i=-\infty}^{\infty} \psi_i^2 \quad (\text{I. 3.19})$$

- **Auto-covariance et auto-corrélation**

$$\gamma_h = \text{Cov}(X_t, X_{t+h}) = E(X_t X_{t+h}) = \sigma_\varepsilon^2 \sum_{i=-\infty}^{\infty} \psi_i \psi_{i+|h|} \quad (\text{I. 3.20})$$

d'où

$$\rho_h = \frac{\sum_{i=-\infty}^{\infty} \psi_i \psi_{i+|h|}}{\sum_{i=-\infty}^{\infty} \psi_i^2} \quad (\text{I. 3.21})$$

la série $\sum_{i=-\infty}^{\infty} \psi_i \psi_{i+|h|} < \infty$ étant convergente à la limite. Si $Y_t = \mu + X_t$ alors $E(Y_t) = \mu$, $\text{Var}(Y_t) = \text{Var}(X_t)$ et $\text{Cov}(Y_t, Y_{t+h}) = \text{Cov}(X_t, X_{t+h})$.

- **Stationnarité**

Les deux processus X_t :

$$X_t = \sum_{i=-\infty}^{\infty} \psi_i \varepsilon_{t-i}, \forall t \in Z$$

et Y_t :

$$Y_t = \mu + \sum_{i=-\infty}^{\infty} \psi_i \varepsilon_{t-i}, \forall t \in Z$$

sont stationnaires du second ordre, car $E(X_t)$ et $\text{Cov}(X_t, X_{t+h})$ et $E(Y_t)$ et $\text{Cov}(Y_t, Y_{t+h})$ ne dépendent pas de t .

I.3.3.2 Processus moyenne mobile d'ordre fini MA(q)

- **Définition :** On appelle processus MA(q) un processus linéaire centré $X_t, t \in Z$ vérifiant une relation :

$$X_t = \sum_{i=0}^q \theta_i \varepsilon_{t-i}, \forall t \in Z \quad (\text{I. 3.22})$$

où ε_t est un bruit blanc de variance σ_ε^2 , et $\theta_0 = 1$.

L'introduction de l'opérateur de retard B ramène (I. 3.22) à la forme :

$$X_t = \boldsymbol{\theta}(B) \varepsilon_t \quad (\text{I. 3.23})$$

où $\boldsymbol{\theta}(B) = \sum_{i=0}^q \theta_i B^i$ est un polynôme de degré q dont les coefficients sont $\{1, -\theta_1, -\theta_2, \dots, -\theta_q\}$ et ε_t est un bruit blanc de variance σ_ε^2 .

Il faut noter que ce processus est un cas spécial du processus MA(∞) avec

$$\begin{cases} \psi_0 = 1, \psi_i = -\theta_i & 1 \leq i \leq q \\ \psi_i = 0 & i < 0 \text{ ou } i > q \end{cases}$$

dans un tel modèle, on suppose de plus que l'influence des ε_t passés se manifeste au travers d'une fonction linéaire. Un tel modèle est appelé moyenne mobile d'ordre q car X_t est une moyenne mobile (en anglais *moving average*) appliquée aux chocs aléatoires $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q}$.

- **Espérance variance et stationnarité**

Il est clair que ce processus est stationnaire d'ordre deux, avec

$$E(X_t) = 0 \quad (\text{I. 3.24})$$

$$\begin{aligned} \text{Var}(X_t) &= \text{Var}(\varepsilon_t) + \theta_1^2 \text{Var}(\varepsilon_{t-1}) + \dots + \theta_q^2 \text{Var}(\varepsilon_{t-q}) \\ &= \sigma_\varepsilon^2 (1 + \theta_1^2 + \dots + \theta_q^2) \end{aligned}$$

$$\text{Var}(X_t) = \sigma_\varepsilon^2 (1 + \sum_{i=1}^q \theta_i^2) \quad (\text{I. 3.25})$$

- **Auto-covariance et auto-corrélation**

$$\gamma_h = \sigma_\varepsilon^2 \sum_{-\infty}^{\infty} \theta_i \theta_{i+h} \quad (\text{I. 3.26})$$

En définissant $\theta_0 = -1$ alors

$$\gamma_h = \sigma_\varepsilon^2 \left(-\theta_h + \sum_{i=1}^{q-h} \theta_i \theta_{i+h} \right) \quad (\text{I. 3.27})$$

$$\gamma_h = \sigma_\varepsilon^2 (-\theta_h + \theta_1 \theta_{h+1} + \dots + \theta_{q-h} \theta_q) \quad 1 \leq h \leq q \quad (\text{I. 3.28})$$

$$\gamma_h = \begin{cases} 0 & , \text{pour } h \geq q + 1 \\ \gamma_{-h} & , \text{pour } h < 0 \end{cases} \quad (\text{I. 3.29})$$

La fonction d'auto-corrélation de X_t peut alors être déduite, elle est donc donnée par :

$$\rho_h = \begin{cases} \frac{(-\theta_h + \sum_{i=1}^{q-h} \theta_i \theta_{i+h})}{1 + \sum_{i=1}^q \theta_i^2} & 1 \leq h \leq q \\ 0 & h > q \end{cases} \quad (\text{I. 3.30})$$

qui est nulle pour $h > q$. Cette propriété est très précieuse pour l'identification du modèle et la détermination de l'ordre q d'un processus MA(q).

– pour $q = 1$, le processus est un MA(1) sa fonction d'auto-corrélation est nulle pour $h > 1$:

$$\rho_h = \begin{cases} \frac{-\theta_1}{(1 + \theta_1^2)} & h = 1 \\ 0 & h > 1 \end{cases} \quad (\text{I. 3.31})$$

- Pour $q = 2$, le processus est un MA(2) et sa fonction d'auto-corrélation est nulle pour $h > 2$:

$$\rho_h = \begin{cases} \frac{(-\theta_1 + \theta_1\theta_2)}{(1 + \theta_1^2 + \theta_2^2)} & h = 1 \\ \frac{-\theta_2}{(1 + \theta_1^2 + \theta_2^2)} & h = 2 \\ 0 & h > 2 \end{cases} \quad (\text{I. 3.32})$$

Les auto-corrélations partielles ϕ_h d'un processus moyenne mobile d'ordre q ont un comportement semblable à celui des auto-corrélations ρ_h d'un processus autorégressif de même ordre : elle s'amortissent à vitesse exponentielle.

I.3.3.3 Processus autorégressif d'ordre fini AR(p)

Dans l'étude d'une série chronologique, il est naturel de penser que la valeur de la série à la date t peut dépendre des valeurs prises aux dates précédentes. Généralement il n'est pas nécessaire de prendre en compte tout le passé de la série et on peut le plus souvent se limiter à p valeurs :

$$X_t = f(X_{t-1}, X_{t-2}, \dots, X_{t-p})$$

De façon générale, un processus autorégressif d'ordre p écrit AR(p) est un processus qui dépend linéairement des p valeurs antérieures tel que :

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t$$

Où ε_t est un bruit blanc.

- **Définition :** Un processus stationnaire centré X_t , $t \in Z$ sera appelé processus autorégressif linéaire d'ordre p : AR(p) s'il existe un bruit blanc ε_t et des réels $\phi_i, i = 1, \dots, p$ tels qu'une relation de récurrence :

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t \quad \forall t \in Z \quad (\text{I. 3.33})$$

où ε_t est un bruit blanc de variance σ_ε^2 .

est vérifiée.

En introduisant le polynôme de retard (I. 3.33) peut être exprimée comme suit :

$$\Phi(B)X_t = \varepsilon_t \quad (\text{I. 3.34})$$

où $\Phi(B) = \sum_{i=0}^p \phi_i B^i$ est appelé « polynôme caractéristique » ou « symbole » du modèle. C'est un polynôme de degré p dont les coefficients sont $(1, -\phi_1, -\phi_2, \dots, -\phi_p)$.

- **Stationnarité**

L'ensemble des paramètres $\phi_1, \phi_2, \dots, \phi_p$ d'un AR(p) ou d'une manière équivalente les racines du polynôme $\Phi(B)$ doivent satisfaire un certain nombre de conditions pour que le processus AR(p) soit stationnaire. Pour une meilleure compréhension considérons d'abord

– le processus AR(1) :

$$\Phi(B)X_t = (1 - \phi_1 B)X_t = \varepsilon_t \quad (\text{I. 3.35})$$

on peut l'écrire

$$X_t = (1 - \phi_1 B)^{-1} \varepsilon_t = \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i} = \psi(B) \varepsilon_t \quad (\text{I. 3.36})$$

d'où

$$\psi(B) = (1 - \phi_1 B)^{-1} = \sum_{i=0}^{\infty} \phi_1^i B^i$$

$$\psi(B) = \Phi^{-1}(B) = \sum_{i=0}^{\infty} \phi_1^i B^i \quad (\text{I. 3.37})$$

pour que X_t soit fini et donc stationnaire, $\psi(B)$ doit converger. Ceci est équivalent à dire que le polynôme caractéristique $\Phi(B) = 1 - \phi_1 B$ (ou de manière équivalente l'équation caractéristique ($\Phi(B) = 0$)) doit avoir toutes ses racines à l'extérieur du cercle unité $|B| = 1$.

pour AR(1) ceci implique

$$\Phi(B) = 0 \Rightarrow 1 - \phi_1 B = 0 \Rightarrow \tilde{B} = \frac{1}{\phi_1}$$

où la racine $|\tilde{B}| = \frac{1}{|\phi_1|} > 1$ d'où $|\phi_1| < 1$.

la solution stationnaire peut alors s'écrire :

$$X_t = \Phi(B)^{-1} \varepsilon_t$$

$$X_t = \frac{1}{1 - \phi_1 B} \varepsilon_t = \psi(B) \varepsilon_t \quad (\text{I. 3.38})$$

Si $\psi(B) = \Phi^{-1}(B)$ doit être convergent alors on doit avoir $|\phi_1| < 1$.

– Pour un AR(p) :

la condition de stationnarité est : si le polynôme caractéristique $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ a toutes ses racines à l'extérieur du cercle unité alors l'équation (I. 3.33) possède une et une seule solution.

Le polynôme d'ordre p peut s'écrire :

$$\Phi(B) = (1 - G_1 B)(1 - G_2 B) \dots (1 - G_p B)$$

et a pour racines :

$$\tilde{B}_1 = \frac{1}{G_1}, \dots, \tilde{B}_p = \frac{1}{G_p}$$

La condition de stationnarité peut alors s'écrire :

$$\tilde{B}_i = \frac{1}{G_i} > 1 \Rightarrow |G_i| < 1, i = 1, \dots, p.$$

La solution stationnaire peut s'écrire :

$$X_t = \Phi(B)^{-1} \varepsilon_t$$

$$\Phi(B)^{-1} = \prod_{i=1}^p (1 - G_i B)^{-1} = \prod_{i=1}^p \left(\sum_{k=0}^{\infty} G_i^k B^k \right) = \sum_{i=1}^p \frac{K_i}{1 - G_i B}$$

d'où,

$$X_t = \sum_{i=1}^p \frac{K_i}{1 - G_i B} \varepsilon_t$$

$$X_t = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k} = \psi(B) \varepsilon_t \quad (\text{I. 3.39})$$

avec

$$\psi_k = \sum_{i=1}^p K_i G_i^k$$

Si $\psi(B) = \Phi(B)^{-1}$ doit être convergent alors on doit avoir $|G_i| < 1, i = 1, \dots, p$.

Pour $p = 1$ et $p = 2$ les conditions de stationnarité des processus AR(1) et AR(2) sont les suivantes :

- AR(1) : $(1 - \phi_1 B)X_t = \varepsilon_t \Rightarrow |\phi_1| < 1$
- AR(2) : $(1 - \phi_1 B - \phi_2 B^2)X_t = \varepsilon_t \Rightarrow \begin{aligned} \phi_1 + \phi_2 &< 1 \\ \phi_2 - \phi_1 &< 1 \\ -1 &< \phi_2 < 1 \end{aligned}$

• Espérance

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t$$

$$E(X_t) = \sum_{i=1}^p \phi_i E(X_{t-i}) + E(\varepsilon_t)$$

$$E(X_t) = \sum_{i=1}^p \phi_i E(X_{t-i})$$

La stationnarité implique que $\sum_{i=1}^p \phi_i \neq 1$ et que l'espérance ne dépend pas du temps, elle est donc constante d'où

$$\mu = \sum_{i=1}^p \phi_i \mu$$

ce qui implique $\mu = 0$ donc

$$E(X_t) = 0 \quad (\text{I. 3.40})$$

- **Auto-covariance et auto-corrélation**

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t$$

$$X_{t+h} = \sum_{i=1}^p \phi_i X_{t+h-i} + \varepsilon_{t+h}$$

En prenant l'espérance mathématique

$$E(X_{t+h} X_t) = \sum_{i=1}^p \phi_i E(X_{t+h-i} X_t) + E(\varepsilon_{t+h} X_t)$$

Nous obtenons la covariance

$$\gamma_h = \sum_{i=1}^p \phi_i \gamma_{h-i} + E(\varepsilon_{t+h} X_t) \quad (\text{I. 3.40})$$

$$E(\varepsilon_{t+h} X_t) = \begin{cases} \sigma_\varepsilon^2 & \text{si } h = 0 \\ 0 & \text{si } h > 0 \end{cases}$$

d'où

$$\gamma_h = \sum_{i=1}^p \phi_i \gamma_{h-i} \quad \text{pour } h > 0 \quad (\text{I. 3.41})$$

en divisant par γ_0 nous obtenons l'auto-corrélation ρ_h

$$\rho_h = \sum_{i=1}^p \phi_i \rho_{h-i} \quad \text{pour } h > 0 \quad (\text{I. 3.42})$$

On appelle ces équations les « équations de Yule-Walker ». Si on connaît $\rho_0, \rho_1, \dots, \rho_{p-1}$ on peut aisément calculer ρ_h pour $h > p$. En introduisant l'opérateur de retard B les équations de Yule-Walker peuvent s'écrire :

$$\Phi(B)\rho_h = 0 \quad \text{pour } p > h > 0 \quad (\text{I. 3.43})$$

où $\rho_h B^i = \rho_{h-i}$. Pour obtenir $\rho_1, \dots, \rho_{p-1}$ lorsque $p > 1$, il suffit de résoudre le système d'équations :

$$\begin{aligned}
\rho_1 &= \phi_1 + \phi_2 \rho_1 + \dots + \phi_p \rho_{p-1} \\
\rho_2 &= \phi_1 \rho_1 + \phi_2 + \dots + \phi_p \rho_{p-2} \\
&\vdots \\
\rho_{p-1} &= \phi_1 \rho_{p-2} + \phi_2 \rho_{p-3} + \dots + \phi_p \rho_1
\end{aligned}$$

Où l'on se sert de l'identité $\rho_i = \rho_{-i}$. Les autres auto-corrélations sont ensuite obtenues par l'équation de récurrence :

$$\rho_h = \sum_{i=1}^p \phi_i \rho_{h-i} \quad \text{pour } h \geq p \quad (\text{I. 3.44})$$

On peut montrer que les auto-corrélations tendent vers zéro d'une manière exponentielle quand h tend vers l'infini

$$h \rightarrow \infty \quad \rho_h \rightarrow 0$$

Pour les auto-corrélations partielles, il faut rappeler que dans un tel processus, X_t et X_{t-p+1} sont indépendants conditionnellement aux valeurs intermédiaires $\{X_{t-1}, X_{t-2}, \dots, X_{t-p}\}$ et donc

$$\phi_h = 0, \text{ pour } h > p \quad (\text{I. 3.45})$$

la valeur à la date t dépend des p dates précédentes et pas des autres. Cette propriété sert à l'identification des modèles et à déterminer l'ordre p d'un processus $AR(p)$ au vu du corrélogramme partiel.

les coefficients ϕ_k avec $1 \leq k \leq p$ du modèle peuvent être estimés à partir de l'estimation des ρ_h en utilisant le système de Yule et Walker :

$$\rho_h = \sum_{i=1}^p \phi_i \rho_{h-i} \quad , h = 1, 2, 3 \dots$$

en tenant compte du fait que $\rho_0 = 1$ et que $\rho_h = \rho_{-h}$, on trouve pour un processus $AR(p)$:

$$\begin{bmatrix} 1 & \rho_1 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \dots & \rho_{p-2} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \rho_{p-1} & \rho_{p-2} & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \cdot \\ \cdot \\ \cdot \\ \phi_p \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \cdot \\ \cdot \\ \cdot \\ \rho_p \end{bmatrix} \quad (\text{I. 3.46})$$

Ces coefficients ϕ_k peuvent alternativement être estimés par les formules récurrentes de Durbin-Levinson ou les méthodes classiques connues telle que celle des moindres carrés ou du maximum de vraisemblance.

- **Variance**

Pour calculer la variance $Var(X_t) = \gamma_0$, on résout l'équation

$$\gamma_0 = \sum_{i=1}^p \phi_i \gamma_{-i} + E(\varepsilon_t X_t)$$

$$\gamma_0 = \sum_{i=1}^p \phi_i \gamma_i + \sigma_\varepsilon^2$$

d'où, comme $\gamma_i = \rho_i \gamma_0$

$$\gamma_0 \left(1 - \sum_{i=1}^p \phi_i \rho_i \right) = \sigma_\varepsilon^2$$

$$\gamma_0 = \frac{\sigma_\varepsilon^2}{1 - \sum_{i=1}^p \phi_i \rho_i} \quad (\text{I. 3.47})$$

- Pour un AR(1) : $X_t = \phi_1 X_{t-1} + \varepsilon_t$

$$\begin{aligned} \rho_1 &= \phi_1 \\ \rho_k &= \phi_1 \rho_{k-1} \\ \rho_2 &= \phi_1 \rho_1 = \phi_1^2 \end{aligned}$$

$$\begin{aligned} \rho_k &= \phi_1^k, \quad k \geq 1 \\ \gamma_0 &= \frac{\sigma_\varepsilon^2}{1 - \phi_1^2} \end{aligned} \quad (\text{I. 3.48})$$

- Pour un AR(2) : $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1 = \frac{\phi_1}{1 - \phi_2} \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 = \frac{\phi_1^2}{1 - \phi_2} + \phi_2 = \frac{\phi_1^2 + \phi_2 - \phi_2^2}{1 - \phi_2} \\ \rho_k &= \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, \quad k \geq 2 \\ \gamma_0 &= \frac{\sigma_\varepsilon^2}{1 - \phi_1 \rho_1 + \phi_2 \rho_2} \end{aligned} \quad (\text{I. 3.49})$$

I.3.3.4 Les processus mixtes ARMA(p,q)

Les deux modèles précédents sont simples, mais ils nécessitent parfois un grand nombre de paramètres pour obtenir un ajustement correct aux données. Aussi il est intéressant de définir une classe de processus mixtes dit processus ARMA au moyen de l'équation:

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \bar{\mu} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

Popularisés par **Box and Jenkins (1976)**, ces processus jouent un très grand rôle dans la modélisation des séries temporelles en permettant la représentation très parcimonieuse de celles-ci. Intuitivement, un AR(1) résume un MA(∞) et un MA(1) résume un AR(∞). Un ARMA(1,1) devrait donc au moyen de deux paramètres simplement représenter raisonnablement bien une large classe de processus. On généralise le modèle en considérant des polynômes d'ordre p et q pour obtenir un processus ARMA(p,q).

- **Définition :** On appelle processus ARMA(p,q) un processus stationnaire X_t , $t \in Z$ vérifiant une relation de récurrence :

$$X_t = \bar{\mu} + \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad \forall t \in Z \quad (\text{I. 3.50})$$

où les ϕ_i et θ_i sont des réels et ε_t est un bruit blanc de variance σ_ε^2 .

La notation des polynômes de retard ramène l'équation (I.3.50) à la forme :

$$\Phi(B)X_t = \bar{\mu} + \Theta(B)\varepsilon_t \quad (\text{I. 3.51})$$

• Conditions de stationnarité

Si le polynôme $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ a toutes ses racines à l'extérieur du cercle unité, l'équation (I.3.51) possède une et une seule solution stationnaire qui peut s'écrire :

$$X_t = \mu + \frac{\Theta(B)}{\Phi(B)} \varepsilon_t = \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \quad (\text{I. 3.52})$$

où

$$\mu = \frac{\bar{\mu}}{\Phi(B)} = \frac{\bar{\mu}}{1 - \sum_{i=1}^p \phi_i} \quad (\text{I. 3.53})$$

$$\frac{\Theta(B)}{\Phi(B)} \equiv \Psi(B) = \sum_{i=0}^{\infty} \psi_i B^i \quad (\text{I. 3.54})$$

les coefficients ψ_i sont obtenus en résolvant l'équation

$$\Phi(B) \Psi(B) = \Theta(B) \quad (\text{I. 3.55})$$

dans ce cas on a aussi

$$E(X_{t-i} \varepsilon_t) = 0, \forall i \geq 1$$

et le calcul des coefficients ψ_i se fait de la manière suivante, si on définit $\theta_0 = -1$:

$$\left(1 - \sum_{k=1}^p \phi_k B^k\right) \left(\sum_{i=0}^{\infty} \psi_i B^i\right) = 1 - \sum_{i=1}^q \theta_i B^i = - \sum_{i=0}^q \theta_i B^i$$

d'où,

$$\Phi(B)\psi_i = \begin{cases} -\theta_i & , i = 0, 1, \dots, q \\ 0 & , i \geq q + 1 \end{cases}$$

où $\psi_i = 0$ si $i < 0$, par conséquent,

$$\psi_i = \begin{cases} \sum_{k=1}^p \phi_k \psi_{i-k} - \theta_i & , i = 0, 1, \dots, q \\ \sum_{k=1}^p \phi_k \psi_{i-k} & , i \geq q + 1 \end{cases} \quad (\text{I. 3.56})$$

et

$$\begin{aligned}\psi_0 &= 1 \\ \psi_1 &= \phi_1 \psi_0 - \theta_1 \\ \psi_2 &= \phi_1 \psi_1 + \phi_2 \psi_0 - \theta_2 \\ &= \phi_1^2 - \phi_1 \theta_1 + \phi_2 - \theta_2 \\ &\vdots \\ &\vdots \\ &\vdots \\ \psi_i &= \sum_{k=1}^p \phi_k \psi_{i-k}, \quad i \geq q+1\end{aligned}$$

Il faut noter que les coefficients ψ_i se comportent comme les auto-corrélations d'un AR(p) sauf pour les coefficients initiaux ψ_0, \dots, ψ_q .

- **Auto-covariance et auto-corrélation**

Supposons que

- 1) Le processus X_t est stationnaire du second ordre avec $\sum_{i=1}^p \phi_i \neq 1$;
- 2) $E(X_{t-i} \varepsilon_t) = 0, \forall i \geq 1$

Par l'hypothèse de stationnarité,

$$E(X_t) = \mu, \forall t$$

d'où

$$\mu = \bar{\mu} + \sum_{i=1}^p \phi_i = \frac{\bar{\mu}}{1 - \sum_{i=1}^p \phi_i}$$

et

$$E(X_t) = \mu = \frac{\bar{\mu}}{1 - \sum_{i=1}^p \phi_i} \quad (\text{I. 3.57})$$

La moyenne est la même que dans le cas d'un AR(p) pur. La partie MA(q) n'a pas d'effet sur la moyenne. Si nous travaillons sur le processus centré \tilde{X}_t tel que :

$$\tilde{X}_t = X_t - \mu \quad (\text{I. 3.58})$$

nous pouvons par conséquent, l'écrire sous la forme :

$$\begin{aligned}\tilde{X}_t &= \sum_{i=1}^p \phi_i \tilde{X}_{t-i} + \varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad \forall t \in Z \\ \tilde{X}_{t+h} &= \sum_{i=1}^p \phi_i \tilde{X}_{t+h-i} + \varepsilon_{t+h} - \sum_{i=1}^q \theta_i \varepsilon_{t+h-i} \quad \forall t \in Z\end{aligned} \quad (\text{I. 3.59})$$

en prenant l'espérance

$$E(\tilde{X}_t \tilde{X}_{t+h}) = \sum_{i=1}^p \phi_i E(\tilde{X}_t \tilde{X}_{t+h-i}) + E(\tilde{X}_t \varepsilon_{t+h}) - \sum_{i=1}^q \theta_i E(\tilde{X}_t \varepsilon_{t+h-i})$$

ce qui donne l'auto-covariance

$$\gamma_h = \sum_{i=1}^p \phi_i \gamma_{h-i} + \gamma_{h_{X\varepsilon}} - \sum_{i=1}^q \theta_i \gamma_{h-i_{X\varepsilon}} \quad (\text{I. 3.60})$$

où

$$\gamma_{h_{X\varepsilon}} = E(\tilde{X}_t \varepsilon_{t+h}) = \begin{cases} 0 & , si h \geq 1 \\ \sigma_\varepsilon^2 & , si h = 0 \\ \neq 0 & , si h \leq 0 \end{cases} \quad (\text{I. 3.61})$$

d'où pour $h \geq q + 1$

$$\gamma_h = \sum_{i=1}^p \phi_i \gamma_{h-i} \quad , \quad h \geq q + 1 \quad (\text{I. 3.62})$$

$$\rho_h = \sum_{i=1}^p \phi_i \rho_{h-i} \quad , \quad h \geq q + 1 \quad (\text{I. 3.63})$$

La variance est donnée par

$$\gamma_0 = \sum_{i=1}^p \phi_i \gamma_i + \sigma_\varepsilon^2 - \sum_{i=1}^q \theta_i \gamma_{-i_{X\varepsilon}}$$

d'où

$$\gamma_0 = \frac{(\sigma_\varepsilon^2 - \sum_{i=1}^q \theta_i \gamma_{-i_{X\varepsilon}})}{1 - \sum_{i=1}^p \phi_i \rho_i} \quad (\text{I. 3.64})$$

En introduisant l'opérateur de retard B l'autocovariance peut être exprimée comme suit :

$$\Phi(B)\gamma_h = \Theta(B)\gamma_{h_{X\varepsilon}} \quad , \quad h \geq 0 \quad (\text{I. 3.65})$$

où

$\gamma_h = \gamma_{-h}$, $B^i \gamma_h = \gamma_{h-i}$ et $B^i \gamma_{h_{X\varepsilon}} = \gamma_{h-i_{X\varepsilon}}$. En particulier,

$$\Phi(B)\gamma_h = 0 \quad , \quad h \geq q + 1 \quad (\text{I. 3.66})$$

$$\Phi(B)\rho_h = 0 \quad , \quad h \geq q + 1 \quad (\text{I. 3.67})$$

Pour calculer la suite des auto-covariances, on résout les équations (I. 3.60) pour $h = 0, 1, \dots, p$ puis on applique l'équation (I. 3.62).

Enfin, il faut retenir que les auto-corrélations d'un ARMA(p,q) se comportent comme celles d'un AR(p), sauf que les valeurs initiales sont modifiées.

– En particulier pour un ARMA(1,1) :

Si l'on travaille sur le processus centré $\tilde{X}_t = X_t - \mu$, ce processus peut être écrit :

$$\tilde{X}_t - \phi_1 \tilde{X}_{t-1} = \varepsilon_t - \theta_1 \varepsilon_{t-1} \quad (I. 3.68)$$

où $|\phi_1| < 1$, alors les covariances

$$\begin{aligned} \gamma_0 &= \phi_1 \gamma_1 + \gamma_{0_{X\varepsilon}} - \theta_1 \gamma_{-1_{X\varepsilon}} \\ \gamma_1 &= \phi_1 \gamma_0 + \gamma_{1_{X\varepsilon}} - \theta_1 \gamma_{0_{X\varepsilon}} \end{aligned}$$

et

$$\begin{aligned} \gamma_{1_{X\varepsilon}} &= 0 \\ \gamma_{0_{X\varepsilon}} &= \sigma_\varepsilon^2 \\ \gamma_{-1_{X\varepsilon}} &= E(\tilde{X}_t \varepsilon_{t-1}) \\ &= \phi_1 E(\tilde{X}_{t-1} \varepsilon_{t-1}) + E(\varepsilon_t \varepsilon_{t-1}) - \theta_1 E(\varepsilon_{t-1}^2) \\ &= \phi_1 \gamma_{0_{X\varepsilon}} - \theta_1 \sigma_\varepsilon^2 \\ &= (\phi_1 - \theta_1) \sigma_\varepsilon^2 \end{aligned}$$

Donc

$$\begin{aligned} \gamma_0 &= \phi_1 \gamma_1 + \sigma_\varepsilon^2 - \theta_1 (\phi_1 - \theta_1) \sigma_\varepsilon^2 \\ &= \phi_1 \gamma_1 + [1 - \theta_1 (\phi_1 - \theta_1)] \sigma_\varepsilon^2 \\ \gamma_1 &= \phi_1 \gamma_0 - \theta_1 \sigma_\varepsilon^2 \\ &= \phi_1 \{ \phi_1 \gamma_1 + [1 - \theta_1 (\phi_1 - \theta_1)] \sigma_\varepsilon^2 \} - \theta_1 \sigma_\varepsilon^2 \end{aligned}$$

$$\gamma_1 = \frac{(1 - \theta_1 \phi_1)(\phi_1 - \theta_1) \sigma_\varepsilon^2}{1 - \phi_1^2}$$

De même en remplaçant γ_1 par sa valeur dans γ_0 on obtient

$$\gamma_0 = \frac{(1 - 2\theta_1\phi_1 + \theta_1^2)\sigma_\varepsilon^2}{1 - \phi_1^2} \quad (I. 3.69)$$

$$\gamma_1 = \frac{(1 - \theta_1\phi_1)(\phi_1 - \theta_1)\sigma_\varepsilon^2}{1 - \phi_1^2} \quad (I. 3.70)$$

$$\gamma_h = \phi_1 \gamma_{h-1} \quad , h \geq 2 \quad (I. 3.71)$$

Les auto-corrélations partielles pour un ARMA(p,q) présentent un calcul plus complexe que celui des deux processus précédents. On peut cependant montrer que ces auto-corrélations partielles sont des fonctions amorties tendant vers zéro en valeur absolue à vitesses exponentielles. Elles ont donc le même comportement que les auto-corrélations du même processus.

- **Inversibilité des processus ARMA(p,q)**

Tout processus AR(p) stationnaire du second ordre peut se mettre sous la forme d'un MA(∞). De même, tout processus ARMA(p,q) stationnaire du second ordre peut aussi s'écrire sous la forme MA(∞). Par analogie, on peut aussi se demander si un processus MA(q) ou ARMA(p,q) peut se mettre sous la forme autorégressive.

Considérons le processus à moyenne mobile ARMA(0,1) ou MA(1) donné par l'équation :

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} \quad , t \in Z \quad (\text{I. 3.72})$$

où, ε_t est un bruit blanc de variance σ_ε^2 . On voit aisément que

$$\begin{aligned} \varepsilon_t &= X_t + \theta_1 \varepsilon_{t-1} \\ &= X_t + \theta_1 (X_{t-1} + \theta_1 \varepsilon_{t-2}) \end{aligned}$$

$$\varepsilon_t = \sum_{i=0}^n \theta_1^i X_{t-i} + \theta_1^{n+1} \varepsilon_{t-n-1} \quad (\text{I. 3.73})$$

et la moyenne quadratique

$$E \left[\left(\sum_{i=0}^n \theta_1^i X_{t-i} - \varepsilon_t \right)^2 \right] = E [(\theta_1^{n+1} \varepsilon_{t-n-1})^2] = \theta_1^{2(n+1)} \sigma_\varepsilon^2 \xrightarrow[n \rightarrow \infty]{} 0 \quad (\text{I. 3.74})$$

pourvu que $|\theta_1| < 1$. Par conséquent, la série $\sum_{i=0}^n \theta_1^i X_{t-i}$ converge en moyenne quadratique vers ε_t si $|\theta_1| < 1$. En d'autres termes, on peut écrire

$$\varepsilon_t = \sum_{i=0}^{\infty} \theta_1^i X_{t-i} \quad , t \in Z \quad (\text{I. 3.75})$$

qui est l'écriture du processus MA(1) sous forme autorégressive.

D'une manière équivalente, en introduisant l'opérateur de retard B , le processus MA(1) s'écrit

$$(1 - \theta_1 B)^{-1} X_t = \varepsilon_t \quad , t \in Z \quad (\text{I. 3.76})$$

où $(1 - \theta_1 B)^{-1} = \sum_{i=0}^{\infty} \theta_1^i B^i$. La condition $|\theta_1| < 1$ assure la convergence de la série $\sum_{i=0}^{\infty} \theta_1^i B^i$, elle est équivalente à ce que les racines de l'équation caractéristique $1 - \theta_1 B = 0$ sortent du cercle unité. Si $\theta_1 = 1$ on aura la série $X_t = \varepsilon_t - \varepsilon_{t-1}$

$$(1 - \theta_1 B)^{-1} X_t = \sum_{i=0}^{\infty} \theta_1^i X_{t-i} = \sum_{i=0}^{\infty} X_{t-i} \quad (\text{I. 3.77})$$

ne converge pas vers zéro quand $i \rightarrow \infty$. De même si $\theta_1 = -1$, la la série $X_t = \varepsilon_t + \varepsilon_{t-1}$

$$(1 - \theta_1 B)^{-1} X_t = \sum_{i=0}^{\infty} \theta_1^i X_{t-i} = \sum_{i=0}^{\infty} (-1)^i X_{t-i} \quad (\text{I. 3.78})$$

ne converge pas non plus vers zéro quand $i \rightarrow \infty$. Ce sont des processus non inversibles.

Théorème (Condition d'inversibilité d'un processus MA(q) : Soit $\{X_t, t \in Z\}$, un processus stationnaire du second ordre, tel que :

$$X_t = \mu + \boldsymbol{\theta}(B)\varepsilon_t \quad (\text{I. 3.79})$$

où $\boldsymbol{\theta}(B) = 1 - \theta_1 B - \dots - \theta_q B^q$. Alors le processus X_t satisfait une équation de la forme :

$$\sum_{i=0}^{\infty} \pi_i X_{t-i} = \bar{\mu} + \varepsilon_t \quad (\text{I. 3.80})$$

ssi les racines du polynôme $\theta(B)$ sortent du cercle unité. De plus lorsque la représentation (I. 3.80) existe, on a :

$$\pi(B) = \theta(B)^{-1} \quad (\text{I. 3.81})$$

$$\bar{\mu} = \theta(B)^{-1} \mu = \frac{\mu}{1 - \sum_{i=1}^q \theta_i} \quad (\text{I. 3.82})$$

Corollaire (Condition d'inversibilité d'un processus ARMA(p,q)) : Soit $\{X_t, t \in Z\}$, un processus stationnaire du second ordre, tel que :

$$\Phi(B)X_t = \mu + \theta(B)\varepsilon_t \quad (\text{I. 3.83})$$

où $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ et $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$. Alors le processus X_t satisfait une équation de la forme :

$$\sum_{i=0}^{\infty} \pi_i X_{t-i} = \bar{\mu} + \varepsilon_t \quad (\text{I. 3.84})$$

ssi les racines du polynôme $\theta(B)$ sortent du cercle unité. De plus lorsque la représentation (I. 3.84) existe, on a :

$$\pi(B) = \theta(B)^{-1} \Phi(B) \quad (\text{I. 3.85})$$

$$\bar{\mu} = \theta(B)^{-1} \bar{\mu} = \frac{\mu}{1 - \sum_{i=1}^q \theta_i} \quad (\text{I. 3.86})$$

Notes

A la fin de cette section, il importe de noter que dans les conditions d'inversibilité et de stationnarité, les processus ARMA(p,q) peuvent avoir deux représentations équivalentes :

1. Celle d'un MA(∞) de X_t en termes de ε_t (appelée représentation causale) obtenue par inversion formelle, suivie par un développement de la fraction correspondante dans une série des puissances :

$$X_t = \frac{\theta(B)}{\Phi(B)} \varepsilon_t = \left(\sum_{i=0}^{\infty} \psi_i B^i \right) \varepsilon_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \quad (\text{I. 3.87})$$

2. Celle d'un AR(∞), de ε_t en termes de X_t (appelée aussi représentation inverse) obtenue de la même manière :

$$\varepsilon_t = \frac{\Phi(B)}{\theta(B)} X_t = \left(\sum_{i=0}^{\infty} \pi_i B^i \right) X_t = \sum_{i=0}^{\infty} \pi_i X_{t-i} \quad (\text{I. 3.88})$$

Il faut noter également que pour ces processus, l'observation et l'analyse des corrélogrammes et corrélogrammes partiels constituent un outil privilégié dans l'identification du modèle.

I.3.3.5 Les processus mixtes intégrés ARIMA(p,d,q)

Les séries chronologiques en hydrologie ne sont pas toujours stationnaires. La condition de stationnarité des processus présentés jusque là n'est évidemment pas toujours convenable. On peut intégrer certains types de non-stationnarités en élargissant le processus ARMA. Ces élargissements consistent en une série d'opérations préalables visant à éliminer les facteurs de non-stationnarités. Le principe de ces modèles est de se ramener à un modèle ARMA(p, q) en opérant des différenciations sur la série .

- **Définition :** On appelle processus ARIMA(p,d,q) un processus X_t pour le quel le processus différencié d'ordre d , $Y_t = (1 - B)^d X_t$, $t \in Z$ vérifie une relation de récurrence ARMA(p,q) :

$$Y_t = \sum_{i=1}^p \phi_i Y_{t-i} - \sum_{i=0}^q \theta_i \varepsilon_{t-i} \quad \forall t \in Z \quad (\text{I. 3.89})$$

où les ϕ_i et θ_i sont des réels et ε_t est un bruit blanc de variance σ_ε^2 .

La notation des polynômes de retard ramène (I. 3.89) à la forme :

$$\Phi(B)Y_t = \Phi(B)(1 - B)^d X_t = \check{\Phi}(B)X_t = \Theta(B)\varepsilon_t \quad , \forall t \in Z \quad (\text{I. 3.90})$$

où $\Phi(B)$ et $\Theta(B)$ sont deux polynômes de degrés respectifs p et q et où

$$\check{\Phi}(B) = \Phi(B)(1 - B)^d \quad (\text{I. 3.91})$$

Il faut noter que ces processus ont à peu près la même définition que les processus ARMA(p,q), la seule différence étant qu'on permet au symbole autorégressif $\check{\Phi}(B)$ d'avoir la racine 1, c-à-d, qu'on relâche l'hypothèse de stationnarité et on travaille avec des séries non stationnaires.

La tendance est un premier facteur de non-stationnarité. Comme on a déjà vu à la section des opérateurs linéaires de différenciation, si un processus X_t admet une tendance polynomiale de degré d , le processus différencié d fois résultant est stationnaire :

$$Y_t = \Delta^d X_t = (I - B)^d X_t$$

Le processus ARIMA appliqué à X_t est équivalent au processus ARMA appliqué sur le processus différencié Y_t :

$$Y_t = \text{ARMA}(p,q) \quad \Longleftrightarrow \quad X_t = \text{ARIMA}(p,d,q)$$

Le « I » de ARIMA signifie « *integrated* » comme réciproque de la différenciation. Evidemment, le degré d n'est généralement pas connu. Pour le déterminer on peut agir par tâtonnements ou avoir recours à des tests de stationnarité: puisqu'un processus ARMA(p, q) est stationnaire, on cherche d tel qu'on puisse accepter l'hypothèse de stationnarité pour le processus $Y_t = \Delta^d X_t$. Dans la pratique $d = 1$, et très rarement $d = 2$.

I.3.3.6 Les processus mixtes saisonniers intégrés SARIMA(p, d, D, q)

La saisonnalité est un autre facteur de non-stationnarité (Langousis & Koutsoyiannis, 2006). Une façon simple d'éliminer une saisonnalité de période s consiste à appliquer l'opérateur Δ_s

$$Y_t = \Delta_s X_t = (I - B^s)X_t = X_t - X_{t-s}$$

- **Définition :** On appelle processus SARIMA(p, d, D, q) un processus X_t pour le quel le processus obtenu en différenciant d fois et en enlevant une saisonnalité d'ordre D : $Y_t = (\sum_{i=0}^{D-1} B^i)(1 - B)^d X_t, t \in Z$ vérifie une relation de récurrence ARMA(p, q) :

$$Y_t = \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{i=0}^q \theta_i \varepsilon_{t-i} \quad \forall t \in Z \quad (I. 3.92)$$

où les ϕ_i et θ_i sont des réels et ε_t est un bruit blanc de variance σ^2 .

La notation des polynômes de retard ramène (I. 3.92) à la forme :

$$\Phi(B)Y_t = \Phi(B) \left(\sum_{i=0}^{D-1} B^i \right) (1 - B)^d X_t = \check{\Phi}(B)X_t = \Theta(B)\varepsilon_t \quad (I. 3.93)$$

Où

$$\check{\Phi}(B) = \Phi(B) \left(\sum_{i=0}^{D-1} B^i \right) (1 - B)^d \quad (I. 3.94)$$

D'une manière similaire, et pratique on peut supposer que l'influence des chocs aléatoires se transmet entre dates distantes d'un nombre entier de périodes selon un processus ARIMA(P, D, Q):

$$\Phi_s(B^s)\Delta_s^D X_t = \Theta_s(B^s)E_t \quad (I. 3.95)$$

et que ces chocs aléatoires eux-mêmes suivent un modèle ARIMA(p, d, q)

$$\Phi(B)\Delta^d E_t = \Theta(B)\varepsilon_t \quad (I. 3.96)$$

où ε_t est un bruit blanc de variance σ_ε^2 .

Un tel processus est noté SARIMA(p, d, q) \times (P, D, Q) $_s$ et son équation générale est la suivante :

$$\Phi(B)\Phi_s(B^s)\Delta^d \Delta_s^D X_t = \Theta(B)\Theta_s(B^s)\varepsilon_t \quad (I. 3.97)$$

où Φ , Φ_s , θ et θ_s sont des polynômes de degrés respectifs p, P, q, Q.

Par exemple un modèle SARIMA(0,1,1) \times (0,1,1)₁₂ a pour objet d'éliminer l'effet annuel d'une série de données mensuelles par un modèle de la forme :

$$\Delta^{12}X_t = (I - \theta_{12}B^{12})E_t$$

puis à poser pour E_t un modèle ARIMA(p, d, q) tel que:

$$\Delta E_t = (I - \theta B)\varepsilon_t$$

On obtient ainsi

$$\Delta\Delta^{12}X_t = (I - \theta_{12}B^{12})(I - \theta B)\varepsilon_t \quad (\text{I. 3.98})$$

Soit

$$X_t - X_{t-1} - X_{t-12} + X_{t-13} = \varepsilon_t - \theta\varepsilon_{t-1} - \theta_{12}\varepsilon_{t-12} + \theta_{12}\theta\varepsilon_{t-13} \quad (\text{I. 3.99})$$

I.3.4. CONCLUSION

Dans ce chapitre nous avons d'abord introduit et explicité le lien entre série chronologiques et processus linéaires, et en faisant intervenir les différents opérateurs linéaires, nous avons également montré que beaucoup de processus stochastiques avec dépendance sont obtenus par des transformations linéaires de bruits blancs.

Ensuite, nous avons introduit le théorème de Wold qui est à la base de la modélisation ARMA, abordée dans la deuxième partie de cette thèse. A cet effet, alors que dans les sections précédentes nous avons défini un processus stochastique non paramétrique se définissant à partir de la distribution jointe des observations ou de ses premiers moments, dans ce chapitre on définit un processus stochastique paramétrique se définissant au contraire à partir d'un mécanisme générateur qui est indexé par des paramètres. Le théorème de Wold permet de caractériser ce mécanisme en montrant que tout processus stationnaire peut être représenté de manière unique par la somme de deux composantes indépendantes, une composante déterministe parfaitement prévisible et une composante stochastique.

Enfin, nous présentons les différents modèles de ces processus stochastiques linéaires qui seront préconisés pour la modélisation des séries chronologiques, dans la deuxième partie de ce travail. Les Modèles ARMA, ARIMA et SARIMA sont alors présentés avec leurs hypothèses et leurs caractéristiques.

CONCLUSION I

Cette première partie, avec ses trois chapitres, constitue un arrière plan théorique de la méthode de modélisation stochastique de Box-Jenkins et celle du filtre de Kalman qui seront abordées respectivement, dans la deuxième et troisième partie de cette thèse. Elle nous a permis d'introduire les notions et les concepts essentiels à la compréhension des mécanismes et des fondements mathématiques sur lesquels reposent les méthodes évoquées plus haut.

Dans le chapitre 1, nous avons introduit les notions mathématiques et les concepts de base qui servent à la compréhension du principe de la modélisation des processus hydrologiques naturels et présenté les instruments mathématiques qui servent à l'analyse des processus stochastiques et leur transmission dans les systèmes linéaires. L'accent a été mis sur l'importance de la notion de bruit blanc gaussien qui est complètement caractérisé par sa fonction d'auto-corrélation et qui présente l'une des hypothèses de base aussi bien pour la méthode de Box-Jenkins que celle du Filtre de Kalman. D'autres notions telles que les modèles Espace-état ou encore le filtrage linéaire ont également été introduites et soulignées.

Dans le chapitre 2, nous avons présenté le concept processus stochastique/série chronologique. Dans ce contexte et dans un premier temps, nous avons abordé les séries chronologiques et insisté sur la particularité et l'intérêt de cet ordre temporel dans lequel elles ont été observées. Nous avons montré que cet ordre n'était pas fortuit, mais qu'il était l'expression de la présence d'une relation temporelle imposée par les lois de la nature. Par conséquent, cette relation d'antériorité attribue aux séries chronologiques une sorte de mémoire qui fait appel à ces lois de la nature qu'on appelle « mécanismes générateurs sous-jacents » et qui sont de nature stochastique. L'établissement de ce lien nous a conduits, dans un deuxième temps à aborder les processus stochastiques non paramétriques et montrer comment ces derniers peuvent être caractérisés. Cela étant fait en insistant sur la notion de stationnarité et d'ergodicité qui sont à la base de cette caractérisation stochastique.

Dans le chapitre 3, nous avons accordé un intérêt particulier à une classe de processus stochastiques linéaires : les processus de Box-Jenkins, en raison de la qualité de leur représentation des processus hydrologiques en général et des apports de rivières en particulier. Dans cette optique, nous avons d'abord introduit et explicité le lien entre série chronologiques et processus linéaires, et en faisant intervenir les différents opérateurs linéaires, nous avons montré que beaucoup de processus stochastiques avec dépendance sont obtenus par des transformations linéaires de bruits blancs. Ensuite, nous avons introduit le théorème de Wold qui est à la base de la modélisation ARMA abordée dans la deuxième partie de cette thèse. A cet effet, on définit un processus stochastique linéaire se définissant non pas par la distribution jointe des observations ou de ses premiers moments, comme c'était le cas dans les chapitres précédents, mais à partir d'un mécanisme générateur indexé par des paramètres. Ce théorème permet la caractérisation du mécanisme générateur et stipule que tout processus stationnaire peut être représenté, de manière unique, par la somme de deux composantes indépendantes : l'une déterministe parfaitement prévisible et l'autre stochastique. Enfin, nous présentons les différents modèles de ces processus stochastiques linéaires, les modèles ARMA, ARIMA et SARIMA sont alors présentés avec leurs hypothèses et leurs caractéristiques respectives.

[Tapez un texte]

PARTIE II

La méthode de Box-Jenkins
pour la modélisation et prédiction simple
des apports liquides au Nord d'Algérie

PREAMBULE II

L'analyse des séries chronologiques ou temporelles touche énormément de domaines de la vie professionnelle, et plus précisément celui du domaine de l'eau où les séries temporelles constituent une branche de l'hydrologie dont l'objet est l'étude des variables hydrologiques au cours du temps.

Le but de l'analyse des séries temporelles qui fait l'objet de cette seconde partie est de s'intéresser à la « dynamique » de la variable observée dans le temps. La dimension temporelle est ici importante car il s'agit de l'analyse d'une chronique historique : des variations d'une même variable au cours du temps, afin de pouvoir comprendre la dynamique.

Parmi les multiples applications d'une telle analyse, il importe dans le contexte présent d'en citer au moins une : la prévision. En effet, la fonction première pour laquelle il est intéressant d'observer l'historique d'une variable est de pouvoir déceler certaines régularités pouvant être représentées par des modèles adéquats. Le but poursuivi est la formulation d'un modèle statistique qui soit une représentation du processus stochastique (inconnu) qui a généré la série observée. On se préoccupant de construire un modèle pour la série temporelle, on veut acquérir de l'information sur le processus théorique qui la génère, qui est de nature dynamique, et que l'on appelle Processus Générateur. Le problème est alors de trouver le modèle pratique qui approchera le plus possible le processus théorique et ensuite de l'estimer. Une fois cette étape franchie, la prévision de la variable étudiée moyennant ce modèle devient une opération automatique.

Dans cette approche, les types de modèles que l'on peut considérer sont nombreux. Dans cette partie, notre attention est particulièrement focalisée sur les modèles stochastiques linéaires de Box-Jenkins car ils se prêtent bien à la représentation des apports de rivières en général. Nous allons nous intéresser à modéliser une dizaine de séries uni-variées au moyen d'un modèle ARMA et utiliser le modèle retenu pour la prévision. La modélisation statistique usuelle suppose que les séries sont stationnaires. Cependant la plupart des séries à traiter sont non stationnaires, c'est à dire par exemple qu'elles croissent dans le temps en affichant une tendance linéaire traduisant une augmentation dans le niveau de la moyenne, ou présentent des variations saisonnières qui traduisent une instabilité au niveau de la variance. Dans ce cas, il est en général toujours possible de trouver une transformation ou un filtre qui puisse les rendre stationnaires, les modèles ARMA deviennent alors ARIMA où le « I » vient de l'anglais (Auto-Regressive Integrated Moving Average) ou SARIMA qui est une ARIMA saisonnière.

La méthode préconisée est présentée dans le premier chapitre tandis que son application à la variable hydrologique des apports liquides est présentée dans le deuxième chapitre, pour une périodicité annuelle, et dans le troisième chapitre, pour une périodicité mensuelle.

Chapitre II.1

LA METHODE DE BOX – JENKINS ET LA MODELISATION STOCHASTIQUE LINEAIRE DES SERIES CHRONOLOGIQUES

II.1.1 INTRODUCTION

Box et Jenkins (1970) ont proposé une procédure unifiée de traitement des processus stochastiques stationnaires et ergodiques qui sert toujours de référence. La méthodologie qu'ils ont promue consiste à modéliser les séries temporelles uni-variées au moyen des processus ARMA (**Anderson, 1976**). Ces processus sont parcimonieux et constituent une bonne approximation de processus plus généraux pourvu que l'on se restreigne au cadre linéaire (**Parzen, 1974 ; Pena et al., 2001 ; Salas, 1993**). Cette méthode de Box-Jenkins vise à formuler un modèle permettant de représenter une série chronologique. Son idée principale est le concept de parcimonie ou de la minimisation du nombre de paramètres (**Akaike, 1974**). En pratique, ces derniers étant inconnus, ils sont remplacés par leur valeur estimée sachant que plus il y a de paramètres plus nombreuses sont les chances de se tromper. Dans cette approche une hypothèse et quatre étapes sont considérées.

- **Hypothèse**

On suppose que la série chronologique X_t est générée à partir d'une autre série ε_t qui suit une loi de Gauss de moyenne nulle et d'écart-type σ_ε , sous la forme:

$$X_t = \mu + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \dots \quad (\text{II. 1.1})$$

avec ε_t qui vérifie les conditions suivantes :

$$\left\{ \begin{array}{l} \varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2), \\ E(\varepsilon_t) = 0, \forall t, \\ \gamma_h = \text{Cov}(\varepsilon_t, \varepsilon_{t+h}) = \begin{cases} \sigma_\varepsilon^2, & \text{si } h = 0 \\ 0 & \text{si } h \neq 0 \end{cases} \end{array} \right. \quad (\text{II. 1.2})$$

La série ε_t est donc un bruit blanc qui est l'exemple parfait d'une série stationnaire puisque son espérance est toujours nulle, sa dispersion est stable et ses observations sont indépendantes. La valeur μ est la moyenne de la série X_t . L'analyse des séries sous l'aspect Box-Jenkins consiste justement à chercher la série ε_t et les coefficients ψ_i qui permettent de passer de la série ε_t à X_t .

- **Etapas**

L'étape 1 cherche à obtenir la stationnarité des données en interprétant le graphique des auto-covariances, sachant qu'un processus non-stationnaire a ses auto-covariances qui décroissent très lentement à l'inverse d'un processus stationnaire. L'étape 2 repose également sur l'interprétation de graphiques pour choisir l'ordre des parties AR et MA. L'étape 3 consiste à estimer les paramètres du modèle choisi et l'étape 4 est une analyse de diagnostic consistant à vérifier si le modèle est bien spécifié au moyen de tests statistiques. Dans le cas où le modèle n'est pas valable, il faut reprendre la procédure à partir de l'étape 2.

II.1.2 STATIONARISATION ET DESAISONALISATION

II.1.2.1 Transformation et différenciation

Les modèles ARMA sont assez restrictifs dans la mesure où ils supposent que les observations sont stationnaires, que les paramètres sont constants et que les erreurs sont de variance constante (**Fuller, 1976 ; Hamilton, 1994**). Supposer que les observations sont stationnaires signifie que la moyenne de la série est constante dans le temps, ainsi que la variance. Si l'on se trouve devant une non-stationnarité en termes de moyenne, la meilleure méthode pour se ramener à la stationnarité est la différenciation, c'est-à-dire de remplacer la série originale par la série des différences adjacentes, sachant qu'une série temporelle qui a besoin d'être différenciée pour atteindre la stationnarité est considérée comme une version intégrée d'une série stationnaire (d'où le terme *Integrated*). D'autre part, la correction d'une non-stationnarité en termes de variance peut être réalisée par des transformations (**Droesbeke et al., 1989 ; Gouriéroux et Monfort, 1990**).

La transformation logarithmique par exemple permet de stabiliser la variance d'une série si la variance de la série originale croît avec les valeurs de celle-ci, si à l'inverse, la variance de la série originale décroît avec le temps, une transformation exponentielle doit être envisagée. Une série non stationnaire est une série qui en général croît avec le temps, on peut donc être tenté de retirer par régression une tendance temporelle de cette série. Il faut noter que ces transformations doivent être réalisées avant la différenciation.

L'autre solution consiste à utiliser un filtre du type $(1 - B)$. On va alors différencier la série. En revenant sur l'exemple précédent, si l'on a tout d'abord pris le logarithme, la série transformée

$$(1 - B)\log X_t = \log X_t - \log X_{t-1} \quad (\text{II. 1.3})$$

représentera les différences logarithmiques de la série. En posant $Y_t = \log X_t$,

- une différenciation d'ordre 1 suppose que la différence entre deux valeurs successives de Y_t est constante.

$$Y_t - Y_{t-1} = \mu + \varepsilon_t \quad (\text{II. 1.4})$$

où μ est la constante du modèle, et représente la différence moyenne en Y_t . Un tel modèle est un ARIMA(0,1,0). Il peut être représenté comme un accroissement linéaire en fonction du temps. Si μ est égal à 0, la série est stationnaire.

- Une différence d'ordre 2 signifie que l'on ne travaille pas sur les différences brutes, mais sur les différences de différence. Ainsi, la seconde différence de Y_t au moment t est égale à

$(Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})$, c'est-à-dire à $Y_t - 2Y_{t-1} + Y_{t-2}$. Un tel modèle est un ARIMA(0,2,0). Il obéira à l'équation de prédiction suivante :

$$Y_t - 2Y_{t-1} + Y_{t-2} = \mu + \varepsilon_t \quad (\text{II. 1.5})$$

Ou encore à

$$Y_t = \mu + 2Y_{t-1} - Y_{t-2} + \varepsilon_t \quad (\text{II. 1.6})$$

De même, si une série présente une forte composante saisonnière, on peut éliminer celle-ci en différenciant la série non pas à l'ordre un, mais à l'ordre de la saisonnalité. Ainsi pour des séries mensuelles, on pourra utiliser $(1 - B^{12})$ pour éliminer l'effet annuel.

Par exemple, pour une série de données mensuelles présentant une tendance apparemment linéaire, on combine les deux filtres Δ et Δ_{12} pour éliminer la tendance et une saisonnalité de période 12. Cela revient à travailler sur la série

$$Y_t = \Delta \Delta_{12} X_t = \Delta(X_t - X_{t-12}) = X_t - X_{t-1} - X_{t-12} + X_{t-13}$$

$$Y_t = X_t - X_{t-1} - X_{t-12} + X_{t-13} \quad (\text{II. 1.7})$$

Dans ce cas, si la série X_t comprend N valeurs, la série transformée Y_t n'en contient plus que $N - 13$, c'est à dire que l'analyse du modèle ARMA(p, q) se fait sur une série raccourcie de plus d'un an. Il faut toutefois être prudent dans l'utilisation de ces transformations, car elles peuvent avoir des conséquences sur le processus étudié et introduire artificiellement des caractéristiques qui ne s'y trouvaient pas.

II.1.2.2 Détermination de l'ordre de différenciation

Une série stationnaire fluctue autour d'une valeur moyenne et sa fonction d'auto-corrélation décline rapidement vers zéro. Si une série présente des auto-corrélations positives pour un grand nombre de décalages (par exemple 10 ou plus), alors elle nécessite d'être différenciée (**King, 1987**). La différenciation tend à introduire des auto-corrélations négatives. Si l'auto-corrélation de décalage 1 est égale à 0 ou négative, la série n'a pas besoin d'être différenciée. Si l'auto-corrélation de décalage 1 est inférieure à -0.5 , la série est sur-différenciée. L'ordre optimal de différenciation est souvent celui pour lequel l'écart-type est minimal. Un accroissement de l'écart-type doit donc être considéré comme un symptôme de sur-différenciation. Un troisième symptôme de sur-différenciation est un changement systématique de signe d'une observation à une autre.

II.1.3 DETERMINATION DU TYPE ET DE L'ORDRE DU MODELE

Après que la série ait été stationnarisée (la saisonnalité et la tendance ont été supprimées) on s'intéresse à l'identification d'un processus stationnaire ARMA(p,q) et on tente de reconnaître le type de processus auquel on a affaire et d'identifier les termes AR et MA nécessaires pour corriger

les auto-corrélations résiduelles. Cette analyse est basée sur l'examen des fonctions d'auto-corrélation et d'auto-corrélation partielle (**Kottegoda, 1980 ; Thomas & Fiering, 1962**). Rappelons que l'auto-corrélation est la corrélation d'une série avec elle-même, selon un décalage défini (appelé aussi « pas »). L'auto-corrélation de décalage 0 est par définition égale à 1. La fonction d'auto-corrélation (ACF) fait correspondre à chaque décalage h l'autocorrélation correspondante ρ_h , de même la fonction d'autocorrélation partielle (PACF) fait correspondre à chaque décalage l'auto-corrélation partielle ϕ_h correspondante (**Hipel & McLeod, 1994 ; Kendall & Ord, 1990**).

A ce stade, il est essentiel d'établir une hypothèse maximale du nombre de coefficients à utiliser, c-à-d. des valeurs p_{max} et q_{max} à partir desquelles travailler sur le modèle ARMA(p,q).

Pour ce faire, on utilise les fonctions ACF (ρ_h) et PACF (ϕ_h) et on les observe visuellement, sachant que leurs propriétés sont les suivantes :

MA(q) : $\rho_h = 0$ si $h > q$
 AR(p) : ρ_h tend graduellement vers 0

Et

AR(p) : $\phi_h = 0$ si $h > p$
 MA(q) : ϕ_h tend graduellement vers 0

II.1.3.1 Paramètres AR

Ainsi pour un AR(p) pur, on observe un seuil de PACF pour $h \leq p$, ϕ_h est non nul et il devient nul pour $h > p$.

Une extinction brutale de l'auto-corrélation partielle associée à un déclin plus progressif de l'auto-corrélation constitue donc la signature caractéristique d'un processus autorégressif. Plus particulièrement, l'auto-corrélation partielle de décalage p est égale au coefficient AR(p) estimé dans un modèle contenant p termes AR.

Le décalage auquel l'auto-corrélation partielle disparaît indique le nombre de termes autorégressifs à inclure.

Généralement, ce schéma est associé à une auto-corrélation de décalage 1 positive, signe que la série demeure sous-différenciée. Une légère sous-différenciation peut donc être compensée par l'ajout d'un terme autorégressif.

II.1.3.2 Paramètres MA

La fonction d'auto-corrélation ACF joue, pour les processus de moyenne mobile, le même rôle que la fonction d'auto-corrélation partielle pour les processus autorégressifs. Si l'auto-corrélation est significative au décalage q mais plus au décalage $q + 1$, ceci indique que q termes de moyenne mobile doivent être ajoutés au modèle.

On voit bien que pour un MA(q) pur, le comportement est le même que AR(p) pur, mais cette fois-ci en utilisant ACF et q comme valeur de coupure.

En général, une signature MA est associée à une auto-corrélation négative au décalage 1, signe que la série est sur-différenciée. Une légère sur-différenciation peut donc être compensée par l'ajout d'un terme de moyenne mobile.

II.1.3.3 Paramètres ARIMA

Pour les ARMA, il faut malheureusement établir un diagnostic en observant séparément les parties ACF et PACF.

II.1.3.4 Paramètres SARIMA(p,d,q)(P,D,Q)_s

L'ARIMA saisonnière multiplicative est une généralisation et une extension de la méthode ARIMA, pour des séries où le même phénomène se répète au cours du temps, à chaque saison. Outre les paramètres non saisonniers, des paramètres saisonniers pour un décalage spécifié (repéré lors de la stationnarisation) doivent être estimés.

Comme pour les paramètres de l'ARIMA simple, ce sont : les paramètres autorégressifs saisonniers (P), de différenciation saisonnière (D), et de moyenne mobile saisonnière (Q). Par exemple, le modèle (0,1,2)(0,1,1)_s décrit un modèle sans paramètre autorégressif, avec 2 paramètres de moyenne mobile normaux et 1 paramètre de moyenne mobile saisonnier. Le calcul de ces paramètres intervient après différenciation une fois de 1 période, et une fois saisonnièrement. Le décalage saisonnier utilisé pour les paramètres saisonniers est souvent déterminé lors de la phase d'identification et doit être explicitement spécifié.

Les recommandations générales concernant la sélection des paramètres à estimer (basés sur les ACF et PACF) s'appliquent également aux modèles saisonniers. La principale différence est que dans les séries saisonnières, les ACF et PACF auront des coefficients assez importants pour de nombreuses périodes saisonnières (en plus de leur composante globale reflétant la composante non saisonnière de la série).

En résumé, nous pouvons dire que les outils principaux utilisés lors de la phase d'identification sont donc les tracés de la série, les corrélogrammes d'auto-corrélation (ACF), et d'auto-corrélation partielle (PACF). La décision n'est pas simple et les cas les plus atypiques requièrent, outre l'expérience, de nombreuses expérimentations avec des modèles différents (avec divers paramètres ARIMA) (**Brillinger, 1975 ; Brillinger, 1976 ; WMO, 1994**). Toutefois, les composantes des séries chronologiques empiriques peuvent souvent être assez bien approchées en utilisant l'un des 5 modèles de base suivants, identifiables par la forme de la ACF et de la PACF. Puisque le nombre de paramètres (à estimer) de chaque type dépasse rarement 2, il est souvent judicieux d'essayer des modèles alternatifs sur les mêmes données.

1. Un paramètre autorégressif (P) : ACF - décomposition exponentielle; PACF-pic à la période 1, pas de corrélation pour les autres périodes.
2. Deux paramètres autorégressifs (P): ACF - une composante de forme sinusoïdale ou un ensemble de décompositions exponentielles; PACF - pics aux périodes 1 et 2, aucune corrélation pour les autres périodes.
3. Un paramètre de moyenne mobile (q): ACF - pic à la période 1, aucune corrélation pour les autres périodes; PACF - exponentielle amortie.
4. Deux paramètres de moyenne mobile (q): ACF - pics aux périodes 1 et 2, aucune corrélation pour les autres périodes; PACF - une composante de forme sinusoïdale ou un ensemble de décompositions exponentielles.

5. Un paramètre autorégressif (P) et un de moyenne mobile (q): ACF - décomposition exponentielle commençant à la période 1 ; PACF - décomposition exponentielle commençant à la période 1.

Le tableau comparatif suivant résume l'information que nous avons accumulée jusque là, sur le comportement des auto-corrélogrammes totaux et partiels des processus exposés.

Tableau II.1.1 comportement des corrélogrammes et corrélogrammes partiels des processus ARMA

Processus	Auto-corrélations ρ_h	Auto-corrélations partielles ϕ_h
AR(p)	Amortie	Nulle pour $h > p$
MA(q)	Nulle pour $h > q$	Amortie
ARMA(p,q)	Amortie	Amortie

II.1.4 ESTIMATION DES PARAMETRES DU MODELE

L'estimation des paramètres d'un modèle ARMA(p; q) lorsque les ordres p et q sont supposés connus peut se réaliser par différentes méthodes dans le domaine temporel :

- Moindres Carrés Ordinaires (modèle sans composante MA, $q = 0$). Dans ce cas, on retrouve les équations de Yule-Walker. En remplaçant les auto-corrélations théoriques par leurs estimateurs, on peut retrouver les estimateurs des MCO des paramètres du modèle par la résolution des équations de Yule Walker.
- Maximum de Vraisemblance approché (**Box & Jenkins 1970**)
- Maximum de Vraisemblance exacte (**Newbold 1974, Harvey et Philips 1979, Harvey (1981)**)

Nous allons présenter ici brièvement la démarche de l'estimation par le maximum de vraisemblance, en nous contentant de montrer comment s'écrit le programme de maximisation de la vraisemblance permettant d'estimer les paramètres d'un modèle ARMA(p, q).

II.1.4.1 Rappels : l'estimateur du maximum de vraisemblance

- **Définition 1 :** soit X une variable aléatoire à valeur dans (X, a) , de loi P_θ . On note $f(x, \theta)$ la densité de P_θ et $f(x_1, \dots, x_n, \theta)$ la densité empirique correspondante. On appelle vraisemblance du paramètre θ l'application $\theta \rightarrow R^+$ définie par

$$\forall \theta \in \Theta \rightarrow L(x_1, \dots, x_n, \theta) = \prod_{i=1}^n f(x_i, \theta) \quad (\text{II.1.8})$$

- **Définition 2 :** Soit $L(x, \theta)$ la vraisemblance au point θ , $\theta \in R$. On appelle estimateur du maximum de vraisemblance pour θ la statistique :

$$\begin{aligned} \hat{\theta} &: X^N \rightarrow R \\ (x_1, \dots, x_n) &\rightarrow \hat{\theta}(x_1, \dots, x_n) \end{aligned}$$

Telle que

$$\forall \theta \in \Theta, L(x, \hat{\theta}) \geq L(x, \theta) \quad (\text{II. 1.9})$$

Le principe de la vraisemblance revient à déterminer la valeur du paramètre θ fonction des observations (x_1, \dots, x_n) , qui assure la plus grande probabilité d'apparition de ces observations (x_1, \dots, x_n) .

- **Corollaire :** Lorsque l'on suppose que (i) l'ensemble X est indépendant de θ et que (ii) la fonction de vraisemblance $L(\cdot)$ est deux fois continument différentiable par rapport à θ , $\forall \theta \in \Theta$ alors l'estimateur du maximum de vraisemblance $\hat{\theta}$ est solution du système :

$$\left(\frac{\partial L(x, \theta)}{\partial \theta} \right)_{\theta=\hat{\theta}} = 0 \quad (\text{II. 1.10})$$

$$\left(\frac{\partial^2 L(x, \theta)}{\partial \theta^2} \right)_{\theta=\hat{\theta}} = 0 \quad (\text{II. 1.11})$$

- **Théorème :** S'il existe un estimateur efficace du paramètre θ (au sens de la borne de Cramer Rao), alors cet estimateur est identique à celui du maximum de vraisemblance $\hat{\theta}$.

II.1.4.2 Application aux modèles ARMA(p,q)

On considère un processus stationnaire X_t satisfaisant une représentation ARMA(p,q) telle que :

$$\Phi(B)X_t = \mu + \Theta(B)\varepsilon_t \quad (\text{II. 1.12})$$

avec $\mu \in R$, $\Phi(B) = \sum_{i=0}^p \phi_i B^i$, $\Theta(B) = \sum_{i=0}^q \theta_i B^i$, $\forall i \in Z$, ε_t est un bruit blanc de variance connue σ_ε^2 .

En plus de la définition standard d'un processus ARMA on fait l'hypothèse de la normalité des résidus afin de spécifier une forme fonctionnelle à la vraisemblance du modèle.

Hypothèse : On suppose que la population des résidus ε_t peut être décrite par un processus bruit blanc gaussien $N(0, \sigma_\varepsilon^2)$.

Alors la vraisemblance associée au vecteur de réalisation (x_1, \dots, x_n) est donnée par :

$$(2\pi\sigma_\varepsilon^2)^{-\frac{N}{2}} \det[(\theta_i, \phi_i)]^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2} \hat{X}[(\theta_i, \phi_i)]^{-1}X\right\} \quad (\text{II. 1.13})$$

II.1.5 VALIDATION

C'est la dernière étape dans la procédure de modélisation, il s'agit de vérifier les hypothèses des modèles ajustés et d'en sélectionner le plus parcimonieux. Le premier test concerne les paramètres estimés le second concerne les résidus des modèles. Les références principales sont (**Gourieroux & Monfort, 1990** ; **Pena et al., 2001**; **Salas, 1993**).

II.1.5.1 Tests de signification des paramètres

Sous l'hypothèse que le bruit blanc est gaussien, on peut montrer que les estimateurs $\hat{\phi}_i$, et $\hat{\theta}_j$ sont approximativement gaussiens. On peut donc effectuer des tests sur les paramètres en utilisant la loi de Student :

$$\frac{\hat{\phi}_i}{\sqrt{\text{Var}(\hat{\phi}_i)}} \rightarrow \mathcal{T} \quad (\text{II. 1.14})$$

Dans le cadre d'un modèle ARMA(p, q), on peut remarquer que le test sur un paramètre ϕ_p :

$$H_0 : \phi_p = 0 \quad \text{contre} \quad H_1 : \phi_p \neq 0$$

est équivalent au test sur les modèles

$$H_0 : \text{ARMA}(p, q) \quad \text{contre} \quad H_1 : \text{ARMA}(p-1, q)$$

puisque p est le paramètre du dernier terme de la partie autorégressive du modèle. On montre que sous l'hypothèse H_0 , \mathcal{T} suit une loi de Student à (n-1) degré de liberté. Au niveau α (en général égal à 5% et $\mathcal{T}_c = 1.96$) :

$$\text{Si } \mathcal{T} < \mathcal{T}_c, \text{ on accepte } H_0 \Rightarrow \phi_p = 0$$

$$\text{Si } \mathcal{T} > \mathcal{T}_c, \text{ on rejette } H_0 \Rightarrow \phi_p \neq 0$$

II.1.5.2 Tests de Bruit Blanc

Dans l'analyse des séries chronologiques par processus, le bruit blanc joue un rôle particulier puisque c'est un processus sans aucune structure. Quand, pour un processus X_t , on a éliminé toute tendance, toute saisonnalité et toute dépendance vis-à-vis du passé, il reste un processus ε_t complètement imprévisible sur lequel il n'y a plus grand chose à dire.

Le test préliminaire concerne justement ce processus. Quand on étudie une série X_t la première hypothèse testée est H_0 : X_t est un bruit blanc.

Si on accepte cette hypothèse, l'analyse de la série est virtuellement achevée: la série étudiée n'a aucune structure.

Si on rejette cette hypothèse et que l'on procède à la modélisation de la série, lorsque le processus est bien estimé, les résidus entre les valeurs observées et les valeurs estimées par le modèle doivent se comporter comme un bruit blanc. On notera par la suite $\hat{\varepsilon}_t$ le résidu d'estimation du modèle. Ce que l'on cherche à savoir ici, c'est si les résidus ne sont pas distribués de façon systématique dans la série (par exemple, systématiquement négatifs dans une première partie de la série puis proches de zéro dans une seconde) ou s'ils ne sont pas auto-corrélés ce qui tendrait à montrer que le modèle ARIMA est inadéquat. L'analyse des résidus de l'ARIMA constitue un test important du modèle. La procédure d'estimation postule que les résidus ne sont pas auto-corrélés et qu'ils sont distribués normalement.

• Test de nullité de la moyenne

Soit T le nombre de données disponibles (après avoir enlevé les retards correspondant aux termes AR et MA): Si le processus $\varepsilon_t, t \in Z$ est i.i.d $(0, \sigma_\varepsilon^2)$, on doit avoir :

$$\bar{\varepsilon}_t = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t \xrightarrow{T \rightarrow \infty} 0 \quad (\text{II. 1.15})$$

Par application du théorème central limite, on montre que :

$$\frac{\bar{\varepsilon}_t}{\hat{\sigma}_\varepsilon} \sqrt{T} \xrightarrow{T \rightarrow \infty} \mathcal{N}(0,1) \quad (\text{II. 1.16})$$

Dès lors, on peut tester la nullité de la moyenne des résidus en construisant l'intervalle de confiance sur $\bar{\varepsilon}_t$ au seuil standard de 95%.

$$P \left\{ \bar{\varepsilon}_t \in \left[\frac{-1,96\hat{\sigma}_\varepsilon}{\sqrt{T}}, \frac{+1,96\hat{\sigma}_\varepsilon}{\sqrt{T}}, \right] \right\} = 0.95 \quad (\text{II. 1.17})$$

• Test d'auto-corrélation

Si les résidus $\varepsilon_t, t \in Z$ obéissent à un bruit blanc, il ne doit pas exister d'auto-corrélation dans la série. On peut alors utiliser les différents tests suivants :

– Etude de la ACF et de la PACF : on doit vérifier qu'il n'existe aucune auto-corrélation ou auto-corrélation partielle significativement non nulle pour le processus étudié. Cette étude est prolongée par les tests du « porte-manteau ».

– Tests du « porte-manteau » : ils reposent sur l'idée que la ACF d'un bruit blanc ne doit pas révéler d'auto-corrélations non nulles. En pratique, on utilise deux tests :

- **Test de Box et Pierce** : On note ρ_h l'auto-corrélation d'ordre h du processus $\varepsilon_t, t \in Z$. Pour un ordre H , le test de Box et Pierce est le test de l'hypothèse $H_0: \rho_1 = \rho_2 = \dots = \rho_H = 0$ contre $H_1: \exists j \in [1, H], \text{ tel que } \rho_j \neq 0$. Pour un ARMA(p,q) la statistique de ce test est :

$$Q_{BP} = T \sum_{h=1}^H \rho_h^2 \xrightarrow{T \rightarrow \infty} \chi^2(H - p - q) \quad (\text{II. 1.18})$$

L'hypothèse H_0 est rejetée au seuil de 5% si Q_{BP} est supérieure au quantile 0.95 de la loi χ^2 correspondant.

- **Test de Ljung-Box** : Ces statistiques, définies pour un ordre H ; correspondent à l'hypothèse nulle $H_0: \rho_h = 0, \forall h \leq H$ et sont construites de la façon suivante :

$$Q_H = T(T + 2) \sum_{h=1}^H \frac{\rho_h^2}{T - h} \xrightarrow{T \rightarrow \infty} \chi^2(H - p - q) \quad (\text{II. 1.19})$$

- **Test d'homoscédasticité**

Ce test est fait pour tester l'égalité des variances. L'hypothèse d'homoscédasticité est :

$$H_0 = \sigma_1^2 = \sigma_2^2 = \dots = \sigma_m^2 \quad (\text{II. 1.20})$$

et se conduit de la façon suivante:

- pour chaque sous-groupe i de la série des résidus, calculer la variance σ_i^2 telle que :

$$\sigma_i^2 = \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 / \nu_i \quad (\text{II. 1.21})$$

$i = 1, 2, \dots, m ; j = 1, 2, \dots, n_i ; n_i$ étant le nombre de valeurs dans le sous groupe $i ; \nu_i = n_i - 1$ et $\bar{y}_i = \sum_{j=1}^{n_i} y_{ij} / n_i$.

- calculer la variance

$$\sigma^2 = \sum_{i=1}^m \nu_i \sigma_i^2 / \nu \quad (\text{II. 1.22})$$

avec $\nu = \sum_{i=1}^m \nu_i$ et la quantité \hat{Q} donnée par :

$$\hat{Q} = \nu \ln \sigma^2 - \sum_{i=1}^m \nu_i \ln \sigma_i^2 \quad (\text{II. 1.23})$$

Sous l'hypothèse H_0 , \hat{Q} est approximativement distribuée selon la loi du Chi-deux à $(m - 1)$ degrés de liberté. Toute fois, l'approximation se voit améliorée si l'on divise \hat{Q} par la constante C telle que:

$$C = 1 + \left(\frac{1}{3(m-1)} \right) \left(\sum_i \frac{1}{v_i} - \frac{1}{v} \right) \quad (\text{II. 1.24})$$

pour donner Q tel que :

$$Q = \frac{\hat{Q}}{C} \quad (\text{II. 1.25})$$

Si $Q < \chi_{0.95}^2(m-1)$, l'hypothèse nulle H_0 d'homoscédasticité est acceptée à 95% de confiance.

- **Tests de normalité**

Pour vérifier si le processus des résidus ε_t est gaussien, plusieurs tests peuvent être utilisés mais le test le plus simple est celui qui consiste à un ajustement à la loi Normale puis appliquer le test de qualité d'adéquation de cet ajustement par le test du Chi-deux. Il ya aussi le test d'Anderson et le test de Jarque et Bera.

- **Test de Jarque et Bera** : Ce dernier est fondé sur la notion de skewness (moment d'ordre 3 et asymétrie) et de Kurtosis (moment d'ordre 4 et queue de distribution). Soit μ_k le moment empirique d'ordre k du processus ε_t , $t \in Z$:

$$\mu_k = \frac{1}{T} \sum_{i=1}^T (\hat{\varepsilon}_t - \bar{\varepsilon}_t)^k \quad (\text{II. 1.26})$$

Les coefficients de la Skewness (Sk) et de la Kurtosis (Ku) sont alors définis par

$$(Sk)^{1/2} = \frac{\mu_3}{\mu_2^{3/2}} \xrightarrow{T \rightarrow \infty} \mathcal{N}(0, \sqrt{6/T}) \quad (\text{II. 1.27})$$

$$Ku = \frac{\mu_4}{\mu_2^2} \xrightarrow{T \rightarrow \infty} \mathcal{N}(3, \sqrt{24/T}) \quad (\text{II. 1.28})$$

On construit alors les statistiques centrées réduites correspondantes à $(Sk)^{1/2}$ et Ku que l'on compare aux seuils d'une loi normale centrée réduite.

$$\frac{(Sk)^{1/2}}{\sqrt{6/T}} \xrightarrow{T \rightarrow \infty} \mathcal{N}(0, 1) \quad (\text{II. 1.29})$$

$$\frac{Ku - 3}{\sqrt{24/T}} \xrightarrow{T \rightarrow \infty} \mathcal{N}(0, 1) \quad (\text{II. 1.30})$$

Si la statistique centrée réduite de $(Sk)^{1/2}$ est inférieure au seuil 1.96 à 5%, on accepte l'hypothèse de symétrie et l'hypothèse de normalité. Si la statistique centrée réduite de Ku est inférieure au seuil 1.96 à 5%, on accepte l'hypothèse de queue de distributions plates et l'hypothèse de normalité.

Le test de Jarque et Bera regroupe ces deux tests en un seul test. On construit la statistique s , si $s \geq \chi_{1-\alpha}^2(2)$ on rejette l'hypothèse H_0 de normalité au seuil $\alpha\%$.

$$s = \frac{T}{6} Sk + \frac{T}{24} (Ku - 3)^2 \xrightarrow{T \rightarrow \infty} \chi^2(2) \quad (\text{II. 1.31})$$

- **Test d'adéquation du Chi-deux** : On commence par diviser la série des résidus en intervalles d'égale fréquence, de façon que chaque intervalle contienne en moyenne le même nombre de résidus. Le calcul du nombre d'intervalles se fait selon l'équation :

$$K = 1 + 3.3 \log N \quad (\text{II. 1.32})$$

Où N est la taille de la série des résidus et K est le nombre de classes (ou d'intervalles). Dans ce cas le nombre de résidus dans chacun des K intervalles sera γ tel que :

$$\gamma = N/K \quad (\text{II. 1.33})$$

Ceci n'est pas le cas en général, car le nombre de résidus est différent pour chaque intervalle. Si n_i est le nombre de résidus dans chaque intervalle i ($i: 1 \dots k$) alors :

$$\sum n_i = \sum \gamma = N \quad (\text{II. 1.34})$$

La variable aléatoire Chi-deux (χ^2) est définie par :

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - \gamma)^2}{\gamma} \quad (\text{Chi - deux calculé}) \quad (\text{II. 1.35})$$

Cette variable aléatoire est tabulée et ses valeurs théoriques sont données pour une probabilité au dépassement q et un nombre de degrés de liberté μ par :

$$\chi_{(q, \mu)}^2 \rightarrow \text{Table} \quad (\text{Chi - deux théorique}) \quad (\text{II. 1.36})$$

Avec :

$$\mu = K - p - 1 \quad (\text{II. 1.37})$$

Où p est le nombre de paramètres définissant la fonction de densité de probabilité, qui est égal à 2 dans ce cas car il s'agit de la loi Normale.

Si pour une probabilité au non dépassement g ($= 1 - q$)

$$\chi_{(g)}^2 \text{ calculé} < \chi_{(g)}^2 \text{ théorique} \quad (\text{II. 1.38})$$

Alors l'ajustement par la loi Normale est accepté à une probabilité g (généralement égale à 0.95) au non dépassement.

II.1.5.3 Critères de comparaison de modèles

Ce test vérifie l'ordre du modèle ajusté, relativement à d'autres modèles éventuels. Il est également utilisé dans la sélection finale du modèle et se base sur le critère de parcimonie.

Le meilleur des modèles ARMA(p, q) est le modèle qui minimise la statistique :

- Le critère de Akaike ou AIC :

$$AIC(p, q) = T \log(\sigma_{\hat{\varepsilon}}^2) + 2(p + q) \quad (\text{II. 1.39})$$

- Le critère d'information bayésien BIC :

$$BIC(p, q) = T \log(\sigma_{\hat{\varepsilon}}^2) - (n - p - q) \log \left[1 - \frac{(p + q)}{T} \right] + (p + q) \log T \\ + \log \left[(p + q)^{-1} \left(\frac{\sigma_X^2}{\sigma_{\hat{\varepsilon}}^2 - 1} \right) \right] \quad (\text{II. 1.40})$$

- Le critère de Schwarz SC:

$$SC(p, q) = T \log(\sigma_{\hat{\varepsilon}}^2) + (p + q) \log T \quad (\text{II. 1.41})$$

II.1.6 PREDICTION

II.1.6.1 Transformation de la série

Lorsque pour identifier le processus étudié à un processus ARMA, on a appliqué différentes transformations (exemple différenciation), il est nécessaire lors de la phase de prédiction de prendre en compte la transformation retenue et de « recolorer la prévision ». Plusieurs cas sont possibles :

- Si le processus contient une tendance déterministe, on extrait cette dernière par régression afin d'obtenir une série stationnaire lors de la phase d'estimation. Ensuite, lors de la phase de prédiction, on adjoint aux prédictions réalisées sur la composante ARMA stationnaire, la projection de la tendance.
- Si la transformation résulte de l'application d'un filtre linéaire (de type par exemple différences premières), on réalise les prédictions sur la série filtrée stationnaire et l'on reconstruit ensuite par inversion du filtre les prédictions sur la série initiale.

II.1.6.2 Prédicteur pour un processus ARMA

On considère un processus ARMA(p,q) tel que vu précédemment (**Lachtermacher & Fuller, 1994 ; Pankratz, 1983**) :

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

Où les paramètres $\phi_i, \theta_j, i = 1 \dots p; j = 1 \dots q$ sont des réels et ε_t est un bruit blanc de variance σ_ε^2 . Appliquons le théorème de Wold au processus $X_t, t \in Z$ et considérons la forme MA(∞) correspondante :

$$X_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}, \quad \psi_0 = 1 \quad (\text{II. 1.42})$$

Il s'ensuit que la meilleure prédiction que l'on peut faire de X_{t+1} compte tenu de toute l'information disponible jusqu'à la date t , notée $\hat{X}_t(1)$ est donnée par :

$$\begin{aligned} \hat{X}_t(1) &= E(X_{t+1} | X_t, X_{t-1}, X_{t-2}, \dots, X_0) \\ &= E(X_{t+1} | \varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_0) \\ \hat{X}_t(1) &= \sum_{i=1}^{\infty} \psi_i \varepsilon_{t+1-i} \end{aligned} \quad (\text{II. 1.43})$$

Alors l'erreur de prédiction est donnée par la réalisation en $t + 1$ de l'innovation qui en t n'est pas connue :

$$X_{t+1} - \hat{X}_t(1) = \varepsilon_{t+1} \quad (\text{II. 1.44})$$

Plus généralement pour une prédiction à un horizon k on a :

$$\hat{X}_t(k) = \sum_{i=k}^{\infty} \psi_i \varepsilon_{t+k-i} \quad (\text{II. 1.45})$$

d'où

$$X_{t+k} - \hat{X}_t(k) = \sum_{i=0}^{k-1} \psi_i \varepsilon_{t+k-i} \quad (\text{II. 1.46})$$

Déterminons un intervalle de confiance sur la prédiction sous l'hypothèse de normalité des résidus ε_t . On montre alors que :

$$\frac{X_{t+k} - \hat{X}_t(k)}{\text{Var}[X_{t+k} - \hat{X}_t(k)]^{1/2}} \xrightarrow{T \rightarrow \infty} \mathcal{N}(0, 1)$$

Or on sait que :

$$E \left\{ [X_{t+k} - \hat{X}_t(k)]^2 \right\} = E \left[\left(\sum_{i=0}^{k-1} \psi_i \varepsilon_{t+k-i} \right)^2 \right] = \sum_{i=0}^{k-1} \psi_i^2 \sigma_\varepsilon^2$$

d'où

$$\frac{X_{t+k} - \hat{X}_t(k)}{\sigma_\varepsilon [\sum_{i=0}^{k-1} \psi_i^2]^{1/2}} \xrightarrow{T \rightarrow \infty} \mathcal{N}(0, 1) \quad (\text{II. 1.47})$$

On peut donc construire un intervalle de confiance sous la forme :

$$IC = \left[\hat{X}_t(k) \mp t^{\alpha/2} \left(\sum_{i=0}^{k-1} \psi_i^2 \right)^{1/2} \sigma_\varepsilon \right] \quad (\text{II. 1.48})$$

Avec $\alpha = 0.05$.

II.1.7 CONCLUSION

Dans ce chapitre nous avons présenté la méthodologie de Box-Jenkins pour la modélisation stochastique des séries chronologiques en hydrologie sachant que la procédure est valable pour tout autre domaine.

Dans ce contexte, nous avons mis la lumière sur les étapes principales de cette méthode à savoir la stationarisation si les données ne sont pas stationnaires, l'identification des paramètres du modèle ajusté en se basant essentiellement sur la structure de dépendance interne des données, l'estimation de ces paramètres et enfin, la validation du modèle. Nous avons également présenté les outils mathématiques indispensables à l'accomplissement de chacune de ces étapes et indiqué le caractère itératif de la procédure en insistant particulièrement sur la stationnarité des séries concernées. Cette stationnarité étant en arrière plan de cette méthodologie, nous avons à cet effet rappelé les quelques artifices mathématiques visant son obtention. Ces artifices qui ont déjà été introduits dans la première partie de cette thèse consistent généralement en des transformations mais cette fois ci elles sont abordées avec des mises en gardes particulières pour éviter les problèmes qui peuvent en résulter.

Enfin, nous avons exposé comment le modèle ARMA résultant peut être mis en œuvre pour faire de la prédiction tout en indiquant les intervalles de confiance de ces dernières.

Chapitre II.2

MODELISATION ARIMA ET PREDICTION DES APPORTS ANNUELS

II.2.1 INTRODUCTION

Dans ce chapitre la méthode de Box-Jenkins (ARIMA) est appliquée aux apports annuels d'un échantillon de 10 stations hydrométriques au Nord d'Algérie. A cet effet, il convient de rappeler que l'usage de cette dernière exige la normalité et la stationnarité des séries chronologiques à étudier. Cette dernière condition signifie que la moyenne de la série et sa variabilité doivent être constantes. En d'autres termes, si la série en question n'est pas stationnaire il faut la transformer pour remplir cette condition. Il convient également à ce stade de rappeler que la méthode de Box-Jenkins se base essentiellement sur la structure de dépendance interne des séries chronologiques. Il va de soit que pour les séries indépendantes le recours à la méthode se trouve complètement non justifié.

II.2.2 MODELISATION ET DISCUSSION DES RESULTATS

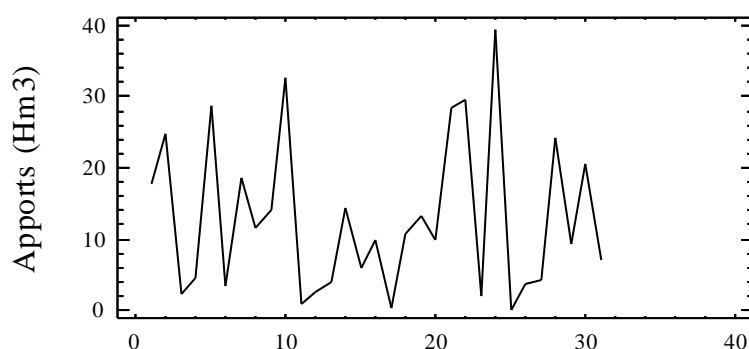
La procédure de la méthode est itérative ce qui fait que le traitement reste identique pour les 10 stations concernées. Cependant, avant de passer au traitement propre des données, un bref aperçu statistique de la série considérée est fourni avec une représentation graphique de sa variabilité. Ceci est fait dans le but de résumer l'information contenue dans ces données.

II.2.2.1 AIN BERDA

Le Tableau II.2.1 suivant résume les statistiques de la série chronologique des apports annuels observés à AIN BERDA entre 1963 et 1993. Il présente les mesures de la tendance centrale, de la variabilité et de la forme de la série chronologique en question. La skewness (coefficient d'asymétrie) et la kurtosis (coefficient d'aplatissement) standardisées sont comprises dans l'intervalle $[-2, +2]$ ce qui indique que la distribution est très probablement normale (Cox, 1981).

Tableau II.2.1 Statistiques des apports annuels observés à AIN BERDA (1963-1993)

Paramètre	valeur
Nombre d'observations	31
Moyenne	12.8748
Variance	119.08
Déviatoin Standard	10.9124
Minimum	0.119
Maximum	39.2679
Std. skewness	1,82264
Std. kurtosis	-0.416708

**Fig. II.2.1** Apports annuels originaux observés à AIN BERDA (1963-1993)

II.2.2.1.1 Stationnarisation

Le tracé de la série chronologique originale tel que montré par la Fig. II.2.1 ne semble présenter aucun signe de non stationnarité. Pour le confirmer, nous avons procédé par le calcul des premières et deuxièmes différences régulières de la série où le calcul de l'écart-type permet de choisir l'ordre optimal de la différenciation.

Le détail du calcul des séries différenciées pour les 5 premières années est montré dans le Tableau II.2.2, tandis que les statistiques des séries différenciées pour l'ensemble des données sont représentées dans le Tableau II.2.3.

Tableau II.2.2 Calcul des séries différenciées (Données des 5 premières années)

N°	Apport	DIFF(Apport,1)	DIFF(Apports,2)
1	17,85		
2	24,8	6,95	
3	2,26	-22,54	-29,49
4	4,4	2,14	24,68
5	28,74	24,34	22,2

Tableau II.2.3 Statistiques descriptives des séries différenciées (1924-1992)

Série	N	Moyenne	Variance	Ecart-type
Apports	31	12,87	119,08	10,91
DIFF(Apports,1)	30	-0,35	304	17,43
DIFF(Apports ,2)	29	-0,7	1007,6	31,74

Notation : DIFF (x,n) : différences régulières de x d'ordre n .

Le minimum d'écart-type est observé pour la série des apports d'origine ce qui confirme la thèse de sa stationnarité et rejette le recours à la différenciation. La série correspondante tel qu'illustré dans la Fig. II.2.1 est donc stationnaire et sa variance est de 119,08 est stable autour de sa moyenne qui est de 12,87. La Fig. II.2.2 quant à elle montre l'ajustement des données à une distribution normale dont la qualité d'adéquation est confirmée par le test du Chi-deux (statistique du test = 14,96, P = 0,2432) (Brockwell & Davis, 1990).

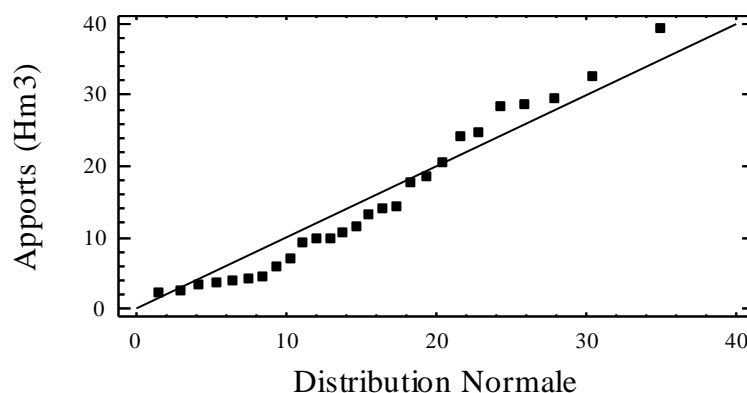


Fig. II.2.2 Ajustement à la loi Normale des apports annuels observés à AIN BERDA (1963-1993)

II.2.2.1.2. Identification

L'examen visuel du tracé de l'ACF et de celui de la PACF des apports considérés, tel que montré par les Figures II.2.3 et 4 ainsi que les Tableaux II.2.4 et 5, indique 1 pic significatif au lag 2 pour l'ACF et 1 pic significatif au lag 2 pour la PACF.

Les résultats du test de Box-Pierce basé sur les 10 premières auto-corrélations montrent une valeur de P supérieure ou égale à 0,1 ce qui veut dire que l'hypothèse nulle d'indépendance ne peut pas être rejetée à 90% de confiance ou plus (la statistique calculée du test est de 4,74 avec une valeur de probabilité P de 0,90 qui est supérieure à 0,1). La série ne présentant aucune structure de dépendance interne, peut donc être considérée comme complètement aléatoire, son traitement s'arrête ici.

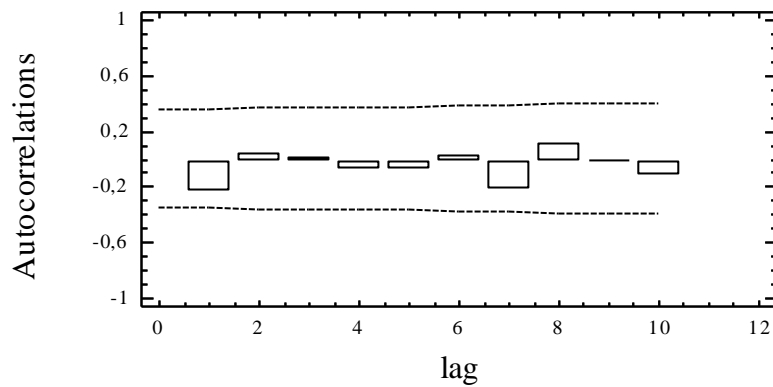


Fig. II.2.3 ACF des apports annuels observés à AIN BERDA (1963-1993)

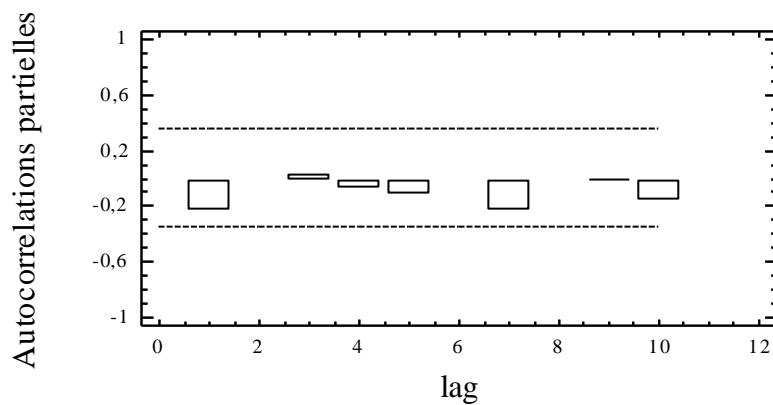


Fig. II.2.4 PACF des apports annuels observés à AIN BERDA (1963-1993)

Tableau II.2.4 ACF des apports annuels observés à AIN BERDA (1963-1993)

Lag	ACF	Ecart-type	-IC à 95%	IC à 95%
1	-0,242828	0,179605	-0,352021	0,352021
2	0,0387402	0,189901	-0,372199	0,372199
3	0,0133398	0,190155	-0,372699	0,372699
4	-0,0861318	0,190186	-0,372758	0,372758
5	-0,0741913	0,19144	-0,375216	0,375216
6	0,025745	0,192365	-0,377029	0,377029
7	-0,227489	0,192476	-0,377247	0,377247
8	0,108747	0,200962	-0,39388	0,39388
9	-0,00917072	0,202852	-0,397583	0,397583
10	-0,123073	0,202865	-0,397609	0,397609

Tableau II.2.5 PACF des apports annuels observés à AIN BERDA (1963-1993)

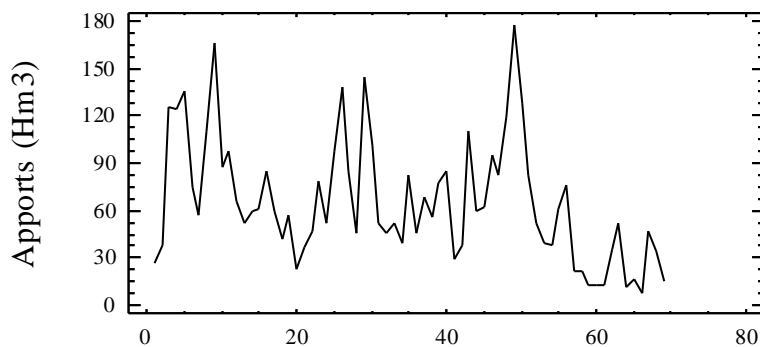
Lag	ACF	Ecart-type	-IC à 95%	IC à 95%
1	-0,242828	0,179605	-0,352021	0,352021
2	-0,0214925	0,179605	-0,352021	0,352021
3	0,0188499	0,179605	-0,352021	0,352021
4	-0,0825294	0,179605	-0,352021	0,352021
5	-0,123277	0,179605	-0,352021	0,352021
6	-0,0221423	0,179605	-0,352021	0,352021
7	-0,241052	0,179605	-0,352021	0,352021
8	-0,0180999	0,179605	-0,352021	0,352021
9	-0,00601551	0,179605	-0,352021	0,352021
10	-0,16077	0,179605	-0,352021	0,352021

II.2.2.2 BENI BAHDEL

Le Tableau II.2.6 suivant résume les statistiques de la série chronologique des apports annuels observés à BENI BAHDEL entre 1924 et 1992. Il présente les mesures de la tendance centrale, de la variabilité et de la forme de la série chronologique en question. La skewness et la kurtosis standardisées sont d'un intérêt particulier car ils sont utilisés pour déterminer si oui ou non la distribution est gaussienne. Une valeur de ces paramètres en dehors de l'intervalle $[-2, +2]$ indique un éloignement significatif de la normalité. Dans notre cas, la kurtosis standard est comprise dans l'intervalle tandis que la skewness ne l'est pas.

Tableau II.2.6 Statistiques des apports annuels observés à BENI BAHDEL (1924-1992)

Paramètre	valeur
Nombre d'observations	69
Moyenne	65,6599
Variance	1502,69
Déviatoin Standard	38,7645
Minimum	8,266
Maximum	177,69
Std. skewness	2,77406
Std. kurtosis	0,553211

**Fig. II.2.5** Apports annuels originaux observés à BENI BAHDEL (1924-1992)

II.2.2.2.1 Stationnarisation

Le tracé de la série chronologique originale tel que montré par la Fig. II.2.5 présente une tendance vers sa fin ce qui indique l'existence d'une non stationnarité. Le calcul des premières ou deuxièmes différences régulières de la série permet d'obtenir cette stationnarité et le calcul de l'écart-type permet de choisir l'ordre optimal de la différenciation.

Le détail du calcul des séries différenciées pour les 5 premières années est montré dans le Tableau II.2.7, tandis que les statistiques des séries différenciées pour l'ensemble des données sont représentées dans le Tableau II.2.8.

Tableau II.2.7 Calcul des séries différenciées
(données des 5 premières années)

N°	Apport	DIFF(Apport,1)	DIFF(Apports,2)
1	26,878		
2	38,631	11,753	
3	125,258	86,627	74,874
4	123,925	-1,333	-87,96
5	135,425	11,5	12,833

Tableau II.2.8 Statistiques descriptives des séries différenciées
(1924-1992)

Série	N	Moyenne	Variance	Ecart-type
Apports	69	65,66	1502,69	38,76
DIFF(Apports,1)	68	-0,17	1304,74	36,76
DIFF(Apports ,2)	67	-0,46	3099	55,66

Notation : DIFF (x,n) : différences régulières de x d'ordre n .

Dans le cas de la série étudiée le minimum d'écart-type est obtenu par les premières différences régulières (d=1). La série correspondante tel qu'illustré dans la Fig. II.2.6 est plus stable car sa variance est de 1304,74 au lieu de 1502,69 pour les données de départ, et la tendance est éliminée de la série. Le résultat est une série stationnaire avec une variance à expliquer qui est de 1304,74. La Fig. II.2.7 quant à elle montre l'ajustement des données différenciées à une distribution normale dont la qualité d'adéquation est confirmée par le test du Chi-deux (statistique du test = 18,4706, P = 0,42508).

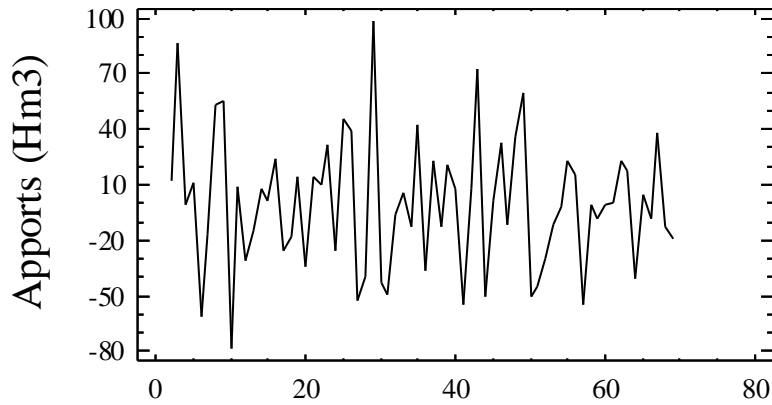


Fig. II.2.6 Différences premières des apports annuels observés à BENI BAHDEL (1924-1992)

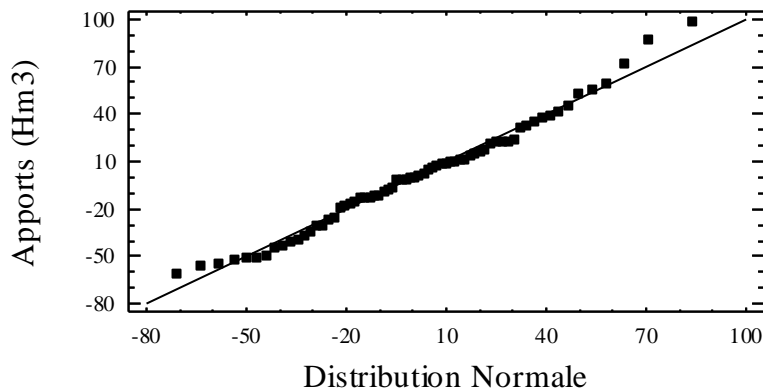


Fig. II.2.7 Ajustement Normal des différences premières des apports annuels observés à BENI BAHDEL (1924-1992)

II.2.2.2.2. Identification

L'examen visuel du tracé de la fonction d'auto-corrélation (ACF) et de celui de la fonction d'auto-corrélation partielle (PACF) des apports différenciés, tel que montré par les Fig. II.2.8 et 9 ainsi que les Tableaux II.2.9 et 10, indique 1 pic significatif au lag 2 pour l'ACF et 1 pic significatif au lag 2 pour la PACF.

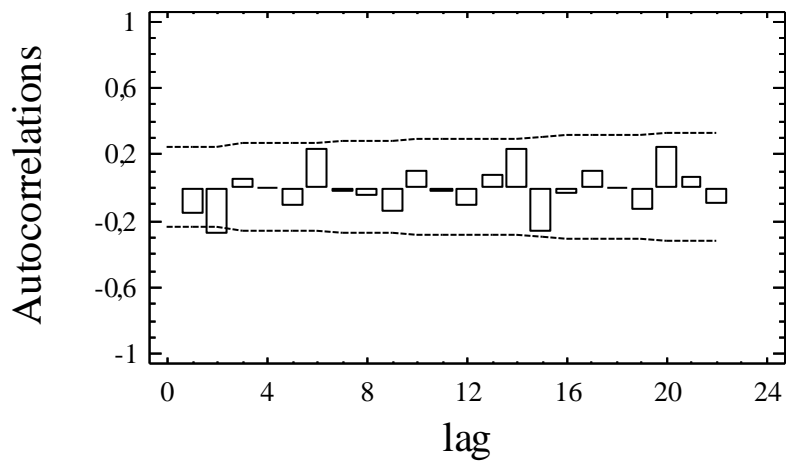


Fig. II.2.8 ACF des apports annuels différenciés observés à BENI BAHDEL (1924-1992)

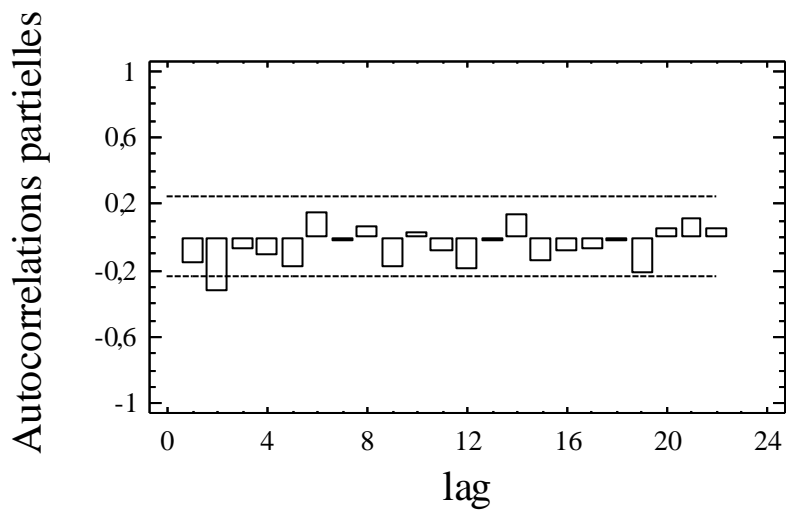


Fig. II.2.9 PACF des apports annuels différenciés observés à BENI BAHDEL (1924-1992)

Tableau II.2.9 ACF des apports annuels stationnaires

Lag	ACF	Ecart-type	-IC à 95%	IC à 95%
1	-0,172898	0,121268	-0,237681	0,237681
2	-0,290805	0,12484	-0,244683	0,244683
3	0,0498469	0,134434	-0,263486	0,263486
4	-0,00134746	0,134705	-0,264018	0,264018
5	-0,121948	0,134705	-0,264018	0,264018
6	0,223973	0,136319	-0,267181	0,267181
7	-0,0242908	0,141627	-0,277585	0,277585
8	-0,0559708	0,141689	-0,277705	0,277705
9	-0,152282	0,142013	-0,278342	0,278342
10	0,0957206	0,144395	-0,283009	0,283009
11	-0,0248371	0,145325	-0,284832	0,284832
12	-0,117777	0,145387	-0,284955	0,284955
13	0,071692	0,146784	-0,287691	0,287691
14	0,230201	0,147298	-0,288699	0,288699
15	-0,277616	0,152497	-0,298889	0,298889
16	-0,0456503	0,159756	-0,313117	0,313117
17	0,0924641	0,159948	-0,313493	0,313493
18	0,00582414	0,160732	-0,31503	0,31503
19	-0,147837	0,160735	-0,315036	0,315036
20	0,24248	0,162722	-0,318931	0,318931
21	0,0607415	0,167952	-0,329181	0,329181
22	-0,107881	0,168275	-0,329813	0,329813

Tableau II.2.10 PACF des apports annuels stationnaires

Lag	PACF	Ecart-type	-IC à 95%	IC à 95%
1	-0,172898	0,121268	-0,237681	0,237681
2	-0,330581	0,121268	-0,237681	0,237681
3	-0,0858835	0,121268	-0,237681	0,237681
4	-0,122645	0,121268	-0,237681	0,237681
5	-0,187145	0,121268	-0,237681	0,237681
6	0,140674	0,121268	-0,237681	0,237681
7	-0,0314734	0,121268	-0,237681	0,237681
8	0,0583464	0,121268	-0,237681	0,237681
9	-0,197826	0,121268	-0,237681	0,237681
10	0,0208234	0,121268	-0,237681	0,237681
11	-0,0973229	0,121268	-0,237681	0,237681
12	-0,209222	0,121268	-0,237681	0,237681
13	-0,0419165	0,121268	-0,237681	0,237681
14	0,137604	0,121268	-0,237681	0,237681
15	-0,15364	0,121268	-0,237681	0,237681
16	-0,0906374	0,121268	-0,237681	0,237681
17	-0,0843104	0,121268	-0,237681	0,237681
18	-0,0184088	0,121268	-0,237681	0,237681
19	-0,229417	0,121268	-0,237681	0,237681
20	0,0457429	0,121268	-0,237681	0,237681
21	0,11034	0,121268	-0,237681	0,237681
22	0,04904	0,121268	-0,237681	0,237681

Les résultats du test de Box-Pierce basé sur les 22 premières auto-corrélations montrent qu'à 90% de confiance la série n'est pas complètement aléatoire (la statistique calculée du test est de 32,259 avec une valeur de probabilité P de 0,07 qui est inférieure à 0,1).

Les comportements de l'ACF et la PACF nous poussent à opter pour les modèles possibles suivants :

(A) ARIMA(2,1,2) avec constante

- (B) ARIMA(2,1,1) avec constante
- (C) ARIMA(2,1,0) avec constante
- (D) ARIMA(1,1,2) avec constante
- (E) ARIMA(0,1,2) avec constante

II.2.2.2.3 Estimation

La présente procédure va estimer les 12 valeurs futures de la série de 69 données. Un modèle ARIMA a été sélectionné. Ce modèle suppose que la meilleure prédiction des valeurs futures est donnée par un modèle paramétrique qui relie les valeurs de données les plus récentes aux valeurs de données précédentes et au bruit précédent. Il convient de rappeler qu'avant que le modèle ne soit ajusté, les différences régulières d'ordre 1 ont été prises.

Pour l'estimation des paramètres de ces modèles, il faut rappeler que 69 données ont été utilisées dont les 20 dernières ont servi à la validation. Les paramètres des modèles ont alors été estimés sur une période de 39 ans.

L'analyse et la comparaison des modèles ajustés tel que présenté dans le Tableau II.2.11 montre que tous les modèles ajustés produisent des résidus complètement indépendants avec une moyenne et une variance qui sont stables entre la période d'estimation et celle de validation. Cependant, le modèle qui produit le minimum de variance des résidus est le modèle (E) c'est lui qui est donc retenu pour la prédiction.

Tableau II.2.11 Comparaison des modèles ajustés.

Modèle	Variance des résidus		AUTO	MOY	VAR
	Période d'estimation	Période de validation			
(A)	1287,39	798,677	OK	OK	OK
(B)	1267,62	822,48	OK	OK	OK
(C)	1284,2	852,257	OK	OK	OK
(D)	1226,28	884,77	OK	OK	OK
(E)	1170,67	809,399	OK	OK	OK

Notation : AUTO :test d'indépendance des résidus ; MOY : test d'égalité de la moyenne ; VAR : test d'homoscédasticité ; OK : le test est positif.

Les Tableaux II.2.12 et 13 résument l'estimation des paramètres du modèle retenu ARIMA(0,1,2) avec et sans constante ainsi que la signification statistique de ses termes. Les termes correspondant à une valeur de $P < 0.05$ sont statistiquement significativement différents de zéro à 95% de confiance.

Tableau II.2.12 Estimation des paramètres du modèle retenu ARIMA(0,1,2) avec constante

Paramètre	Estimation	Ecart- type	t	P
MA(1)	0,459571	0,133983	3,43007	0,001053
MA(2)	0,462918	0,130167	3,55633	0,000708
Moyenne	0,199681	0,687542	0,290427	0,772414
Constante	0,199681			

La variance estimée du bruit blanc = 1238,96 avec 65 DDL et l'écart- type estimé du bruit blanc = 35,1989.

Tableau II.2.13 Estimation des paramètres du modèle retenu ARIMA(0,1,2) sans constante

Paramètre	Estimation	Ecart- type	t	P
MA(1)	0,466639	0,138751	3,36314	0,001286
MA(2)	0,472472	0,135975	3,47468	0,000909

La variance estimée du bruit blanc = 1220,21 avec 66 DDL et l'écart type estimé du bruit blanc = 34,9315.

La valeur de P pour les termes MA(1) et MA(2) est inférieure à 0.05 donc ces derniers sont significativement différents de zéro à 95% de confiance avec des valeurs différentes de un (1). La valeur de P pour la moyenne (constante) est très supérieure à 0.05 donc elle est statistiquement non significativement différente de zéro et peut donc être éliminée. L'estimation de l'écart-type du bruit blanc est de 34,9315 et l'estimation de la variance des résidus est de 1220,21.

Le Tableau II.2.14 résume la performance du modèle ARIMA(0,1,2) à ajuster les données considérées et ce par l'analyse des erreurs qu'il produit.

Tableau II.2.14 Statistiques des erreurs du modèle ARIMA(0,1,2)

Statistique	Période d'estimation	Période de validation
EQM	1170,67	809,399
EAM	27,4819	23,7702
EM	3,58874	-21,5266

Nous y trouvons :

- 1) L'erreur quadratique moyenne (EQM)
- 2) L'erreur absolue moyenne (EAM)
- 3) L'erreur moyenne (EM)

Chacune de ces statistiques est basée sur l'erreur de prédiction une année à l'avance qui est la différence entre la valeur de la donnée au temps t et sa prédiction au temps $t - 1$. Les deux premières mesurent la magnitude des erreurs alors que la dernière mesure le biais. Ces statistiques d'erreurs sont calculées pour la période d'estimation et celle de validation. Pour le modèle considéré ces statistiques d'erreurs ne sont pas différentes d'une manière considérable pour les deux

périodes d'estimation et de validation, il s'ensuit que le modèle ajusté peut être considéré comme performant et donc bon à faire des prévisions des valeurs futures.

II.2.2.1.4 Validation

Il s'agit de vérifier l'hypothèse de la méthode à savoir que les résidus forment bien un bruit blanc gaussien de moyenne nulle. Pour cela, le test des auto-corrélogrammes (ACF et PACF) et celui du Portmanteau (Box-Pierce) sont utilisés pour tester l'indépendance des résidus, le test t de student est utilisé pour tester la nullité de leur moyenne et le test d'adéquation du Chi-deux à 95% de confiance est utilisé pour tester leur normalité.

- **Test des ACF et PACF**

L'auto-corrélogramme total et partiel des résidus du modèle ont été tracés tel que montré dans la Fig. II.2.10 et Fig. II.2.11.

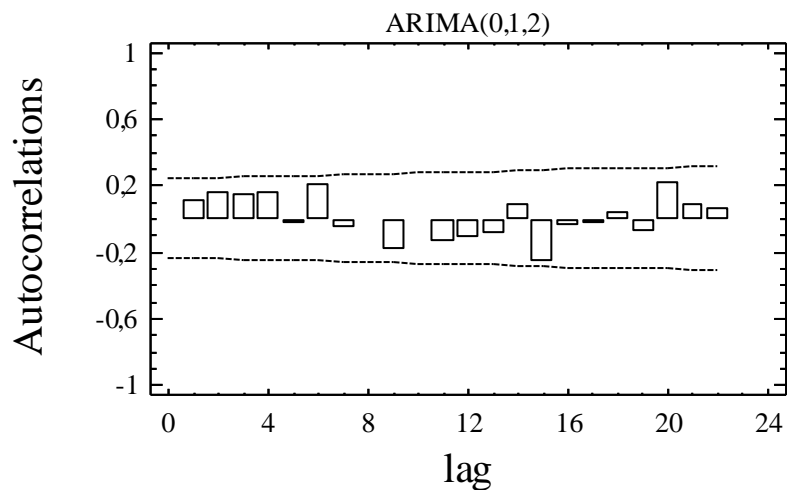


Fig. II.2.10 ACF des résidus du modèle ARIMA(0,1,2)

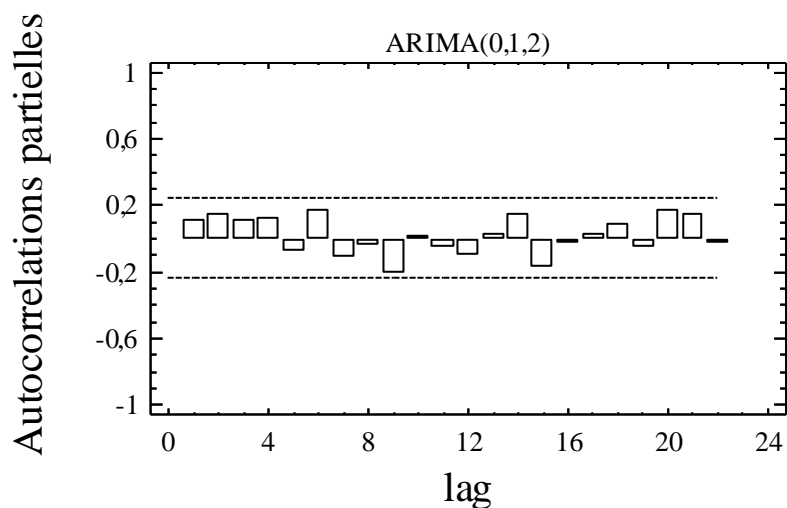


Fig. II.2.11 PACF des résidus du modèle ARIMA(0,1,2)

Tous les coefficients sont inclus dans l'intervalle de confiance (IC) tracé à 95% de confiance ce qui indique que ces derniers sont statistiquement non significatifs, d'où leur indépendance. Leurs valeurs sont reportées dans le Tableau II.2.15 et Tableau II.2.16.

Tableau II.2.15 ACF des résidus du modèle ARIMA(0,1,2)

Lag	Auto-corrélation	Ecart-type	limite inférieure IC 95%	limite supérieure IC 95%
1	0,112425	0,121268	-0,237681	0,237681
2	0,160056	0,122791	-0,240666	0,240666
3	0,141545	0,125822	-0,246606	0,246606
4	0,154885	0,128142	-0,251154	0,251154
5	-0,0243581	0,130866	-0,256493	0,256493
6	0,20269	0,130933	-0,256624	0,256624
7	-0,0579734	0,135468	-0,265514	0,265514
8	-0,00781779	0,135833	-0,266228	0,266228
9	-0,194832	0,135839	-0,266241	0,266241
10	-0,00689348	0,139889	-0,274177	0,274177
11	-0,138411	0,139894	-0,274187	0,274187
12	-0,121359	0,141893	-0,278106	0,278106
13	-0,0927206	0,143411	-0,281082	0,281082
14	0,081839	0,14429	-0,282804	0,282804
15	-0,266396	0,144971	-0,284139	0,284139
16	-0,0502468	0,152	-0,297915	0,297915
17	-0,036957	0,152244	-0,298393	0,298393
18	0,0360405	0,152376	-0,298652	0,298652
19	-0,0871261	0,152501	-0,298897	0,298897
20	0,220584	0,153231	-0,300328	0,300328
21	0,087133	0,157832	-0,309346	0,309346
22	0,0578123	0,158538	-0,310729	0,310729

Tableau II.2.16 PACF des résidus du modèle ARIMA(0,1,2)

Lag	Auto-corrélation partielle	Ecart-type	limite inférieure IC 95%	limite supérieure IC 95%
1	0,112425	0,121268	-0,237681	0,237681
2	0,149304	0,121268	-0,237681	0,237681
3	0,11338	0,121268	-0,237681	0,237681
4	0,11421	0,121268	-0,237681	0,237681
5	-0,0866041	0,121268	-0,237681	0,237681
6	0,169694	0,121268	-0,237681	0,237681
7	-0,11619	0,121268	-0,237681	0,237681
8	-0,0440805	0,121268	-0,237681	0,237681
9	-0,220218	0,121268	-0,237681	0,237681
10	0,00980342	0,121268	-0,237681	0,237681
11	-0,0557539	0,121268	-0,237681	0,237681
12	-0,103855	0,121268	-0,237681	0,237681
13	0,0283118	0,121268	-0,237681	0,237681
14	0,1387	0,121268	-0,237681	0,237681
15	-0,180125	0,121268	-0,237681	0,237681
16	-0,0402466	0,121268	-0,237681	0,237681
17	0,0217007	0,121268	-0,237681	0,237681
18	0,0821783	0,121268	-0,237681	0,237681
19	-0,0573294	0,121268	-0,237681	0,237681
20	0,167391	0,121268	-0,237681	0,237681
21	0,143541	0,121268	-0,237681	0,237681
22	-0,0281806	0,121268	-0,237681	0,237681

- **Test du Portmanteau**

Ce test est basé sur les 22 premières auto-corrélations des résidus, les résultats du test montrent une valeur de P supérieure ou égale à 0.10, donc on ne peut pas rejeter l'hypothèse nulle que les résidus sont indépendants avec un intervalle de confiance de 90% ou plus.

Test de Box-Pierce

Hypothèse nulle : résidus indépendants

Valeur calculée de la statistique $Q = 24,3376$

Valeur de P = 0.2279

- **Test de nullité de la moyenne**

La moyenne des résidus est de -3,79, leur médiane est de -11,59 et leur variance est de 1030,72 avec un écart-type de 32,10. Les résultats du test t de Student montrent une valeur $P = 0,33$ qui est supérieure à 0.05 ce qui implique que l'hypothèse nulle de nullité de la moyenne des résidus ne peut pas être rejetée à 95% de confiance.

Test t

Hypothèse nulle : moyenne = 0

Valeur calculée de la statistique $t = -0,97$

Valeur de P = 0.33

- **Test de Normalité**

Afin de vérifier si les résidus du modèle ARIMA(0,1,2) appartiennent à une distribution Normale, un ajustement à la loi de Gauss a été effectué tel que présenté dans la Fig. II.2.12. Les résultats du test du Chi-deux montrent une valeur de P supérieure à 0,01 donc l'hypothèse nulle de Normalité ne peut pas être rejetée à 95% de confiance.

test du Chi-deux

Valeur de la statistique du test = 33,91.

Valeur de P = 0,0129

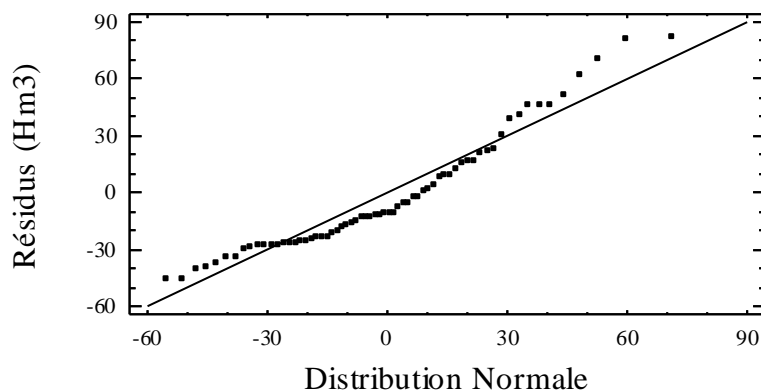


Fig. II.2.12 Ajustement des résidus du modèle ARIMA(0,1,2) à La loi de Gauss

Les résultats des tests précédents appliqués aux résidus du modèle ARIMA(0,1,2) ont permis la vérification des hypothèses de la méthode de Box-Jenkins.

En fin, ce modèle peut être formulé de la manière suivante :

$$(1 - B)X_t = (1 - 0,466639 B - 0,472472B^2)\varepsilon_t$$

Où ε_t est un bruit blanc gaussien de moyenne nulle et de variance 1220,21. Ainsi la variance expliquée par le modèle peut être estimée à seulement 6,5%. Ce pourcentage qui exprime la contribution de la partie structurée au sein de la série étudiée reste insignifiant devant la contribution de la partie aléatoire. Ce résultat concorde avec celui du test d'indépendance de Box-Pierce déjà effectué. La prépondérance de la variabilité aléatoire de la série considérée fait des prédictions procurées par le modèle dénudées de tous sens pratique même si sur le plan mathématique elles sont tout à fait réalisables.

II.2.2.2.5 Prédiction

Les résultats des prédictions obtenues par le modèle retenu ARIMA(0,1,2) sont regroupés dans le Tableau II.2.17 pour la période suivant la fin de la série chronologique de départ et dans le Tableau II.2.18 pour la période où les données sont disponibles avec les résidus correspondants. Pour cette dernière période des intervalles de confiances à 95% ont été calculés pour les prédictions obtenues. Ces limites de confiance indiquent où va être située la vraie valeur à une période déterminée dans le future, dans l'hypothèse que le modèle ajusté est adéquat aux données en question. Ces prédictions représentées dans la Fig. II.2.13 en trait continu suivent l'allure générale des données observées représentées en points carrés. Pour la période où les données ne sont pas disponibles (après l'indice 69 sur l'axe des x) les prédictions obtenues pour les 12 ans futurs sont représentées avec leurs limites de confiance tracées à 95%.

Tableau II.2.17 Prédictions des apports annuels à BENI BAHDEL pour les 12 ans futurs

Période	Prédiction	- IC 95%	+IC 95%
70,0	39,9893	-29,7539	109,732
71,0	52,7737	-26,2695	131,817
72,0	52,7737	-26,3835	131,931
73,0	52,7737	-26,4973	132,045
74,0	52,7737	-26,611	132,158
75,0	52,7737	-26,7245	132,272
76,0	52,7737	-26,8378	132,385
77,0	52,7737	-26,951	132,498
78,0	52,7737	-27,064	132,611
79,0	52,7737	-27,1769	132,724
80,0	52,7737	-27,2896	132,837
81,0	52,7737	-27,4021	132,949

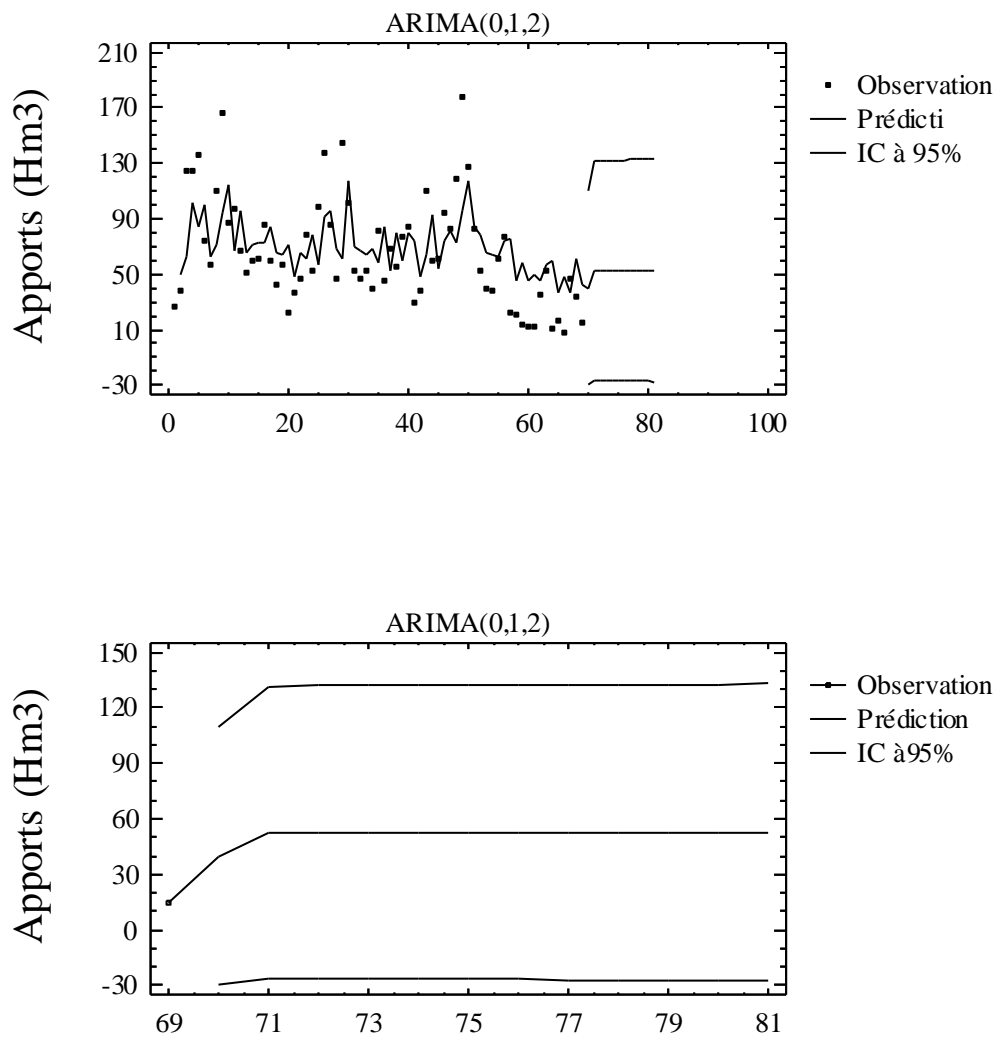


Fig. II.2.13 Prédictions des apports annuels à BENI BAHDEL obtenues par le modèle ARIMA(0,1,2)

Tableau II.2.18 Prédictions avec ARIMA(0,1,2)

Période	Donnée	Prédiction	Résidu
1,0	26,878		
2,0	38,631	49,6806	-11,0496
3,0	125,258	62,6064	62,6516
4,0	123,925	101,243	22,6821
5,0	135,425	83,7395	51,6855
6,0	74,449	100,59	-26,1409
7,0	57,263	62,2274	-4,96445
8,0	110,844	71,9304	38,9136
9,0	166,39	95,031	71,359
10,0	87,607	114,706	-27,0986
11,0	97,046	66,5371	30,5089
12,0	66,543	95,6127	-29,0697
13,0	51,64	65,6935	-14,0535
14,0	59,77	71,9325	-12,1625
15,0	61,5	72,0854	-10,5854
16,0	85,11	72,186	12,924
17,0	59,55	84,0804	-24,5304
18,0	41,82	64,8906	-23,0706
19,0	56,531	64,1756	-7,64457
20,0	22,526	70,9985	-48,4725
21,0	36,363	48,757	-12,394
22,0	46,876	65,0484	-18,1724
23,0	78,218	61,2117	17,0063
24,0	52,114	78,8682	-26,7542
25,0	98,026	56,5636	41,4624
26,0	137,62	91,3186	46,3014
27,0	85,32	96,4241	-11,1041
28,0	46,172	68,6255	-22,4535
29,0	144,722	61,8961	82,8259
30,0	101,488	116,681	-15,1929
31,0	52,316	69,4447	-17,1287
32,0	46,239	67,4871	-21,2481
33,0	52,17	64,247	-12,077
34,0	39,906	67,8447	-27,9387
35,0	81,91	58,6493	23,2607
36,0	45,634	84,2559	-38,6219
37,0	68,346	52,6665	15,6795
38,0	56,005	79,2771	-23,2721
39,0	76,788	59,4565	17,3315
40,0	84,645	79,6959	4,94913
41,0	29,455	74,1469	-44,6919
42,0	37,706	47,9717	-10,2657
43,0	110,15	63,612	46,538
44,0	60,006	93,2838	-33,2778
45,0	61,899	53,5468	8,35216
46,0	95,018	73,7244	21,2936
47,0	82,911	81,1354	1,77559
48,0	118,497	72,0218	46,4752
49,0	177,69	95,971	81,719
50,0	127,378	117,599	V9,7795
51,0	82,332	84,2046	V-1,87257
52,0	51,852	78,5853	V-26,7333
53,0	39,723	65,2115	V-25,4885
54,0	38,001	64,2477	V-26,2467
55,0	60,675	62,2913	V-1,61631
56,0	76,456	73,83	V2,62596
57,0	22,179	75,9943	V-53,8153
58,0	21,372	46,0506	V-24,6786
59,0	13,007	58,3142	V-45,3072
60,0	12,558	45,809	V-33,251
61,0	12,581	49,4806	V-36,8996
62,0	35,128	45,51	V-10,382
63,0	52,56	57,4066	V-4,84665
64,0	11,517	59,7268	V-48,2098
65,0	16,4202	36,3035	V-19,8833
66,0	8,266	48,4763	V-40,2103
67,0	46,454	36,424	V10,03
68,0	34,303	60,7718	V-26,4688
69,0	14,857	41,9155	V-27,0585

Notation : V : période de validation

II.2.2.3 BOUCHEGOUF

Le Tableau II.2.19 suivant résume les statistiques de la série chronologique des apports annuels observés à BOUCHEGOUF entre 1948 et 1993. Il présente les mesures de la tendance centrale, de la variabilité et de la forme de la série chronologique en question. La skewness et la kurtosis standardisées sont comprises dans l'intervalle $[-2, +2]$ ce qui indique une distribution normale.

Tableau II.2.19 Statistiques des apports annuels observés à BOUCHEGOUF (1948-1993)

Paramètre	valeur
Nombre d'observations	34
Moyenne	115.948
Variance	5016.59
Déviatation Standard	70.8279
Minimum	18.494
Maximum	236.37
Std. skewness	1.00022
Std. kurtosis	-1.43564

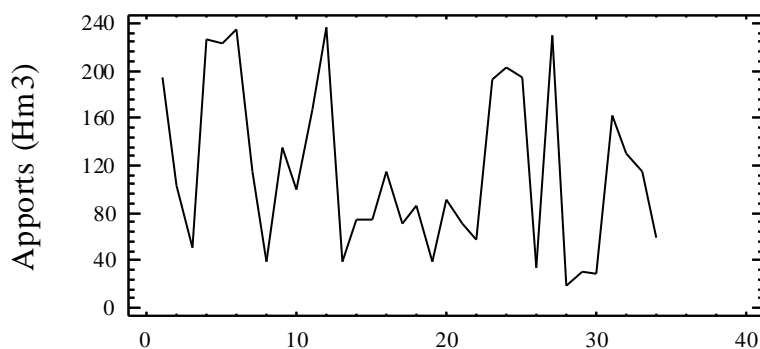


Fig. II.2.14 Apports annuels originaux observés à BOUCHEGOUF (1948-1993)

II.2.2.3.1 Stationnarisation

Le tracé de la série chronologique originale tel que montré par la Fig. II.2.18 ne semble présenter aucun signe de non stationnarité. Pour le confirmer, nous avons procédé par le calcul des premières et deuxièmes différences régulières de la série où le calcul de l'écart-type permet de choisir l'ordre optimal de la différenciation.

Le détail du calcul des séries différenciées pour les 5 premières années est montré dans le Tableau II.2.20, tandis que les statistiques des séries différenciées pour l'ensemble des données sont représentées dans le Tableau II.2.21.

Tableau II.2.20 Calcul des séries différenciées
(Données des 5 premières années)

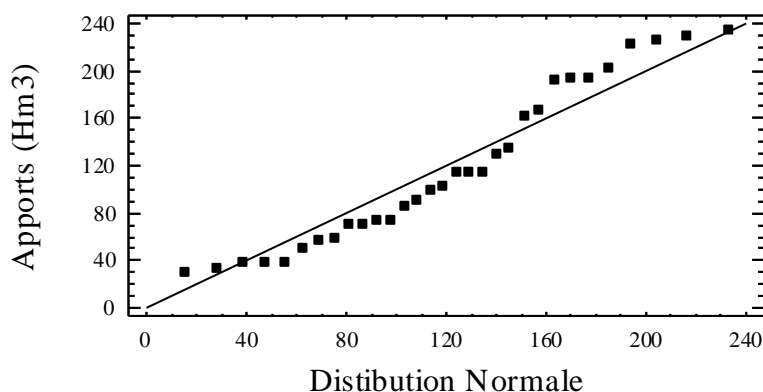
N°	Apport	DIFF(Apport,1)	DIFF(Apports,2)
1	194,26		
2	103,2	-91,06	
3	50,31	-52,89	38,17
4	226,72	176,41	229,3
5	223,49	-3,23	-179,64

Tableau II.2.21 Statistiques descriptives des séries différenciées
(1948-1993)

Série	N	Moyenne	Variance	Ecart-type
Apports	34	115,95	5016,59	70,82
DIFF(Apports,1)	33	-4,1	8771	93,65
DIFF(Apports ,2)	32	1,08	24900,3	157,79

Notation : DIFF (x,n) : différences régulières de x d'ordre n .

Le minimum d'écart-type est observé pour la série des apports d'origine ce qui confirme la thèse de sa stationnarité et rejette le recours à la différenciation. La série correspondante tel qu'illustré dans la Fig. II.2.18 est donc stationnaire et sa variance est de 5016,59 est stable autour de sa moyenne qui est de 115,95. La Fig. II.2.15 quant à elle montre l'ajustement des données à une distribution normale dont la qualité d'adéquation est confirmée par le test du Chi-deux (statistique du test = 15,88 ; P = 0,2555).

**Fig. II.2.15** Ajustement à la loi Normale des apports annuels observés à BOUCHEGOUF (1948-1993)

II.2.2.3.2. Identification

L'examen visuel du tracé de l'ACF et de celui de la PACF des apports considérés, tel que montré par les Fig. 16 et 17 ainsi que les Tableaux 22 et 23, indique 1 pic significatif au lag 2 pour l'ACF et 1 pic significatif au lag 2 pour la PACF.

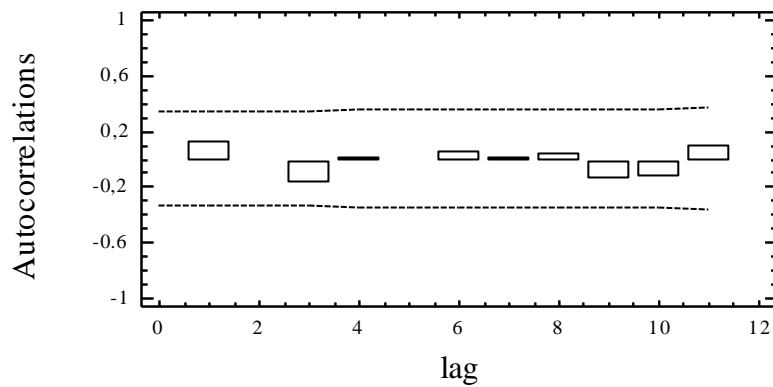


Fig. II.2.16 ACF des apports annuels observés à BOUCHEGOUF (1948-1993)

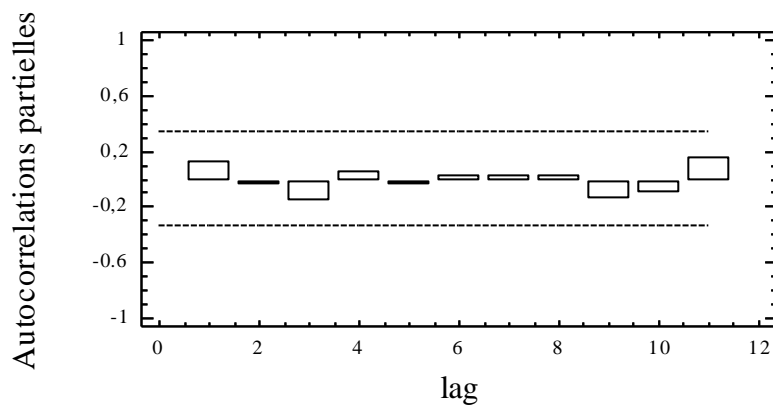


Fig. II.2.17 PACF des apports annuels observés à BOUCHEGOUF (1948-1993)

Tableau II.2.22 ACF des apports annuels observés à BOUCHEGOUF (1948-1993)

Lag	ACF	Ecart-type	-IC à 95%	IC à 95%
1	0,122196	0,171499	-0,336132	0,336132
2	-0,0259011	0,174041	-0,341114	0,341114
3	-0,178162	0,174154	-0,341336	0,341336
4	0,0029711	0,179434	-0,351686	0,351686
5	-0,0157582	0,179436	-0,351689	0,351689
6	0,0455102	0,179477	-0,351768	0,351768
7	0,0139426	0,179816	-0,352433	0,352433
8	0,030447	0,179848	-0,352495	0,352495
9	-0,151439	0,179999	-0,352792	0,352792
10	-0,144575	0,183708	-0,360062	0,360062
11	0,0982086	0,187025	-0,366562	0,366562

Tableau II.2.23 PACF des apports annuels observés à BOUCHEGOUF (1948-1993)

Lag	PACF	Ecart-type	-IC à 95%	IC à 95%
1	0,122196	0,171499	-0,336132	0,336132
2	-0,041452	0,171499	-0,336132	0,336132
3	-0,172671	0,171499	-0,336132	0,336132
4	0,0471296	0,171499	-0,336132	0,336132
5	-0,0313296	0,171499	-0,336132	0,336132
6	0,0219808	0,171499	-0,336132	0,336132
7	0,0144814	0,171499	-0,336132	0,336132
8	0,0205787	0,171499	-0,336132	0,336132
9	-0,15231	0,171499	-0,336132	0,336132
10	-0,110785	0,171499	-0,336132	0,336132
11	0,146753	0,171499	-0,336132	0,336132

Les résultats du test de Box-Pierce basé sur les 11 premières auto-corrélations montrent une valeur de P supérieure ou égale à 0,1 ce qui veut dire que l'hypothèse nulle d'indépendance ne peut pas être rejetée à 90% de confiance ou plus (la statistique calculée du test est de 3,54 avec une valeur de probabilité P de 0,98 qui est supérieure à 0,1). La série ne présentant aucune structure de dépendance interne peut donc être considérée comme complètement aléatoire, son traitement s'arrête ici.

II.2.2.4 BOUHNIFIA

Le Tableau II.2.24 suivant résume les statistiques de la série chronologique des apports annuels observés à BOUHNIFIA entre 1934 et 1992. Il présente les mesures de la tendance centrale, de la variabilité et de la forme de la série chronologique en question. La skewness et la kurtosis standardisées sont d'un intérêt particulier car ils sont utilisés pour déterminer si oui ou non la distribution est gaussienne. Une valeur de ces paramètres en dehors de l'intervalle $[-2, +2]$ indique un éloignement significatif de la normalité. Dans notre cas, la kurtosis standard est comprise dans l'intervalle tandis que la skewness ne l'est pas.

Tableau II.2.24 Statistiques des apports annuels observés à BOUHNIFIA (1934-1992)

Paramètre	valeur
Nombre d'observations	59
Moyenne	121.23
Variance	4049.08
Déviatiion Standard	63.63
Minimum	22.05
Maximum	296.2
Std. skewness	2.71
Std. kurtosis	0.72

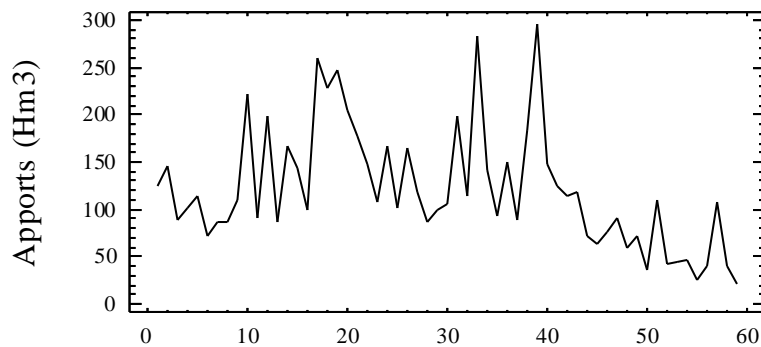


Fig. II.2.18 Apports annuels originaux observés à BOUHNIFIA (1934-1992)

II.2.2.4.1 Stationnarisation

Le tracé de la série chronologique originale tel que montré par la Fig. II.2.18 présente une tendance vers sa fin ce qui indique l'existence d'une non stationnarité. Le calcul des premières ou deuxièmes différences régulières de la série permet d'obtenir cette stationnarité et le calcul de l'écart-type permet de choisir l'ordre optimal de la différenciation.

Le détail du calcul des séries différenciées pour les 5 premières années est montré dans le Tableau II.2.25, tandis que les statistiques des séries différenciées pour l'ensemble des données sont représentées dans le Tableau II.2.26.

Tableau II.2.25 Calcul des séries différenciées (Données des 5 premières années)

N °	Apport	DIFF(Apport,1)	DIFF(Apports,2)
1	123,7		
2	145,3	21,6	
3	89,3	-56	-77,6
4	100,7	11,4	67,4
5	114,5	13,8	2,4

Tableau II.2.26 Statistiques descriptives des séries différenciées (1934-1992)

Série	N	Moyenne	Variance	Ecart-type
Apports	59	121,23	4049,08	63,63
DIFF(Apports,1)	58	-1,75	4529,43	67,3
DIFF(Apports,2)	57	-0,68	14121,7	118,79

Notation : DIFF (x,n) : différences régulières de x d'ordre n .

Dans le cas de la série étudiée le minimum d'écart-type est observé pour la série des apports originaux avec une valeur de 63,63. La série correspondante est donc celle qui présente le plus de stabilité, sa variance est de 4049,08 par rapport aux données différenciées, c'est la variance à expliquer. La Fig. II.2.19 montre l'ajustement de ces données à la distribution Normale, la qualité

d'adéquation de cet ajustement est confirmée par le test du Chi-deux (statistique du test = 17,94 avec une probabilité $P = 0,39$).

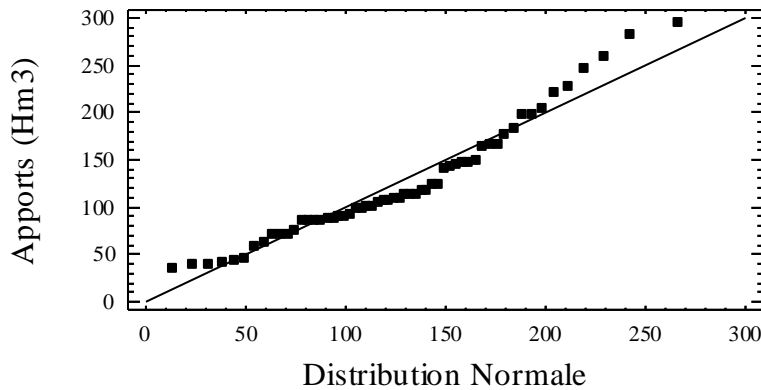


Fig. II.2.19 Ajustement Normal des apports annuels originaux observés à BOUHNIFIA (1934-1992)

II.2.2.4.2. Identification

L'examen visuel du tracé de la fonction d'auto-corrélation (ACF) et de celui de la fonction d'auto-corrélation partielle (PACF) des apports observés, tel que montré par les Fig. II.2.20 et 21 ainsi que les Tableaux II.2.27 et 28, indique 1 pic significatif aux lag 1, 2 et 3 pour l'ACF et 1 pic significatif aux lag 1 et 2 pour la PACF.

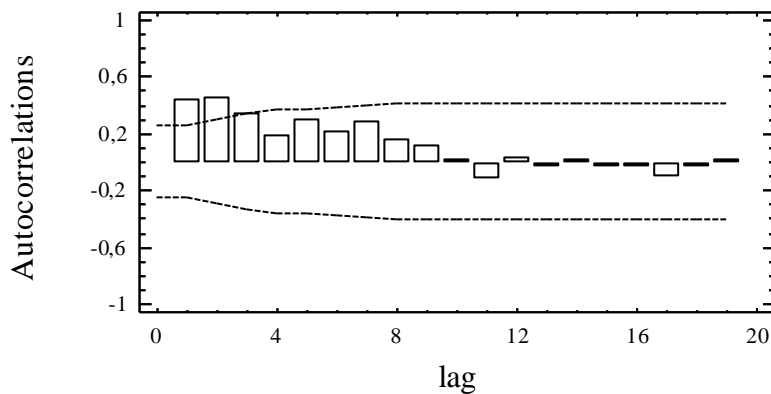


Fig. II.2.20 des apports annuels observés à BOUHNIFIA (1934-1992)

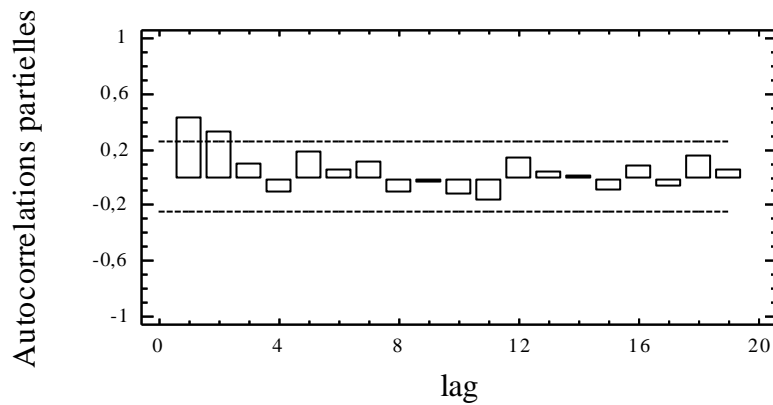


Fig. II.2.21 PACF des apports annuels observés à BOUHNIFIA (1934-1992)

Tableau II.2.27 ACF des apports annuels stationnaires

Lag	ACF	Ecart-type	-IC à 95%	IC à 95%
1	0,428994	0,130189	-0,255166	0,255166
2	0,44932	0,152275	-0,298454	0,298454
3	0,337302	0,173295	-0,339654	0,339654
4	0,177536	0,184087	-0,360805	0,360805
5	0,298533	0,186966	-0,366448	0,366448
6	0,204271	0,194878	-0,381955	0,381955
7	0,274995	0,198474	-0,389003	0,389003
8	0,150355	0,20483	-0,401461	0,401461
9	0,108797	0,206693	-0,405111	0,405111
10	0,0105907	0,207661	-0,407009	0,407009
11	-0,133548	0,20767	-0,407027	0,407027
12	0,0313	0,209121	-0,40987	0,40987
13	-0,0434745	0,2092	-0,410025	0,410025
14	0,0170809	0,209353	-0,410325	0,410325
15	-0,0485878	0,209377	-0,410372	0,410372
16	-0,0427228	0,209568	-0,410746	0,410746
17	-0,11014	0,209715	-0,411035	0,411035
18	-0,0370397	0,210693	-0,412952	0,412952
19	0,0116757	0,210804	-0,413169	0,413169

Tableau II.2.28 PACF des apports annuels stationnaires

Lag	PACF	Ecart-type	-IC à 95%	IC à 95%
1	0,428994	0,130189	-0,255166	0,255166
2	0,325118	0,130189	-0,255166	0,255166
3	0,0928328	0,130189	-0,255166	0,255166
4	-0,115796	0,130189	-0,255166	0,255166
5	0,186077	0,130189	-0,255166	0,255166
6	0,0528543	0,130189	-0,255166	0,255166
7	0,108025	0,130189	-0,255166	0,255166
8	-0,116168	0,130189	-0,255166	0,255166
9	-0,0482019	0,130189	-0,255166	0,255166
10	-0,132301	0,130189	-0,255166	0,255166
11	-0,18514	0,130189	-0,255166	0,255166
12	0,13763	0,130189	-0,255166	0,255166
13	0,0432058	0,130189	-0,255166	0,255166
14	0,00344743	0,130189	-0,255166	0,255166
15	-0,115471	0,130189	-0,255166	0,255166
16	0,0848466	0,130189	-0,255166	0,255166
17	-0,0723164	0,130189	-0,255166	0,255166
18	0,146092	0,130189	-0,255166	0,255166
19	0,04605	0,130189	-0,255166	0,255166

Les résultats du test de Box-Pierce basé sur les 19 premières auto-corrélations montrent qu'à 99% de confiance la série n'est pas complètement aléatoire (la statistique calculée du test est de 47,85 avec une valeur de probabilité P de 0,0002 qui est inférieure à 0,01).

Les comportements de l'ACF et la PACF nous poussent à opter pour les modèles possibles suivants :

- (A) ARIMA(2,0,3) avec constante
- (B) ARIMA(1,0,3) avec constante
- (C) ARIMA(1,0,2) avec constante
- (D) ARIMA(1,0,1) avec constante

II.2.2.4.3 Estimation

Pour l'estimation des paramètres de ces modèles, il faut rappeler que 59 données ont été utilisées dont les 20 dernières ont servi à la validation. Les paramètres des modèles ont alors été estimés sur une période de 39 ans.

L'analyse et la comparaison des modèles ajustés tel que présenté dans le Tableau II.2.29 montre que tous les modèles ajustés produisent des résidus complètement indépendants avec une moyenne et une variance qui sont stables entre la période d'estimation et celle de validation. Cependant, le modèle qui produit le minimum de variance des résidus est le modèle (A) mais en contre partie sa performance est la plus médiocre et son coefficient AR(2) est statistiquement non significatif. Le modèle (D) présente quant à lui la meilleure performance mais là aussi les coefficients AR(1) et MA(1) sont statistiquement non significatifs et la variance de ses résidus est la plus élevée. Le modèle (C) présente une performance passable mais ses coefficients AR(1) et MA(1) sont statistiquement non significatifs. Le modèle (B) présente une performance passable et présente l'avantage d'avoir des coefficients significatifs avec une variance des résidus plus petite, c'est lui qui est donc retenu pour la prédiction.

Tableau II.2.29 Comparaison des modèles ajustés

Modèle	Variance des résidus		AUTO	MOY	VAR
	Periode d'estimation	Période de validation			
(A)	3039,6	5352,2	OK	OK	OK
(B)	3118,6	4164,2	OK	OK	OK
(C)	3122,4	4218,5	OK	OK	OK
(D)	3615,4	3798,2	OK	OK	OK

Notation : AUTO :test d'indépendance des résidus ; MOY : test d'égalité de la moyenne ; VAR : test d'homoscédasticité ; OK : le test est positif.

Les Tableaux II.2.30 résume l'estimation des paramètres du modèle retenu ARIMA(1,0,3) avec constante ainsi que la signification statistique de ses termes. Les termes correspondant à une valeur de $P < 0.05$ sont statistiquement significativement différents de zéro à 95% de confiance.

Tableau II.2.30 Estimation des paramètres du modèle ARIMA(1,0,3)

Paramètre	Estimation	Ecart- type	t	P
AR(1)	-0,783491	0,167725	-4,67127	0,000020
MA(1)	-1,12162	0,188258	-5,95787	0,000000
MA(2)	-0,819156	0,18176	-4,50679	0,000036
MA(3)	-0,675301	0,148436	-4,54945	0,000031
Moyenne	150,498	17,0929	8,8047	0,000000
Constante	268,411			

La variance estimée du bruit blanc = 3031,81 avec 54 DDL et l'écart- type estimé du bruit blanc = 55,06.

La valeur de P pour tous les termes est inférieure à 0.01 donc ces derniers sont significativement différents de zéro à 99% de confiance avec des valeurs différentes de un (1). L'estimation de l'écart-type du bruit blanc est de 55,06 et l'estimation de la variance des résidus est de 3031,81.

Le Tableau II.2.31 résume la performance du modèle ARIMA(1,0,3) à ajuster les données considérées et ce par l'analyse des erreurs qu'il produit. La différence de ces statistiques d'erreurs n'est pas trop importante entre la période d'estimation et celle de validation. Ces différences expriment une performance moyenne du modèle même si c'est la meilleure par rapport aux modèles ajustés. Il s'ensuit que le modèle ajusté peut être considéré comme moyennement performant et donc les prévisions obtenues peuvent être considérées en conséquence.

Tableau II.2.31 Statistiques des erreurs du modèle ARIMA(1,0,3)

Statistique	Période d'estimation	Période de validation
EQM	3118,66	4164,25
EAM	42,054	56,5265
EM	31,5666	113,576

Notation : EQM : erreur quadratique moyenne ; EAM : erreur absolue moyenne ; EM:erreur moyenne.

II.2.2.4.4 Validation

Il s'agit de vérifier l'hypothèse de la méthode à savoir que les résidus forment bien un bruit blanc gaussien de moyenne nulle. Pour cela, le test des auto-corrélogrammes (ACF et PACF) et celui du Portmanteau (Box-Pierce) sont utilisés pour tester l'indépendance des résidus, le test t de student est utilisé pour tester la nullité de leur moyenne et le test d'adéquation du Chi-deux à 95% de confiance est utilisé pour tester leur normalité.

- **Test des ACF et PACF**

L'auto-corrélogramme total et partiel des résidus du modèle ont été tracés tel que montré dans la Fig. II.2.22 et Fig. II.2.23.

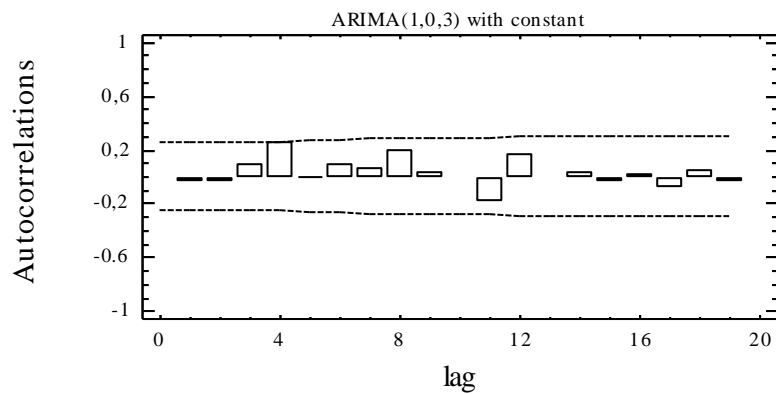


Fig. II.2.22 ACF des résidus du modèle ARIMA(1,0,3)

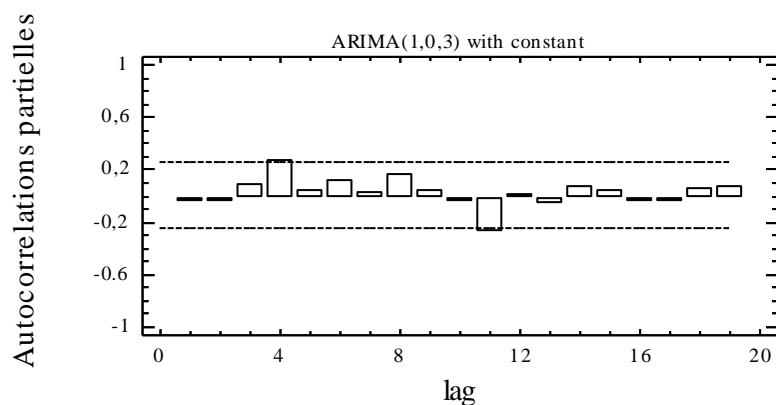


Fig. II.2.23 PACF des résidus du modèle ARIMA(1,0,3)

Tous les coefficients de l'ACF sont inclus dans l'intervalle de confiance (IC) tracé à 95% de confiance ce qui indique que ces derniers sont statistiquement non significatifs, d'où leur indépendance. Leurs valeurs sont reportées dans le Tableau II.2.32 et Tableau II.2.33.

Tableau II.2.32 ACF des résidus du modèle ARIMA(1,0,3)

Lag	Auto-corrélation partielle	Ecart-type	limite inférieure IC 95%	limite supérieure IC 95%
1	-0,0349097	0,130189	-0,255166	0,255166
2	-0,0446435	0,130347	-0,255477	0,255477
3	0,085048	0,130606	-0,255984	0,255984
4	0,259332	0,131542	-0,257817	0,257817
5	-0,00108147	0,139939	-0,274276	0,274276
6	0,0916714	0,139939	-0,274277	0,274277
7	0,0545673	0,140953	-0,276264	0,276264
8	0,198588	0,141311	-0,276965	0,276965
9	0,0284394	0,145965	-0,286086	0,286086
10	-0,0195839	0,146059	-0,28627	0,28627
11	-0,192468	0,146103	-0,286357	0,286357
12	0,156885	0,150339	-0,29466	0,29466
13	-0,0109565	0,153089	-0,300049	0,300049
14	0,0261348	0,153102	-0,300075	0,300075
15	-0,0386056	0,153178	-0,300223	0,300223
16	0,00865435	0,153342	-0,300546	0,300546
17	-0,0859284	0,153351	-0,300562	0,300562
18	0,0392246	0,154165	-0,302158	0,302158
19	-0,0312941	0,154334	-0,302489	0,302489

Tableau II.2.33 PACF des résidus du modèle ARIMA(1,0,3)

Lag	Auto-corrélation partielle	Ecart-type	limite inférieure IC 95%	limite supérieure IC 95%
1	-0,0349097	0,130189	-0,255166	0,255166
2	-0,0459182	0,130189	-0,255166	0,255166
3	0,0820879	0,130189	-0,255166	0,255166
4	0,265734	0,130189	-0,255166	0,255166
5	0,0316036	0,130189	-0,255166	0,255166
6	0,115062	0,130189	-0,255166	0,255166
7	0,024483	0,130189	-0,255166	0,255166
8	0,155855	0,130189	-0,255166	0,255166
9	0,0342717	0,130189	-0,255166	0,255166
10	-0,0587546	0,130189	-0,255166	0,255166
11	-0,277438	0,130189	-0,255166	0,255166
12	0,013473	0,130189	-0,255166	0,255166
13	-0,0623201	0,130189	-0,255166	0,255166
14	0,0621714	0,130189	-0,255166	0,255166
15	0,0368494	0,130189	-0,255166	0,255166
16	-0,0447658	0,130189	-0,255166	0,255166
17	-0,0491578	0,130189	-0,255166	0,255166
18	0,0487086	0,130189	-0,255166	0,255166
19	0,0627108	0,130189	-0,255166	0,255166

- **Test du Portmanteau**

Ce test est basé sur les 22 premières auto-corrélations des résidus, les résultats du test montrent une valeur de P supérieure ou égale à 0.10, donc on ne peut pas rejeter l'hypothèse nulle que les résidus sont indépendants avec un intervalle de confiance de 90% ou plus.

Test de Box-Pierce

Hypothèse nulle : résidus indépendants

Valeur calculée de la statistique $Q = 12,01$

Valeur de P = 0.6779

- **Test de nullité de la moyenne**

La moyenne des résidus est de -5,6336, leur médiane est de -12,59 et leur variance est de 3015,51 avec un écart-type de 54,9136. Les résultats du test t de Student montrent une valeur P= 0,33 qui est supérieure à 0.05 ce qui implique que l'hypothèse nulle de nullité de la moyenne des résidus ne peut pas être rejetée à 95% de confiance.

Test t

Hypothèse nulle : moyenne = 0

Valeur calculée de la statistique $t = -0,97$

Valeur de P = 0.33

- **Test de Normalité**

Afin de vérifier si les résidus du modèle ARIMA(1,0,3) appartiennent à une distribution Normale, un ajustement à la loi de Gauss a été effectué tel que présenté dans la Fig. II.2.24. Les

résultats du test du Chi-deux montrent une valeur de P supérieure à 0,05 donc l'hypothèse nulle de Normalité ne peut pas être rejetée à 95% de confiance.

test du Chi-deux

Valeur de la statistique du test = 25,4068.

Valeur de P = 0,085968

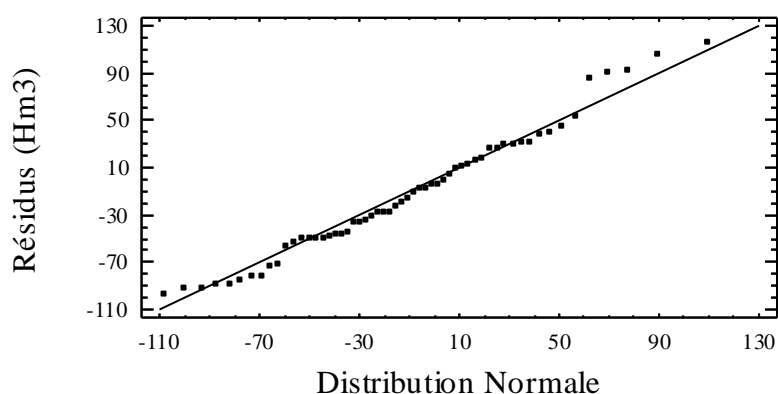


Fig. II.2.24 Ajustement des résidus du modèle ARIMA(1,0,3) à La loi de Gauss

Les résultats des tests précédents appliqués aux résidus du modèle ARIMA(1,0,3) ont permis la vérification des hypothèses de la méthode de Box-Jenkins.

En fin, ce modèle peut être formulé de la manière suivante :

$$(1 + 0,783491B)(X_t - 150,49) = (1 + 1,12162B + 0,819156B^2 - 0,675301B^3)\varepsilon_t + 268,41$$

Où ε_t est un bruit blanc gaussien de moyenne nulle et de variance 3015,51. Ainsi la variance expliquée par le modèle peut être estimée à seulement 25,5%. Ce pourcentage qui exprime la contribution de la partie structurée au sein de la série étudiée, représente le un quart de la variabilité totale ce qui fait que la contribution de la partie aléatoire est trois fois plus importante. Cette considération doit être prise en compte lors de l'utilisation du modèle pour la prédiction.

II.2.2.4.5 Prédiction

Les résultats des prédictions obtenues par le modèle retenu ARIMA(1,0,3) sont regroupés dans le Tableau II.2.34 pour la période suivant la fin de la série chronologique de départ et dans le Tableau II.2.35 pour la période où les données sont disponibles avec les résidus correspondants. Pour cette dernière période des intervalles de confiances à 95% ont été calculés pour les prédictions obtenues. Ces limites de confiance indiquent où va être située la vraie valeur à une période déterminée dans le future, dans l'hypothèse que le modèle ajusté est adéquat aux données en question. Ces prédictions représentées dans la Fig. II.2.25 en trait continu suivent l'allure générale des données observées représentées en points carrés. Pour la période où les données ne sont pas disponibles (après l'indice 59 sur l'axe des x) les prédictions obtenues pour les 12 ans futurs sont représentées avec leurs limites de confiance tracées à 95% .

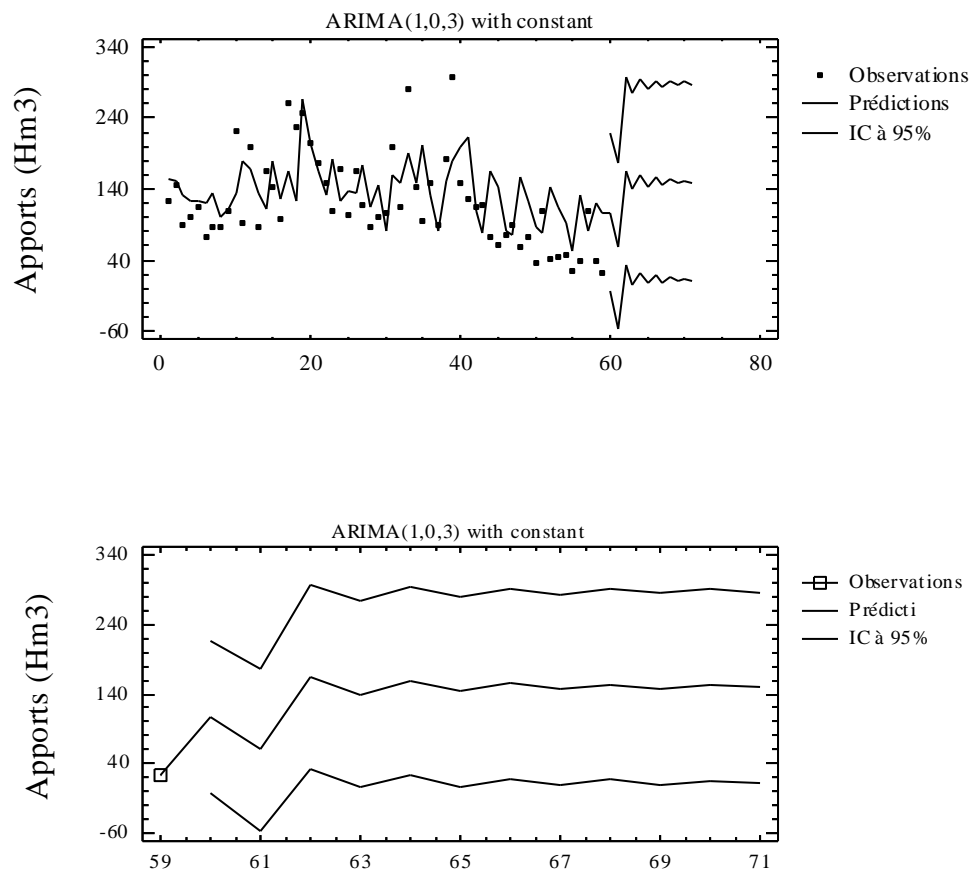


Fig. II.2.25 Prédictions des apports annuels à BOUHNIFIA obtenues par le modèle ARIMA(1,0,3)

Tableau II.2.34 Prédictions des apports annuels à BOUHNIFIA pour les 12 ans futurs

Période	Donnée	Prédiction	Résidu
60,0	107,557	-2,8355	217,95
61,0	59,3599	-57,1727	175,892
62,0	164,534	32,9162	296,152
63,0	139,5	5,21922	273,782
64,0	159,114	23,2238	295,004
65,0	143,747	6,87834	280,616
66,0	155,787	18,3212	293,253
67,0	146,354	8,52272	284,185
68,0	153,745	15,6898	291,799
69,0	147,954	9,762	286,146
70,0	152,491	14,2148	290,767
71,0	148,936	10,6085	287,264

Tableau II.2.35 Prédications avec ARIMA(1,0,3)

Période	Donnée	Prédiction	Résidu
1,0	123,7	154,303	-30,6031
2,0	145,3	152,6	-7,2996
3,0	89,3	132,829	-43,5287
4,0	100,7	122,977	-22,2773
5,0	114,5	123,941	-9,44105
6,0	71,5	120,469	-48,9689
7,0	87,6	134,69	-47,0898
8,0	85,8	100,472	-14,6721
9,0	109,1	113,089	-3,98872
10,0	221,2	134,64	86,5598
11,0	91,4	179,015	-87,6148
12,0	197,8	166,742	31,0576
13,0	85,9	134,956	-49,0555
14,0	166,6	112,363	54,2374
15,0	143,4	179,505	-36,1046
16,0	98,9	126,865	-27,9648
17,0	259,8	166,61	93,1904
18,0	228,0	122,095	105,905
19,0	247,9	266,013	-18,1127
20,0	204,1	203,552	0,547574
21,0	177,5	165,795	11,7045
22,0	147,2	130,687	16,5132
23,0	108,4	181,561	-73,1606
24,0	167,5	122,854	44,6463
25,0	102,1	138,474	-36,3743
26,0	165,3	134,786	30,5142
27,0	118,2	173,479	-55,2792
28,0	86,7	114,233	-27,5329
29,0	99,6	144,925	-45,3254
30,0	106,7	79,6541	27,0459
31,0	199,3	159,427	39,8734
32,0	114,0	148,531	-34,531
33,0	281,9	191,29	90,6104
34,0	141,5	147,816	-6,31596
35,0	94,1	201,369	-107,269
36,0	149,4	130,386	19,0138
37,0	89,8	80,5493	9,25068
38,0	183,0	151,566	31,4337
39,0	296,2	180,707	115,493
40,0	148,5	197,877	V-49,3765
41,0	124,7	212,515	V-87,8154
42,0	114,2	109,76	V4,43965
43,0	117,4	78,6378	V38,7622
44,0	72,8	164,241	V-91,4408
45,0	62,6	143,562	V-80,9619
46,0	76,4	79,8284	V-3,4284
47,0	90,41	76,6368	V13,7732
48,0	59,7	155,542	V-95,8421
49,0	71,5	123,106	V-51,6059
50,0	36,03	85,3011	V-49,2711
51,0	109,85	77,9232	V31,9268
52,0	41,86	142,944	V-101,084
53,0	43,5	115,117	V-71,6166
54,0	46,26	92,7595	V-46,4995
55,0	25,71	53,0849	V-27,3749
56,0	40,11	131,111	V-91,0005
57,0	108,72	81,0923	V27,6277
58,0	39,46	121,188	V-81,7281
59,0	22,052	107,006	V-84,9538

Notation : V : période de validation

II.2.2.5 CHEFFIA

Le Tableau II.2.36 suivant résume les statistiques de la série chronologique des apports annuels observés à CHEFFIA entre 1966 et 1992. Il présente les mesures de la tendance centrale, de la variabilité et de la forme de la série chronologique en question. La skewness et la kurtosis standardisées sont comprises dans l'intervalle $[-2, +2]$ ce qui indique une distribution normale.

Tableau II.2.36 Statistiques des apports annuels observés à CHEFFIA (1966-1992)

Paramètre	valeur
Nombre d'observations	27
Moyenne	135,661
Variance	8735,7
Déviatiion Standard	93,465
Minimum	17,88
Maximum	368,75
Stnd. skewness	1,98532
Stnd. kurtosis	0,383692

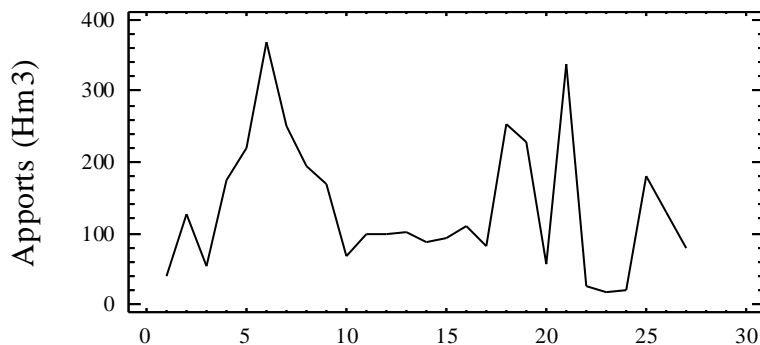


Fig. II.2.26 Apports annuels originaux observés à CHEFFIA (1966-1992)

II.2.2.5.1 Stationnarisation

Le tracé de la série chronologique originale tel que montré par la Fig. II.2.26 ne semble présenter aucun signe de non stationnarité. Pour le confirmer, nous avons procédé par le calcul des premières et deuxièmes différences régulières de la série où le calcul de l'écart-type permet de choisir l'ordre optimal de la différenciation.

Le détail du calcul des séries différenciées pour les 5 premières années est montré dans le Tableau II.2.37, tandis que les statistiques des séries différenciées pour l'ensemble des données sont représentées dans le Tableau II.2.20.

Tableau II.2.37 Calcul des séries différenciées
(Données des 5 premières années)

N°	Apport	DIFF(Apport,1)	DIFF(Apports,2)
1	40,64		
2	125,61	84,97	
3	54,58	-71,03	-156
4	173,44	118,86	189,89
5	219,97	46,53	-72,33

Tableau II.2.38 Statistiques descriptives des séries différenciées
(1966-1992)

Série	N	Moyenne	Variance	Ecart-type
Apports	27	135,66	8735,7	93,46
DIFF(Apports,1)	26	1,44	13956,5	118,13
DIFF(Apports,2)	25	-5,51	41733,8	204,28

Notation : DIFF (x,n) : différences régulières de x d'ordre n .

Le minimum d'écart-type est observé pour la série des apports d'origine. La série correspondante est donc stationnaire et n'a pas besoin d'être différenciée. Sa variance est de 8735,7 est stable autour de sa moyenne qui est de 135,66. La Fig. II.2.27 quant à elle montre l'ajustement des données à une distribution normale dont la qualité d'adéquation est confirmée par le test du Chi-deux (statistique du test = 10,85 ; P = 0,4557).

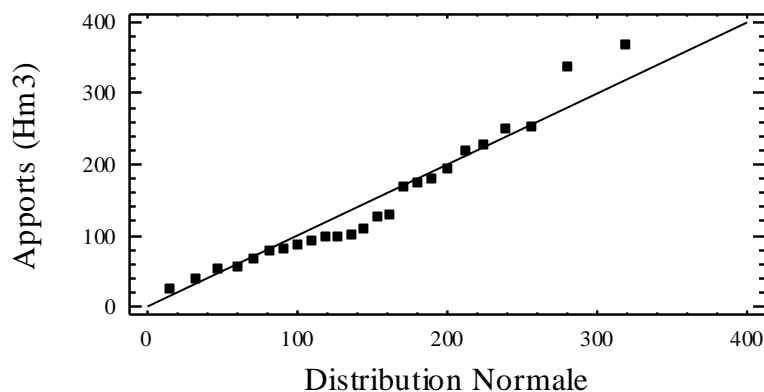


Fig. II.2.27 Ajustement à la loi Normale des apports annuels observés à CHEFFIA (1966-1992)

II.2.2.5.2. Identification

L'examen visuel du tracé de l'ACF et de celui de la PACF des apports considérés, tel que montré par les Fig. II.2.28 et 29 ainsi que les Tableaux II.2.39 et 40, n'indique aucun pic significatif aussi bien pour l'ACF que pour la PACF.

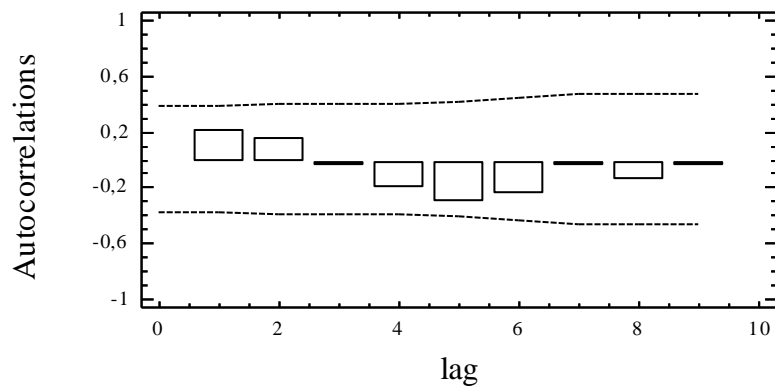


Fig. II.2.28 ACF des apports annuels observés à CHEFFIA (1966-1992)

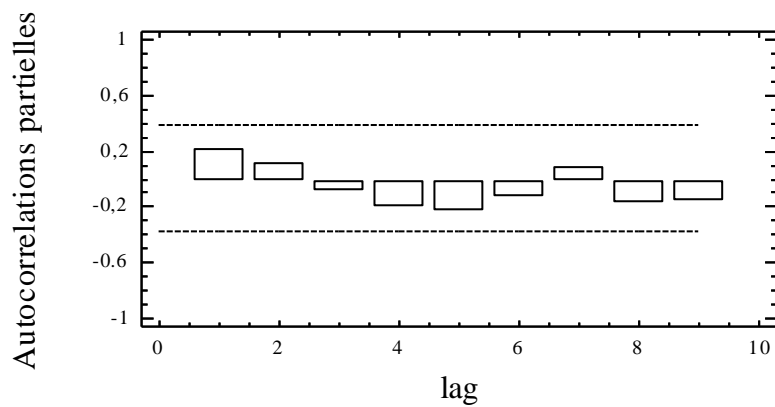


Fig. II.2.29 PACF des apports annuels observés à CHEFFIA (1966-1992)

Tableau II.2.39 ACF des apports annuels observés à CHEFFIA (1966-1992)

Lag	ACF	Ecart-type	-IC à 95%	IC à 95%
1	0,204649	0,19245	-0,377196	0,377196
2	0,150402	0,200348	-0,392676	0,392676
3	-0,042949	0,204487	-0,400788	0,400788
4	-0,207756	0,204821	-0,401442	0,401442
5	-0,309214	0,212483	-0,416459	0,416459
6	-0,252574	0,228542	-0,447934	0,447934
7	-0,0452135	0,238656	-0,467758	0,467758
8	-0,144967	0,238973	-0,46838	0,46838
9	-0,0502801	0,242208	-0,47472	0,47472

Tableau II.2.40 PACF des apports annuels observés à CHEFFIA (1966-1992)

Lag	PACF	Ecart-type	-IC à 95%	IC à 95%
1	0,204649	0,19245	-0,377196	0,377196
2	0,113264	0,19245	-0,377196	0,377196
3	-0,0987725	0,19245	-0,377196	0,377196
4	-0,21247	0,19245	-0,377196	0,377196
5	-0,239888	0,19245	-0,377196	0,377196
6	-0,131788	0,19245	-0,377196	0,377196
7	0,0781151	0,19245	-0,377196	0,377196
8	-0,175783	0,19245	-0,377196	0,377196
9	-0,161128	0,19245	-0,377196	0,377196

Les résultats du test de Box-Pierce basé sur les 10 premières auto-corrélations montrent une valeur de P supérieure ou égale à 0,1 ce qui veut dire que l'hypothèse nulle d'indépendance ne peut pas être rejetée à 90% de confiance ou plus (la statistique calculée du test est de 7,95 avec une valeur de probabilité P de 0,53 qui est supérieure à 0,1). La série ne présentant aucune structure de dépendance interne peut donc être considérée comme complètement aléatoire, son traitement s'arrête ici.

II.2.2.6 KSOB

Le Tableau II.2.41 suivant résume les statistiques de la série chronologique des apports annuels observés à KSOB entre 1966 et 1992. Il présente les mesures de la tendance centrale, de la variabilité et de la forme de la série chronologique en question. La skewness standardisée est en dehors de l'intervalle $[-2, +2]$, tandis que la kurtosis y est comprise ce qui nécessite une vérification de la normalité des données.

Tableau II.2.41 Statistiques des apports annuels observés à KSOB (1966-1992)

Paramètre	valeur
Nombre d'observations	27
Moyenne	32,986
Variance	509,51
Déviatation Standard	22,5723
Minimum	22,5723
Maximum	87,671
Std. skewness	2,48647
Std. kurtosis	0,528756

II.2.2.6.1 Stationnarisation

Le tracé de la série chronologique originale tel que montré par la Fig. II.2.30 présente une tendance vers sa fin ce qui peut éventuellement indiquer l'existence d'une non stationnarité. Si c'est bien le cas, le calcul des premières ou deuxièmes différences régulières de la série permet d'obtenir cette stationnarité et le calcul de l'écart-type permet de choisir l'ordre optimal de la différenciation.

Le détail du calcul des séries différenciées pour les 5 premières années est montré dans le Tableau II.2.42, tandis que les statistiques des séries différenciées pour l'ensemble des données sont représentées dans le Tableau II.2.43.

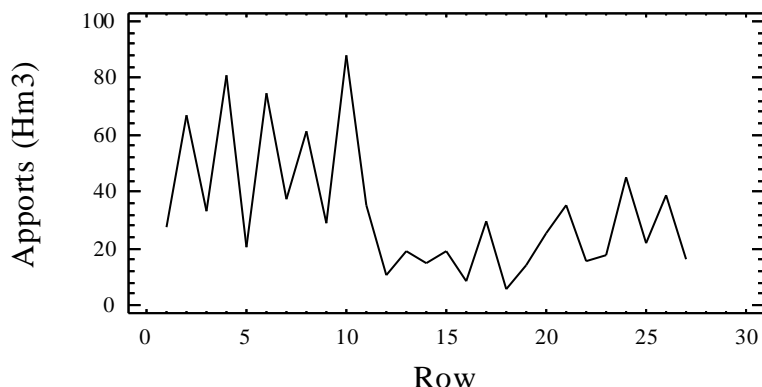


Fig. II.2.30 Apports annuels originaux observés à KSOB (1966-1992)

Tableau II.2.42 Calcul des séries différenciées
(Données des 5 premières années)

	Apport	DIFF(Apport,1)	DIFF(Apports,2)
1	27,682		
2	66,527	38,845	
3	33,197	-33,33	-72,175
4	80,707	47,51	80,84
5	20,59	-60,117	-107,627

Tableau II.2.43 Statistiques descriptives des séries différenciées
(1966-1992)

Série	N	Moyenne	Variance	Ecart-type
Apports	27	32,98	509,51	22,57
DIFF(Apports,1)	26	-0,42	1012,52	31,82
DIFF(Apports ,2)	25	-2,43	3607,86	60,06

Notation : DIFF (x,n) : différences régulières de x d'ordre n .

Dans le cas de la série étudiée le minimum d'écart-type est observé pour les apports originaux malgré le décalage dans le niveau de la moyenne observé vers la fin de la série. La série correspondante tel qu'illustré dans la Fig. II.2.30 est donc plus stable avec une variance de 509,51. La Fig. II.2.31 quant à elle montre l'ajustement de ces données à une distribution normale, dont la qualité d'adéquation est confirmée par le test du Chi-deux à plus de 90% de confiance (statistique du test = 16,03 et P = 0,1397).

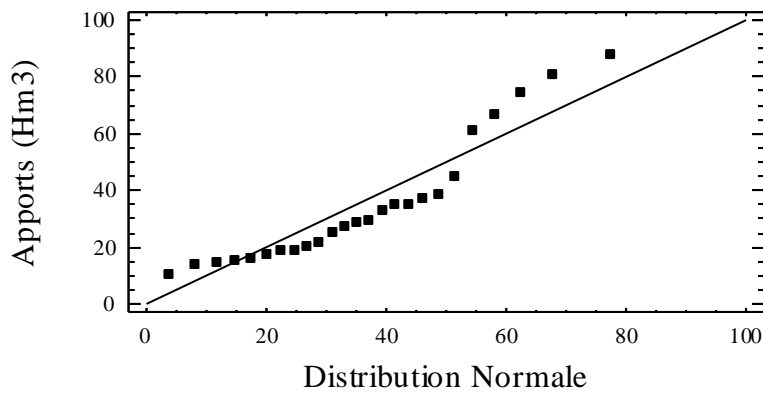


Fig. II.2.31 Ajustement Normal des apports annuels observés à KSOB (1966-1992)

II.2.2.6.2. Identification

L'examen visuel du tracé de la fonction d'auto-corrélation (ACF) et de celui de la fonction d'auto-corrélation partielle (PACF) des apports observés, tel que montré par les Fig. II.2.32 et 33 ainsi que les Tableaux II.2.44 et 45, indique 1 pic significatif au lag 2 pour l'ACF et 1 pic significatif au lag 2 également pour la PACF.

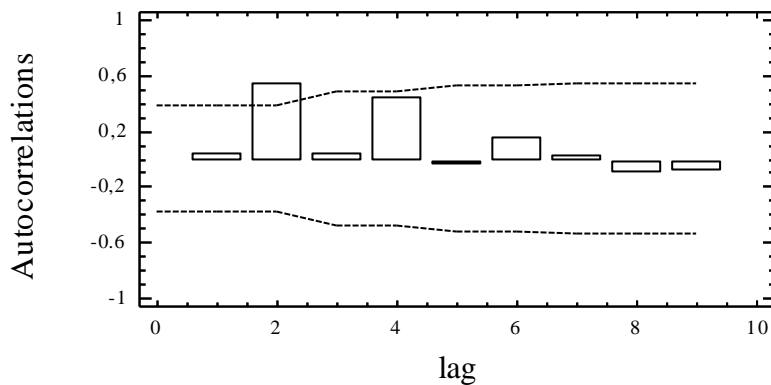


Fig. II.2.32 ACF des apports annuels observés à KSOB (1966-1992)

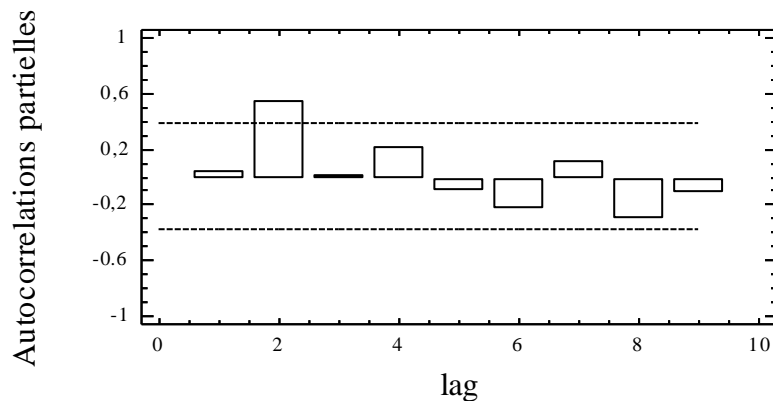


Fig. II.2.33 PACF des apports annuels observés à KSOB (1966-1992)

Tableau II.2.44 ACF des apports annuels stationnaires

Lag	ACF	Ecart-type	-IC à 95%	IC à 95%
1	0,0331717	0,19245	-0,377196	0,377196
2	0,547012	0,192662	-0,377611	0,377611
3	0,0363886	0,243481	-0,477215	0,477215
4	0,443851	0,243683	-0,47761	0,47761
5	-0,0459988	0,271982	-0,533076	0,533076
6	0,148651	0,27227	-0,53364	0,53364
7	0,0228451	0,275259	-0,539499	0,539499
8	-0,112891	0,275329	-0,539637	0,539637
9	-0,101005	0,277038	-0,542986	0,542986

Tableau II.2.45 PACF des apports annuels stationnaires

Lag	PACF	Ecart-type	-IC à 95%	IC à 95%
1	0,0331717	0,19245	-0,377196	0,377196
2	0,546513	0,19245	-0,377196	0,377196
3	0,0143189	0,19245	-0,377196	0,377196
4	0,205997	0,19245	-0,377196	0,377196
5	-0,115597	0,19245	-0,377196	0,377196
6	-0,23371	0,19245	-0,377196	0,377196
7	0,1065	0,19245	-0,377196	0,377196
8	-0,303862	0,19245	-0,377196	0,377196
9	-0,124332	0,19245	-0,377196	0,377196

Les résultats du test de Box-Pierce basé sur les 9 premières auto-corrélations montrent qu'à 90% de confiance la série n'est pas complètement aléatoire (la statistique calculée du test est de 14,75 avec une valeur de probabilité P de 0,09 qui est inférieure à 0,1).

Les comportements de l'ACF et la PACF nous poussent à opter pour les modèles possibles suivants :

- (A) ARIMA(2,0,2) avec constante
- (B) ARIMA(2,0,1) avec constante
- (C) ARIMA(2,0,0) avec constante

(D) ARIMA(1,0,1) avec constante

(E) ARIMA(1,0,2) avec constante

II.2.2.6.3 Estimation

La présente procédure va estimer les 12 valeurs futures de la série des données en question. Un modèle ARIMA a été sélectionné. Pour l'estimation des paramètres de ce modèle, il faut rappeler que 27 données ont été utilisées dont les 10 dernières ont servi à la validation. Les paramètres des modèles ont alors été estimés sur une période de 17 ans.

L'analyse et la comparaison des modèles ajustés tel que présenté dans le Tableau II.2.46 montre que tous les modèles ajustés produisent des résidus complètement indépendants avec une moyenne et une variance qui sont stables entre la période d'estimation et celle de validation. Cependant, le modèle qui produit le minimum de variance des résidus est le modèle (C) c'est lui qui est donc retenu pour la prédiction.

Tableau II.2.46 Comparaison des modèles ajustés

Modèle	Variance des résidus		AUTO	MOY	VAR
	Periode d'estimation	Période de validation			
(A)	500,377	195,864	OK	OK	OK
(B)	491,412	219,965	OK	OK	OK
(C)	477,72	191,004	OK	OK	OK
(D)	693,204	385,861	OK	OK	OK
(E)	585,653	247,766	OK	OK	OK

Notation : AUTO :test d'indépendance des résidus ; MOY : test d'égalité de la moyenne ; VAR : test d'homoscédasticité ; OK : le test est positif.

Les Tableaux II.2.47 et 48 résument l'estimation des paramètres du modèle retenu ARIMA(2,0,0) avec constante ainsi que la signification statistique de ses termes. Les termes correspondant à une valeur de $P < 0.05$ sont statistiquement significativement différents de zéro à 95% de confiance.

Tableau II.2.47 Estimation des paramètres du modèle retenu ARIMA(2,0,0) avec constante

Paramètre	Estimation	Ecart- type	t	P
AR(1)	-0,02248	0,210716	-0,106726	0,915894
AR(2)	0,656932	0,216353	3,03639	0,005691
Moyenne	36,185	13,168	2,74794	0,011201
Constante	13,2277			

La variance estimée du bruit blanc = 487,318 avec 24 DDL et l'écart- type estimé du bruit blanc = 22,0753.

La valeur de P pour les termes AR(2) et la moyenne est inférieure à 0.05 donc ces derniers sont

significativement différents de zéro à 95% de confiance avec une valeur du paramètre autorégressif différente de un (1). La valeur de P pour le terme AR(1) est très supérieure à 0.1 donc elle est statistiquement non significativement différente de zéro à 90% de confiance et plus et peut donc être éliminée. L'estimation de l'écart-type du bruit blanc est de 22,0753 et l'estimation de la variance des résidus est de 487,318.

Le Tableau II.2.48 résume la performance du modèle ARIMA(2,0,0) à ajuster les données considérées et ce par l'analyse des erreurs qu'il produit. La différence de ces statistiques d'erreurs n'est pas trop importante entre la période d'estimation et celle de validation. Ces différences expriment une bonne performance du modèle. Ceci veut dire que le modèle ajusté peut être considéré comme assez performant pour la prédiction des données considérées.

Tableau II.2.48 Statistiques des erreurs du modèle ARIMA(2,0,0)

Statistique	Période d'estimation	Période de validation
EQM	477,72	191,004
EAM	13,878	12,4063
EM	-0,977937	-3,46315

Notation : EQM : erreur quadratique moyenne ; EAM : erreur absolue moyenne ; EM: erreur moyenne.

II.2.2.6.4 Validation

Il s'agit de vérifier l'hypothèse de la méthode à savoir que les résidus forment bien un bruit blanc gaussien de moyenne nulle. Pour cela, le test des auto-corrélogrammes (ACF et PACF) et celui du Portmanteau (Box-Pierce) sont utilisés pour tester l'indépendance des résidus, le test t de Student est utilisé pour tester la nullité de leur moyenne et le test d'adéquation du Chi-deux à 95% de confiance est utilisé pour tester leur normalité.

- **Test des ACF et PACF**

L'auto-corrélogramme total et partiel des résidus du modèle ont été tracés tel que montré dans la Fig. II.2.34 et Fig. II.2.35.

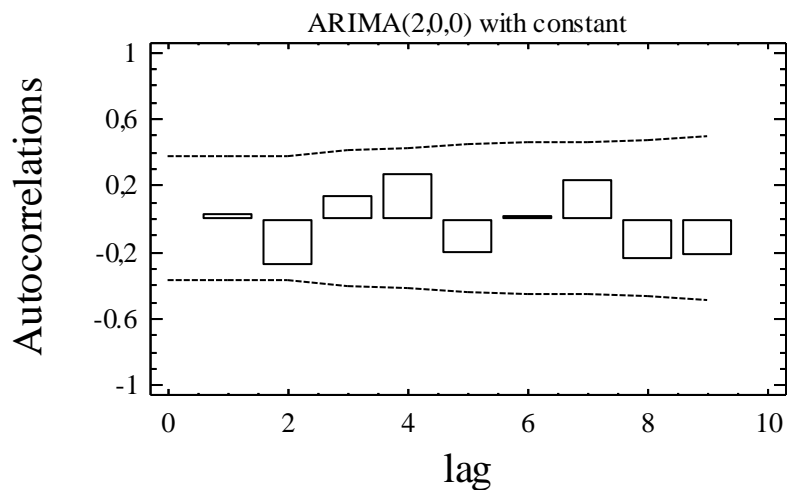


Fig. II.2.34 ACF des résidus du modèle ARIMA(2,0,0)

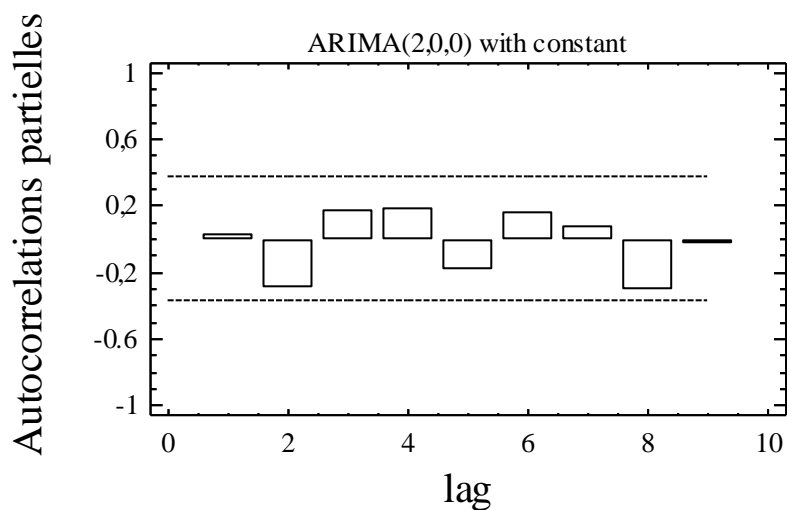


Fig. II.2.35 PACF des résidus du modèle ARIMA(2,0,0)

Tous les coefficients sont inclus dans l'intervalle de confiance (IC) tracé à 95% de confiance ce qui indique que ces derniers sont statistiquement non significatifs, d'où leur indépendance. Leurs valeurs sont reportées dans le Tableau II.2.49 et Tableau II.2.50.

Tableau II.2.49 ACF des résidus du modèle ARIMA(2,0,0)

Lag	Auto-corrélation	Ecart-type	limite inférieure IC 95%	limite supérieure IC 95%
1	0,028253	0,19245	-0,377196	0,377196
2	-0,293916	0,192604	-0,377497	0,377497
3	0,134799	0,208555	-0,408761	0,408761
4	0,262736	0,211757	-0,415037	0,415037
5	-0,220014	0,223505	-0,438063	0,438063
6	0,00678004	0,231387	-0,453512	0,453512
7	0,230025	0,231395	-0,453526	0,453526
8	-0,252761	0,239714	-0,469832	0,469832
9	-0,229074	0,24939	-0,488796	0,488796

Tableau II.2.50 PACF des résidus du modèle ARIMA(2,0,0)

Lag	Auto-corrélation partielle	Ecart-type	limite inférieure IC 95%	limite supérieure IC 95%
1	0,028253	0,19245	-0,377196	0,377196
2	-0,29495	0,19245	-0,377196	0,377196
3	0,168683	0,19245	-0,377196	0,377196
4	0,178071	0,19245	-0,377196	0,377196
5	-0,188083	0,19245	-0,377196	0,377196
6	0,157468	0,19245	-0,377196	0,377196
7	0,0705785	0,19245	-0,377196	0,377196
8	-0,311212	0,19245	-0,377196	0,377196
9	-0,0260842	0,19245	-0,377196	0,377196

- **Test du Portmanteau**

Ce test est basé sur les 9 premières auto-corrélations des résidus, les résultats du test montrent une valeur de P supérieure ou égale à 0.10, donc on ne peut pas rejeter l'hypothèse nulle que les résidus sont indépendants avec un intervalle de confiance de 90% ou plus.

Test de Box-Pierce

Hypothèse nulle : résidus indépendants

Valeur calculée de la statistique $Q = 10,587$

Valeur de P = 0,157674

- **Test de nullité de la moyenne**

La moyenne des résidus est de -1,98, leur médiane est de -2,27 et leur variance est de 307,92 avec un écart-type de 17,54. Les résultats du test t de Student montrent une valeur P= 0,56 qui est supérieure à 0.1 ce qui implique que l'hypothèse nulle de nullité de la moyenne des résidus ne peut pas être rejetée à 90% de confiance et plus.

Test t

Hypothèse nulle : moyenne = 0

Valeur calculée de la statistique $t = -0,5875$

Valeur de P = 0,56

- **Test de Normalité**

Afin de vérifier si les résidus du modèle ARIMA(2,0,0) appartiennent à une distribution Normale, un ajustement à la loi de Gauss a été effectué tel que présenté dans la Fig. II.2.36. Les résultats du test du Chi-deux montrent une valeur de P supérieure à 0,1 donc l'hypothèse nulle de Normalité ne peut pas être rejetée à 90% de confiance ou plus.

test du Chi-deux

Valeur de la statistique du test = 8,77.

Valeur de P = 0,642

Les résultats des tests précédents appliqués aux résidus du modèle ARIMA(2,0,0) ont permis la vérification des hypothèses de la méthode de Box-Jenkins.

En fin, ce modèle peut être formulé de la manière suivante :

$$(1 - 0,656932B^2)(X_t - 36,18) = \varepsilon_t + 13,22$$

Où ε_t est un bruit blanc gaussien de moyenne nulle et de variance 307,92. Ainsi la variance expliquée par le modèle peut être estimée à seulement 39,5%. Ce pourcentage qui exprime la contribution de la partie structurée au sein de la série étudiée est pratiquement inférieur à la contribution de la partie aléatoire, cette considération doit être prise en compte lors de l'utilisation du modèle pour la prédiction.

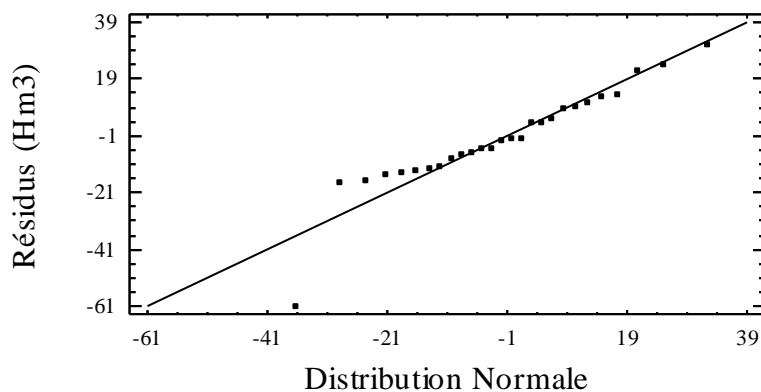


Fig. II.2.36 Ajustement des résidus du modèle ARIMA(2,0,0) à La loi de Gauss

II.2.2.6.5 Prédiction

Les résultats des prédictions obtenues par le modèle retenu ARIMA(2,0,0) sont regroupés dans le Tableau II.2.51 pour la période où les données sont disponibles avec les résidus correspondants et dans le Tableau II.2.52 pour la période suivant la fin de la série chronologique de départ. Pour cette dernière période des intervalles de confiances à 95% ont été calculés pour les prédictions obtenues. Ces limites de confiance indiquent où va être située la vraie valeur à une période déterminée dans le future, dans l'hypothèse que le modèle ajusté est adéquat aux données en question. Ces prédictions sont présentées dans la Fig. II.2.37 en trait continu suivent l'allure générale des données observées (en points carrés). Pour la période où les données ne sont pas disponibles les prédictions obtenues pour les 12 ans futurs sont présentées avec leurs limites de confiance tracées à 95%.

Tableau II.2.51 Prédictions des apports annuels à KSOB
par le modèle ARIMA(2,0,0)

Période	Donnée	Prédiction	Résidu
1	27,682	31,7656	-4,08357
2	66,527	49,5962	16,9308
3	33,197	29,9167	3,28026
4	80,707	56,1848	24,5222
5	20,59	33,2208	-12,6308
6	74,605	65,7836	8,82136
7	37,156	25,0761	12,0799
8	61,01	61,4025	-0,392489
9	28,775	36,2646	-7,48959
10	87,671	52,66	35,011
11	35,04	30,1593	4,88074
12	10,759	70,0336	-59,2746
13	18,959	36,0046	-17,0456
14	14,849	19,8692	-5,02022
15	19,082	25,3485	-6,26649
16	8,481	22,5533	-14,0723
17	29,697	25,5725	4,12449
18	5,65	18,1312	V-12,4812
19	14,466	32,6095	V-18,1435
20	25,656	16,614	V9,042
21	35,258	22,1539	V13,1041
22	15,25	29,289	V-14,039
23	17,459	36,0468	V-18,5878
24	45,423	22,8532	V22,5698
25	21,538	23,6755	V-2,13752
26	38,572	42,5831	V-4,01112
27	16,562	26,5092	V-9,9472

Notation : V : période de validation

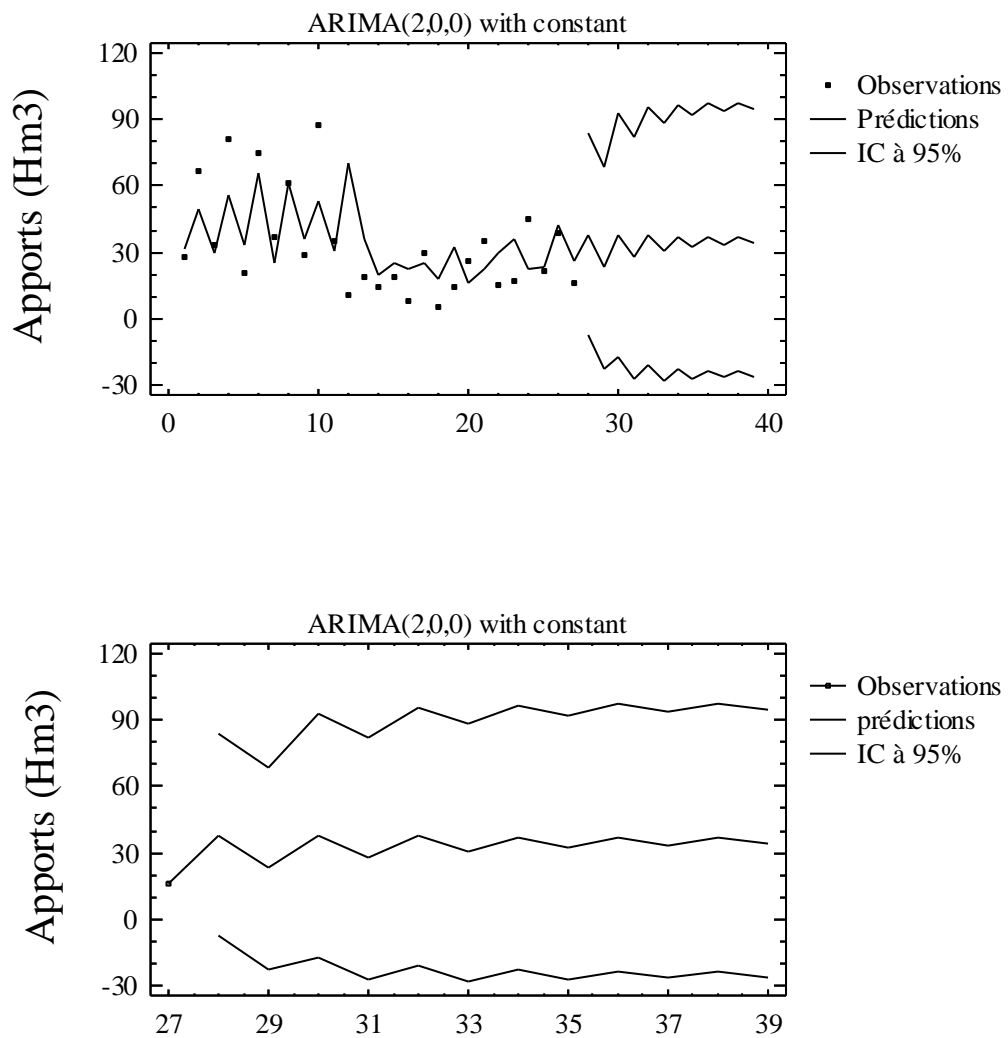


Fig. II.2.37 Prédictions des apports annuels à KSOB obtenues par le modèle ARIMA(2,0,0)

Tableau II.2.52 Prédictions des apports annuels à KSOB pour les 12 ans futurs

Période	Prédiction	- IC 95%	+IC 95%
28	38,1944	-7,36685	83,7556
29	23,2488	-22,3239	68,8216
30	37,7959	-16,7394	92,3312
31	27,6506	-26,9014	82,2025
32	37,4352	-20,5675	95,4379
33	30,5503	-27,4676	88,5682
34	37,133	-22,3184	96,5844
35	32,4621	-27,0007	91,9249
36	36,8915	-23,183	96,9659
37	33,7234	-26,3587	93,8055
38	36,7045	-23,6417	97,0506
39	34,5562	-25,7946	94,907

II.2.2.7 MEFROUCHE

Les statistiques de la série chronologique des apports annuels observés à MEFROUCHE entre 1943 et 1992 sont résumées dans le Tableau II.2.53 suivant. Il présente les mesures de la tendance centrale, de la variabilité et de la forme de la série chronologique en question. La skewness et la kurtosis standardisées sont comprises dans l'intervalle $[-2, +2]$ ce qui indique la normalité des données en question.

Tableau II.2.53 Statistiques des apports annuels observés à MEFROUCHE (1943-1992)

Paramètre	valeur
Nombre d'observations	50
Moyenne	13,4835
Variance	84,4757
Déviatoin Standard	9,19107
Minimum	0,0
Maximum	33,915
Std. skewness	1,43122
Std. kurtosis	-0,575731

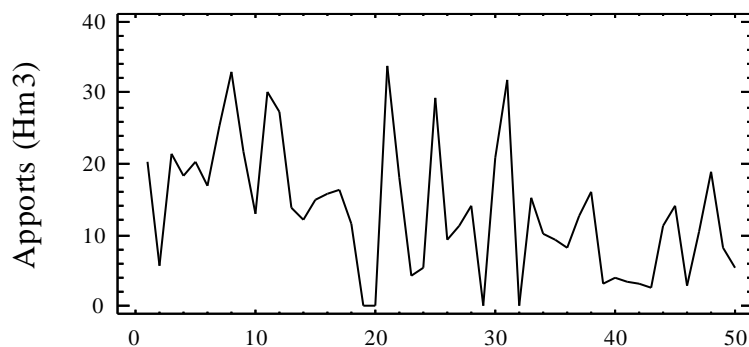


Fig. II.2.38 Apports annuels originaux observés à MEFROUCHE (1943-1992)

II.2.2.7.1 Stationnarisation

Le tracé de la série chronologique originale tel que montré par la Fig. II.2.38 semble présenter une tendance ce qui peut être à l'origine d'une non stationnarité. Le calcul des premières ou deuxièmes différences régulières de la série permet d'obtenir cette stationnarité et le calcul de l'écart-type permet de choisir l'ordre optimal de la différenciation.

Le détail du calcul des séries différenciées pour les 5 premières années est montré dans le Tableau II.2.54, tandis que les statistiques des séries différenciées pour l'ensemble des données sont représentées dans le Tableau II.2.55.

Tableau II.2.54 Calcul des séries différenciées
(Données des 5 premières années)

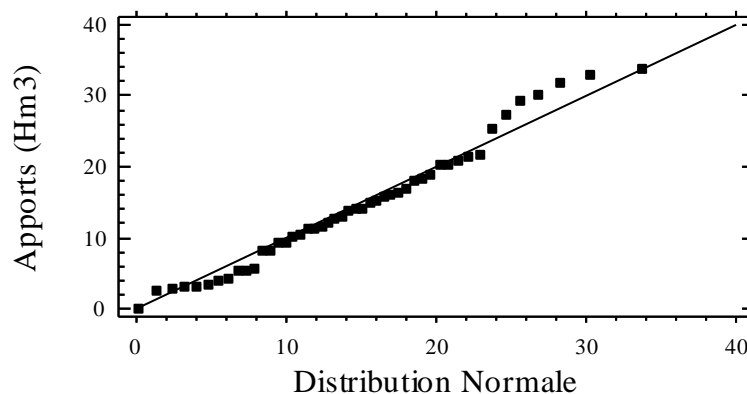
N°	Apport	DIFF(Apport,1)	DIFF(Apports,2)
1	20,176		
2	5,771	-14,405	
3	21,345	15,574	29,979
4	18,325	-3,02	-18,594
5	20,15	1,825	4,845

Tableau II.2.55 Statistiques descriptives des séries différenciées
(1943-1992)

Série	N	Moyenne	Variance	Ecart-type
Apports	50	13,48	84,47	9,19
DIFF(Apports,1)	49	-0,30	140,03	11,83
DIFF(Apports ,2)	48	0,24	379,11	19,47

Notation :DIFF (x,n) : différences régulières de x d'ordre n .

Dans le cas de la série étudiée le minimum d'écart-type est observé pour les apports originaux malgré le décalage dans le niveau de la moyenne observé dans le tracé chronologique des données. La série correspondante tel qu'illustré dans la Fig. II.2.38 est donc stable avec une variance de 84,47. La Fig. II.2.39 quant à elle montre l'ajustement de ces données à une distribution normale, dont la qualité d'adéquation est confirmée par le test du Chi-deux à plus de 90% de confiance (statistique du test = 17,68 et P = 0,2798).

**Fig. II.2.39** Ajustement Normal des apports annuels observés à
MEFROUCHE (1943-1992)

II.2.2.7.2. Identification

L'examen visuel du tracé de la fonction d'auto-corrélation (ACF) et de celui de la fonction d'auto-corrélation partielle (PACF) des apports observés, tel que montré par les Fig. II.2.40 et 41 ainsi que les Tableaux II.2.56 et 57, indique 2 pic significatifs au lag 3 et 4 pour l'ACF et 1 pic significatif au lag 3 pour la PACF.

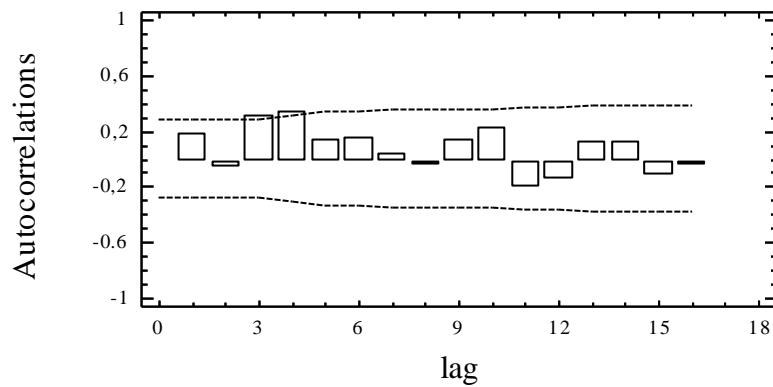


Fig. II.2.40 ACF des apports annuels observés à MEFROUCHE (1943-1992)

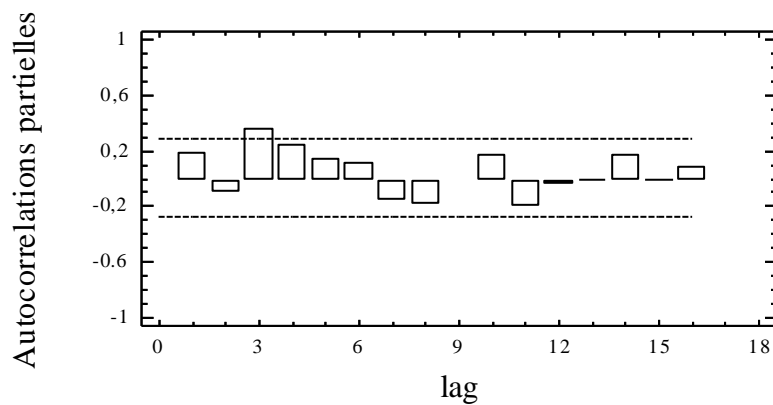


Fig. II.2.41 PACF des apports annuels observés à MEFROUCHE (1943-1992)

Tableau II.2.56 ACF des apports annuels stationnaires

Lag	ACF	Ecart-type	-IC à 95%	IC à 95%
1	0,17426	0,141421	-0,277181	0,277181
2	-0,0692425	0,145653	-0,285474	0,285474
3	0,306392	0,146309	-0,286762	0,286762
4	0,339682	0,158624	-0,310897	0,310897
5	0,143968	0,17256	-0,338211	0,338211
6	0,15202	0,174945	-0,342887	0,342887
7	0,0352232	0,177568	-0,348027	0,348027
8	-0,0401717	0,177707	-0,348301	0,348301
9	0,138477	0,177889	-0,348657	0,348657
10	0,220531	0,180032	-0,352857	0,352857
11	-0,210822	0,185356	-0,363292	0,363292
12	-0,145987	0,190091	-0,372573	0,372573
13	0,124836	0,192321	-0,376942	0,376942
14	0,126252	0,193934	-0,380105	0,380105
15	-0,119132	0,195571	-0,383314	0,383314
16	-0,0369632	0,197017	-0,386148	0,386148

Tableau II.2.57 PACF des apports annuels stationnaires

Lag	PACF	Ecart-type	-IC à 95%	IC à 95%
1	0,17426	0,141421	-0,277181	0,277181
2	-0,102729	0,141421	-0,277181	0,277181
3	0,351886	0,141421	-0,277181	0,277181
4	0,233897	0,141421	-0,277181	0,277181
5	0,143801	0,141421	-0,277181	0,277181
6	0,112096	0,141421	-0,277181	0,277181
7	-0,167003	0,141421	-0,277181	0,277181
8	-0,201788	0,141421	-0,277181	0,277181
9	-0,0162297	0,141421	-0,277181	0,277181
10	0,162294	0,141421	-0,277181	0,277181
11	-0,206934	0,141421	-0,277181	0,277181
12	-0,0397957	0,141421	-0,277181	0,277181
13	-0,0091627	0,141421	-0,277181	0,277181
14	0,167124	0,141421	-0,277181	0,277181
15	-0,0101706	0,141421	-0,277181	0,277181
16	0,0773303	0,141421	-0,277181	0,277181

Les résultats du test de Box-Pierce basé sur les 16 premières auto-corrélations montrent qu'à 90% de confiance la série n'est pas complètement aléatoire (la statistique calculée du test est de 23,58 avec une valeur de probabilité P de 0,09 qui est inférieure à 0,1).

Les comportements de l'ACF et la PACF nous poussent à opter pour les modèles possibles suivants :

- (A) ARIMA(3,0,4) avec constante
- (B) ARIMA(3,0,3) avec constante
- (C) ARIMA(2,0,3) avec constante
- (D) ARIMA(3,0,2) avec constante
- (E) ARIMA(3,0,1) avec constante

II.2.2.7.3 Estimation

Pour l'estimation des paramètres de ces modèles, il faut rappeler que 50 données ont été utilisées dont les 20 dernières ont servi à la validation. Les paramètres des modèles ont alors été estimés sur une période de 30 ans.

L'analyse et la comparaison des modèles ajustés tel que présenté dans le Tableau II.2.58 montre que tous les modèles ajustés sont plus ou moins performants en produisant des erreurs quadratiques moyennes pas très différentes entre la période de validation que celle d'estimation. D'un autre côté les résidus produits par tous les modèles sauf (C) et (D) ne sont pas indépendants ce qui les exclue de l'analyse. Le modèle (C) qui produit le minimum de variance des résidus observe une instabilité au niveau de la moyenne, alors que le (D) est instable au niveau de la variance et présente en plus des paramètres non significativement différents de zéro, il est donc également écarté ce qui laisse la place au modèle (C).

Tableau II.2.58 Comparaison des modèles ajustés

Modèle	Variance des résidus		AUTO	MOY	VAR
	Période d'estimation	Période de validation			
(A)	190,365	104,719	NON	OK	OK
(B)	138,261	74,1016	NON	OK	OK
(C)	73,4124	88,74	OK	NON	OK
(D)	77,3815	92,1671	OK	OK	NON
(E)	249,38	85,85	NON	OK	OK

Notation : AUTO :test d'indépendance des résidus ; MOY : test d'égalité de la moyenne ; VAR : test d'homoscédasticité ; OK : le test est positif.

Le Tableaux II.2.59 résume l'estimation des paramètres du modèle retenu ARIMA(2,0,3) avec constante ainsi que la signification statistique de ses termes. Les termes correspondant à une valeur de $P < 0.05$ sont statistiquement significativement différents de zéro à 95% de confiance.

Tableau II.2.59 Estimation des paramètres du modèle retenu ARIMA(2,0,3) avec constante

Paramètre	Estimation	Ecart- type	t	P
AR(1)	0,516727	0,26333	1,96228	0,056074
AR(2)	-0,586194	0,272258	-2,15309	0,036834
MA(1)	0,510188	0,247381	2,06235	0,045109
MA(2)	-0,317108	0,284567	-1,11435	0,271178
MA(3)	-0,54414	0,192894	-2,82093	0,000000
Moyenne	16,1042	1,91189	8,42314	0,000000
Constante	17,2229			

La variance estimée du bruit blanc = 71,6702 avec 44 DDL et l'écart- type estimé du bruit blanc = 8,46582.

La valeur de P pour tous les termes, sauf MA(2), est inférieure à 0.05 ce qui veut dire que ces derniers sont significativement différents de zéro à 95% de confiance avec des valeurs différentes de un (1). Pour MA(2) la valeur de P est très supérieure à 0.05 donc elle est statistiquement non significativement différente de zéro et peut donc être éliminée. L'estimation de l'écart-type du bruit blanc est de 8,46582 et l'estimation de la variance des résidus est de 71,6702.

Le Tableau II.2.60 résume la performance du modèle ARIMA(2,0,3) à ajuster les données considérées et ce par l'analyse des erreurs qu'il produit. La différence de ces statistiques d'erreurs n'est pas trop importante entre la période d'estimation et celle de validation. Ces différences expriment une performance moyenne et le modèle ajusté peut être considéré comme moyennement performant pour faire des prédictions.

Tableau II.2.60 Statistiques des erreurs du modèle ARIMA(2,0,3)

Statistique	Période d'estimation	Période de validation
EQM	73,4124	88,74
EAM	5,97294	8,38489
EM	0,150209	-5,55289

Notation : EQM : erreur quadratique moyenne ; EAM : erreur absolue moyenne ; EM:erreur moyenne.

II.2.2.7.4 Validation

Il s'agit de vérifier l'hypothèse de la méthode à savoir que les résidus forment bien un bruit blanc gaussien de moyenne nulle. Pour cela, le test des auto-corrélogrammes (ACF et PACF) et celui du Portmanteau (Box-Pierce) sont utilisés pour tester l'indépendance des résidus, le test t de student est utilisé pour tester la nullité de leur moyenne et le test d'adéquation du Chi-deux à 95% de confiance est utilisé pour tester leur normalité.

- **Test des ACF et PACF**

L'auto-corrélogramme total et partiel des résidus du modèle ont été tracés tel que montré dans la Fig. II.2.42 et Fig. 43.

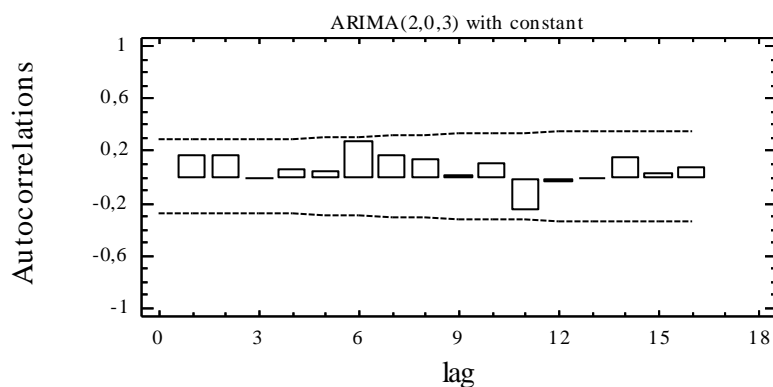


Fig. II.2.42 ACF des résidus du modèle ARIMA(2,0,3)

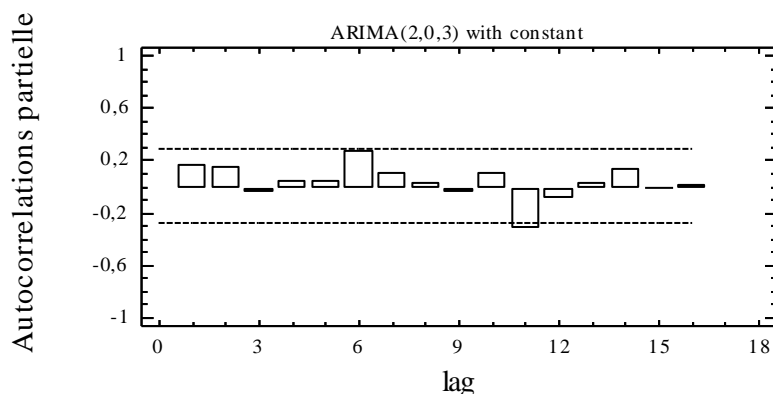


Fig. II.2.43 PACF des résidus du modèle ARIMA(2,0,3)

Pratiquement tous les coefficients sont inclus dans l'intervalle de confiance (IC) tracé à 95% de confiance ce qui indique que ces derniers sont statistiquement non significatifs, d'où leur indépendance. Leurs valeurs sont reportées dans le Tableau II.2.61 et Tableau II.2.62.

Tableau II.2.61 ACF des résidus du modèle ARIMA(2,0,3)

Lag	Auto-corrélation	Ecart-type	limite inférieure IC 95%	limite supérieure IC 95%
1	0,1564	0,141421	-0,277181	0,277181
2	0,163146	0,144839	-0,28388	0,28388
3	-0,00568316	0,148469	-0,290995	0,290995
4	0,056698	0,148474	-0,291003	0,291003
5	0,0404343	0,148906	-0,291851	0,291851
6	0,275513	0,149125	-0,292281	0,292281
7	0,166691	0,15898	-0,311596	0,311596
8	0,130599	0,162438	-0,318373	0,318373
9	0,00352831	0,164525	-0,322463	0,322463
10	0,104176	0,164526	-0,322466	0,322466
11	-0,262587	0,16584	-0,325041	0,325041
12	-0,0515043	0,173957	-0,34095	0,34095
13	-0,00512875	0,174262	-0,341547	0,341547
14	0,150171	0,174265	-0,341553	0,341553
15	0,0282869	0,176834	-0,346589	0,346589
16	0,0673645	0,176924	-0,346766	0,346766

Tableau II.2.62 PACF des résidus du modèle ARIMA(2,0,3)

Lag	Auto-corrélation partielle	Ecart-type	limite inférieure IC 95%	limite supérieure IC 95%
1	0,1564	0,141421	-0,277181	0,277181
2	0,142162	0,141421	-0,277181	0,277181
3	-0,0521077	0,141421	-0,277181	0,277181
4	0,043346	0,141421	-0,277181	0,277181
5	0,0376412	0,141421	-0,277181	0,277181
6	0,262198	0,141421	-0,277181	0,277181
7	0,09762	0,141421	-0,277181	0,277181
8	0,0242735	0,141421	-0,277181	0,277181
9	-0,0412799	0,141421	-0,277181	0,277181
10	0,0920112	0,141421	-0,277181	0,277181
11	-0,327827	0,141421	-0,277181	0,277181
12	-0,103453	0,141421	-0,277181	0,277181
13	0,0300697	0,141421	-0,277181	0,277181
14	0,126543	0,141421	-0,277181	0,277181
15	-0,00954747	0,141421	-0,277181	0,277181
16	0,0069371	0,141421	-0,277181	0,277181

- **Test du Portmanteau**

Ce test est basé sur les 16 premières auto-corrélations des résidus, les résultats du test montrent une valeur de P supérieure ou égale à 0.10, donc on ne peut pas rejeter l'hypothèse nulle que les résidus sont indépendants avec un intervalle de confiance de 90% ou plus.

Test de Box-Pierce

Hypothèse nulle : résidus indépendants

Valeur calculée de la statistique $Q = 14,35$

Valeur de $P = 0.2139$

- **Test de nullité de la moyenne**

La moyenne des résidus est de -1,48, leur médiane est de -1,30 et leur variance est de 73,28 avec un écart-type de 8,56. Les résultats du test t de Student montrent une valeur $P = 0,22$ qui est supérieure à 0.05 ce qui implique que l'hypothèse nulle de nullité de la moyenne des résidus ne peut pas être rejetée à 95% de confiance.

Test t

Hypothèse nulle : moyenne = 0

Valeur calculée de la statistique $t = -1,22$

Valeur de $P = 0.22$

- **Test de Normalité**

Afin de vérifier si les résidus du modèle ARIMA(2,0,3) appartiennent à une distribution Normale, un ajustement à la loi de Gauss a été effectué tel que présenté dans la Fig. II.2.44. Les résultats du test du Chi-deux montrent une valeur de P supérieure à 0,1 donc l'hypothèse nulle de Normalité ne peut pas être rejetée à 90% de confiance ou plus.

test du Chi-deux

Valeur de la statistique du test = 6,88.

Valeur de $P = 0,96$

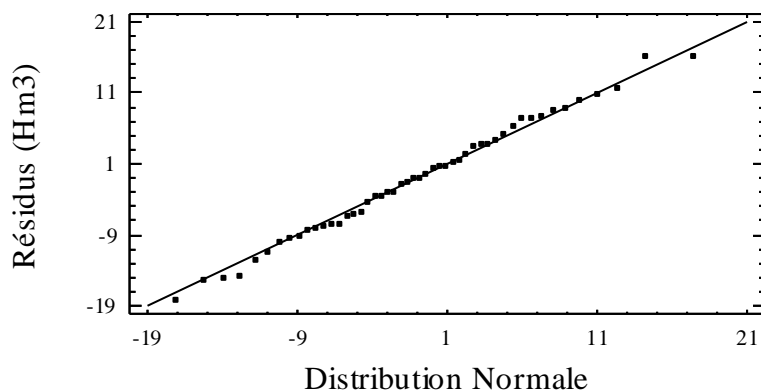


Fig. II.2.44 Ajustement des résidus du modèle ARIMA(2,0,3) à La loi de Gauss

Les résultats des tests précédents appliqués aux résidus du modèle ARIMA(2,0,3) ont permis la vérification des hypothèses de la méthode de Box-Jenkins.

En fin, ce modèle peut être formulé de la manière suivante :

$$(1 - 0,516727B + 0,586194B^2)(X_t - 16,1042) \\ = (1 - 0,510188B + 0,54414B^3)\varepsilon_t + 17,2229$$

Où ε_t est un bruit blanc gaussien de moyenne nulle et de variance 73,28. Ainsi la variance expliquée par le modèle peut être estimée à seulement 13,24 %. Ce pourcentage qui exprime la contribution de la partie structurée au sein de la série étudiée représente presque le un sixième de la variabilité totale ce qui fait que la contribution de la partie aléatoire est six fois plus importante. Cette considération doit être prise en compte lors de l'utilisation du modèle à la prédiction.

II.2.2.7.5 Prédiction

Les résultats des prédictions obtenues par le modèle retenu ARIMA(2,0,3) sont regroupés dans le Tableau II.2.63 pour la période où les données sont disponibles avec les résidus correspondants et dans le Tableau II.2.64 pour la période suivant la fin de la série chronologique de départ. Pour cette dernière période des intervalles de confiance à 95% ont été calculés pour les prédictions obtenues. Ces limites de confiance indiquent où va être située la vraie valeur à une période déterminée dans le future, dans l'hypothèse que le modèle ajusté est adéquat aux données en question. On ne peut pas dire que les prédictions qui sont représentées dans la Fig. II.2.45 en trait continu suivent l'allure générale des données observées, cependant elles peuvent être considérées comme une trajectoire possible pour ces dernières. Pour la période où les données ne sont pas disponibles, les prédictions obtenues pour les 12 ans futurs sont représentées avec leurs limites de confiance tracées à 95% .

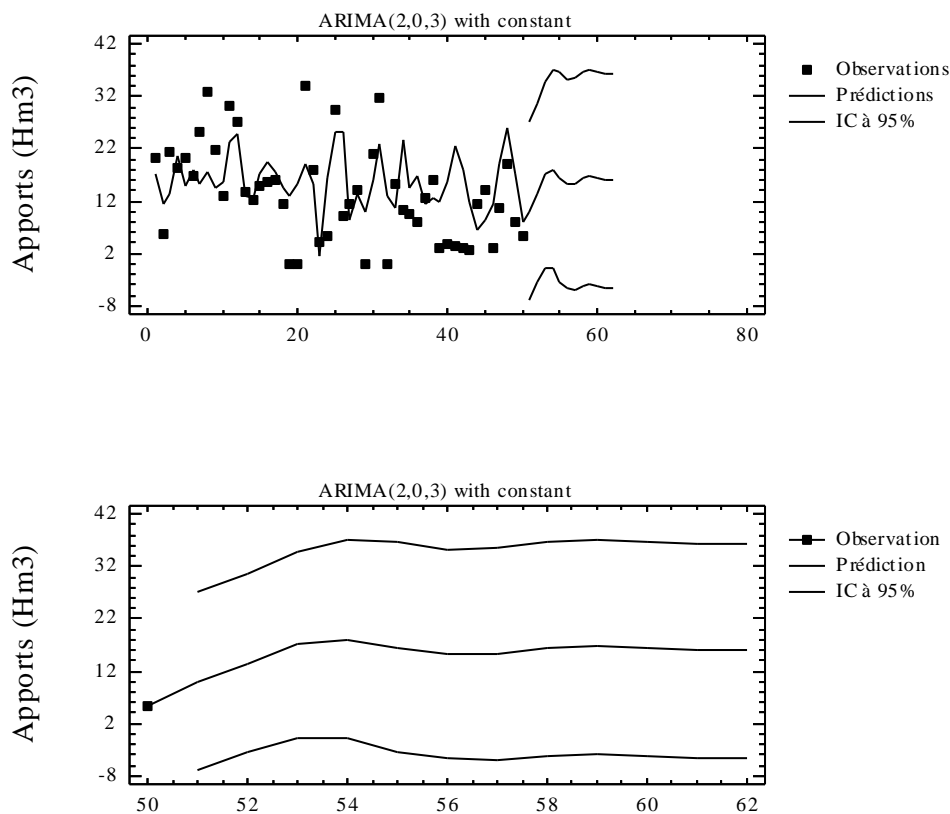


Fig. II.2.45 Prédiction des apports annuels à MEFROUCHE obtenues par le modèle ARIMA(2,0,3)

Tableau II.2.63 Prédications avec ARIMA(2,0,3)

Période	Donnée	Prédiction	Résidu
1	20,176	17,2207	2,95529
2	5,771	11,384	-5,61304
3	21,345	13,275	8,07001
4	18,325	20,5804	-2,2554
5	20,15	14,835	5,31497
6	16,928	17,8573	-0,929271
7	25,389	15,0905	10,2985
8	32,86	17,7622	15,0978
9	21,639	14,377	7,26203
10	12,948	15,8284	-2,88045
11	30,118	23,2165	6,90146
12	27,236	24,7127	2,5233
13	13,84	12,9752	0,864783
14	12,077	12,5231	-0,44609
15	14,806	17,2253	-2,4193
16	15,665	19,3575	-3,69246
17	16,219	17,5121	-1,29314
18	11,424	14,5933	-3,16933
19	0,0	12,8161	-12,8161
20	0,0	15,3562	-15,3562
21	33,915	19,2688	14,6462
22	17,923	15,432	2,49103
23	4,115	1,62102	2,49398
24	5,443	16,33	-10,887
25	29,347	25,324	4,02304
26	9,283	25,0488	-15,7658
27	11,398	8,21184	3,18616
28	14,096	13,235	0,861034
29	0,0	9,81747	-9,81747
30	20,833	15,9754	4,8576
31	31,843	22,8649	V8,97813
32	0,0	13,0826	V-13,0826
33	15,14	10,7215	V4,41848
34	10,251	23,5286	V-13,2776
35	9,437	14,7013	V-5,26433
36	8,053	16,9698	V-8,91677
37	12,567	11,5071	V1,05985
38	16,038	12,7631	V3,27488
39	2,994	11,9567	V-8,96274
40	3,962	15,5565	V-11,5945
41	3,5	22,3703	V-18,8703
42	3,199	17,7826	V-14,5836
43	2,619	11,9717	V-9,35269
44	11,335	6,57993	V4,75507
45	14,051	8,2174	V5,8336
46	2,876	11,2814	V-8,40535
47	10,5	19,198	V-8,69798
48	19,012	25,9091	V-6,89711
49	8,116	17,0788	V-8,96278
50	5,415	7,92456	V-2,50956

Notation : V : période de validation

Tableau II.2.64 Prédications des apports annuels à MEFROUCHE pour les 12 ans futurs

Période	Prédiction	- IC 95%	+IC 95%
51	9,94858	-7,1132	27,0104
52	13,5165	-3,54562	30,5787
53	17,0099	-0,644283	34,664
54	18,089	-0,856938	37,035
55	16,5989	-3,33904	36,5368
56	15,1963	-4,75855	35,1511
57	15,345	-5,01964	35,7097
58	16,2441	-4,18482	36,673
59	16,6215	-3,8659	37,1088
60	16,2894	-4,27238	36,8512
61	15,8966	-4,66515	36,4584
62	15,8883	-4,69908	36,4757

II.2.2.8 MIREBECK

Le Tableau II.2.65 suivant résume les statistiques de la série chronologique des apports annuels observés à MIREBECK entre 1968 et 1993. Il présente les mesures de la tendance centrale, de la variabilité et de la forme de la série chronologique en question. La skewness et la kurtosis standardisées sont comprises dans l'intervalle $[-2, +2]$ ce qui indique que la distribution des données est très probablement normale.

Tableau II.2.65 Statistiques des apports annuels observés à MIREBECK (1968-1993)

Paramètre	valeur
Nombre d'observations	27
Moyenne	328,868
Variance	72451,0
Déviation Standard	269,167
Minimum	37,81
Maximum	1057,09
Std. skewness	2,73628
Std. kurtosis	1,21306

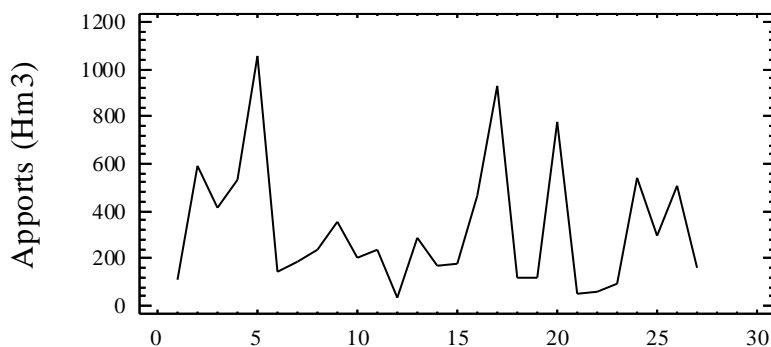


Fig. II.2.46 Apports annuels originaux observés à MIREBECK (1968-1993)

II.2.2.8.1 Stationnarisation

Le tracé de la série chronologique originale tel que montré par la Fig. II.2.46 ne semble présenter aucun signe de non stationnarité. Pour le confirmer, nous avons procédé par le calcul des premières et deuxièmes différences régulières de la série où le calcul de l'écart-type permet de choisir l'ordre optimal de la différenciation en cas de non stationnarité.

Le détail du calcul des séries différenciées pour les 5 premières années est montré dans le Tableau II.2.66, tandis que les statistiques des séries différenciées pour l'ensemble des données sont représentées dans le Tableau II.2.67.

Tableau II.2.66 Calcul des séries différenciées
(Données des 5 premières années)

N°	Apport	DIFF(Apport,1)	DIFF(Apports,2)
1	112,77		
2	592,3	479,53	
3	413,37	-178,93	-658,46
4	533,09	119,72	298,65
5	1057,09	524	404,28

Tableau II.2.67 Statistiques descriptives des séries différenciées
(1968-1993)

Série	N	Moyenne	Variance	Ecart-type
Apports	27	328,86	7251,0	269,16
DIFF(Apports,1)	26	1,98	154567	393,4
DIFF(Apports ,2)	25	-32,86	441263	664,25

Notation : DIFF (x,n) : différences régulières de x d'ordre n .

Le minimum d'écart-type est observé pour la série des apports d'origine ce qui confirme la thèse de sa stationnarité et rejette le recours à la différenciation. La série correspondante tel qu'illustré dans la Fig. II.2.46 est donc stationnaire et sa variance qui est de 7251,0 est stable autour de sa moyenne qui est de 328,86. La Fig. II.2.47 quant à elle montre l'ajustement des données à une distribution normale dont la qualité d'adéquation est confirmée par le test du Chi-deux (statistique du test = 15, P = 0,1824).

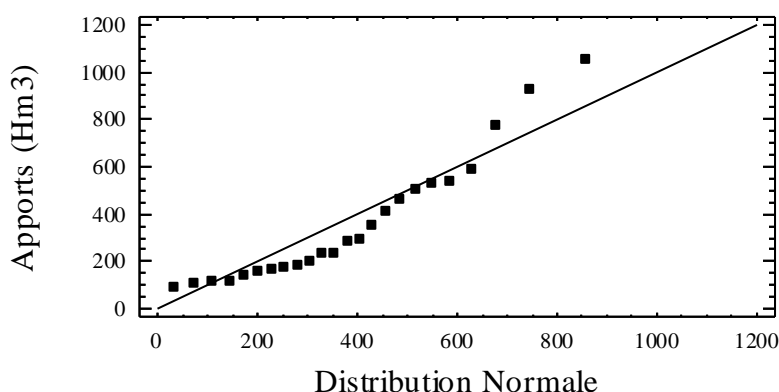


Fig. II.2.47 Ajustement à la loi Normale des apports annuels observés à MIREBECK (1968-1993)

II.2.2.8.2. Identification

L'examen visuel du tracé de l'ACF et de celui de la PACF des apports considérés, tel que montré

par les Fig. II.2.48 et 49 ainsi que les Tableaux II.2.68 et 69, n'indique aucun pic significatif aussi bien pour l'ACF que pour la PACF.

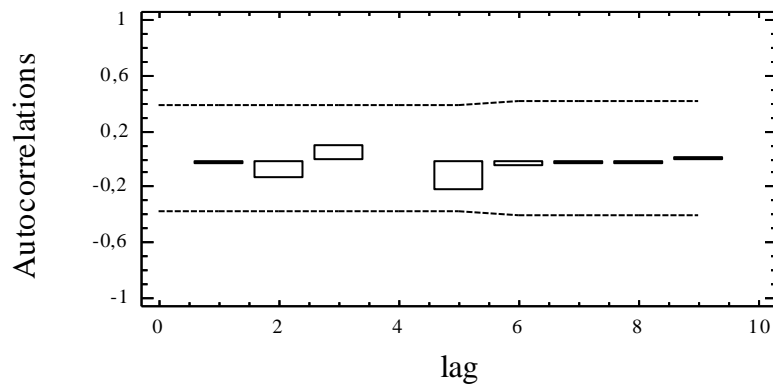


Fig. II.2.48 ACF des apports annuels observés à MIREBECK (1968-1993)

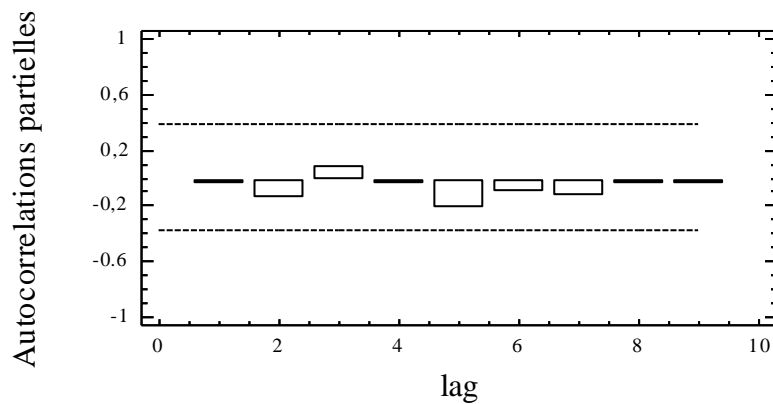


Fig. II.2.49 PACF des apports annuels observés à MIREBECK (1968-1993)

Tableau II.2.68 ACF des apports annuels observés à MIREBECK (1968-1993)

Lag	ACF	Ecart-type	-IC à 95%	IC à 95%
1	-0,0466006	0,19245	-0,377196	0,377196
2	-0,15209	0,192868	-0,378014	0,378014
3	0,090798	0,19726	-0,386622	0,386622
4	-0,015057	0,198801	-0,389644	0,389644
5	-0,233847	0,198844	-0,389727	0,389727
6	-0,0683926	0,208781	-0,409204	0,409204
7	-0,0549951	0,209609	-0,410827	0,410827
8	-0,0407721	0,210143	-0,411873	0,411873
9	0,0111827	0,210436	-0,412447	0,412447

Tableau II.2.69 PACF des apports annuels observés à MIREBECK (1968-1993)

Lag	PACF	Ecart-type	-IC à 95%	IC à 95%
1	-0,0466006	0,19245	-0,377196	0,377196
2	-0,154597	0,19245	-0,377196	0,377196
3	0,0774251	0,19245	-0,377196	0,377196
4	-0,031534	0,19245	-0,377196	0,377196
5	-0,21738	0,19245	-0,377196	0,377196
6	-0,110009	0,19245	-0,377196	0,377196
7	-0,13893	0,19245	-0,377196	0,377196
8	-0,0577831	0,19245	-0,377196	0,377196
9	-0,0329505	0,19245	-0,377196	0,377196

Les résultats du test de Box-Pierce basé sur les 9 premières auto-corrélations montrent une valeur de P supérieure ou égale à 0,1 ce qui veut dire que l'hypothèse nulle d'indépendance ne peut pas être rejetée à 90% de confiance ou plus (la statistique calculée du test est de 2,64 avec une valeur de probabilité P de 0,97). La série ne présentant aucune structure de dépendance interne peut donc être considérée comme complètement aléatoire et son traitement s'arrête ici.

II.2.2.9 PIERRE DU CHAT

Les statistiques de la série chronologique des apports annuels observés à PIERRE DU CHAT entre 1952 et 1993 sont résumées dans le Tableau II.2.70 suivant. Il présente les mesures de la tendance centrale, de la variabilité et de la forme de la série chronologique en question. La skewness et la kurtosis standardisées sont comprises dans l'intervalle $[-2, +2]$ ce qui indique la normalité des données en question.

Tableau II.2.70 Statistiques des apports annuels observés à PIERRE DU CHAT (1952-1993)

Paramètre	valeur
Nombre d'observations	42
Moyenne	188,234
Variance	19487,2
Déviation Standard	139,597
Minimum	20,9348
Maximum	564,56
Std. skewness	2,8471
Std. kurtosis	0,723881

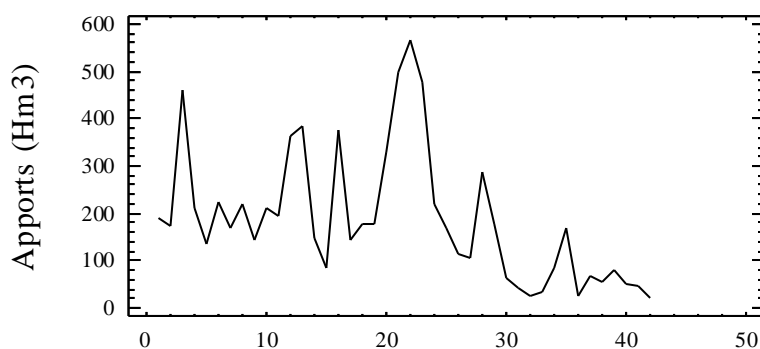


Fig. II.2.50 Apports annuels originaux observés à PIERRE DU CHAT (1952-1993)

II.2.2.9.1 Stationnarisation

Le tracé de la série chronologique originale tel que montré par la Fig. II.2.50 semble présenter une tendance ce qui peut être à l'origine d'une non stationnarité. Le calcul des premières ou deuxièmes différences régulières de la série permet d'obtenir cette stationnarité et le calcul de l'écart-type permet de choisir l'ordre optimal de la différenciation.

Le détail du calcul des séries différenciées pour les 5 premières années est montré dans le Tableau II.2.71, tandis que les statistiques des séries différenciées pour l'ensemble des données sont représentées dans le Tableau II.2.72.

Tableau II.2.71 Calcul des séries différenciées
(Données des 5 premières années)

N°	Apport	DIFF(Apport,1)	DIFF(Apports,2)
	190,88		
1	174,55	-16,33	
2	458,8	284,25	300,58
3	211,68	-247,12	-531,37
4	136,29	-75,39	171,73
5			

Tableau II.2.72 Statistiques descriptives des séries différenciées
(1952-1993)

Série	N	Moyenne	Variance	Ecart-type
Apports	42	188,23	19487,2	139,59
DIFF(Apports,1)	41	-4,14	15644,7	125,07
DIFF(Apports ,2)	40	-0,25	37893,5	194,66

Notation : DIFF (x,n) : différences régulières de x d'ordre n .

Dans le cas de la série étudiée le minimum d'écart-type est observé pour les premières différences régulières des apports. La série correspondante tel qu'illustré dans la Fig. II.2.51 est donc plus stable avec une variance de 15644,7. La Fig. II.2.52 quant à elle montre l'ajustement de ces données à une distribution normale, dont la qualité d'adéquation est confirmée par le test du Chi-deux à plus de 90% de confiance (statistique du test = 16,63 et P = 0,2761).

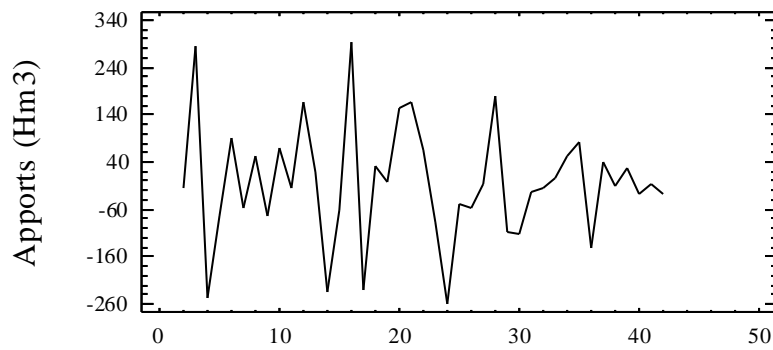


Fig. II.2.51 Premières différences régulières des apports annuels observés à PIERRE DU CHAT (1952-1993)

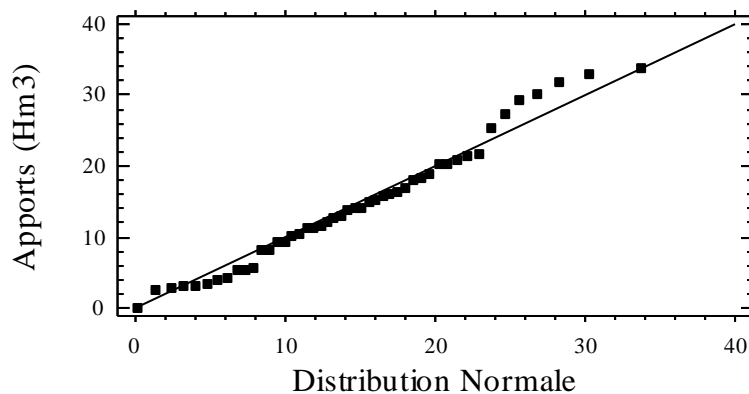


Fig. II.2.52 Ajustement Normal des premières différences régulières des apports annuels observés à PIERRE DU CHAT (1952-1993)

II.2.2.9.2. Identification

L'examen visuel du tracé de la fonction d'auto-corrélation (ACF) et de celui de la fonction d'auto-corrélation partielle (PACF) des premières différences régulières des apports observés, tel que montré par les Fig. II.2.53 et 54 ainsi que les Tableaux II.2.73 et 74, n'indique aucun pic significatif pour l'ACF et 1 pic pas très significatif au lag 2 pour la PACF.

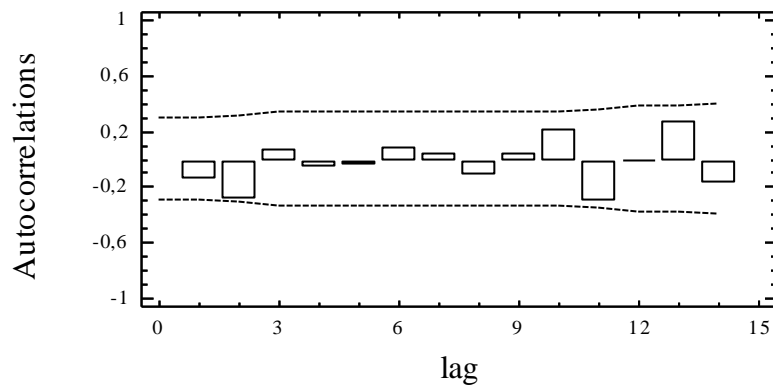


Fig. II.2.53 ACF des premières différences régulières des apports annuels observés à PIERRE DU CHAT (1952-1993)

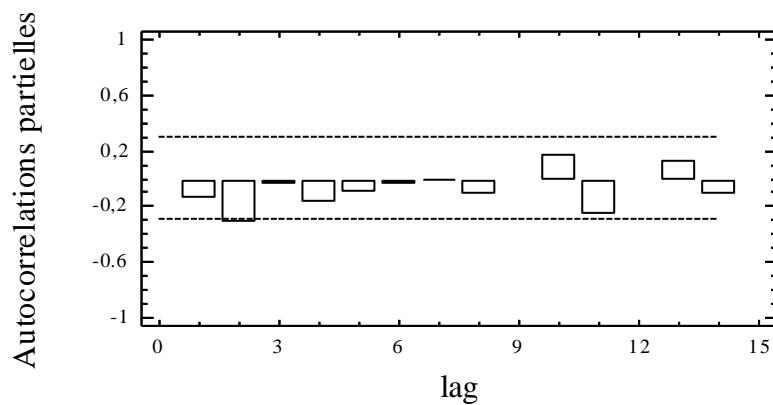


Fig. II.2.54 PACF des premières différences régulières des apports annuels observés à PIERRE DU CHAT (1952-1993)

Tableau II.2.73 ACF des apports annuels stationnaires

Lag	ACF	Ecart-type	-IC à 95%	IC à 95%
1	-0,151863	0,154303	-0,30243	0,30243
2	-0,293287	0,157822	-0,309326	0,309326
3	0,0591316	0,170305	-0,333792	0,333792
4	-0,0663821	0,170793	-0,334749	0,334749
5	-0,043026	0,171406	-0,335951	0,335951
6	0,0807979	0,171663	-0,336454	0,336454
7	0,0306429	0,172566	-0,338225	0,338225
8	-0,116485	0,172696	-0,338478	0,338478
9	0,0334204	0,174557	-0,342125	0,342125
10	0,203701	0,174709	-0,342424	0,342424
11	-0,30822	0,180275	-0,353333	0,353333
12	-0,00517712	0,192413	-0,377124	0,377124
13	0,269751	0,192417	-0,37713	0,37713
14	-0,176797	0,201219	-0,394383	0,394383

Tableau II.2.74 PACF des apports annuels stationnaires

Lag	PACF	Ecart-type	-IC à 95%	IC à 95%
1	-0,151863	0,154303	-0,30243	0,30243
2	-0,323817	0,154303	-0,30243	0,30243
3	-0,056039	0,154303	-0,30243	0,30243
4	-0,183776	0,154303	-0,30243	0,30243
5	-0,107992	0,154303	-0,30243	0,30243
6	-0,0315806	0,154303	-0,30243	0,30243
7	-0,00234289	0,154303	-0,30243	0,30243
8	-0,119156	0,154303	-0,30243	0,30243
9	-0,0156938	0,154303	-0,30243	0,30243
10	0,171913	0,154303	-0,30243	0,30243
11	-0,264046	0,154303	-0,30243	0,30243
12	-0,0218937	0,154303	-0,30243	0,30243
13	0,118765	0,154303	-0,30243	0,30243
14	-0,119672	0,154303	-0,30243	0,30243

Les résultats du test de Box-Pierce basé sur les 14 premières auto-corrélations montrent qu'à 90% de confiance et plus l'on ne peut pas rejeter l'hypothèse que la série des premières différences régulières est complètement aléatoire et qu'elle ne présente pas de structure de corrélation interne (la statistique calculée du test est de 16,02 avec une valeur de probabilité P de 0,311 qui est supérieure à 0,1), par conséquent son traitement s'arrête ici.

II.2.2.10 REMCHI

Le Tableau II.2.75 suivant résume les statistiques de la série chronologique des apports annuels observés à REMCHI entre 1948 et 1993. Il présente les mesures de la tendance centrale, de la variabilité et de la forme de la série chronologique en question. La skewness standardisées sort légèrement de l'intervalle $[-2, +2]$, cependant la kurtosis standardisées y est comprise, chose qui indique une forte probabilité de la normalité des données en question.

Tableau II.2.75 Statistiques des apports annuels observés à REMCHI (1948-1993)

Paramètre	valeur
Nombre d'observations	43
Moyenne	95,7662
Variance	5941,34
Déviatoin Standard	77,0801
Minimum	8,5094
Maximum	281,8
Std. skewness	2,55414
Std. kurtosis	-0,113263

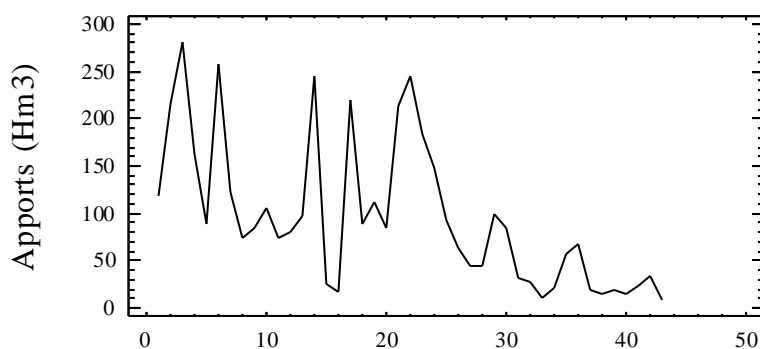


Fig. II.2.55 Apports annuels originaux observés à REMCHI (1948-1993)

II.2.2.10.1 Stationnarisation

Le tracé de la série chronologique originale tel que montré par la Fig. II.2.55 présente une tendance vers sa fin ce qui peut indiquer l'existence d'une non stationnarité au sein de la série de données. Le calcul des premières et deuxièmes différences régulières de la série permet d'obtenir la stationnarité requise et le calcul de l'écart-type permet de choisir l'ordre optimal de la différenciation.

Le détail du calcul des séries différenciées pour les 5 premières années est montré dans le Tableau II.2.76, tandis que les statistiques des séries différenciées pour l'ensemble des données sont présentées dans le Tableau II.2.77.

Tableau II.2.76 Calcul des séries différenciées (Données des 5 premières années)

N°	Apport	DIFF(Apport,1)	DIFF(Apports,2)
1	118,35		
2	214,89	96,54	
3	281,8	66,91	-29,63
4	163,61	-118,19	-185,1
5	88,92	-74,69	43,5

Tableau II.2.77 Statistiques descriptives des séries différenciées (1948-1993)

Série	N	Moyenne	Variance	Ecart-type
Apports	43	95,76	5941,00	77,08
DIFF(Apports,1)	42	-2,61	6146,72	78,4
DIFF(Apports ,2)	41	-2,95	16238	127,43

Notation : DIFF (x,n) : différences régulières de x d'ordre n .

Dans le cas de la série étudiée le minimum d'écart-type est observé pour la série des apports originaux. La série correspondante n'a donc pas besoin d'être différenciée et sa variance de 5941 est stable autour de sa moyenne qui est de 95,76. La Fig. II.2.56 quant à elle montre l'ajustement des données à une distribution normale dont la qualité d'adéquation est rejetée par le test du Chi-deux (statistique du test = 35,67 et $P = 0,0011$). Une transformation mathématique pour des fins de normalisation s'impose : la racine carrée. La Fig. II.2.57 montre l'ajustement de la nouvelle série à la loi Normale. Les résultats du test du Chi-deux confirment l'adéquation de cet ajustement (avec une Probabilité $P = 0,1346$ et une valeur de la statistique du test de 19,86).

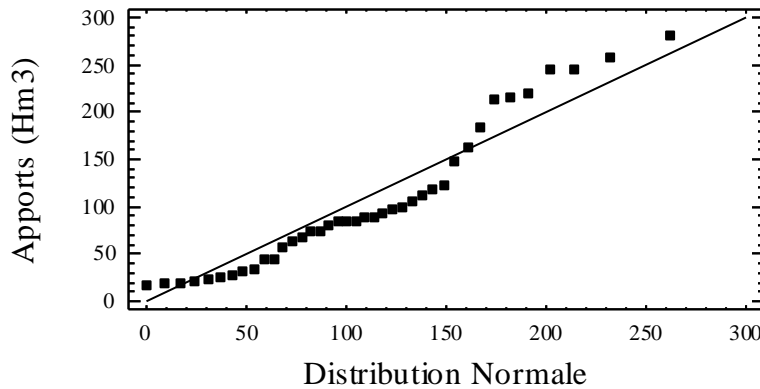


Fig. II.2.56 Ajustement Normal des apports annuels observés à REMCHI (1948-1993)

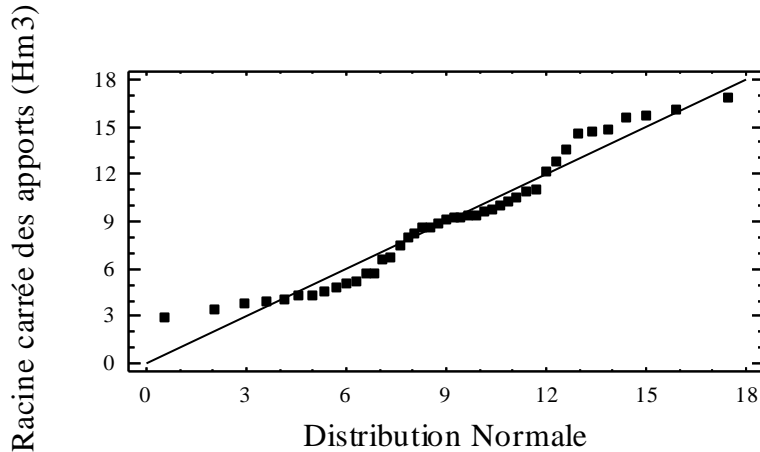


Fig. II.2.57 Ajustement Normal de la racine carrée des apports annuels observés à REMCHI (1948-1993)

II.2.2.10.2. Identification

L'examen visuel du tracé de la fonction d'auto-corrélation (ACF) et de celui de la fonction d'auto-corrélation partielle (PACF) de la racine carrée des apports observés, tel que montré par les Fig. II.2.58 et 59 ainsi que les Tableaux II.2.78 et 79, indique 1 pic significatif au lag 1 et 3 pour l'ACF et 1 pic significatif au lag 1 pour la PACF.

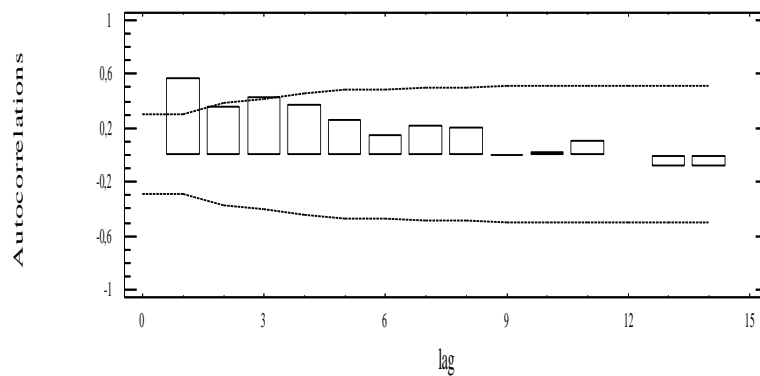


Fig. II.2.58 ACF de la racine carrée des apports annuels observés à REMCHI (1952-1993)

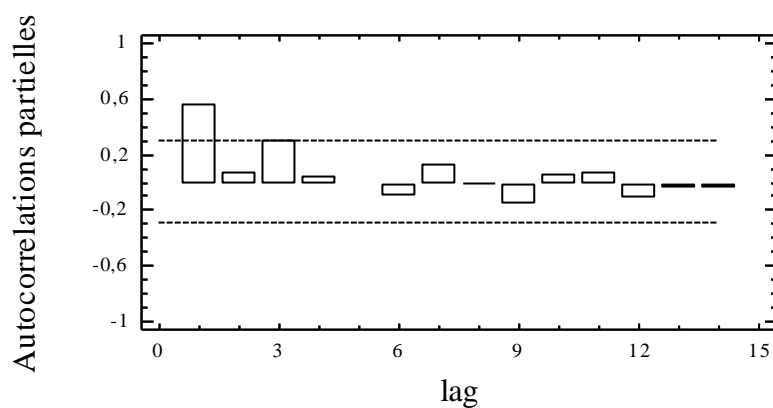


Fig. II.2.59 PACF de la racine carrée des apports annuels observés à REMCHI (1952-1993)

Tableau II.2.78 ACF de la racine carrée des apports annuels

Lag	ACF	Ecart-type	-IC à 95%	IC à 95%
1	0,556917	0,152499	-0,298892	0,298892
2	0,352797	0,194118	-0,380465	0,380465
3	0,420569	0,208497	-0,408646	0,408646
4	0,370007	0,227371	-0,445564	0,445564
5	0,253063	0,240968	-0,472289	0,472289
6	0,144719	0,247071	-0,484251	0,484251
7	0,205118	0,249034	-0,488099	0,488099
8	0,191077	0,252933	-0,49574	0,49574
9	0,00510381	0,256268	-0,502277	0,502277
10	0,0200768	0,25627	-0,502281	0,502281
11	0,103904	0,256307	-0,502353	0,502353
12	-0,0136262	0,257284	-0,504269	0,504269
13	-0,0981842	0,257301	-0,504302	0,504302
14	-0,0958343	0,258171	-0,506007	0,506007

Tableau II.2.79 PACF de la racine carrée des apports annuels

Lag	PACF	Ecart-type	-IC à 95%	IC à 95%
1	0,556917	0,152499	-0,298892	0,298892
2	0,061811	0,152499	-0,298892	0,298892
3	0,293669	0,152499	-0,298892	0,298892
4	0,042477	0,152499	-0,298892	0,298892
5	-0,0187228	0,152499	-0,298892	0,298892
6	-0,113301	0,152499	-0,298892	0,298892
7	0,120122	0,152499	-0,298892	0,298892
8	-0,00560498	0,152499	-0,298892	0,298892
9	-0,171159	0,152499	-0,298892	0,298892
10	0,0441126	0,152499	-0,298892	0,298892
11	0,0621159	0,152499	-0,298892	0,298892
12	-0,120367	0,152499	-0,298892	0,298892
13	-0,0466549	0,152499	-0,298892	0,298892
14	-0,0566803	0,152499	-0,298892	0,298892

Les résultats du test de Box-Pierce basé sur les 14 premières auto-corrélations montrent qu'à 99% de confiance la série n'est pas complètement aléatoire (la statistique calculée du test est de 40,51 avec une valeur de probabilité P de 0,0002 qui est inférieure à 0,01).

Les comportements de l'ACF et la PACF nous poussent à opter pour les modèles possibles suivants :

- (A) ARIMA(1,0,3) avec constante
- (B) ARIMA(1,0,2) avec constante
- (C) ARIMA(1,0,1) avec constante
- (D) ARIMA(2,0,3) avec constante
- (E) ARIMA(0,0,3) avec constante

II.2.2.10.3 Estimation

Pour l'estimation des paramètres de ces modèles, il faut rappeler que 42 données ont été utilisées dont les 12 dernières ont servi à la validation. Les paramètres des modèles ont alors été estimés sur une période de 30 ans. Il convient de rappeler qu'avant que le modèle ne soit ajusté, la racine carrée a été choisie comme transformation pour normaliser la distribution des données considérées.

L'analyse et la comparaison des modèles ajustés tel que présenté dans le Tableau II.2.80 montre que tous les modèles ajustés produisent des résidus non complètement indépendants sauf le (A), ceci est fait avec une moyenne et une variance stables entre la période d'estimation et celle de validation. C'est ce dernier modèle qui est retenu.

Tableau II.2.80 Comparaison des modèles ajustés

Modèle	Variance des résidus		AUTO	MOY	VAR
	Période d'estimation	Période de validation			
(A)	5211,81	2379,47	OK	OK	OK
(B)	8403,16	9846,84	NON	OK	OK
(C)	17285,6	43841,1	NON	OK	NON
(D)	10229,4	3298,31	NON	OK	OK
(E)	4898,83	3507,19	NON	OK	OK

Notation : AUTO :test d'indépendance des résidus ; MOY : test d'égalité de la moyenne ; VAR : test d'homoscédasticité ; OK : le test est positif.

Les Tableaux II.2.81 résume l'estimation des paramètres du modèle retenu ARIMA(1,0,3) avec constante ainsi que la signification statistique de ses termes. Les termes correspondant à une valeur de $P < 0.05$ sont statistiquement significativement différents de zéro à 95% de confiance.

Tableau II.2.81 Estimation des paramètres du modèle retenu ARIMA(1,0,3) avec constante

Paramètre	Estimation	Ecart- type	t	P
AR(1)	0,491819	0,431962	1,13857	0,262010
MA(1)	0,245782	0,39138	0,627989	0,533767
MA(2)	0,161447	0,195783	0,824623	0,414732
MA(3)	-0,476016	0,178579	-2,66558	0,011222
Moyenne	10,5976	1,04679	10,1239	0,000000
Constante	5,38551			

La variance estimée du bruit blanc = 10,9155 avec 38 DDL et l'écart- type estimé du bruit blanc = 3,30.

La valeur de t pour les termes MA(1) et MA(2) est inférieure à 1 donc ces derniers sont non significativement différents de zéro à 95% de confiance et peuvent donc être éliminés. La valeur de t pour le reste des termes est en valeur absolue supérieure à 1 par conséquent ses derniers sont statistiquement significativement différents de zéro. L'estimation de l'écart-type du bruit blanc est de 3,3 et l'estimation de la variance des résidus est de 10,9155.

Le Tableau II.2.82 résume la performance du modèle ARIMA(1,0,3) à ajuster les données considérées et ce par l'analyse des erreurs qu'il produit. La différence de ces statistiques d'erreurs est plus ou moins importante entre la période d'estimation et celle de validation. Ces différences expriment une performance moyenne du modèle ajusté pour faire des prédictions dans le futur.

Tableau II.2.82 Statistiques des erreurs du modèle
ARIMA(1,0,3)

Statistique	Période d'estimation	Période de validation
EQM	5211,81	2379,47
EAM	53,6642	42,3533
EM	6,02511	37,922

Notation : EQM : erreur quadratique moyenne ; EAM : erreur absolue moyenne ; EM: erreur moyenne.

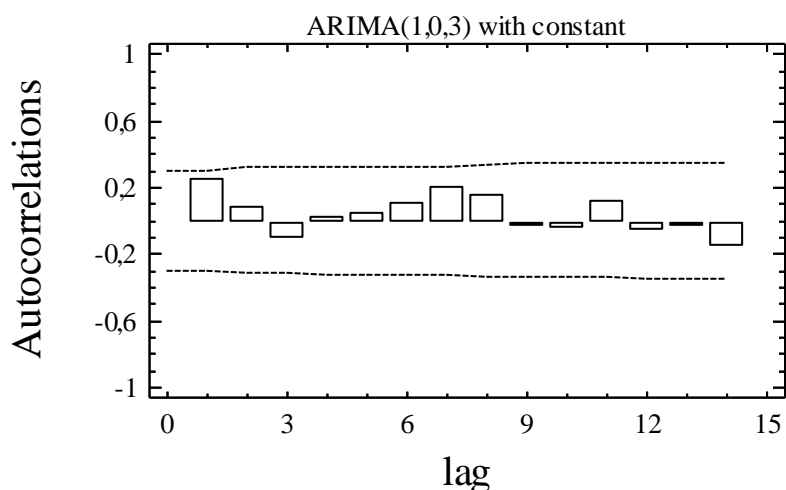
II.2.2.10.4 Validation

Il s'agit de vérifier l'hypothèse de la méthode à savoir que les résidus forment bien un bruit blanc gaussien de moyenne nulle. Pour cela, le test des auto-corrélogrammes (ACF et PACF) et celui du Portmanteau (Box-Pierce) sont utilisés pour tester l'indépendance des résidus, le test t de student est utilisé pour tester la nullité de leur moyenne et le test d'adéquation du Chi-deux à 95% de confiance est utilisé pour tester leur normalité.

- **Test des ACF et PACF**

L'auto-corrélogramme total et partiel des résidus du modèle ont été tracés tel que montré dans la Fig. II.2.60 et Fig. II.2.61.

Tous les coefficients sont inclus dans l'intervalle (IC) tracé à 95% de confiance ce qui indique que ces derniers sont statistiquement non significatifs, d'où leur indépendance. Leurs valeurs sont reportées dans le Tableau II.2.83 et Tableau II.2.84.

**Fig. II.2.60** ACF des résidus du modèle ARIMA(1,0,3)

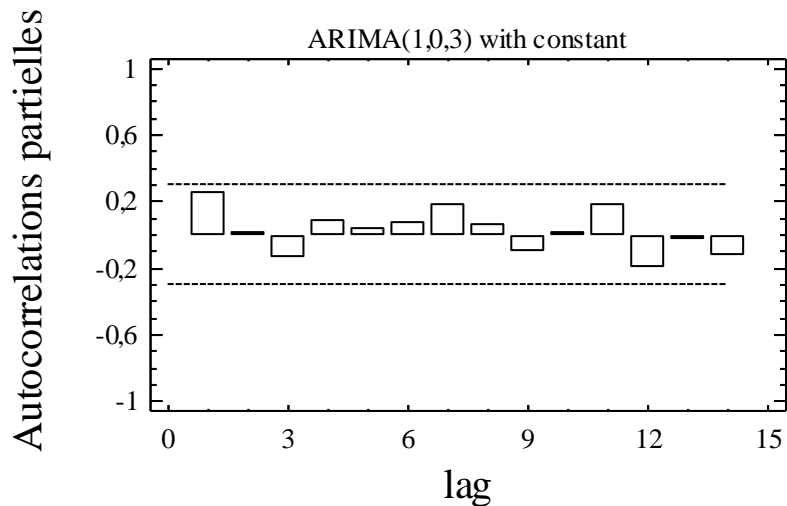


Fig. II.2.61 PACF des résidus du modèle ARIMA(1,0,3)

Tableau II.2.83 ACF des résidus du modèle ARIMA(1,0,3)

Lag	Autocorrelation	Ecart-type	limite inférieure IC 95%	limite supérieure IC 95%
1	0,254594	0,152499	-0,298892	0,298892
2	0,0811342	0,162082	-0,317676	0,317676
3	-0,107774	0,163024	-0,319521	0,319521
4	0,0175499	0,164672	-0,322753	0,322753
5	0,0405716	0,164716	-0,322838	0,322838
6	0,108959	0,164948	-0,323293	0,323293
7	0,200811	0,166614	-0,326557	0,326557
8	0,150928	0,17215	-0,337409	0,337409
9	-0,0341884	0,1752	-0,343387	0,343387
10	-0,0450747	0,175355	-0,343691	0,343691
11	0,122744	0,175625	-0,344219	0,344219
12	-0,063451	0,177609	-0,348107	0,348107
13	-0,0391721	0,178135	-0,349139	0,349139
14	-0,161593	0,178335	-0,349531	0,349531

Tableau II.2.84 PACF des résidus du modèle ARIMA(1,0,3)

Lag	Auto-corrélation partielle	Ecart-type	limite inférieure IC 95%	limite supérieure IC 95%
1	0,254594	0,152499	-0,298892	0,298892
2	0,0174472	0,152499	-0,298892	0,298892
3	-0,141739	0,152499	-0,298892	0,298892
4	0,0835808	0,152499	-0,298892	0,298892
5	0,0363729	0,152499	-0,298892	0,298892
6	0,0708798	0,152499	-0,298892	0,298892
7	0,178504	0,152499	-0,298892	0,298892
8	0,0628598	0,152499	-0,298892	0,298892
9	-0,104868	0,152499	-0,298892	0,298892
10	0,0151133	0,152499	-0,298892	0,298892
11	0,17426	0,152499	-0,298892	0,298892
12	-0,208058	0,152499	-0,298892	0,298892
13	-0,039739	0,152499	-0,298892	0,298892
14	-0,13077	0,152499	-0,298892	0,298892

- **Test du Portmanteau**

Ce test est basé sur les 14 premières auto-corrélations des résidus, les résultats du test montrent une valeur de P supérieure ou égale à 0.10, donc on ne peut pas rejeter l'hypothèse nulle que les résidus sont indépendants avec un intervalle de confiance de 90% ou plus.

Test de Box-Pierce

Hypothèse nulle : résidus indépendants

Valeur calculée de la statistique $Q = 9,0250$

Valeur de P = 0.5297

- **Test de nullité de la moyenne**

La moyenne des résidus est de -0,91 leur médiane est de -1,07 et leur variance est de 9,83 avec un écart-type de 3,13. Les résultats du test t de Student montrent une valeur P= 0,061 qui est supérieure à 0.05 ce qui implique que l'hypothèse nulle de nullité de la moyenne des résidus ne peut pas être rejetée à 95% de confiance.

Test t

Hypothèse nulle : moyenne = 0

Valeur calculée de la statistique $t = -1,92$

Valeur de P = 0.061

- **Test de Normalité**

Afin de vérifier si les résidus du modèle ARIMA(1,0,3) appartiennent à une distribution Normale, un ajustement à la loi de Gauss a été effectué tel que présenté dans la Fig. II.2.8. Les résultats du test du Chi-deux montrent une valeur de P supérieure à 0,01 donc l'hypothèse nulle de Normalité ne peut pas être rejetée à 95% de confiance.

test du Chi-deux

Valeur de la statistique du test = 15,90.

Valeur de P = 0,319

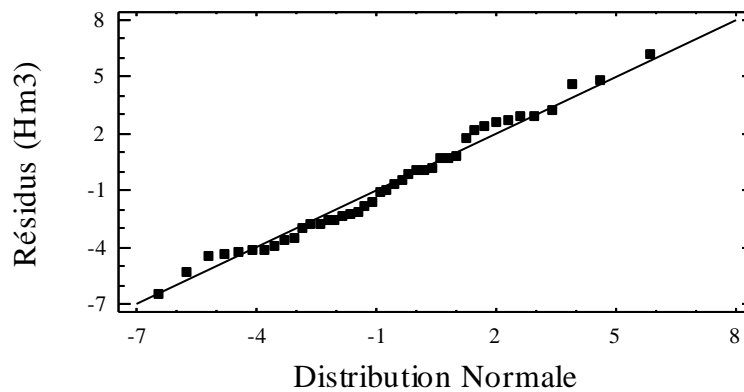


Fig. II.2.62 Ajustement des résidus du modèle ARIMA(1,0,3) à La loi de Gauss

Les résultats des tests précédents appliqués aux résidus du modèle ARIMA(1,0,3) ont permis la vérification des hypothèses de la méthode de Box-Jenkins.

En fin, ce modèle peut être formulé de la manière suivante :

$$(1 - 0,491819B)(X_t - 10,5976) = (1 + 0,476016B^3)\varepsilon_t + 5,38$$

Où ε_t est un bruit blanc gaussien de moyenne nulle et de variance 9,83. Ainsi le modèle explique seulement 36,25% d'une variance totale de 15,42. Ce pourcentage qui exprime la contribution de la partie structurée au sein de la série étudiée, représente un peu plus que le tiers de la variabilité totale ce qui fait que la contribution de la partie aléatoire est deux fois plus importante. Cette considération doit être prise en compte lors de l'utilisation du modèle à la prédiction.

II.2.2.10.5 Prédiction

Les résultats des prédictions obtenues par le modèle retenu ARIMA(1,0,3) sont regroupés dans le Tableau II.2.85 pour la période suivant la fin de la série chronologique de départ et dans le Tableau II.2.86 pour la période où les données sont disponibles avec les résidus correspondants. Pour cette dernière période des intervalles de confiances à 95% ont été calculés pour les prédictions obtenues. Ces limites de confiance indiquent où va être située la vraie valeur à une période déterminée dans le future, dans l'hypothèse que le modèle ajusté est adéquat aux données en question. Ces prédictions présentées dans la Fig. II.2.63 en trait continu suivent l'allure générale des données observées représentées en points carrés. Pour la période où les données ne sont pas disponibles les prédictions obtenues pour les 12 ans futurs sont représentées avec leurs limites de confiance tracées à 95%.

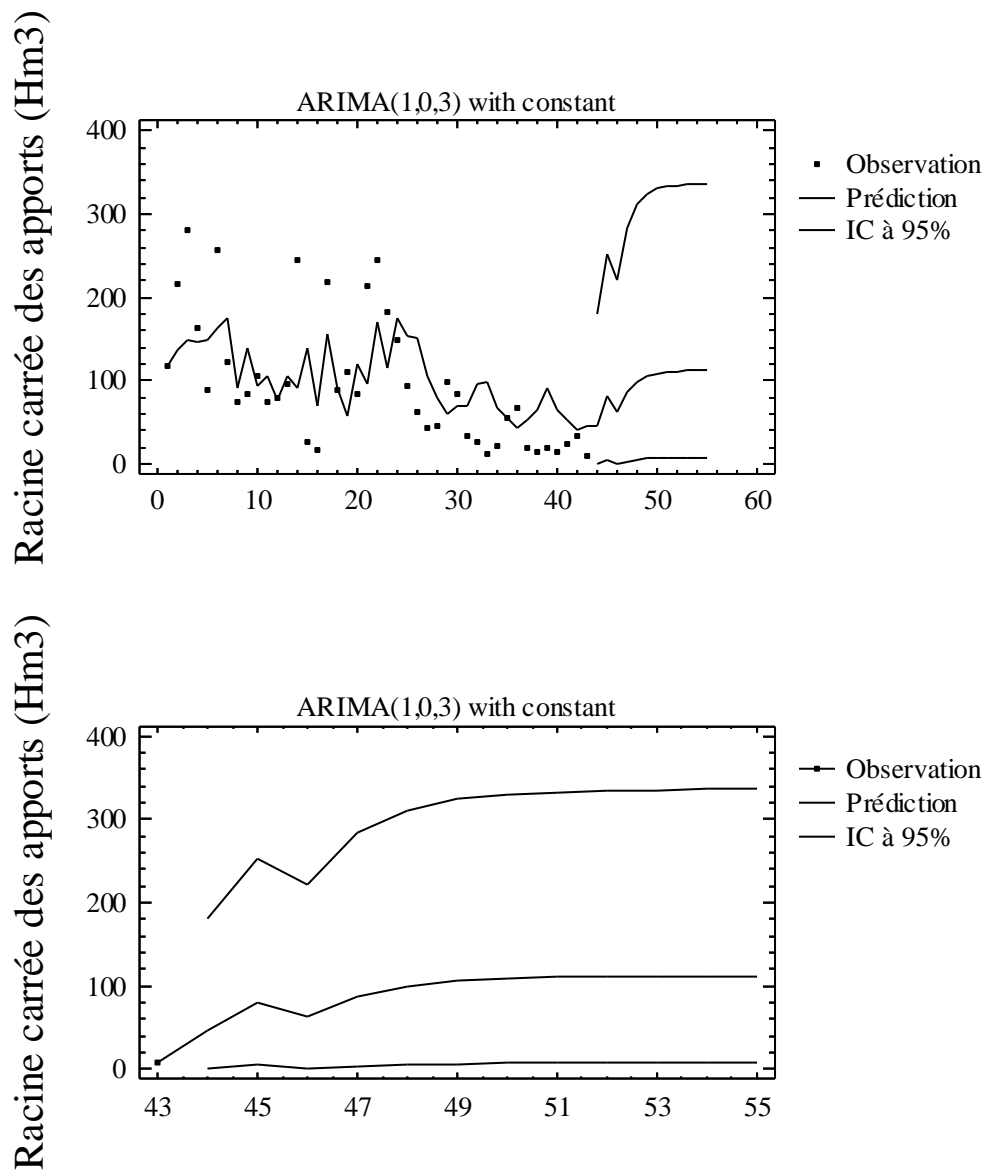


Fig. II.2.63 Prédictions des apports annuels à REMCHI obtenues par le modèle ARIMA(1,0,3)

Tableau II.2.85 Prédictions des apports annuels à REMCHI pour les 12 ans futurs

Période	Prédiction	- IC 95%	+IC 95%
44	45,5743	0,0039101	180,613
45	80,9907	4,45918	252,406
46	63,2853	1,12803	220,473
47	86,4534	3,09759	283,453
48	99,1709	5,16448	311,324
49	105,746	6,56394	324,163
50	109,057	7,36168	330,25
51	110,704	7,78269	333,187
52	111,518	7,99696	334,618
53	111,92	8,10412	335,318
54	112,118	8,15726	335,661
55	112,215	8,1835	335,83

Tableau II.2.86 Prédictions avec ARIMA(1,0,3)

Période	Donnée	Prédiction	Résidu
1	118,35	116,779	1,57149
2	214,89	136,774	78,1158
3	281,8	148,38	133,42
4	163,61	145,578	18,0321
5	88,92	147,998	-59,0777
6	257,77	163,096	94,6738
7	122,65	175,868	-53,2185
8	73,81	90,9598	-17,1498
9	84,56	138,293	-53,733
10	106,5	93,1247	13,3753
11	74,25	105,269	-31,019
12	79,54	75,6657	3,87434
13	96,45	106,129	-9,67855
14	245,73	90,5566	155,173
15	25,27	138,381	-113,111
16	16,58	68,7271	-52,1471
17	218,77	154,869	63,9014
18	88,0	91,3471	-3,34714
19	111,0	58,6118	52,3882
20	83,91	121,104	-37,1939
21	212,68	95,9413	116,739
22	243,94	170,293	73,6465
23	182,96	116,306	66,6542
24	147,82	175,001	-27,1814
25	93,69	153,97	-60,28
26	62,88	151,161	-88,2808
27	44,3	105,869	-61,5693
28	44,78	80,2393	-35,4593
29	99,12	59,9438	39,1762
30	84,57	70,1271	14,4429
31	32,44	68,3992	-35,9592
32	27,49	94,878	V-67,388
33	11,43	97,5459	V-86,1159
34	21,19	66,3453	V-45,1553
35	56,2096	55,1737	V1,03588
36	68,2652	42,7135	V25,5517
37	18,5676	53,6908	V-35,1232
38	14,8293	64,0143	V-49,185
39	18,3114	92,3272	V-74,0158
40	15,274	64,5323	V-49,2583
41	23,1157	51,933	V-28,8173
42	33,2133	41,8438	V-8,63049
43	8,5094	46,4716	V-37,9622

Notation : V : période de validation

II.2.3 RECAPITULATIF ET CONCLUSIONS

L'application de la méthode de Box-Jenkins à la modélisation et la prédiction des apports liquides au pas annuel observés en 10 stations hydrométriques au Nord d'Algérie, a conduit aux principaux résultats récapitulés dans le Tableau II.2.87.

Ces résultats récapitulent le travail de modélisation effectué au pas annuel, ils montrent d'une part, le modèle obtenu pour chaque série chronologique avec le pourcentage de la variance expliquée et d'autre, les statistiques des erreurs de prédictions entre la période d'estimation et la période de validation.

[Tapez un texte]

Tableau II.2.87 Récapitulatif des principaux résultats obtenus à l'échelle annuelle.

N°	Station	N	Transf.	σ^2 (totale)	Modèle	σ_ε^2 (estimée)	σ_ε^2 (calculée)	σ^2 (expliquée en %)	TEM	TEV
01	Ain berda	31	d=0	119,08	-	-	-	-	-	-
02	Béni bahdel	69	d=1	1304,74	ARIMA(0,1,2)	1220,21	1030,72	21,00	OK	OK
03	Bouchegouf	34	d=0	5016,59	-	-	-	-	-	-
04	Bouhnifia	59	d=0	4049,08	ARIMA(1,0,3)	3031,81	3015,51	25,00	OK	OK
05	Cheffia	27	d=0	8735,7	-	-	-	-	-	-
06	Ksob	27	d=0	509,51	ARIMA(2,0,0)	487,31	307,92	39,50	OK	OK
07	Mefrouche	50	d=0	84,47	ARIMA(2,0,3)	71,67	73,28	13,24	NON	OK
08	Mirebeck	27	d=0	72451	-	-	-	-	-	-
09	Pierre du chat	42	d=1	15644,7	-	-	-	-	-	-
10	Remchi	43	√	15,42	ARIMA(1,0,3)	10,83	9,83	36,25	OK	OK
Moyenne				1192,64		887,45		26,99		

Notation : N : nombre d'observations ; Transf. : transformation mathématique ; σ^2 (totale) : variance totale ; σ_ε^2 (calculée) : variance calculée des résidus ; ; σ^2 (expliquée en %) : variance expliquée par le modèle en pourcentage ; TEM : test d'égalité de la moyenne ; TEV : test d'égalité de la variance.

Tableau II.2.88 Statistiques des erreurs des modèles ajustés à l'échelle annuelle.

N°	Station	Modèle	EQM		EAM		EM	
			Période d'estimation	Période de validation	Période d'estimation	Période de validation	Période d'estimation	Période de validation
01	Béni bahdel	ARIMA(0,1,2)	1170,67	809,39	27,48	23,77	3,58	-21,52
02	Bouhnifia	ARIMA(1,0,3)	3118,66	4164,25	42,05	56,52	31,56	113,57
03	Ksob	ARIMA(2,0,0)	477,72	191,00	13,878	12,40	-0,97	-3,46
04	Mefrouche	ARIMA(2,0,3)	73,41	88,74	5,97	8,38	0,15	0,15
05	Remchi	ARIMA(1,0,3)	5211,81	2379,47	53,66	42,35	6,02	37,92
Moyenne			2010,45	1526,57	28,61	28,68	8,07	25,33

Notation : EQM : erreur quadratique moyenne ; EAM : erreur absolue moyenne ; EM : erreur moyenne.

Il est clair que la partie structurée des séries chronologiques considérées est presque deux fois plus faible que la partie aléatoire qui représente à elle seule, plus de 60 % de la variabilité totale et les statistiques des erreurs de prédiction qui sont plus faibles en moyenne pendant la période de validation pour l'erreur quadratique moyenne, sont insensiblement supérieures pour l'erreur absolue moyenne et nettement supérieures pour l'erreur moyenne. Ce résultat est très logique et il fallait s'y attendre par ce qu'il s'agit de données annuelles qui sont généralement indépendantes et ne présentent donc pas de structure de corrélation prononcée au sein des données ; alors que la méthode de Box-Jenkins se base essentiellement sur cette structure de dépendance interne des données pour la détermination des modèles.

Cela étant, les principales conclusions que nous pouvons tirer peuvent être présentées comme suit :

1. L'application de la méthode de Box-Jenkins a révélé l'ampleur de la partie aléatoire dans la variabilité des séries de données considérées en confirmant le caractère normal communément admis à propos d'une telle variable. En effet, parmi les stations hydrométriques étudiées 50% présentent une variabilité chronologique strictement aléatoire et la partie structurée des 50% de stations restantes, explique une variabilité moyenne de seulement 27 % de la variabilité totale. Cependant, on a pu identifier des composantes reconnaissables non négligeables telles que la tendance, et observé l'absence de composante cyclique et saisonnière. Concernant la distribution de probabilité, il faut noter que sur la totalité des 10 stations hydrométriques seulement une ne présentait pas une distribution Normale, ce qui confirme encore une fois le caractère Normal de ces données.
2. Sur la totalité des stations hydrométriques étudiées uniquement 20% ont été le siège d'une instabilité dans la variance ou la moyenne ce qui confirme le caractère stationnaire des séries chronologiques des apports annuels.
3. La modélisation des apports annuels par le recours aux techniques ARIMA a permis l'obtention de modèles adéquats qui sont parcimonieux et simples à utiliser.
4. La performance des modèles ajustés est soulignée par la vérification de deux aspects entre la période d'estimation et celle de validation, le premier concerne la stationnarité, tandis que le second concerne l'adéquation des modèles ajustés. En effet, le test d'égalité de la moyenne entre la période d'estimation et celle de validation ainsi que le test d'homoscédasticité entre ces deux périodes sont vérifiés pour pratiquement toute les stations ce qui renforce la stationnarité des données obtenues et joue en faveur du premier aspect. D'un autre côté, les statistiques d'erreurs entre la période d'estimation et celle de validation, comme c'est indiqué par le Tableau II.2.88 joue en faveur du second aspect de cette performance. En effet, les résultats obtenus montrent que la moyenne de l'erreur quadratique moyenne (EQM) des cinq modèles ajustés est visiblement meilleure sur la période de validation comparativement à la période d'estimation, sachant que l'EQM forme à elle seule, un critère assez puissant dans la mesure de la performance de n'importe quel modèle, sinon l'erreur absolue moyenne (EAM) ou encore l'erreur moyenne (EM). Ceci traduit la validité de ces modèles dans la formulation du processus étudié et donne une indication sur la performance de ces derniers à être mis en œuvre pour la prévision. L'égalité de la moyenne de l'EAM sur les deux périodes de validation et d'estimation, vient confirmer cette performance même si la moyenne de l'EM sur la période d'estimation est nettement meilleure que celle de la période de validation.

5. L'utilisation du modèle ARIMA développé pour chacune des séries d'apports considérés a permis l'obtention de prédictions acceptables dans la limite des performances de ces modèles.
6. Ce travail présente un intérêt particulier dans le fait que la technique qu'il utilise pour la modélisation des séries chronologiques, (technique ARIMA) peut être appliquée pour toute autre variable aléatoire particulièrement en hydrologie.

Chapitre II.3

MODELISATION ARIMA ET PREDICTION DES APPORTS MENSUELS

II.3.1 INTRODUCTION

A l'instar du chapitre précédent qui était consacré à l'application de la méthode de Box-Jenkins (ARIMA) aux données annuelles, le présent chapitre constitue une extension de l'application de la dite méthode aux apports mensuels des 10 stations hydrométriques considérées, c'est-à-dire en que l'on fait intervenir l'effet de la saisonnalité des données en question.

A cet effet, il convient également de rappeler que les conditions pré requises de la méthode sont la normalité et la stationnarité des séries chronologiques, ce qui implique l'introduction de quelques transformations mathématiques des données en cas de défaut de stationnarité ou de normalité d'une part, et d'autre que la méthode est basée essentiellement sur la structure de dépendance interne des séries chronologiques, ce qui veut dire que pour les séries indépendantes le recours à la méthode se trouve complètement non justifié.

Les résultats escomptés de ce travail est la mise en équation de la variabilité stochastique du processus hydrologique en question avec une prise en charge de la composante saisonnière mensuelle. Ceci est réalisé par l'étude d'un échantillon de 10 séries chronologiques dont l'analyse offre des modèles stochastiques qui vont servir à la prévision.

II.3.2 APPLICATION ET DISCUSSION DES RESULTATS

Le principe de la méthode de Box-Jenkins, qui est mis en œuvre dans ce chapitre pour des données saisonnières, reste le même avec les mêmes étapes vues précédemment dans le (chapitre II.2) à savoir: (1) transformations des données afin de satisfaire l'hypothèse de stationnarité ou de normalité, (2) identification des modèles par observation des fonctions d'auto-corrélation et d'auto-corrélation partielle, (3) estimations des paramètres des modèles, et (4) validation des modèles par des tests sur les erreurs produites par rapport aux données observées. Cependant la variabilité saisonnière à l'échelle mensuelle est introduite, ce qui n'était pas le cas des données annuelles.

Les modèles Autorégressifs et de moyenne mobile intégrés saisonniers (SARIMA) obtenus sont alors utilisés à la prédiction qui devient alors, une opération mathématique automatique.

II.3.2.1 AIN BERDA

Les statistiques de la tendance centrale, de la variabilité et de la forme de la série chronologique des apports mensuels observés à AIN BERDA durant les 372 mois allant de Septembre 1963 à Aout 1993 sont résumées dans le Tableau II.3.1 et la variabilité temporelle de cette série chronologique est illustrée dans la Fig. II.3.1.

Tableau II.3.1 Statistiques des apports mensuel observés à AIN BERDA (Sep1963 à Aout 1993)

Paramètre	Valeur
Nombre d'observations	372
Moyenne	1,0729
Variance	5,62568
Déviatoin Standard	2,37185
Minimum	0,0
Maximum	18,51
Std. skewness	28,8728
Std. kurtosis	65,8714

La moyenne de la série est de 1,0729 alors que sa variance autour de cette moyenne est de 5,62568 avec un écart-type de 2,37185. La skewness et la kurtosis standardisées indiquent un éloignement remarquable de la distribution normale.

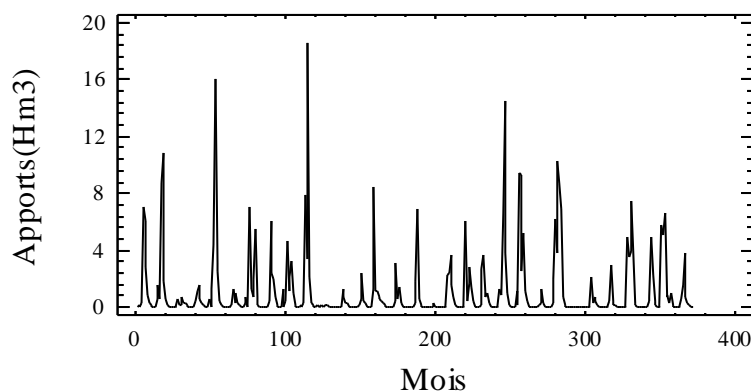


Fig. II.3.1 Apports mensuels observés à AIN BERDA

II.3.2.1.1 Stationnarisation

Comme il a été mentionné précédemment, la stationnarité d'une série implique une moyenne et une variance constantes ou stables dans le temps. Une instabilité de la moyenne peut être provoquée par une tendance, tout comme une instabilité de la variance peut être provoquée par une saisonnalité ou une sur-différenciation (Cohn & Lins, 2005 ; Johnston, 1984). Un moyen de vérifier la stationnarité est l'examen graphique de la série étudiée afin de déceler l'existence de la tendance et de la saisonnalité. Ces deux composantes peuvent être éliminées par le calcul des différences régulières ou saisonnières, dans ce cas le calcul de l'écart-type des séries différenciées nous indiquera laquelle des séries est la plus stable : la série qui correspond au minimum d'écart-type est la plus stable.

Le détail du calcul des séries différenciées pour les 20 premiers mois est montré dans le Tableau II.3.2, tandis que les statistiques des séries différenciées pour l'ensemble des données est représenté dans le Tableau II.3.3.

Tableau II.3.2 Calcul des séries différenciées (données des 20 premiers mois)

N°	Apport	DIFF(Apport,1)	SDIFF(Apports,12) = S	DIFF(S,1)
1	0			
2	0,16	0,16		
3	0,02	-0,14		
4	0,32	0,3		
5	7,08	6,76		
6	6,06	-1,02		
7	2,82	-3,24		
8	0,87	-1,95		
9	0,31	-0,56		
10	0,16	-0,15		
11	0,03	-0,13		
12	0,02	-0,01		
13	0	-0,02	0	
14	0,57	0,57	0,41	0,41
15	1,61	1,04	1,59	1,18
16	0,53	-1,08	0,21	-1,38
17	8,7	8,17	1,62	1,41
18	10,82	2,12	4,76	3,14
19	1,8	-9,02	-1,02	-5,78
20	0,59	-1,21	-0,28	0,74

Le Tableau II.3.3 résume les statistiques descriptives des séries différenciées. Dans le cas de la série étudiée, la stationnarité est obtenue en éliminant la saisonnalité ($S=12$) et ce effectuant les premières différences saisonnières ($D=1$). L'écart-type des premières différences régulières des premières différences saisonnières $DIFF(SDIFF(Apports,12),1)$ étant supérieur à celui des premières différences saisonnières ($SDIFF(Apports,12)$), il s'en suit qu'il n'y a pas lieu de faire des différences régulières ($d=0$) est la série peut être traitée ainsi en considérant qu'elle est stationnaire pour $S=12$ et $D=1$.

Tableau II.3.3 Statistiques descriptives des séries différenciées pour les 372 mois.

Série	N	Moyenne	Variance	Ecart-type
Apports	372	1.07	5.62	2.37
DIFF(Apports,1)	371	0	6.28	2.5
SDIFF(Apports,12)	360	-0.029	10.00	3.16
DIFF(SDIFF(Apports,12),1)	359	0	12.48	3.53

Notation: DIFF (x,n) : différences régulières de x d'ordre n ; SDIFF(x,n,s) :différences saisonnières s de x d'ordre n.

II.3.2.1.2 Identification

Les outils de base utilisés dans cette étape sont la fonction d'auto-corrélation (ACF) et la fonction d'auto-corrélation partielle (PACF) qui expriment les différentes corrélations pouvant exister entre les observations de la série chronologique en question. Il importe de rappeler que l'auto-corrélation entre deux observations séparées de k périodes reflète le degrés de corrélation entre cette paire d'observations dans la série chronologique séparées par un retard de k unités de temps. Alors que l'auto-corrélation partielle est définie comme étant l'auto-corrélation entre n'importe qu'elle paire

d'observations séparées par un retard de k unités de temps, élimination faite des effets des observations intermédiaires.

L'idée centrale de l'identification consiste en la détermination de l'allure ou le comportement des fonctions d'auto-corrélation (ACF) et d'auto-corrélation partielle (PACF) de la série chronologique stationnaire par comparaison à des comportements théoriques communs de fonctions d'auto-corrélation et d'auto-corrélation partielle de modèles stationnaires. Cette comparaison se fait dans le but d'identifier et d'assimiler les modèles qui se rapprochent le plus de point de vue caractéristiques (ACF et PACF) de la série en question.

Dans le cas de la station courante, les calculs de l'ACF et de la PACF des apports stationnaires sont reportés dans les Tableaux II.3.4 et II.3.5 respectivement et leurs tracés sont illustrés par les Figures II.3.2 et II.3.3.

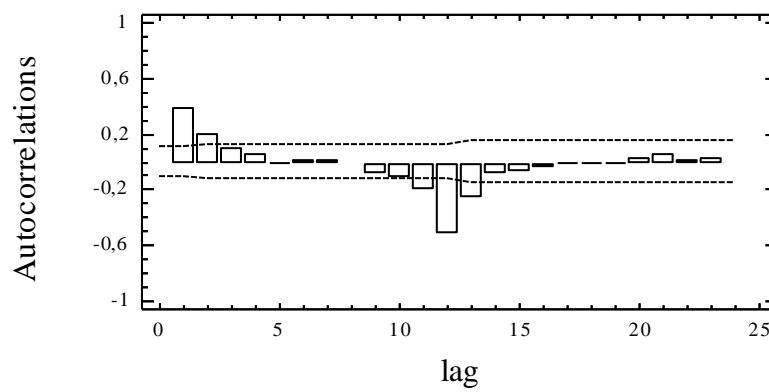


Fig. II.3.2 ACF des apports mensuels

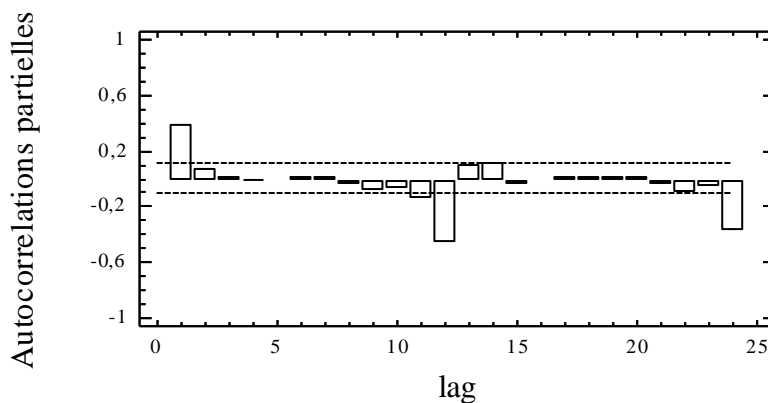


Fig. II.3.3 PACF des apports mensuels

Tableau II.3.4 ACF des apports mensuels stationnaires

Lag	Auto-corrélation	Ecart-type	-IC à 95%	IC à 95%
1	0,377647	0,0527046	-0,103299	0,103299
2	0,201474	0,0597503	-0,117109	0,117109
3	0,0999273	0,0616085	-0,120751	0,120751
4	0,045131	0,0620571	-0,12163	0,12163
5	-0,000276939	0,0621482	-0,121808	0,121808
6	0,00649513	0,0621482	-0,121808	0,121808
7	0,00749	0,0621501	-0,121812	0,121812
8	-0,0268091	0,0621526	-0,121817	0,121817
9	-0,0932801	0,0621847	-0,12188	0,12188
10	-0,128423	0,0625722	-0,122639	0,122639
11	-0,21259	0,0633001	-0,124066	0,124066
12	-0,528207	0,0652532	-0,127894	0,127894
13	-0,26291	0,0762102	-0,14937	0,14937
14	-0,098712	0,0786893	-0,154228	0,154228
15	-0,0818166	0,0790325	-0,154901	0,154901
16	-0,0572447	0,0792675	-0,155362	0,155362
17	-0,0141086	0,0793822	-0,155587	0,155587
18	-0,0091174	0,0793892	-0,1556	0,1556
19	-0,000967519	0,0793921	-0,155606	0,155606
20	0,0235495	0,0793921	-0,155606	0,155606
21	0,0546073	0,0794115	-0,155644	0,155644
22	0,00314142	0,0795157	-0,155848	0,155848
23	0,0228693	0,0795161	-0,155849	0,155849
24	-0,0177115	0,0795344	-0,155885	0,155885

Tableau II.3.5 PACF des apports mensuels stationnaires

Lag	Auto-corrélation	Ecart-type	-IC à 95%	IC à 95%
1	0,377647	0,0527046	-0,103299	0,103299
2	0,0686475	0,0527046	-0,103299	0,103299
3	0,00367954	0,0527046	-0,103299	0,103299
4	-0,00580069	0,0527046	-0,103299	0,103299
5	-0,0256311	0,0527046	-0,103299	0,103299
6	0,0138572	0,0527046	-0,103299	0,103299
7	0,00654236	0,0527046	-0,103299	0,103299
8	-0,0371707	0,0527046	-0,103299	0,103299
9	-0,0873152	0,0527046	-0,103299	0,103299
10	-0,0746011	0,0527046	-0,103299	0,103299
11	-0,14548	0,0527046	-0,103299	0,103299
12	-0,466672	0,0527046	-0,103299	0,103299
13	0,0943215	0,0527046	-0,103299	0,103299
14	0,104819	0,0527046	-0,103299	0,103299
15	-0,0560781	0,0527046	-0,103299	0,103299
16	-0,0267333	0,0527046	-0,103299	0,103299
17	0,00982688	0,0527046	-0,103299	0,103299
18	0,00556243	0,0527046	-0,103299	0,103299
19	0,0123988	0,0527046	-0,103299	0,103299
20	0,00416736	0,0527046	-0,103299	0,103299
21	-0,0331825	0,0527046	-0,103299	0,103299
22	-0,113897	0,0527046	-0,103299	0,103299
23	-0,0707657	0,0527046	-0,103299	0,103299
24	-0,379598	0,0527046	-0,103299	0,103299

L'aspect des auto-corrélogrammes de la série stationnaire considérée laisserait supposer les modèles suivants :

- (A) SARIMA(1,0,1)(1,1,2)₁₂ avec constante
- (B) SARIMA(1,0,1)(1,1,1)₁₂ avec constante
- (C) SARIMA(1,0,1)(0,1,1)₁₂ avec constante
- (D) SARIMA(1,0,0)(1,1,1)₁₂ avec constante
- (E) SARIMA(1,0,0)(0,1,1)₁₂ avec constante

II.3.2.1.3 Estimation des modèles ajustés

Il s'agira alors d'estimer les différents paramètres des modèles ajustés et de vérifier leurs performances en vue de choisir celui qui pourrait mieux traduire le comportement de la série chronologique en question. Dans notre cas, on a utilisé le logiciel SGPLUS.

Pour l'estimation des paramètres de ces modèles, il faut rappeler que 372 données ont été utilisées dont les 100 dernières ont servi à la validation. Les paramètres des modèles ont alors été estimés sur une période de 272 mois.

L'analyse et la comparaison des modèles ajustés tel que présenté dans le Tableau II.3.6, montre que tous les modèles ajustés produisent des résidus complètement indépendants. D'un autre côté, il faut noter que la moyenne de la variance des résidus entre la période d'estimation et celle de validation ne diffère pas considérablement entre les modèles ajustés. Ces derniers produisent une variance des résidus égales au centième près pour la période d'estimation et au dixième près pour la période de validation, tandis que la variance de la période d'estimation est presque le double de celle de la période de validation et ce pour tous les modèles. En outre, le modèle qui produit le minimum de cette variance est le modèle (E), c'est lui qui est donc retenu pour la prédiction.

Tableau II.3.6 Comparaison des modèles ajustés

Modèle	Variance des résidus		Indépendance des résidus
	Période d'estimation	Période de validation	
(A)	4,78745	2,8154	OK
(B)	4,77015	2,72221	OK
(C)	4,75473	2,76159	OK
(D)	4,77016	2,72221	OK
(E)	4,74558	2,79328	OK

Le Tableau II.3.7 résume l'estimation des paramètres du modèle retenu SARIMA(1,0,0)(0,1,1)₁₂ et la signification statistique de ses termes. Les termes correspondant à une valeur de $P < 0.05$ sont statistiquement significativement différents de zéro à 95% de confiance.

Les paramètres du modèle sont stables, et la valeur du terme AR(1) est différente de Un (1). La valeur de P pour les termes AR(1) et SMA(1) est inférieure à 0.05 donc elle est significativement différente de zéro. La valeur de P pour la moyenne est supérieure à 0.05 donc elle est statistiquement non significative et peut donc être éliminée, la constante aussi. L'estimation de l'écart-type du bruit blanc est de 2,2071 et la moyenne de la variance des résidus est de 4,87127, ce qui veut dire que le présent modèle explique 51.3% de la variance de la série modélisée qui est de 10.

Tableau II.3.7 Estimation des paramètres du modèle retenu SARIMA(1,0,0)(0,1,1)₁₂

Paramètre	Estimation	Ecart- type	t	P
AR(1)	0,281059	0,0600852	4,67767	0,000004
SMA(1)	0,956323	0,0130409	73,3327	0,000000
Moyenne	-0,0155172	0,0272899	-0,568608	0,569980
Constante	-0,011156			

Notation: SAR(n): Autorégressif saisonnier d'ordre n; SMA(n) : Moyenne mobile saisonnière d'ordre n.

avec une variance estimée du bruit blanc = 4,87127 avec 357 DDL et un écart type estimé du bruit blanc = 2,2071

Le Tableau II.3.8 résume la performance du modèle SARIMA(1,0,0)(0,1,1)₁₂ à ajuster les données considérées et ce par l'analyse des erreurs qu'il produit.

Tableau II.3.8 Statistiques des erreurs du modèle SARIMA(1,0,0)(0,1,1)₁₂

Statistique	Période d'estimation	Période de validation
EQM	4,74558	2,79328
EAM	1,05718	0,969362
EM	0,00910869	0,176949

Nous y trouvons :

- 4) L'erreur quadratique moyenne (EQM)
- 5) L'erreur absolue moyenne (EAM)
- 6) L'erreur moyenne (EM)

Chacune de ces statistiques est basée sur l'erreur de prédiction un mois à l'avance qui est la différence entre la valeur de la donnée au temps t et sa prédiction au temps $t - 1$. Les deux premières mesurent la magnitude des erreurs alors que la dernière mesure le biais. Ces statistiques d'erreurs sont calculées pour la période d'estimation et celle de validation. Pour le modèle considéré ces statistiques d'erreurs ne sont pas différentes d'une manière considérable entre la période d'estimation et celle de validation sauf pour l'EQM qui exprime la variance des résidus, ce qui signifie un défaut dans la variance. Il s'ensuit que le modèle ajusté peut être utilisé à faire des prévisions avec une mise en garde au sujet de la variance.

Le modèle retenu pour la prédiction est alors le SARIMA(1,0,0)(0,1,1)₁₂ sans constante qui peut être formulé comme suit:

$$(1 - 0,281059B)(1 - B^{12})X_t = (1 - 0,956323B^{12})\varepsilon_t$$

Où ε_t est une séquence bruit blanc avec moyenne nulle et une variance estimée de 4,87.

II.3.2.1.4 Validation du modèle SARIMA(1,0,0)(0,1,1)₁₂

Il est important de réaliser cette étape afin de vérifier et de s'assurer de la performance des modèles retenus dans l'étape d'identification et au besoin suggérer certaines améliorations. Il s'agit de vérifier que les résidus forment bien un bruit blanc de moyenne nulle. Les tests des autocorrélogrammes et celui du Portmanteau (Box-Pierce) sont utilisés pour tester l'indépendance des résidus et le test t de Student est utilisé pour tester la nullité de leur moyenne.

- **Test des ACF et PACF**

L'auto-corrélogramme total et partiel des résidus du modèle ont été tracés tel que montré dans la Fig. II.3.4 et Fig. II.3.5.

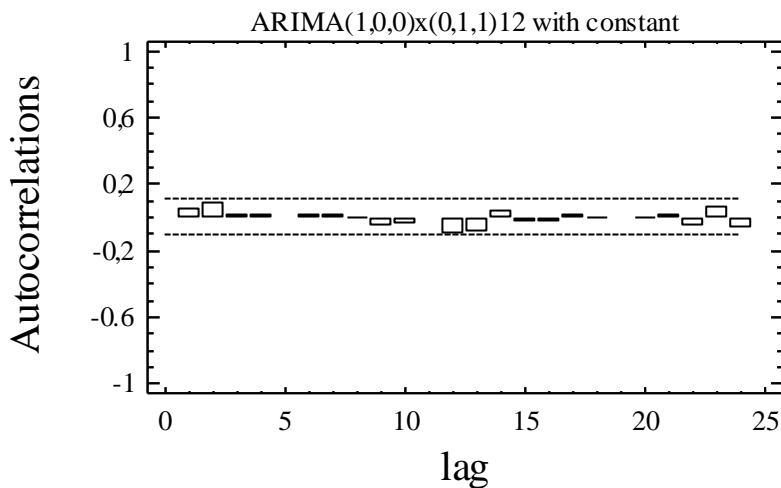


Fig. II.3.4 ACF des résidus du modèle SARIMA(1,0,0)(0,1,1)₁₂

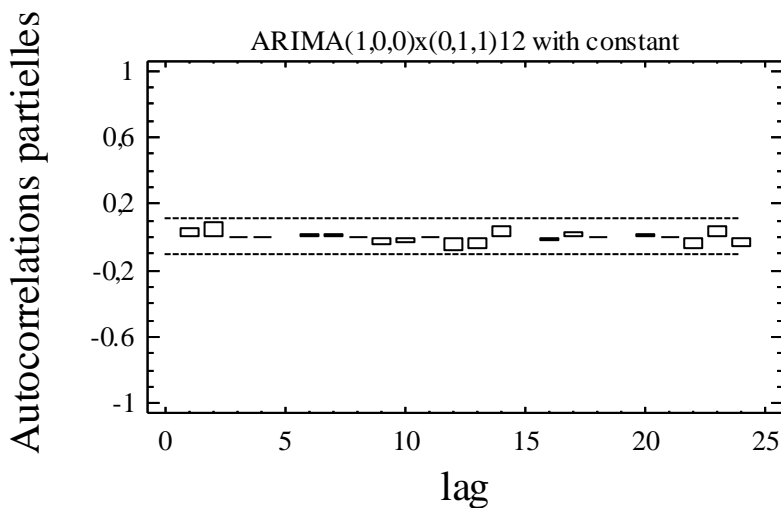


Fig. II.3.5 PACF des résidus du modèle SARIMA(1,0,0)(0,1,1)₁₂

Tous les coefficients sont inclus dans l'intervalle de confiance (IC) tracé à 95% de confiance ce qui implique leur indépendance. Leurs valeurs sont reportées dans le Tableau II.3.9 et Tableau II.3.10.

Tableau II.3.9 ACF des résidus du modèle SARIMA(1,0,0)(0,1,1)₁₂

Lag	Auto-corrélation	Ecart-type	limite inférieure IC 95%	limite supérieure IC 95%
1	0,0421772	0,0527046	-0,103299	0,103299
2	0,0896875	0,0527983	-0,103483	0,103483
3	0,00613841	0,0532198	-0,104309	0,104309
4	0,0138809	0,0532218	-0,104313	0,104313
5	-0,00804744	0,0532318	-0,104333	0,104333
6	0,0124803	0,0532352	-0,104339	0,104339
7	0,00973796	0,0532433	-0,104355	0,104355
8	0,00343201	0,0532483	-0,104365	0,104365
9	-0,0586071	0,0532489	-0,104366	0,104366
10	-0,0485975	0,0534278	-0,104717	0,104717
11	-0,00843971	0,0535504	-0,104957	0,104957
12	-0,104426	0,0535541	-0,104964	0,104964
13	-0,0917174	0,0541168	-0,106067	0,106067
14	0,0337468	0,0545469	-0,10691	0,10691
15	-0,0251411	0,0546048	-0,107024	0,107024
16	-0,0280536	0,054637	-0,107087	0,107087
17	0,0118105	0,054677	-0,107165	0,107165
18	-0,00159778	0,0546841	-0,107179	0,107179
19	-0,00822543	0,0546842	-0,107179	0,107179
20	0,00481297	0,0546876	-0,107186	0,107186
21	0,015625	0,0546888	-0,107188	0,107188
22	-0,057604	0,0547012	-0,107213	0,107213
23	0,0597283	0,0548694	-0,107542	0,107542
24	-0,0698197	0,0550497	-0,107896	0,107896

Tableau II.3.10 PACF des résidus du modèle SARIMA(1,0,0)(0,1,1)₁₂

Lag	Auto-corrélation partielle	Ecart-type	limite inférieure IC 95%	limite supérieure IC 95%
1	0,0421772	0,0527046	-0,103299	0,103299
2	0,0880652	0,0527046	-0,103299	0,103299
3	-0,00103542	0,0527046	-0,103299	0,103299
4	0,00584166	0,0527046	-0,103299	0,103299
5	-0,00934296	0,0527046	-0,103299	0,103299
6	0,011527	0,0527046	-0,103299	0,103299
7	0,0104267	0,0527046	-0,103299	0,103299
8	0,000520063	0,0527046	-0,103299	0,103299
9	-0,0610225	0,0527046	-0,103299	0,103299
10	-0,0453278	0,0527046	-0,103299	0,103299
11	0,00571178	0,0527046	-0,103299	0,103299
12	-0,0970872	0,0527046	-0,103299	0,103299
13	-0,0850956	0,0527046	-0,103299	0,103299
14	0,0582179	0,0527046	-0,103299	0,103299
15	-0,0127913	0,0527046	-0,103299	0,103299
16	-0,0322286	0,0527046	-0,103299	0,103299
17	0,0194949	0,0527046	-0,103299	0,103299
18	-0,0000787006	0,0527046	-0,103299	0,103299
19	-0,0113113	0,0527046	-0,103299	0,103299
20	0,00694636	0,0527046	-0,103299	0,103299
21	0,00549024	0,0527046	-0,103299	0,103299
22	-0,080565	0,0527046	-0,103299	0,103299
23	0,0649631	0,0527046	-0,103299	0,103299
24	-0,0726829	0,0527046	-0,103299	0,103299

- **Test du Portmanteau**

Ce test est basé sur les 24 premières auto-corrélations des résidus, les résultats du test montrent une valeur de $P = 0,698273$ qui est supérieure ou égale à 0.10, donc on ne peut pas rejeter l'hypothèse nulle que les résidus sont indépendants avec un intervalle de confiance de 90% ou plus.

Test de Box-Pierce

Hypothèse nulle : résidus indépendants

Valeur calculée de la statistique $Q = 18,1297$

Valeur de $P = 0,698273$

- **Test de nullité de la moyenne**

La moyenne des résidus est de 0,0540954, leur médiane est de -0,0773425 et leur variance est de 4,17. La Figure II.3.6 montre la variabilité temporelle de ces derniers autour de leur moyenne. Les résultats du test t de Student montrent une valeur $P = 0,615474$ qui est supérieure à 0.05 ce qui implique que l'hypothèse nulle de nullité de la moyenne des résidus ne peut pas être rejetée à 95% de confiance.

Test t de Student

Hypothèse nulle : moyenne = 0,0

Alternative: non nulle.

Valeur calculée de la statistique $t = 0,502714$

Valeur de $P = 0,615474$

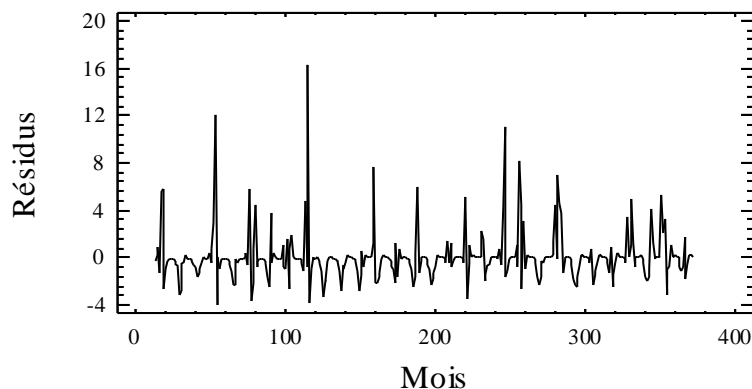


Fig. II.3.6 Résidus du modèle SARIMA(1,0,0)(0,1,1)₁₂

Les résultats de ces tests, ainsi que les calculs des résidus montrent bien que les résidus du modèle SARIMA(1,0,0)(0,1,1)₁₂ forment bel et bien une séquence bruit blanc de moyenne nulle et de variance 4.17. Dans ce cas, on peut dire que le modèle courant explique 58,3% de la variance de départ qui était de 10.

II.3.2.1.5 Prédiction

En permettant au modèle retenu SARIMA(1,0,0)(0,1,1)₁₂ de faire de la prédiction en arrière nous pouvons obtenir les prédictions correspondant aussi bien à la période où les données sont disponibles qu'à celle où ces dernières ne le sont pas. Tableau II.3.11 présente les prédictions des 24 données futures pour la période où les données ne sont pas disponibles, avec les résidus correspondants. Les prédictions de ces 24 données futures sont présentées avec leurs intervalles de confiance à 95%. Ces limites de confiance indiquent où va être située la vraie valeur à une période déterminée dans le futur, dans l'hypothèse que le modèle ajusté est adéquat aux données en question. Tandis que les prédictions correspondant à la période où les données sont disponibles avec les résidus correspondants, sont présentées en Annexe A.II.3.1

Les prédictions obtenues par le modèle, tel que présenté par la Fig. II.3.7, suivent en général l'allure des données observées qui sont représentées par des points carrés sur la dite figure. Pour la période où les données ne sont pas disponibles, les prédictions obtenues pour les 24 mois futurs suivant la fin de la série sont représentées avec leurs limites tracées à 95% de confiance.

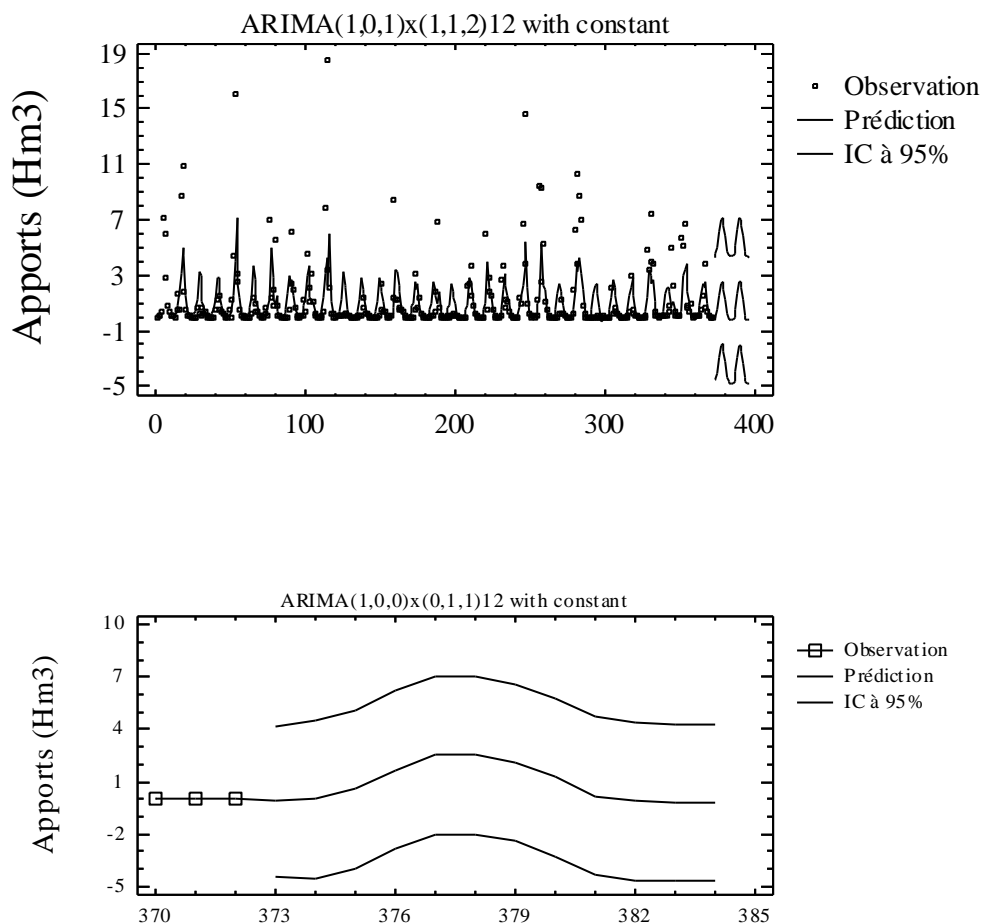


Fig. II.3.7 Prédiction des apports mensuels à AIN BERDA obtenues par le modèle SARIMA(1,0,0)(0,1,1)₁₂

Tableau II.3.11 Prévisions des apports mensuels pour les 24 mois futurs à AIN BERDA

Période	Prédiction	- IC 95%	+IC 95%
373	-0,119273	-4,45982	4,22128
374	-0,0248147	-4,53355	4,48392
375	0,568987	-3,95276	5,09074
376	1,68671	-2,83607	6,20948
377	2,51429	-2,00856	7,03715
378	2,52989	-1,99297	7,05276
379	2,08768	-2,43518	6,61055
380	1,27811	-3,24475	5,80098
381	0,183328	-4,33954	4,70619
382	-0,125289	-4,64815	4,39758
383	-0,191261	-4,71413	4,3316
384	-0,197677	-4,72054	4,32519
385	-0,185988	-4,71282	4,34085
386	-0,0547215	-4,58187	4,47243
387	0,549425	-3,97775	5,0766
388	1,67005	-2,85712	6,19723
389	2,49846	-2,02872	7,02563
390	2,51428	-2,01289	7,04146
391	2,07214	-2,45504	6,59931
392	1,26259	-3,26459	5,78976
393	0,167809	-4,35937	4,69499
394	-0,140807	-4,66798	4,38637
395	-0,206778	-4,73396	4,3204
396	-0,213194	-4,74037	4,31398

II.3.2.2 BENI BAHDEL

Les statistiques de la tendance centrale, de la variabilité et de la forme de la série chronologique des apports mensuels observés à BENI BAHDEL durant les 360 mois allant de Septembre 1924 à Aout 1993 sont résumées dans le Tableau II.3.12 et la variabilité temporelle de cette série chronologique est illustrée par la Fig. II.3.8.

Tableau II.3.12 Statistiques des apports mensuels observés à BENI BAHDEL (Sep1924 à Aout 1993)

Paramètre	Valeur
Nombre d'observations	828
Moyenne	5,47166
Variance	57,2245
Déviatoin Standard	7,56469
Minimum	0,042
Maximum	98,18
Std. skewness	56,523
Std. kurtosis	227,517

La moyenne de la série est de 5,47166 alors que sa variance autour de cette moyenne est de 57,2245 avec un écart-type de 7,56469. La skewness et la kurtosis standardisées indiquent un éloignement remarquable de la distribution normale.

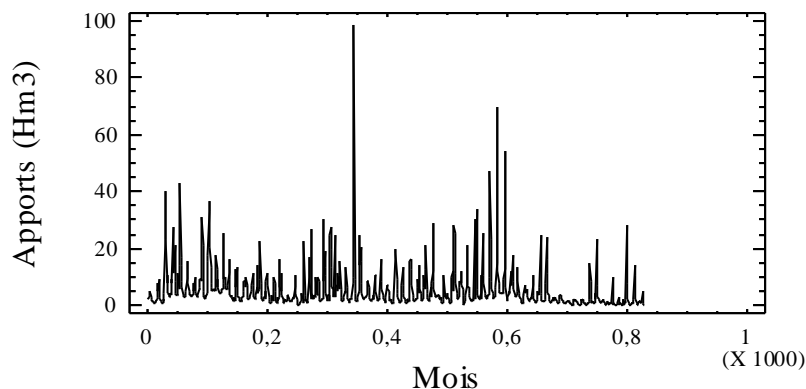


Fig. II.3.8 Apports mensuels observés à BENI BAHDEL

II.3.2.2.1 Stationnarisation

Le détail du calcul des séries différenciées pour les 20 premiers mois est montré dans le Tableau II.3.13, tandis que les statistiques des séries différenciées pour l'ensemble des données sont représentées dans le Tableau II.3.14.

Tableau II.3.13 Calcul des séries différenciées (données des 20 premiers mois)

N°	Apport	DIFF(Apport,1)	SDIFF(Apports,12) = S	DIFF(S,1)
1	1,892			
2	3,081	1,189		
3	3,903	0,822		
4	5,181	1,278		
5	3,241	-1,94		
6	2,02	-1,221		
7	1,835	-0,185		
8	1,57	-0,265		
9	1,531	-0,039		
10	0,768	-0,763		
11	1,102	0,334		
12	0,754	-0,348		
13	1,562	0,808	-0,33	
14	1,833	0,271	-1,248	-0,918
15	2,197	0,364	-1,706	-0,458
16	2,593	0,396	-2,588	-0,882
17	7,615	5,022	4,374	6,962
18	5,139	-2,476	3,119	-1,255
19	9,426	4,287	7,591	4,472
20	2,945	-6,481	1,375	-6,216

Le Tableau II.3.14 résume les statistiques descriptives des séries différenciées. Dans le cas de la série étudiée, la stationnarité est obtenue en éliminant la tendance et la saisonnalité ($S=12$) et ce effectuant les premières différences régulières ($d=1$) des premières différences saisonnières ($D=1$). La variance des premières différences régulières des premières différences saisonnières $\text{DIFF}(\text{SDIFF}(\text{Apports},12),1)$ étant de 10.18 et la série est stationnaire pour $S=12$, $d=1$ et $D=1$.

Tableau II.3.14 Statistiques descriptives des séries différenciées pour les 828 mois.

Série	N	Moyenne	Variance	Ecart-type
Apports	828	5.47	57.22	7.56
DIFF(Apports,1)	827	-0.002	64.02	8.00
SDIFF(Apports,12)	816	-0.014	65.35	8.08
DIFF(SDIFF(Apports,12),1)	815	-0.000	103.73	10.18

Notation :DIFF (x,n) : différences régulières de x d'ordre n ; SDIFF(x,n,s) :différences saisonnières s de x d'ordre n, N :taille.

II.3.2.2.2 Identification

Dans le cas de la station courante, les calculs de l'ACF et de la PACF des apports stationnaires sont reportés dans les Tableaux II.3.15 et 16 respectivement et leurs tracés sont illustrés par les Figures II.3.9 et 10.

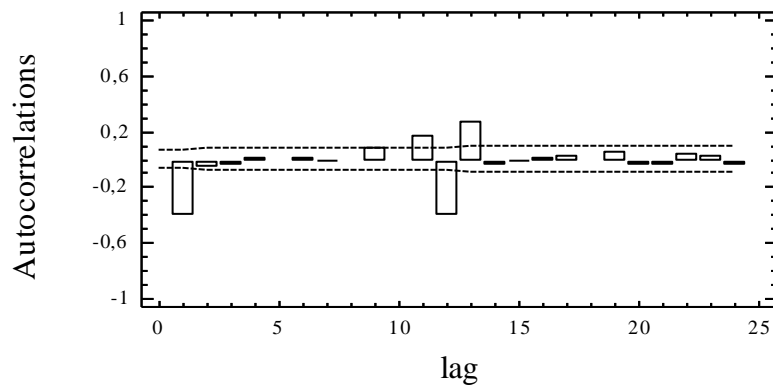


Fig. II.3.9 ACF des apports mensuels

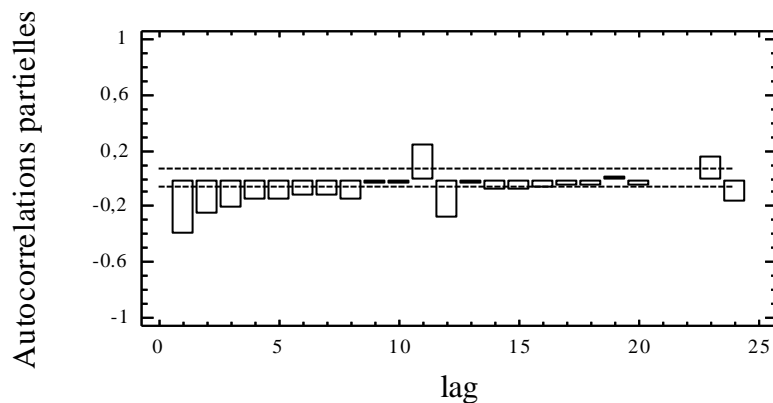


Fig. II.3.10 PACF des apports mensuels

Tableau II.3.15 ACF des apports mensuels stationnaires

Lag	Auto-corrélation	Ecart-type	-IC à 95%	IC à 95%
1	-0,410251	0,0350285	-0,0686547	0,0686547
2	-0,0582501	0,0404971	-0,079373	0,079373
3	-0,0316833	0,0405998	-0,0795742	0,0795742
4	0,013813	0,0406301	-0,0796337	0,0796337
5	-0,0239008	0,0406358	-0,0796449	0,0796449
6	0,0106145	0,0406531	-0,0796787	0,0796787
7	-0,0094014	0,0406565	-0,0796854	0,0796854
8	-0,0240016	0,0406592	-0,0796906	0,0796906
9	0,0772504	0,0406765	-0,0797247	0,0797247
10	-0,015473	0,0408562	-0,0800767	0,0800767
11	0,160577	0,0408633	-0,0800908	0,0800908
12	-0,417177	0,0416304	-0,0815942	0,0815942
13	0,263612	0,0464777	-0,0910947	0,0910947
14	-0,0468429	0,0482774	-0,0946221	0,0946221
15	-0,00467224	0,0483331	-0,0947313	0,0947313
16	0,00642639	0,0483337	-0,0947324	0,0947324
17	0,0235106	0,0483347	-0,0947345	0,0947345
18	-0,014835	0,0483487	-0,094762	0,094762
19	0,0438553	0,0483543	-0,0947729	0,0947729
20	-0,0358427	0,0484031	-0,0948685	0,0948685
21	-0,0340467	0,0484357	-0,0949323	0,0949323
22	0,0351518	0,048465	-0,0949899	0,0949899
23	0,0189771	0,0484963	-0,0950512	0,0950512
24	-0,0427488	0,0485054	-0,095069	0,095069

Tableau II.3.16 PACF des apports mensuels stationnaires

Lag	Auto-corrélation	Ecart-type	-IC à 95%	IC à 95%
1	-0,410251	0,0350285	-0,0686547	0,0686547
2	-0,272403	0,0350285	-0,0686547	0,0686547
3	-0,225777	0,0350285	-0,0686547	0,0686547
4	-0,164263	0,0350285	-0,0686547	0,0686547
5	-0,162934	0,0350285	-0,0686547	0,0686547
6	-0,133927	0,0350285	-0,0686547	0,0686547
7	-0,131091	0,0350285	-0,0686547	0,0686547
8	-0,161006	0,0350285	-0,0686547	0,0686547
9	-0,054011	0,0350285	-0,0686547	0,0686547
10	-0,0357126	0,0350285	-0,0686547	0,0686547
11	0,23721	0,0350285	-0,0686547	0,0686547
12	-0,292663	0,0350285	-0,0686547	0,0686547
13	-0,0331685	0,0350285	-0,0686547	0,0686547
14	-0,0900683	0,0350285	-0,0686547	0,0686547
15	-0,100055	0,0350285	-0,0686547	0,0686547
16	-0,0769107	0,0350285	-0,0686547	0,0686547
17	-0,0640048	0,0350285	-0,0686547	0,0686547
18	-0,0603196	0,0350285	-0,0686547	0,0686547
19	0,00425295	0,0350285	-0,0686547	0,0686547
20	-0,0623091	0,0350285	-0,0686547	0,0686547
21	-0,025315	0,0350285	-0,0686547	0,0686547
22	-0,0226931	0,0350285	-0,0686547	0,0686547
23	0,157217	0,0350285	-0,0686547	0,0686547
24	-0,178068	0,0350285	-0,0686547	0,0686547

L'aspect des auto-corrélogrammes de la série stationnaire considérée laisserait supposer les modèles suivants :

- (A) SARIMA(1,1,1)(1,1,1)₁₂ avec constante
- (B) SARIMA(1,1,1)(2,1,1)₁₂ sans constante
- (C) SARIMA(0,1,2)(1,1,1)₁₂ avec constante
- (D) SARIMA(1,0,2)(1,1,2)₁₂ avec constante
- (E) SARIMA(0,1,1)(1,1,1)₁₂ avec constante

II.3.2.2.3 Estimation des modèles ajustés

Pour l'estimation des paramètres de ces modèles, il faut rappeler que 828 données ont été utilisées dont les 200 dernières ont servi à la validation. Les paramètres des modèles ont alors été estimés sur une période de 628 mois.

L'analyse et la comparaison des modèles ajustés tel que présenté dans le Tableau II.3.17, montre que tous les modèles ajustés produisent des résidus complètement indépendants sauf le modèle (E). D'un autre côté, il faut noter que la moyenne de la variance des résidus produits par tous les modèles ne diffère pas de trop aussi bien pour la période d'estimation que celle de la validation. En outre, le modèle qui produit le minimum de cette variance est le modèle (B), c'est lui qui est donc retenu pour la prédiction.

Tableau II.3.17 Comparaison des modèles ajustés

Modèle	Variance des résidus		Indépendance des résidus
	Période d'estimation	Période de validation	
(A)	47,6524	15,1351	OK
(B)	47,341	15,8532	OK
(C)	47,8047	15,1801	OK
(D)	47,8157	15,6002	OK
(E)	50,1371	16,856	NON

Le Tableau II.3.18 résume l'estimation des paramètres du modèle retenu SARIMA(1,1,1)(2,1,1)₁₂ et la signification statistique de ses termes. Les termes correspondant à une valeur de $P < 0.05$ sont statistiquement significativement différents de zéro à 95% de confiance.

Les paramètres du modèle sont stables, et la valeur du terme AR(1) est différente de Un (1). La valeur de P pour les termes AR(1), MA(1), SAR(1) et SMA(1) est inférieure à 0.05 donc ces derniers sont significativement différents de zéro à 95% de confiance et le terme SAR(2) étant inférieur à 0.1 peut être considéré différent de zéro à 90% de confiance. La valeur de P pour la moyenne est très supérieure à 0.05 donc elle est statistiquement non significative elle a donc été éliminée, la constante aussi. L'estimation de l'écart-type du bruit blanc est de 6,90988 et la moyenne de la variance des résidus est de 47,7465 ce qui veut dire que le présent modèle explique 54% de la variance de la série modélisée qui est de 103.73.

Tableau II.3.18 Estimation des paramètres du modèle retenu SARIMA(1,1,1)(2,1,1)₁₂

Paramètre	Estimation	Ecart- type	t	P
AR(1)	0,239977	0,0404411	5,93399	0,000000
MA(1)	0,970295	0,00848082	114,411	0,000000
SAR(1)	0,205772	0,0412438	4,98917	0,000001
SAR(2)	-0,0731539	0,0415078	-1,76241	0,077999
SMA(1)	0,965876	0,00819984	117,792	0,000000

SAR(n): Autorégressif saisonnier d'ordre n; SMA(n) : Moyenne mobile saisonnière d'ordre n. avec

La variance estimée du bruit blanc = 47,7465 avec 810 DDL et

L'écart type estimé du bruit blanc = 6,90988

Le Tableau II.3.19 résume la performance du modèle SARIMA(1,1,1)(2,1,1)₁₂ à ajuster les données considérées et ce par l'analyse des erreurs qu'il produit.

Tableau II.3.19 Statistiques des erreurs du modèle SARIMA(1,1,1)(2,1,1)₁₂

Statistique	Période d'estimation	Période de validation
EQM	47,341	15,8532
EAM	3,30888	2,46675
EM	0,138925	-0,398649

Nous y trouvons :

- 7) L'erreur quadratique moyenne (EQM)
- 8) L'erreur absolue moyenne (EAM)
- 9) L'erreur moyenne (EM)

Chacune de ces statistiques est basée sur l'erreur de prédiction un mois à l'avance qui est la différence entre la valeur de la donnée au temps t et sa prédiction au temps $t - 1$. Les deux premières mesurent la magnitude des erreurs alors que la dernière mesure le biais. Ces statistiques d'erreurs sont calculées pour la période d'estimation et celle de validation. Pour le modèle considéré ces statistiques d'erreurs ne sont pas différentes d'une manière considérable pour les deux périodes d'estimation et de validation. Il s'ensuit que le modèle ajusté peut être considéré comme assez performant et donc bon à faire des prévisions des valeurs futures.

Le modèle retenu pour la prédiction est alors le SARIMA(1,1,1)(2,1,1)₁₂ sans constante qui peut être formulé comme suit:

$$(1 - 0,239977B)(1 - 0,205772B^{12} + 0,0731539B^{13})(1 - B)(1 - B^{12})X_t = (1 - 0,970295B)(1 - 0,965876B^{12})\varepsilon_t$$

ε_t est une séquence bruit blanc de moyenne nulle et de variance estimée égale à 47,7465.

II.3.2.2.4 Validation du modèle SARIMA(1,1,1)(2,1,1)₁₂

Il s'agit de vérifier que les résidus forment bien un bruit blanc de moyenne nulle. Les tests des autocorrélogrammes et celui du Portmanteau (Box-Pierce) sont utilisés pour tester l'indépendance des résidus et le test t de student est utilisé pour tester la nullité de leur moyenne.

- **Test des ACF et PACF**

L'auto-corrélogramme total et partiel des résidus du modèle ont été tracés tel que montré dans la Fig. II.3.11 et Fig. II.3.12

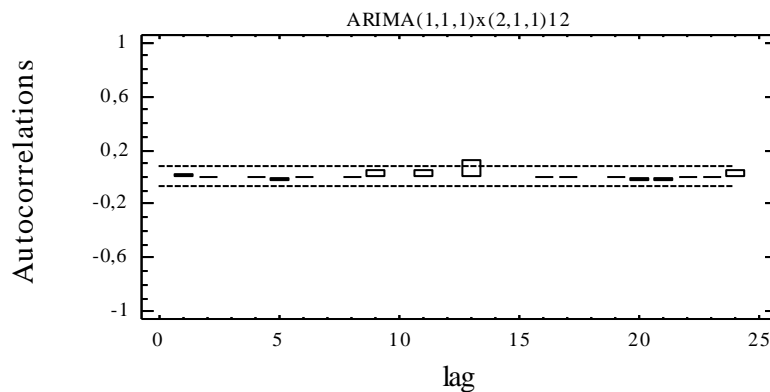


Fig. II.3.11 ACF des résidus du modèle SARIMA(1,1,1)(2,1,1)₁₂

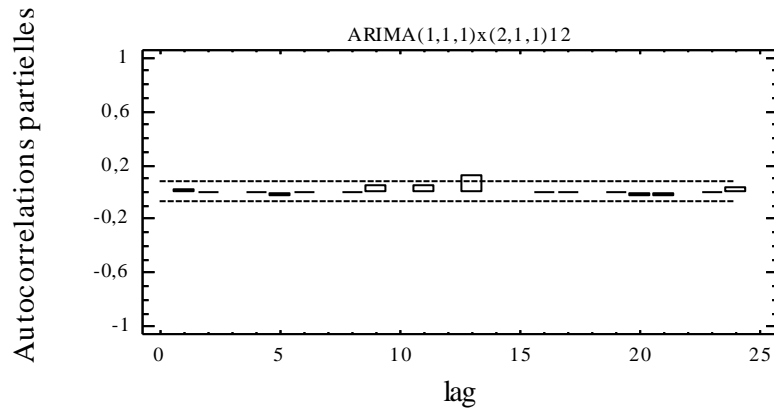


Fig. II.3.12 PACF des résidus du modèle SARIMA(1,1,1)(2,1,1)₁₂

Tous les coefficients sont inclus dans l'intervalle de confiance (IC) tracé à 95% de confiance ce qui implique leur indépendance. Leurs valeurs sont reportées dans le Tableau II.3.20 et Tableau II.3.21.

Tableau II.3.20 ACF des résidus du modèle SARIMA(1,1,1)(2,1,1)₁₂

Lag	Auto-corrélation	Ecart-type	limite inférieure IC 95%	limite supérieure IC 95%
1	0,00869666	0,0350285	-0,0686547	0,0686547
2	0,00295636	0,0350311	-0,0686599	0,0686599
3	-0,00969702	0,0350314	-0,0686605	0,0686605
4	0,00101131	0,0350347	-0,0686669	0,0686669
5	-0,0258978	0,0350348	-0,0686667	0,0686667
6	-0,00442852	0,0350582	-0,068713	0,068713
7	-0,0140358	0,0350589	-0,0687144	0,0687144
8	0,00201529	0,0350658	-0,0687279	0,0687279
9	0,0437353	0,035066	-0,0687282	0,0687282
10	-0,011016	0,0351328	-0,0688592	0,0688592
11	0,0482386	0,0351371	-0,0688675	0,0688675
12	-0,0129437	0,0352182	-0,0690266	0,0690266
13	0,116881	0,0352241	-0,069038	0,069038
14	-0,0207645	0,0356968	-0,0699645	0,0699645
15	-0,0134375	0,0357116	-0,0699936	0,0699936
16	-0,00279088	0,0357178	-0,0700057	0,0700057
17	0,00218871	0,0357181	-0,0700062	0,0700062
18	-0,0144758	0,0357182	-0,0700066	0,0700066
19	0,00142319	0,0357254	-0,0700207	0,0700207
20	-0,0278713	0,0357255	-0,0700208	0,0700208
21	-0,0505325	0,0357522	-0,0700731	0,0700731
22	-0,00730022	0,0358397	-0,0702446	0,0702446
23	-0,00500366	0,0358415	-0,0702482	0,0702482
24	0,0421413	0,0358424	-0,0702499	0,0702499

Tableau II.3.21 PACF des résidus du modèle SARIMA(1,1,1)(2,1,1)₁₂

Lag	Auto-corrélation partielle	Ecart-type	limite inférieure IC 95%	limite supérieure IC 95%
1	0,00869666	0,0350285	-0,0686547	0,0686547
2	0,00288095	0,0350285	-0,0686547	0,0686547
3	-0,00974853	0,0350285	-0,0686547	0,0686547
4	0,00117189	0,0350285	-0,0686547	0,0686547
5	-0,0258639	0,0350285	-0,0686547	0,0686547
6	-0,00408144	0,0350285	-0,0686547	0,0686547
7	-0,0138038	0,0350285	-0,0686547	0,0686547
8	0,00177715	0,0350285	-0,0686547	0,0686547
9	0,0438008	0,0350285	-0,0686547	0,0686547
10	-0,0127407	0,0350285	-0,0686547	0,0686547
11	0,0482141	0,0350285	-0,0686547	0,0686547
12	-0,0137569	0,0350285	-0,0686547	0,0686547
13	0,117221	0,0350285	-0,0686547	0,0686547
14	-0,0206803	0,0350285	-0,0686547	0,0686547
15	-0,0134073	0,0350285	-0,0686547	0,0686547
16	0,0031397	0,0350285	-0,0686547	0,0686547
17	0,000777819	0,0350285	-0,0686547	0,0686547
18	-0,00938042	0,0350285	-0,0686547	0,0686547
19	0,00152653	0,0350285	-0,0686547	0,0686547
20	-0,0302298	0,0350285	-0,0686547	0,0686547
21	-0,0500035	0,0350285	-0,0686547	0,0686547
22	-0,0195872	0,0350285	-0,0686547	0,0686547
23	-0,000110713	0,0350285	-0,0686547	0,0686547
24	0,0312905	0,0350285	-0,0686547	0,0686547

- **Test du Portmanteau**

Ce test est basé sur les 24 premières auto-corrélations des résidus, les résultats du test montrent une valeur de $P = 0,359086$ qui est supérieure ou égale à 0.10, donc on ne peut pas rejeter l'hypothèse nulle que les résidus sont indépendants avec un intervalle de confiance de 90% ou plus.

Test de Box-Pierce

Hypothèse nulle : résidus indépendants

Valeur calculée de la statistique $Q = 20,604$

Valeur de $P = 0,359086$

- **Test de nullité de la moyenne**

La moyenne des résidus est de 0,0070055 et leur médiane est de -0,421062. La variance autour de cette moyenne est de 39,37 avec un écart type de 6,27. La Figure II.3.13 montre la variabilité temporelle de ces derniers autour de leur moyenne. Les résultats du test t de Student montrent une valeur $P = 0,974567$ qui est supérieure à 0.05 ce qui implique que l'hypothèse nulle de nullité de la moyenne des résidus ne peut pas être rejetée à 95% de confiance.

Test t de Student

Hypothèse nulle : moyenne = 0,0

Alternative: non nulle.

Valeur calculée de la statistique $t = 0,0318732$

Valeur de $P = 0,974567$

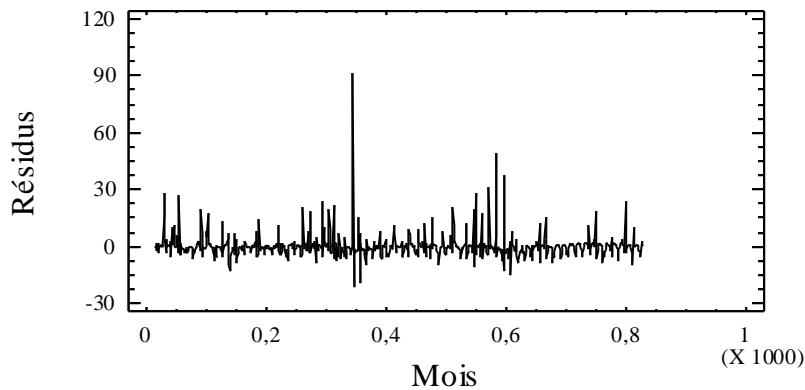


Fig. II.3.13 Résidus du modèle SARIMA(1,1,1)(2,1,1)₁₂

Les résultats de ces tests montrent bien que les résidus du modèle SARIMA(1,1,1)(2,1,1)₁₂ forment bel et bien une séquence bruit blanc de moyenne nulle et de variance 39,37. Le pourcentage de la variance expliquée par le modèle est de 62% d'une variance totale de 103,73.

II.3.2.2.5 Prédiction

En permettant au modèle retenu SARIMA(1,1,1)(2,1,1)₁₂ de faire de la prédiction en arrière nous pouvons obtenir les prédictions correspondant aussi bien à la période où les données sont disponibles qu'à celle où ces dernières ne le sont pas. Le Tableau II.3.22 présente les prédictions des 24 données futures pour la période où les données ne sont pas disponibles, avec les résidus correspondants. Les prédictions de ces 24 données futures sont présentées avec leurs intervalles de confiance à 95%. Ces limites de confiance indiquent où va être située la vraie valeur à une période déterminée dans le future, dans l'hypothèse que le modèle ajusté est adéquat aux données en question. Tandis que les prédictions correspondant à la période où les données sont disponibles avec les résidus correspondants, sont présentées en annexe A.II.3.2.

Les prédictions obtenues par le modèle, tel que présenté par la Fig. II.3.14, suivent en général l'allure des données observées qui sont représentées par des points carrés sur la dite figure. Pour la période où les données ne sont pas disponibles, les prédictions obtenues pour les 24 mois futurs suivant la fin de la série sont représentées avec leurs limites tracées à 95% de confiance.

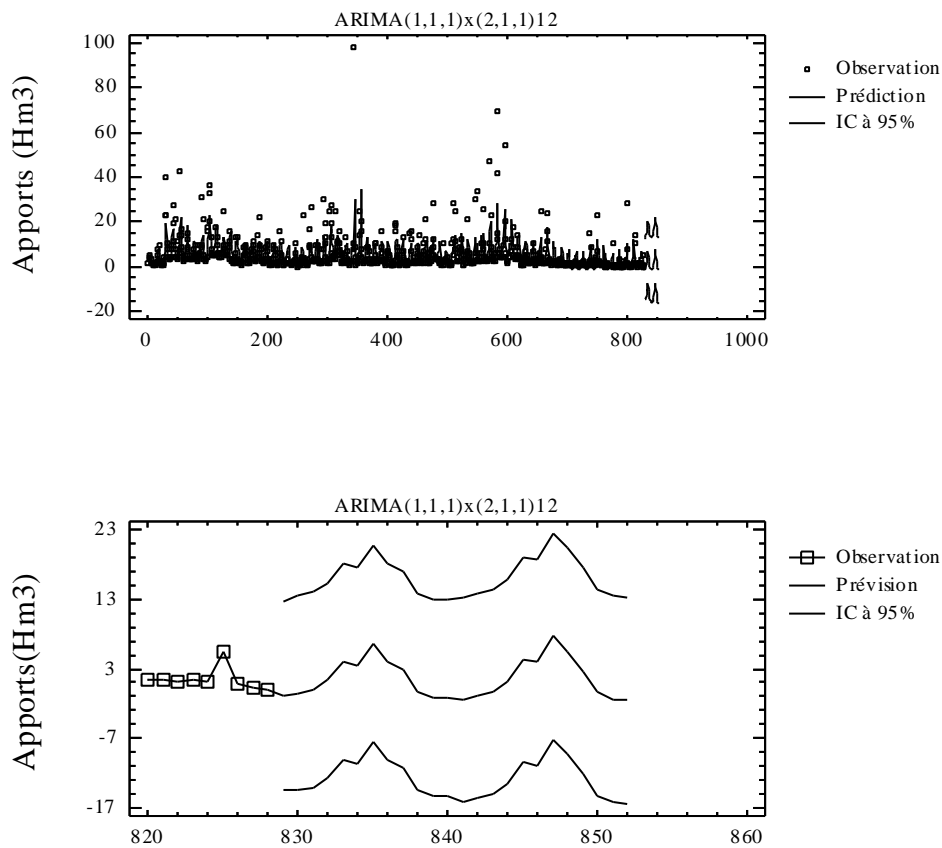


Fig. II.3.14 Prédictions des apports mensuels à BENI BAHDEL obtenues par le modèle SARIMA(1,1,1)(2,1,1)₁₂

Tableau II.3.22 Prévisions des apports mensuels à BENI BAHDEL pour les 24 mois futurs

Période	Prédiction	- IC 95%	+IC 95%
829	0,249854	12,6924	-14,3939
830	0,970351	13,4209	-14,6331
831	0,374523	14,0853	-14,085
832	-0,795815	15,4025	-12,8035
833	-3,30422	18,1087	-10,1205
834	-2,54111	17,652	-10,5979
835	-10,2541	20,7953	-7,47461
836	10,3599	18,2266	-10,0631
837	6,19341	17,0382	-11,2714
838	-0,103892	13,9521	-14,3772
839	1,09937	13,1002	-15,2489
840	1,14746	12,9985	-15,3704
841	0,77885	13,2727	-16,0854
842	0,25805	13,8201	-15,6722
843	1,03014	14,3289	-15,2108
844	-0,344629	15,9002	-13,6725
845	-2,97075	19,1901	-10,4127
846	-1,98632	18,7208	-10,9113
847	-3,11248	22,4972	-7,16409
848	-5,71012	20,4169	-9,27336
849	2,30439	17,5457	-12,1735
850	0,429024	14,473	-15,2753
851	1,08472	13,4843	-16,2929
852	0,790042	13,4065	-16,3996

II.3.2.3 BOUCHEGOUF

Les statistiques de la tendance centrale, de la variabilité et de la forme de la série chronologique des apports mensuels observés à BOUCHEGOUF durant les 408 mois allant de Septembre 1948 à Aout 1993 sont résumées dans le Tableau II.3.23 et la variabilité temporelle de cette série chronologique est illustrée par la Fig. II.3.15.

Tableau II.3.23 Statistiques des apports mensuels observés à BOUCHEGOUF (Sep1948 à Aout 1993)

Paramètre	Valeur
Nombre d'observations	408
Moyenne	9,66325
Variance	248,308
Déviatoin Standard	15,7578
Minimum	0,03
Maximum	110,28
Std. skewness	22,9601
Std. kurtosis	38,6696

La moyenne de la série est de 9,66325 alors que sa variance autour de cette moyenne est de 248,308 avec un écart-type de 15,7578. La skewness et la kurtosis standardisées indiquent un éloignement remarquable de la distribution normale.

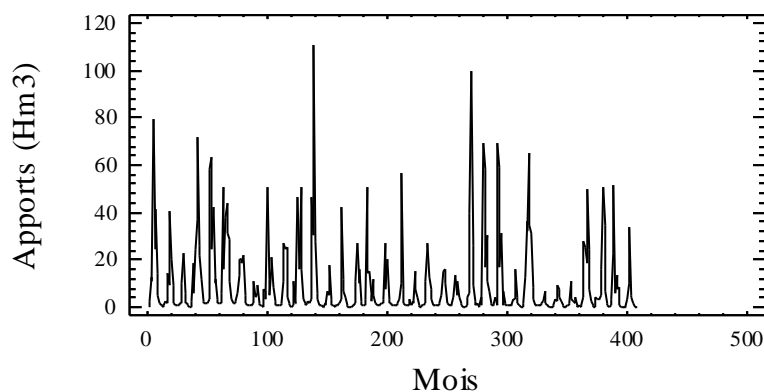


Fig. II.3.15 Apports mensuels observés à BOUCHEGOUF

II.3.2.3.1 Stationnarisation

Le détail du calcul des séries différenciées pour les 20 premiers mois est montré dans le Tableau II.3.24, tandis que les statistiques descriptives des séries différenciées pour l'ensemble des données sont représentées dans le Tableau II.3.25.

Tableau II.3.24 Calcul des séries différenciées (données des 20 premiers mois)

N°	Apport	DIFF(Apport,1)	SDIFF(Apports,12) = S	DIFF(S,1)
1	0,43			
2	2,31	1,88		
3	12,44	10,13		
4	10,74	-1,7		
5	79,35	68,61		
6	24,9	-54,45		
7	41,05	16,15		
8	14,36	-26,69		
9	5,27	-9,09		
10	1,71	-3,56		
11	0,99	-0,72		
12	0,71	-0,28		
13	0,4	-0,31	-0,03	
14	0,97	0,57	-1,34	-1,31
15	2,47	1,5	-9,97	-8,63
16	1,8	-0,67	-8,94	1,03
17	14,44	12,64	-64,91	-55,97
18	9,51	-4,93	-15,39	49,52
19	40,36	30,85	-0,69	14,7
20	20,76	-19,6	6,4	7,09

Dans le cas de la série étudiée, la stationnarité est obtenue en éliminant la saisonnalité ($S=12$) et ce effectuant les premières différences saisonnières ($D=1$). La variance des premières différences saisonnières $SDIFF(Apports,12)$ étant de 339.98 et la série est stationnaire pour $S=12$, $d=0$ et $D=1$.

Tableau II.3.25 Statistiques descriptives des séries différenciées pour les 408 mois.

Série	N	Moyenne	Variance	Ecart-type
Apports	408	9.66	248.308	15.75
DIFF(Apports,1)	407	-0.0007	261.634	16.17
SDIFF(Apports,12)	396	-0.341	339.98	18.43
DIFF(SDIFF(Apports,12),1)	395	-0.0001	493.14	22.2

Notation : DIFF (x,n) : différences régulières de x d'ordre n ; SDIFF(x,n,s) :différences saisonnières s de x d'ordre n.

II.3.2.3.2 Identification

Dans le cas de la station courante, les calculs de la ACF et de la PACF des apports stationnaires sont reportés dans les Tableaux II.3.26 et 27 respectivement et leurs tracés sont illustrés par les Figures II.3.16 et 17.

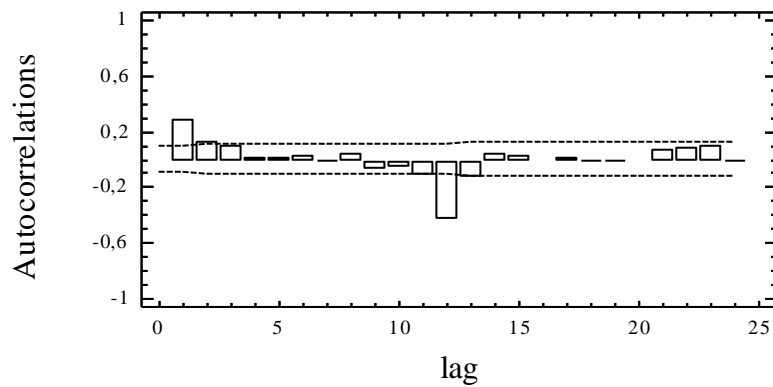


Fig. II.3.16 ACF des apports mensuels

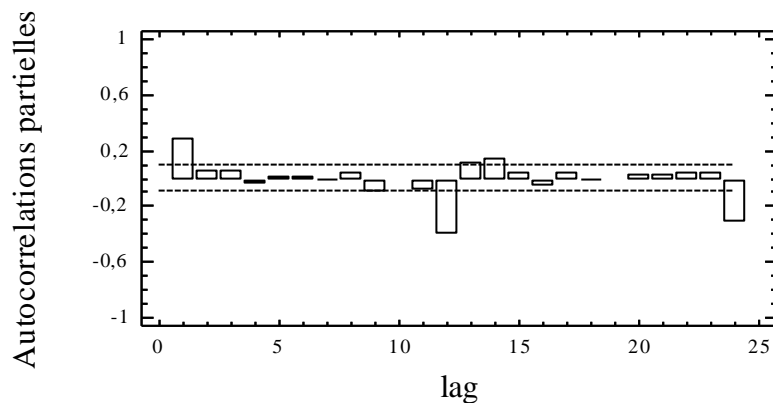


Fig. II.3.17 PACF des apports mensuels

Tableau II.3.26 ACF des apports mensuels stationnaires

Lag	Auto-corrélation	Ecart-type	-IC à 95%	IC à 95%
1	0,276591	0,0502519	-0,0984921	0,0984921
2	0,122876	0,0539595	-0,105759	0,105759
3	0,0879665	0,0546615	-0,107135	0,107135
4	0,0127953	0,0550179	-0,107833	0,107833
5	0,0135669	0,0550254	-0,107848	0,107848
6	0,0150291	0,0550338	-0,107865	0,107865
7	-0,00315512	0,0550442	-0,107885	0,107885
8	0,0309802	0,0550446	-0,107886	0,107886
9	-0,0826057	0,0550887	-0,107972	0,107972
10	-0,0703385	0,0554006	-0,108583	0,108583
11	-0,116615	0,0556256	-0,109024	0,109024
12	-0,447628	0,0562396	-0,110228	0,110228
13	-0,136616	0,0646132	-0,12664	0,12664
14	0,0299067	0,0653386	-0,128061	0,128061
15	0,0176017	0,0653731	-0,128129	0,128129
16	-0,0209395	0,0653851	-0,128153	0,128153
17	0,00905928	0,065402	-0,128186	0,128186
18	-0,0104101	0,0654052	-0,128192	0,128192
19	-0,0115154	0,0654094	-0,1282	0,1282
20	-0,0223167	0,0654145	-0,12821	0,12821
21	0,0679001	0,0654337	-0,128248	0,128248
22	0,0828908	0,0656114	-0,128596	0,128596
23	0,0920755	0,0658753	-0,129113	0,129113
24	-0,0133445	0,0661995	-0,129749	0,129749

Tableau II.3.27 PACF des apports mensuels stationnaires

Lag	Autocorrélation	Ecart-type	-IC à 95%	IC à 95%
1	0,276591	0,0502519	-0,0984921	0,0984921
2	0,0502145	0,0502519	-0,0984921	0,0984921
3	0,0453749	0,0502519	-0,0984921	0,0984921
4	-0,0297724	0,0502519	-0,0984921	0,0984921
5	0,00920575	0,0502519	-0,0984921	0,0984921
6	0,00843033	0,0502519	-0,0984921	0,0984921
7	-0,0093927	0,0502519	-0,0984921	0,0984921
8	0,0343403	0,0502519	-0,0984921	0,0984921
9	-0,108073	0,0502519	-0,0984921	0,0984921
10	-0,0283174	0,0502519	-0,0984921	0,0984921
11	-0,0906716	0,0502519	-0,0984921	0,0984921
12	-0,418715	0,0502519	-0,0984921	0,0984921
13	0,108863	0,0502519	-0,0984921	0,0984921
14	0,139254	0,0502519	-0,0984921	0,0984921
15	0,0330181	0,0502519	-0,0984921	0,0984921
16	-0,0691752	0,0502519	-0,0984921	0,0984921
17	0,0311774	0,0502519	-0,0984921	0,0984921
18	-0,00664805	0,0502519	-0,0984921	0,0984921
19	-0,0211003	0,0502519	-0,0984921	0,0984921
20	0,0147984	0,0502519	-0,0984921	0,0984921
21	0,0191194	0,0502519	-0,0984921	0,0984921
22	0,0400961	0,0502519	-0,0984921	0,0984921
23	0,034519	0,0502519	-0,0984921	0,0984921
24	-0,323252	0,0502519	-0,0984921	0,0984921

L'aspect des auto-corrélogrammes de la série stationnaire considérée laisserait supposer les modèles suivants :

- (A) SARIMA(1,0,1)(1,1,1)₁₂ avec constante
- (B) SARIMA(1,0,0)(0,1,1)₁₂ sans constante
- (C) SARIMA(1,0,0)(0,1,0)₁₂ avec constante
- (D) SARIMA(1,0,1)(1,1,0)₁₂ avec constante
- (E) SARIMA(1,0,1)(0,1,1)₁₂ avec constante

II.3.2.3.3 Estimation des modèles ajustés

Pour l'estimation des paramètres de ces modèles, il faut rappeler que 408 données ont été utilisées dont les 100 dernières ont servi à la validation. Les paramètres des modèles ont alors été estimés sur une période de 308 mois.

L'analyse et la comparaison des modèles ajustés tel que présenté dans le Tableau II.3.28, montre que les modèles ajustés (A), (B) et (E) produisent des résidus complètement indépendants tandis que les modèles (C) et (D) leurs résidus ne le sont pas. D'un autre côté, il faut noter que la moyenne de la variance des résidus produits par tous les modèles ne diffère pas beaucoup entre la période d'estimation et celle de validation. En outre, le modèle qui produit le minimum de cette variance avec le minimum de paramètres est le modèle (B), c'est lui qui est donc retenu pour la prédiction.

Tableau II.3.28 Comparaison des modèles ajustés

Modèle	Variance des résidus		Indépendance des résidus
	Période d'estimation	Période de validation	
(A)	178,223	104,35	OK
(B)	177,656	103,259	OK
(C)	336,671	258,034	NON
(D)	274,159	162,414	NON
(E)	177,749	101,806	OK

Le Tableau II.3.29 résume l'estimation des paramètres du modèle retenu SARIMA(1,0,0)(0,1,1)₁₂ et la signification statistique de ses termes. Les termes correspondant à une valeur de $P < 0.05$ sont statistiquement significativement différents de zéro à 95% de confiance.

Le paramètre AR(1) du modèle est stable. La valeur de P pour les termes AR(1) et SMA(1) est inférieure à 0.05 donc ces derniers sont significativement différents de zéro à 95% de confiance. La valeur de P pour la moyenne est très supérieure à 0.05 donc elle est statistiquement non significative elle a donc été éliminée, la constante aussi. L'estimation de l'écart-type du bruit blanc est de 13,8591 et la variance des résidus est de 192,076 ce qui veut dire que le présent modèle explique 47,7% de la variance de la série modélisée qui est de 339.98.

Tableau II.3.29 Estimation des paramètres du modèle retenu SARIMA(1,0,0)(0,1,1)₁₂

Paramètre	Estimation	Ecart- type	t	P
AR(1)	0,28448	0,0561124	5,06982	0,000001
SMA(1)	0,951109	0,010445	91,0586	0,000000

Notation : SAR(n): Autorégressif saisonnier d'ordre n; SMA(n) : Moyenne mobile saisonnière d'ordre n.

Avec une variance estimée du bruit blanc = 192,076 avec 394 DDL et un écart type estimé du bruit blanc = 13,8591

Le Tableau II.3.30 résume la performance du modèle SARIMA(1,0,0)(0,1,1)₁₂ à ajuster les données considérées et ce par l'analyse des erreurs qu'il produit.

Tableau II.3.30 Statistiques des erreurs du modèle SARIMA(1,0,0)(0,1,1)₁₂

Statistique	Période d'estimation	Période de validation
EQM	177,656	103,259
EAM	7,16911	6,00749
EM	-0,0528137	-0,975215

Nous y trouvons :

- 1) L'erreur quadratique moyenne (EQM)
- 2) L'erreur absolue moyenne (EAM)
- 3) L'erreur moyenne (EM)

Chacune de ces statistiques est basée sur l'erreur de prédiction un mois à l'avance qui est la différence entre la valeur de la donnée au temps t et sa prédiction au temps $t - 1$. Les deux premières mesurent la magnitude des erreurs alors que la dernière mesure le biais. Ces statistiques d'erreurs sont calculées pour la période d'estimation et celle de validation. Pour le modèle considéré ces statistiques d'erreurs ne sont pas différentes d'une manière considérable pour les deux périodes d'estimation et de validation. Il s'ensuit que le modèle ajusté peut être considéré comme assez performant et donc bon à faire des prévisions des valeurs futures.

Le modèle retenu pour la prédiction est alors le SARIMA(1,0,0)(0,1,1)₁₂ sans constante qui peut être formulé comme suit:

$$(1 - 0,28448B)(1 - B^0)(1 - B^{12})X_t = (1 - 0,951109B^{12})\varepsilon_t$$

II.3.2.3.4 Validation du modèle

Il s'agit de vérifier que les résidus forment bien un bruit blanc de moyenne nulle. Les tests des au-corrélogrammes et celui du Portmanteau (Box-Pierce) sont utilisés pour tester l'indépendance des résidus et le test t de student est utilisé pour tester la nullité de leur moyenne.

- **Test des ACF et PACF**

L'autocorrélogramme total et partiel des résidus du modèle ont été tracés tel que montré dans la Fig. II.3.18 et Fig. II.3.19.

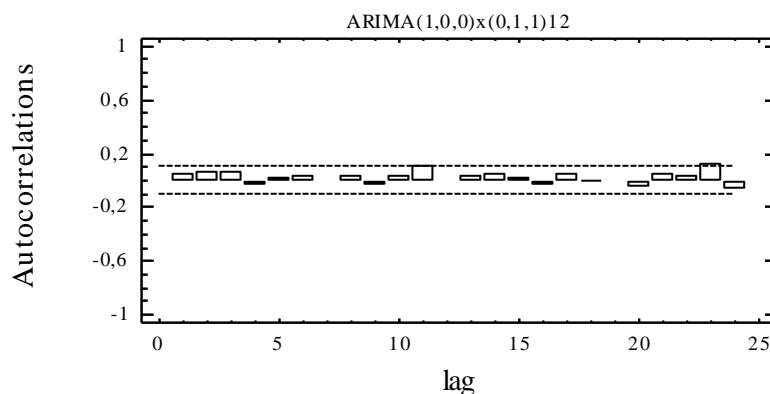


Fig. II.3.18 ACF des résidus du modèle SARIMA(1,0,0)(0,1,1)₁₂

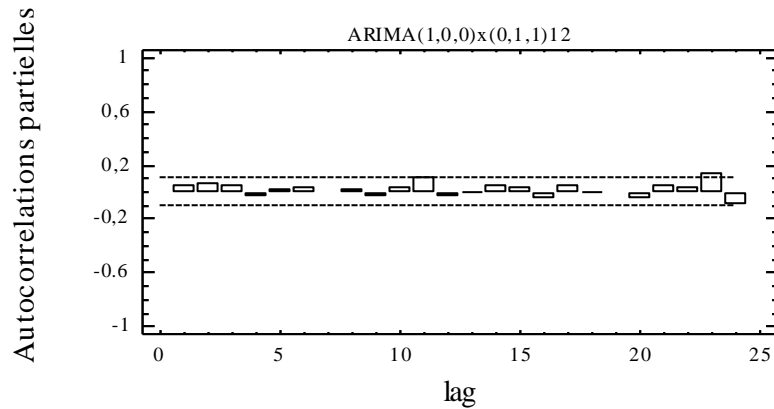


Fig. II.3.19 PACF des résidus du modèle SARIMA(1,0,0)(0,1,1)₁₂

Tous les coefficients sont inclus dans l'intervalle de confiance (IC) tracé à 95% de confiance ce qui implique leur indépendance. Leurs valeurs sont reportées dans le Tableau II.3.31 et Tableau II.3.32.

Tableau II.3.31 ACF des résidus du modèle SARIMA(1,0,0)(0,1,1)₁₂

Lag	Auto-corrélation	Ecart-type	limite inférieure IC 95%	limite supérieure IC 95%
1	0,0391622	0,0502519	-0,0984921	0,0984921
2	0,0656136	0,0503289	-0,098643	0,098643
3	0,0559033	0,0505445	-0,0990655	0,0990655
4	-0,0263226	0,0507003	-0,099371	0,099371
5	0,0210833	0,0507348	-0,0994387	0,0994387
6	0,0228041	0,050757	-0,099482	0,099482
7	-0,0142406	0,0507828	-0,0995327	0,0995327
8	0,0228289	0,0507929	-0,0995525	0,0995525
9	-0,0410343	0,0508188	-0,0996032	0,0996032
10	0,0250295	0,0509024	-0,0997671	0,0997671
11	0,110135	0,0509335	-0,099828	0,099828
12	-0,0213041	0,0515314	-0,101	0,101
13	0,0237274	0,0515536	-0,101043	0,101043
14	0,0470929	0,0515812	-0,101097	0,101097
15	0,0199235	0,0516896	-0,10131	0,10131
16	-0,0472446	0,051709	-0,101348	0,101348
17	0,0408822	0,0518179	-0,101561	0,101561
18	-0,00727474	0,0518993	-0,101721	0,101721
19	-0,0110491	0,0519019	-0,101726	0,101726
20	-0,0572574	0,0519078	-0,101738	0,101738
21	0,0375118	0,0520671	-0,10205	0,10205
22	0,0318886	0,0521353	-0,102183	0,102183
23	0,12084	0,0521845	-0,10228	0,10228
24	-0,069293	0,0528864	-0,103656	0,103656

Tableau II.3.32 PACF des résidus du modèle SARIMA(1,0,0)(0,1,1)₁₂

Lag	Auto-corrélation	Ecart-type	limite inférieure IC 95%	limite supérieure IC 95%
1	0,0391622	0,0502519	-0,0984921	0,0984921
2	0,0641784	0,0502519	-0,0984921	0,0984921
3	0,0512748	0,0502519	-0,0984921	0,0984921
4	-0,0345687	0,0502519	-0,0984921	0,0984921
5	0,0164743	0,0502519	-0,0984921	0,0984921
6	0,0227356	0,0502519	-0,0984921	0,0984921
7	-0,0151092	0,0502519	-0,0984921	0,0984921
8	0,0181983	0,0502519	-0,0984921	0,0984921
9	-0,0422795	0,0502519	-0,0984921	0,0984921
10	0,028372	0,0502519	-0,0984921	0,0984921
11	0,111098	0,0502519	-0,0984921	0,0984921
12	-0,0277319	0,0502519	-0,0984921	0,0984921
13	0,00567879	0,0502519	-0,0984921	0,0984921
14	0,0409012	0,0502519	-0,0984921	0,0984921
15	0,0262159	0,0502519	-0,0984921	0,0984921
16	-0,0657905	0,0502519	-0,0984921	0,0984921
17	0,0377215	0,0502519	-0,0984921	0,0984921
18	-0,00128793	0,0502519	-0,0984921	0,0984921
19	-0,0142177	0,0502519	-0,0984921	0,0984921
20	-0,0571108	0,0502519	-0,0984921	0,0984921
21	0,0420822	0,0502519	-0,0984921	0,0984921
22	0,0295015	0,0502519	-0,0984921	0,0984921
23	0,127401	0,0502519	-0,0984921	0,0984921
24	-0,0967427	0,0502519	-0,0984921	0,0984921

- **Test du Portmanteau**

Ce test est basé sur les 24 premières auto-corrélations des résidus, les résultats du test montrent une valeur de $P = 0,390168$ qui est supérieure ou égale à 0.10, donc on ne peut pas rejeter l'hypothèse nulle que les résidus sont indépendants avec un intervalle de confiance de 90% ou plus.

Test de Box-Pierce

Hypothèse nulle : résidus indépendants

Valeur calculée de la statistique $Q = 23,2062$

Valeur de $P = 0,390168$

- **Test de nullité de la moyenne**

La moyenne des résidus est de -0,285744, leur médiane est de -0,49401 et la variance est de 158,28. La Figure II.3.20 montre la variabilité temporelle de ces derniers autour de leur moyenne. Les résultats du test t de Student montrent une valeur $P = 0,651547$ qui est supérieure à 0.05 ce qui implique que l'hypothèse nulle de nullité de la moyenne des résidus ne peut pas être rejetée à 95% de confiance.

Test t de Student

Hypothèse nulle : moyenne = 0,0

Alternative: non nulle.

Valeur calculée de la statistique $t = -0,451958$

Valeur de $P = 0,651547$

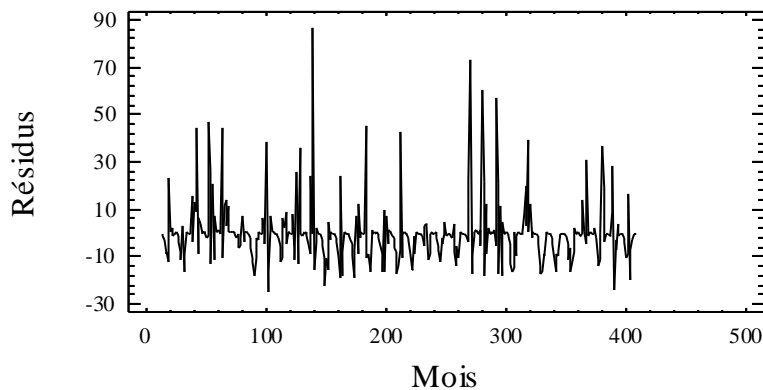


Fig. II.3.20 Résidus du modèle SARIMA(1,0,0)(0,1,1)₁₂

Les résultats de ces tests et les calculs des résidus montrent bien que les résidus du modèle SARIMA(1,0,0)(0,1,1)₁₂ forment bel et bien une séquence bruit blanc de moyenne nulle et de variance 158,28. Dans ce cas le modèle explique 53,4% de la variance totale.

II.3.2.3.5 Prédiction

En permettant au modèle retenu SARIMA(1,0,0)(0,1,1)₁₂ de faire de la prédiction en arrière nous pouvons obtenir les prédictions correspondant aussi bien à la période où les données sont disponibles qu'à celle où ces dernières ne le sont pas. Le Tableau II.3.33 présente les prédictions des 24 données futures pour la période où les données ne sont pas disponibles, avec les résidus correspondants. Les prédictions de ces 24 données futures sont présentées avec leurs intervalles de confiance à 95%. Ces limites de confiance indiquent où va être située la vraie valeur à une période déterminée dans le futur, dans l'hypothèse que le modèle ajusté est adéquat aux données en question. Tandis que les prédictions correspondant à la période où les données sont disponibles avec les résidus correspondants sont présentées en Annexe A.II.3.3.

Les prédictions obtenues par le modèle, tel que présenté par la Fig. II.3.21, suivent en général l'allure des données observées qui sont représentées par des points carrés sur la dite figure. Pour la période où les données ne sont pas disponibles, les prédictions obtenues pour les 24 mois futurs suivant la fin de la série sont représentées avec leurs limites tracées à 95% de confiance.

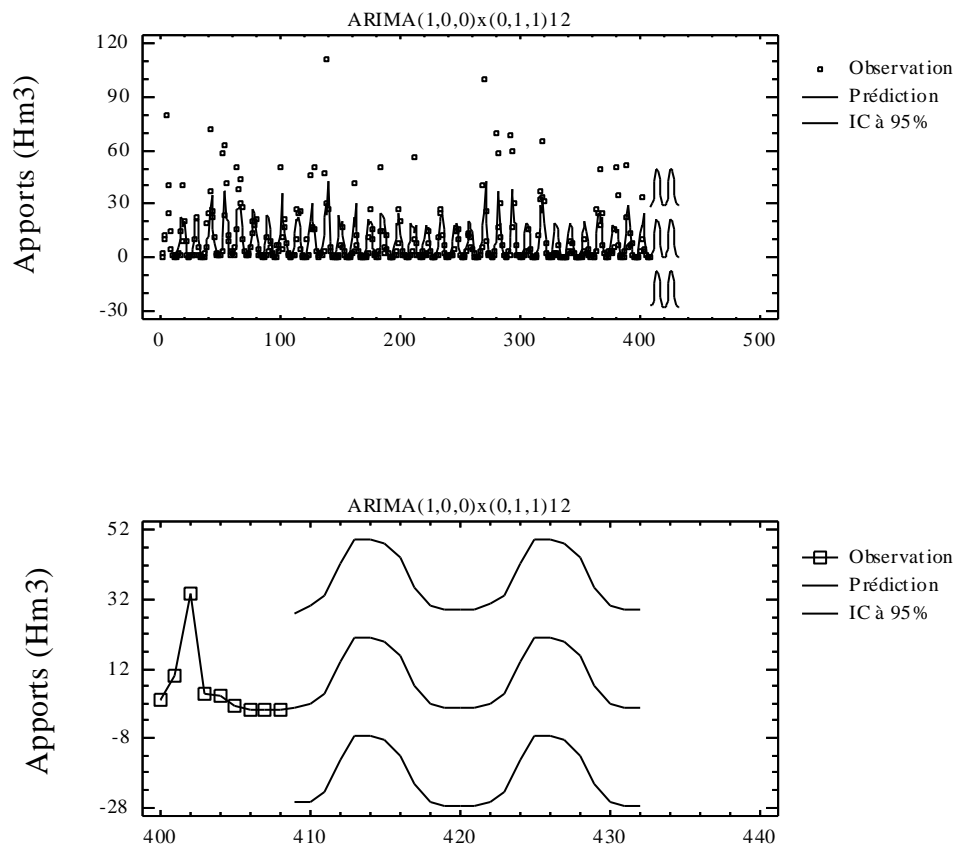


Fig. II.3.21 Prédictions des apports mensuels à BOUCHEGOUF obtenues par le modèle SARIMA(1,0,0)(0,1,1)₁₂

Tableau II.3.33 Prévisions des apports mensuels à BOUCHEGOUF pour les 24 mois futurs

Période	Prédiction	- IC 95%	+IC 95%
409	0,658104	-26,5891	27,9053
410	2,10024	-26,228	30,4285
411	4,70442	-23,7095	33,1184
412	13,999	-14,4219	42,4199
413	21,1228	-7,29868	49,5442
414	20,9754	-7,4461	49,3969
415	19,9026	-8,51891	48,3241
416	15,4847	-12,9368	43,9063
417	7,01135	-21,4101	35,4329
418	1,83974	-26,5818	30,2612
419	0,789193	-27,6323	29,2107
420	0,585309	-27,8362	29,0068
421	0,794173	-27,6585	29,2469
422	2,13895	-26,3163	30,5942
423	4,71544	-23,74	33,1709
424	14,0021	-14,4534	42,4575
425	21,1237	-7,33179	49,5791
426	20,9757	-7,47979	49,4311
427	19,9027	-8,55278	48,3581
428	15,4848	-12,9707	43,9402
429	7,01136	-21,4441	35,4668
430	1,83974	-26,6157	30,2952
431	0,789194	-27,6663	29,2446
432	0,58531	-27,8701	29,0408

II.3.2.4 BOUHNIFIA

Les statistiques de la tendance centrale, de la variabilité et de la forme de la série chronologique des apports mensuels observés à BOUHNIFIA durant les 708 mois allant de Septembre 1934 à Aout 1993 sont résumées dans le Tableau II.3.34 et la variabilité temporelle de cette série chronologique est illustrée par la Fig. II.3.22.

Tableau II.3.34 Statistiques des apports mensuel observés à BOUHNIFIA (Sep1934 à Aout 1993)

Paramètre	Valeur
Nombre d'observations	708
Moyenne	10,1176
Variance	146,124
Déviatoin Standard	12,0882
Minimum	0,12
Maximum	176,0
Std. skewness	73,0469
Std. kurtosis	381,094

La moyenne de la série est de 10,1176 alors que sa variance autour de cette moyenne est de 146,124 avec un écart-type de 12,0882. La skewness et la kurtosis standardisées indiquent un éloignement remarquable de la distribution normale.

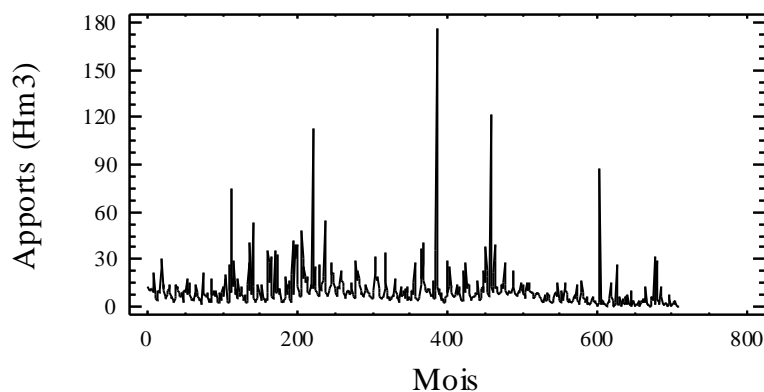


Fig. II.3.22 Apports mensuels observés à BOUHNIFIA

II.3.2.4.1 Stationnarisation

Le détail du calcul des séries différenciées pour les 20 premiers mois est montré dans le Tableau II.3.35, tandis que les statistiques des séries différenciées pour l'ensemble des données sont représentées dans le Tableau II.3.36.

Tableau II.3.35 Calcul des séries différenciées (données des 20 premiers mois)

N°	Apport	DIFF(Apport,1)	SDIFF(Apports,12) = S	DIFF(S,1)
1	12,3			
2	10,2	-2,1		
3	10,6	0,4		
4	11,5	0,9		
5	11,2	-0,3		
6	9,8	-1,4		
7	10,5	0,7		
8	9,6	-0,9		
9	21,6	12		
10	7,6	-14		
11	5,1	-2,5		
12	3,7	-1,4		
13	3,6	-0,1	-8,7	
14	10	6,4	-0,2	8,5
15	9,5	-0,5	-1,1	-0,9
16	9,2	-0,3	-2,3	-1,2
17	9,6	0,4	-1,6	0,7
18	24,6	15	14,8	16,4
19	30	5,4	19,5	4,7
20	11,8	-18,2	2,2	-17,3

Le Tableau II.3.36 résume les statistiques descriptives des séries différenciées. Dans le cas de la série étudiée, la stationnarité est obtenue en éliminant la tendance et la saisonnalité (S=12) et ce effectuant les premières différences régulières (d=1) des premières différences saisonnières (D=1). La variance des premières différences régulières des premières différences saisonnières DIFF(SDIFF(Apports,12),1) étant de 443.5 et la série est stationnaire pour S=12, d=1 et D=1.

Tableau II.3.36 Statistiques descriptives des séries différenciées pour les 708 mois.

Série	N	Moyenne	Variance	Ecart-type
Apports	708	10,11	146.12	12.08
DIFF(Apports,1)	707	-0.017	223.84	14.96
SDIFF(Apports,12)	696	-0.146	255.77	15.99
DIFF(SDIFF(Apports,12),1)	695	0.002	443.5	21.05

Notation : DIFF (x,n) : différences régulières de x d'ordre n ; SDIFF(x,n,s) :différences saisonnières s de x d'ordre n.

II.3.2.4.2 Identification

Dans le cas de la station courante, les calculs de la ACF et de la PACF des apports stationnaires sont reportés dans les Tableaux II.3.37 et II.3.38 respectivement et leurs tracés sont illustrés par les Figures II.3.23 et II.3.24.

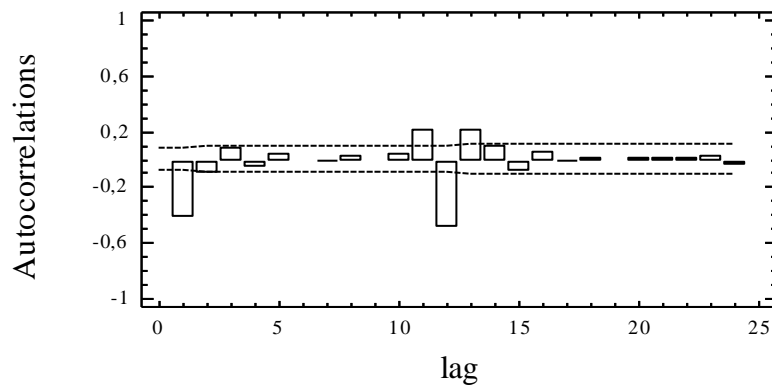


Fig. II.3.23 ACF des apports mensuels

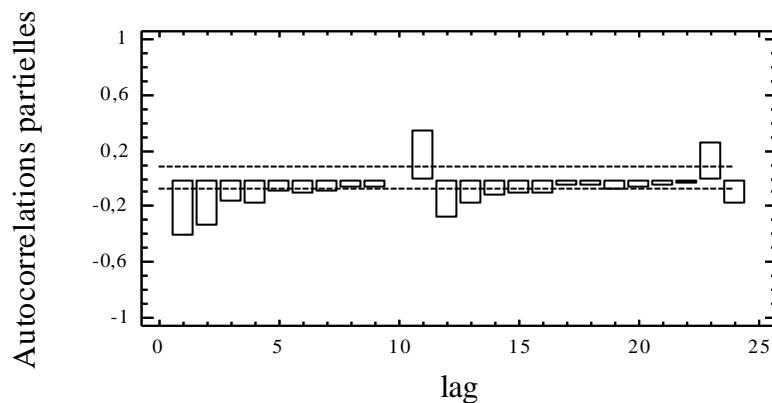


Fig. II.3.24 PACF des apports mensuels

Tableau II.3.37 ACF des apports mensuels stationnaires

Lag	Auto-corrélation	Ecart-type	-IC à 95%	IC à 95%
1	-0,433387	0,0379322	-0,0743458	0,0743458
2	-0,101515	0,0444899	-0,0871987	0,0871987
3	0,0830246	0,0448219	-0,0878495	0,0878495
4	-0,0640858	0,0450427	-0,0882822	0,0882822
5	0,0407213	0,0451737	-0,0885389	0,0885389
6	-0,025917	0,0452265	-0,0886424	0,0886424
7	-0,00490757	0,0452478	-0,0886843	0,0886843
8	0,0239423	0,0452486	-0,0886858	0,0886858
9	-0,0169343	0,0452668	-0,0887215	0,0887215
10	0,0331782	0,0452759	-0,0887393	0,0887393
11	0,203348	0,0453109	-0,0888079	0,0888079
12	-0,494532	0,0466055	-0,0913452	0,0913452
13	0,213981	0,0536269	-0,105107	0,105107
14	0,0869659	0,0548417	-0,107488	0,107488
15	-0,0886018	0,0550397	-0,107876	0,107876
16	0,0499628	0,0552446	-0,108278	0,108278
17	-0,00750388	0,0553096	-0,108405	0,108405
18	0,0118558	0,055311	-0,108408	0,108408
19	-0,0212852	0,0553147	-0,108415	0,108415
20	0,00288336	0,0553265	-0,108438	0,108438
21	0,00510866	0,0553267	-0,108439	0,108439
22	0,00121909	0,0553274	-0,10844	0,10844
23	0,021104	0,0553274	-0,10844	0,10844
24	-0,029835	0,055339	-0,108463	0,108463

Tableau II.3.38 PACF des apports mensuels stationnaires

Lag	Auto-corrélation	Ecart-type	-IC à 95%	IC à 95%
1	-0,433387	0,0379322	-0,0743458	0,0743458
2	-0,356252	0,0379322	-0,0743458	0,0743458
3	-0,184796	0,0379322	-0,0743458	0,0743458
4	-0,200179	0,0379322	-0,0743458	0,0743458
5	-0,115618	0,0379322	-0,0743458	0,0743458
6	-0,126292	0,0379322	-0,0743458	0,0743458
7	-0,112354	0,0379322	-0,0743458	0,0743458
8	-0,0797559	0,0379322	-0,0743458	0,0743458
9	-0,0761593	0,0379322	-0,0743458	0,0743458
10	-0,017804	0,0379322	-0,0743458	0,0743458
11	0,338954	0,0379322	-0,0743458	0,0743458
12	-0,291096	0,0379322	-0,0743458	0,0743458
13	-0,189465	0,0379322	-0,0743458	0,0743458
14	-0,144471	0,0379322	-0,0743458	0,0743458
15	-0,128829	0,0379322	-0,0743458	0,0743458
16	-0,129821	0,0379322	-0,0743458	0,0743458
17	-0,0686352	0,0379322	-0,0743458	0,0743458
18	-0,0672906	0,0379322	-0,0743458	0,0743458
19	-0,0963141	0,0379322	-0,0743458	0,0743458
20	-0,0861983	0,0379322	-0,0743458	0,0743458
21	-0,0686746	0,0379322	-0,0743458	0,0743458
22	-0,0312402	0,0379322	-0,0743458	0,0743458
23	0,25156	0,0379322	-0,0743458	0,0743458
24	-0,188623	0,0379322	-0,0743458	0,0743458

L'aspect des auto-corrélogrammes de la série stationnaire considérée laisserait supposer les modèles suivants :

- (A) SARIMA(1,1,1)(0,1,1)₁₂ sans constante
- (B) SARIMA(1,1,1)(1,1,1)₁₂ avec constante
- (C) SARIMA(1,1,1)(2,1,1)₁₂ sans constante
- (D) SARIMA(1,1,1)(2,1,2)₁₂ avec constante
- (E) SARIMA(1,1,1)(1,1,2)₁₂ avec constante

II.3.2.4.3 Estimation des modèles ajustés

Pour l'estimation des paramètres de ces modèles, il faut rappeler que 708 données ont été utilisées dont les 200 dernières ont servi à la validation. Les paramètres des modèles ont alors été estimés sur une période de 508 mois.

L'analyse et la comparaison des modèles ajustés tel que présenté dans le Tableau II.3.39, montre que tous les modèles ajustés produisent des résidus complètement indépendants. En outre, le modèle qui produit le minimum de cette variance est le modèle (C) mais le terme SAR(2) est non significativement différent de zéro ce qui nous pousse à le réduire, dans ce cas on tombe sur le modèle (B) dont le terme SAR(1) est à réduire car non significatif. Si on réduit le terme SAR(1) de ce modèle on tombe sur le modèle (A) dont tous les termes sont significativement différents de zéro. Les modèles (D) et (E) présentent également des termes SAR et SMA qui incitent à retrouver le modèle (A). C'est le (A) qui est donc retenu pour la prédiction.

Tableau II.3.39 Comparaison des modèles ajustés

Modèle	Variance des résidus		Indépendance des résidus
	Période d'estimation	Période de validation	
(A)	155,908	64,0364	OK
(B)	155,442	65,6462	OK
(C)	154,409	67,8985	OK
(D)	155,4	67,0217	OK
(E)	156,526	64,612	OK

Le Tableau II.3.40 résume l'estimation des paramètres du modèle retenu SARIMA(1,1,1)(0,1,1)₁₂ et la signification statistique de ses termes. Les termes correspondant à une valeur de $P < 0.05$ sont statistiquement significativement différents de zéro à 95% de confiance.

Les paramètres du modèle sont stables, et la valeur du terme AR(1) est différente de Un (1). La valeur de P pour les termes AR(1), MA(1) et SMA(1) est inférieure à 0.05 donc ces derniers sont significativement différents de zéro à 95% de confiance. La valeur de P pour la moyenne est très supérieure à 0.05 donc elle est statistiquement non significative elle a donc été éliminée, la constante aussi. L'estimation de l'écart-type du bruit blanc est de 12,4403 et la moyenne de la variance des résidus est de 154,762 ce qui veut dire que le présent modèle explique 65% de la variance de la série modélisée qui est de 443.5.

Tableau II.3.40 Estimation des paramètres du modèle retenu SARIMA(1,1,1)(0,1,1)₁₂

Paramètre	Estimation	Ecart- type	t	P
AR(1)	0,135019	0,0466008	2,89734	0,003763
MA(1)	0,959937	0,0127462	75,3115	0,000000
SMA(1)	0,968824	0,00880741	110,001	0,000000

Notation : SAR(n): Autorégressif saisonnier d'ordre n; SMA(n) : Moyenne mobile saisonnière d'ordre n.

La variance estimée du bruit blanc = 154,762 avec 692 DDL et l'écart type estimé du bruit blanc = 12,4403

Le Tableau II.3.41 résume la performance du modèle SARIMA(1,1,1)(0,1,1)₁₂ à ajuster les données considérées et ce par l'analyse des erreurs qu'il produit.

Tableau II.3.41 Statistiques des erreurs du modèle SARIMA(1,1,1)(0,1,1)₁₂

Statistique	Période d'estimation	Période de validation
EQM	155,908	64,0364
EAM	5,57763	4,13991
EM	1,20376	-0,106058

Nous y trouvons :

- 1) L'erreur quadratique moyenne (EQM)
- 2) L'erreur absolue moyenne (EAM)
- 3) L'erreur moyenne (EM)

Chacune de ces statistiques est basée sur l'erreur de prédiction un mois à l'avance qui est la différence entre la valeur de la donnée au temps t et sa prédiction au temps $t - 1$. Les deux premières mesurent la magnitude des erreurs alors que la dernière mesure le biais. Ces statistiques d'erreurs sont calculées pour la période d'estimation et celle de validation. Pour le modèle considéré les statistiques d'erreurs EAM et EM ne sont pas différentes d'une manière considérable pour les deux périodes d'estimation et de validation, par contre la EQM l'est. Il s'ensuit que le modèle ajusté peut être utilisé à faire des prévisions, cependant celles ci auront éventuellement une variance sous- estimée.

Le modèle retenu pour la prédiction est alors le SARIMA(1,1,1)(0,1,1)₁₂ sans constante qui peut être formulé comme suit:

$$(1 - 0,135019B)(1 - B)(1 - B^{12})X_t = (1 - 0,959937B)(1 - 0,968824B^{12})\varepsilon_t$$

ε_t est un bruit blanc de moyenne nulle et de variance égale à 154.762.

II.3.2.4.4 Validation du modèle SARIMA(1,1,1)(0,1,1)₁₂

Il s'agit de vérifier que les résidus forment bien un bruit blanc de moyenne nulle. Les tests des autocorrélogrammes et celui du Portmanteau (Box-Pierce) sont utilisés pour tester l'indépendance des résidus et le test t de student est utilisé pour tester la nullité de leur moyenne.

- **Test des ACF et PACF**

L'autocorrélogramme total et partiel des résidus du modèle ont été tracés tel que montré dans la Fig. II.3.25 et Fig. II.3.26.

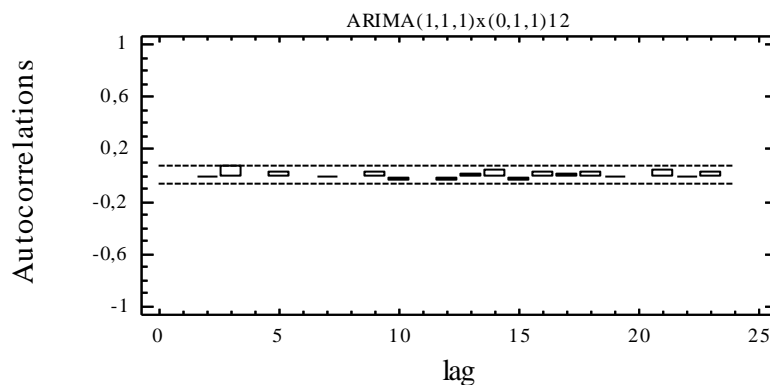


Fig. II.3.25 ACF des résidus du modèle SARIMA(1,1,1)(0,1,1)₁₂

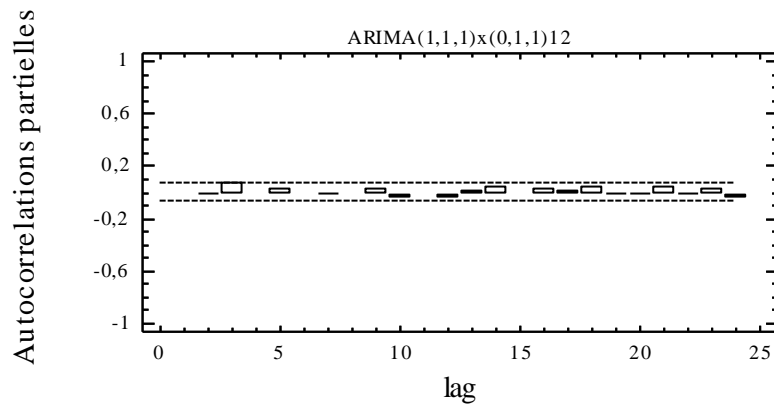


Fig. II.3.26 PACF des résidus du modèle SARIMA(1,1,1)(0,1,1)₁₂

Tous les coefficients sont inclus dans l'intervalle de confiance (IC) tracé à 95% de confiance ce qui implique leur indépendance. Leurs valeurs sont reportées dans le Tableau II.3.42 et Tableau II.3.43.

Tableau II.3.42 ACF des résidus du modèle SARIMA(1,1,1)(0,1,1)₁₂

Lag	Auto-corrélation	Ecart-type	limite inférieure IC 95%	limite supérieure IC 95%
1	-0,0207234	0,0379322	-0,0743458	0,0743458
2	-0,0125776	0,0379484	-0,0743777	0,0743777
3	0,0724037	0,0379544	-0,0743895	0,0743895
4	-0,0240808	0,0381527	-0,074778	0,074778
5	0,0272239	0,0381745	-0,0748208	0,0748208
6	-0,0189186	0,0382025	-0,0748756	0,0748756
7	-0,00454643	0,0382159	-0,074902	0,074902
8	-0,0183073	0,0382167	-0,0749035	0,0749035
9	0,0239985	0,0382293	-0,0749282	0,0749282
10	-0,0325746	0,038251	-0,0749707	0,0749707
11	-0,0238803	0,0382909	-0,0750489	0,0750489
12	-0,0433426	0,0383123	-0,0750909	0,0750909
13	0,00326941	0,0383828	-0,075229	0,075229
14	0,0434559	0,0383832	-0,0752298	0,0752298
15	-0,0352955	0,0384539	-0,0753684	0,0753684
16	0,0224377	0,0385005	-0,0754598	0,0754598
17	0,0100994	0,0385193	-0,0754966	0,0754966
18	0,0285576	0,0385231	-0,0755041	0,0755041
19	-0,00471425	0,0385536	-0,0755638	0,0755638
20	-0,0153259	0,0385544	-0,0755654	0,0755654
21	0,0430732	0,0385632	-0,0755826	0,0755826
22	-0,00961075	0,0386323	-0,0757181	0,0757181
23	0,0203204	0,0386358	-0,0757249	0,0757249
24	-0,0289239	0,0386511	-0,075755	0,075755

Tableau II.3.43 PACF des résidus du modèle SARIMA(1,1,1)(0,1,1)₁₂

Lag	Auto-corrélation partielle	Ecart-type	limite inférieure IC 95%	limite supérieure IC 95%
1	-0,0207234	0,0379322	-0,0743458	0,0743458
2	-0,0130127	0,0379322	-0,0743458	0,0743458
3	0,0719131	0,0379322	-0,0743458	0,0743458
4	-0,0214065	0,0379322	-0,0743458	0,0743458
5	0,02826	0,0379322	-0,0743458	0,0743458
6	-0,0238063	0,0379322	-0,0743458	0,0743458
7	-0,00133125	0,0379322	-0,0743458	0,0743458
8	-0,0236851	0,0379322	-0,0743458	0,0743458
9	0,0277806	0,0379322	-0,0743458	0,0743458
10	-0,0339035	0,0379322	-0,0743458	0,0743458
11	-0,0202498	0,0379322	-0,0743458	0,0743458
12	-0,0506727	0,0379322	-0,0743458	0,0743458
13	0,00834573	0,0379322	-0,0743458	0,0743458
14	0,0420254	0,0379322	-0,0743458	0,0743458
15	-0,0250296	0,0379322	-0,0743458	0,0743458
16	0,0192678	0,0379322	-0,0743458	0,0743458
17	0,00647616	0,0379322	-0,0743458	0,0743458
18	0,0319661	0,0379322	-0,0743458	0,0743458
19	-0,00975473	0,0379322	-0,0743458	0,0743458
20	-0,0131308	0,0379322	-0,0743458	0,0743458
21	0,0374546	0,0379322	-0,0743458	0,0743458
22	-0,00809454	0,0379322	-0,0743458	0,0743458
23	0,0166631	0,0379322	-0,0743458	0,0743458
24	-0,0308093	0,0379322	-0,0743458	0,0743458

- **Test du Portmanteau**

Ce test est basé sur les 24 premières auto-corrélations des résidus, les résultats du test montrent une valeur de $P = 0,874712$ qui est supérieure ou égale à 0.10, donc on ne peut pas rejeter l'hypothèse nulle que les résidus sont indépendants avec un intervalle de confiance de 90% ou plus.

Test de Box-Pierce

Hypothèse nulle : résidus indépendants

Valeur calculée de la statistique $Q = 13,8797$

Valeur de $P = 0,874712$

- **Test de nullité de la moyenne**

La moyenne des résidus est de 0,826831, leur médiane est de -0,853599 et leur variance est de 128,29. La Figure II.3.27 montre la variabilité temporelle de ces derniers autour de leur moyenne. Les résultats du test t de Student montrent une valeur $P = 0,054302$ qui est supérieure à 0.05 ce qui implique que l'hypothèse nulle de nullité de la moyenne des résidus ne peut pas être rejetée à 95% de confiance.

Test t de Student

Hypothèse nulle : moyenne = 0,0

Alternative: non nulle.

Valeur calculée de la statistique $t = 1,92442$

Valeur de $P = 0,054302$

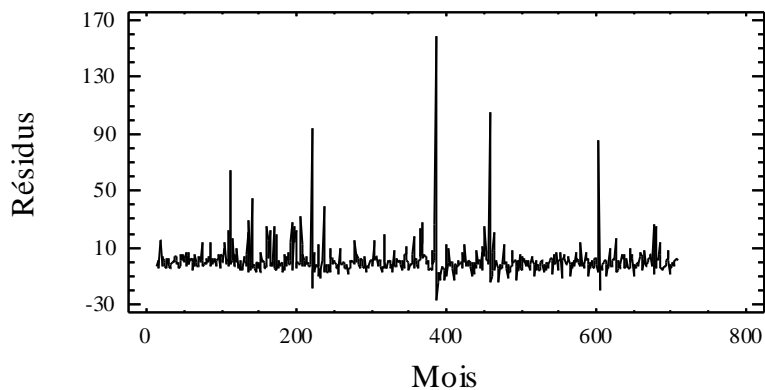


Fig. II.3.27 Résidus du modèle SARIMA(1,1,1)(0,1,1)₁₂

Les résultats de ces tests montrent bien que les résidus du modèle SARIMA(1,1,1)(0,1,1)₁₂ forment bel et bien une séquence bruit blanc de moyenne nulle et de variance 128.29. Dans ce cas, le modèle explique 71% de la variance totale.

II.3.2.4.5 Prédiction

En permettant au modèle retenu SARIMA(1,1,1)(0,1,1)₁₂ de faire de la prédiction en arrière nous pouvons obtenir les prédictions correspondant aussi bien à la période où les données sont disponibles qu'à celle où ces dernières ne le sont pas. Le Tableau II.3.44 présente les prédictions des 24 données futures pour la période où les données ne sont pas disponibles, avec les résidus correspondants. Les prédictions de ces 24 données futures sont présentées avec leurs intervalles de confiance à 95%. Ces limites de confiance indiquent où va être située la vraie valeur à une période déterminée dans le future, dans l'hypothèse que le modèle ajusté est adéquat aux données en question. Tandis que les prédictions correspondant à la période où les données sont disponibles avec les résidus correspondants sont présentées en annexe A.II.3.4.

Les prédictions obtenues par le modèle, tel que présenté par la Fig. II.3.28, suivent en général l'allure des données observées qui sont représentées par des points carrés sur la dite figure. Pour la période où les données ne sont pas disponibles, les prédictions obtenues pour les 24 mois futurs suivant la fin de la série sont représentées avec leurs limites tracées à 95% de confiance.

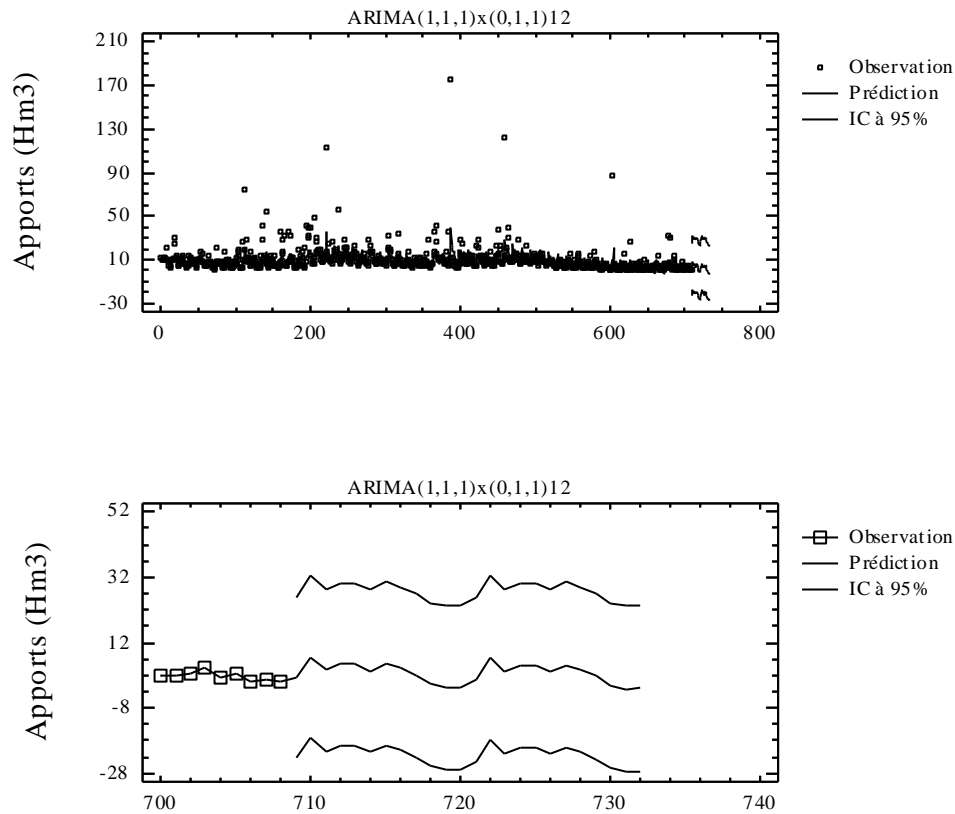


Fig. II.3.28 Prédications des apports mensuels à BOUHNIFIA obtenues par le modèle SARIMA(1,1,1)(0,1,1)₁₂

Tableau II.3.44 Prévisions des apports mensuels à BOUHNIFIA pour les 24 mois futurs

Période	Prédiction	- IC 95%	+IC 95%
709	1,16228	-23,2204	25,5449
710	7,66863	-17,0849	32,4222
711	3,64192	-21,1603	28,4441
712	5,41291	-19,4177	30,2435
713	5,5036	-19,353	30,3602
714	3,40301	-21,4793	28,2853
715	5,50245	-19,4055	30,4104
716	4,10678	-20,8267	29,0403
717	2,02461	-22,9345	26,9837
718	-0,802366	-25,787	24,1823
719	-1,86629	-26,8764	23,1438
720	-1,68598	-26,7216	23,3496
721	0,566804	-24,54	25,6736
722	7,25617	-17,8824	32,3947
723	3,25418	-21,9119	28,4203
724	5,02851	-20,1646	30,2216
725	5,11965	-20,1004	30,3397
726	3,01911	-22,2278	28,266
727	5,11856	-20,1552	30,3923
728	3,72289	-21,5777	29,0235
729	1,64073	-23,6866	26,9681
730	-1,18625	-26,5404	24,1679
731	-2,25018	-27,631	23,1307
732	-2,06987	-27,4774	23,3377

II.3.2.5 CHEFFIA

Les statistiques de la tendance centrale, de la variabilité et de la forme de la série chronologique des apports mensuels observés à CHEFFIA durant les 324 mois allant de Septembre 1966 à Aout 1992 sont résumées dans le Tableau II.3.45 et la variabilité temporelle de cette série chronologique est illustrée par la Fig. II.3.29.

Tableau II.3.45 Statistiques des apports mensuel observés à CHEFFIA (Sep1966 à Aout 1992)

Paramètre	Valeur
Nombre d'observations	324
Moyenne	11,3084
Variance	328,047
Déviatiion Standard	18,1121
Minimum	0,001
Maximum	116,931
Stnd. skewness	23,2793
Stnd. kurtosis	42,5552

La moyenne de la série est de 11,3084 alors que sa variance autour de cette moyenne est de 328,047 avec un écart-type de 18,1121. La skewness et la kurtosis standardisées indiquent un éloignement remarquable de la distribution normale.

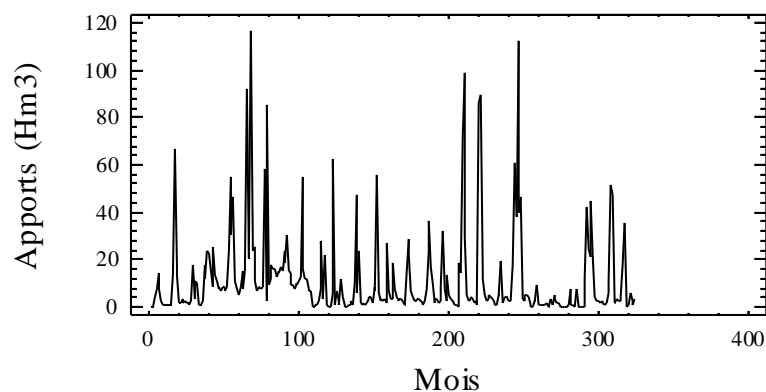


Fig. II.3.29 Apports mensuels observés à CHEFFIA

II.3.2.5.1 Stationnarisation

Le détail du calcul des séries différenciées pour les 20 premiers mois est montré dans le Tableau II.3.46, tandis que les statistiques des séries différenciées pour l'ensemble des données sont représentées dans le Tableau II.3.47.

Tableau II.3.46 Calcul des séries différenciées (données des 20 premiers mois)

N°	Apport	DIFF(Apport,1)	SDIFF(Apports,12) = S	DIFF(S,1)
1	0,412			
2	0,104	-0,308		
3	0,25	0,146		
4	4,56	4,31		
5	7,323	2,763		
6	14,13	6,807		
7	7,963	-6,167		
8	2,357	-5,606		
9	1,33	-1,027		
10	0,65	-0,68		
11	0,666	0,016		
12	0,892	0,226		
13	0,788	-0,104	0,376	
14	0,638	-0,15	0,534	0,158
15	0,812	0,174	0,562	0,028
16	13,713	12,901	9,153	8,591
17	67,171	53,458	59,848	50,695
18	18,191	-48,98	4,061	-55,787
19	13,256	-4,935	5,293	1,232
20	2,137	-11,119	-0,22	-5,513

Dans le cas de la série étudiée, la stationnarité est obtenue en éliminant la tendance et la saisonnalité ($S=12$) et ce effectuant les premières différences régulières ($d=1$) des premières différences saisonnières ($D=1$). La variance des premières différences régulières des premières différences saisonnières $DIFF(SDIFF(Apports,12),1)$ est de 849.91 et la série est stationnaire pour $S=12$, $d=1$ et $D=1$.

Tableau II.3.47 Statistiques descriptives des séries différenciées pour les 324 mois.

Série	N	Moyenne	Variance	Ecart-type
Apports	324	11.30	328.04	18.11
DIFF(Apports,1)	323	0.008	407.85	20.19
SDIFF(Apports,12)	312	0.120	557.67	23.61
DIFF(SDIFF(Apports,12),1)	311	-0.002	849.91	29.15

Notation : DIFF (x,n) : différences régulières de x d'ordre n ; SDIFF(x,n,s) : différences saisonnières s de x d'ordre n.

II.3.2.5.2 Identification

Dans le cas de la station courante, les calculs de l'ACF et de la PACF des apports stationnaires sont reportés dans les Tableaux II.3.48 et 49 respectivement et leurs tracés sont illustrés par les Figures II.3.30 et 31.

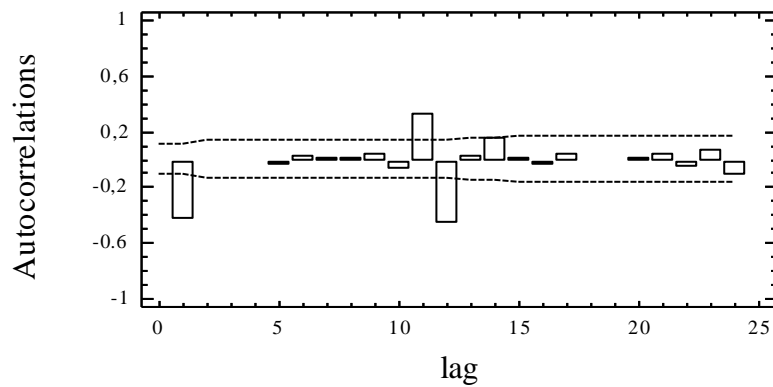


Fig. II.3.30 ACF des apports mensuels

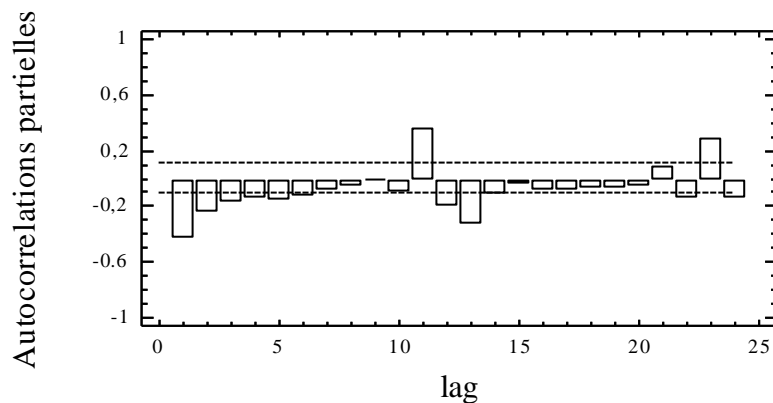


Fig. II.3.31 PACF des apports mensuels

Tableau II.3.48 ACF des apports mensuels stationnaires

Lag	Autocorrélation	Ecart-type	-IC à 95%	IC à 95%
1	-0,434709	0,0567048	-0,11114	0,11114
2	-0,0154146	0,0665634	-0,130462	0,130462
3	-0,0166319	0,0665749	-0,130485	0,130485
4	-0,0165181	0,0665882	-0,130511	0,130511
5	-0,0343466	0,0666014	-0,130537	0,130537
6	0,0183409	0,0666583	-0,130648	0,130648
7	0,00932061	0,0666746	-0,13068	0,13068
8	0,0066934	0,0666788	-0,130688	0,130688
9	0,0297936	0,0666809	-0,130692	0,130692
10	-0,0795147	0,0667237	-0,130776	0,130776
11	0,323075	0,0670277	-0,131372	0,131372
12	-0,468796	0,0718606	-0,140845	0,140845
13	0,0277769	0,0811003	-0,158954	0,158954
14	0,158462	0,0811309	-0,159014	0,159014
15	0,00374978	0,0821201	-0,160953	0,160953
16	-0,0354991	0,0821206	-0,160954	0,160954
17	0,0422146	0,0821699	-0,16105	0,16105
18	-0,0155592	0,0822396	-0,161187	0,161187
19	-0,0162684	0,0822491	-0,161206	0,161206
20	0,00899238	0,0822594	-0,161226	0,161226
21	0,0365965	0,0822626	-0,161232	0,161232
22	-0,0597269	0,0823149	-0,161335	0,161335
23	0,0658312	0,0824542	-0,161608	0,161608
24	-0,119466	0,082623	-0,161938	0,161938

Tableau II.3.49 PACF des apports mensuels stationnaires

Lag	Auto-corrélation	Ecart-type	-IC à 95%	IC à 95%
1	-0,434709	0,0567048	-0,11114	0,11114
2	-0,252009	0,0567048	-0,11114	0,11114
3	-0,17718	0,0567048	-0,11114	0,11114
4	-0,147763	0,0567048	-0,11114	0,11114
5	-0,166012	0,0567048	-0,11114	0,11114
6	-0,132149	0,0567048	-0,11114	0,11114
7	-0,0982381	0,0567048	-0,11114	0,11114
8	-0,0706576	0,0567048	-0,11114	0,11114
9	-0,011734	0,0567048	-0,11114	0,11114
10	-0,101843	0,0567048	-0,11114	0,11114
11	0,353991	0,0567048	-0,11114	0,11114
12	-0,210685	0,0567048	-0,11114	0,11114
13	-0,345173	0,0567048	-0,11114	0,11114
14	-0,123181	0,0567048	-0,11114	0,11114
15	-0,0415568	0,0567048	-0,11114	0,11114
16	-0,0955836	0,0567048	-0,11114	0,11114
17	-0,0992758	0,0567048	-0,11114	0,11114
18	-0,0835065	0,0567048	-0,11114	0,11114
19	-0,0734939	0,0567048	-0,11114	0,11114
20	-0,071731	0,0567048	-0,11114	0,11114
21	0,0847309	0,0567048	-0,11114	0,11114
22	-0,150208	0,0567048	-0,11114	0,11114
23	0,284796	0,0567048	-0,11114	0,11114
24	-0,154192	0,0567048	-0,11114	0,11114

L'aspect des auto-corrélogrammes de la série stationnaire considérée laisserait supposer les modèles suivants :

- (A) SARIMA(0,1,1)(0,1,1)₁₂ sans constante
- (B) SARIMA(0,1,2)(0,1,1)₁₂ avec constante
- (C) SARIMA(0,1,1)(1,1,1)₁₂ sans constante
- (D) SARIMA(1,1,1)(1,1,1)₁₂ sans constante
- (E) SARIMA(1,1,1)(0,1,1)₁₂ sans constante

II.3.2.5.3 Estimation des modèles ajustés

Pour l'estimation des paramètres de ces modèles, il faut rappeler que 324 données ont été utilisées dont les 100 dernières ont servi à la validation. Les paramètres des modèles ont alors été estimés sur une période de 224 mois.

L'analyse et la comparaison des modèles ajustés tel que présenté dans le Tableau II.3.50, montre que tous les modèles ajustés produisent des résidus complètement indépendants. D'un autre côté, il faut noter que la moyenne de la variance des résidus ne diffère pas beaucoup entre la période d'estimation et celle de validation, ceci est valable pour tous les modèles ajustés. En outre, le modèle qui produit le minimum de cette variance est le modèle (E), les autres modèles présentent tous des termes qui sont non significatifs et quand ils sont réduits on est renvoyé vers le modèle (A) ou (E). Le choix entre ces deux modèles est fait par rapport aux résidus. L'ACF des résidus du modèle (A) présente un terme au 1er décalage qui sort de l'intervalle de confiance, par contre le modèle (E) produit des résidus qui sont tous inclus dans l'intervalle tracé à 95% de confiance c'est lui qui est donc retenu pour la prédiction.

Tableau II.3.50 Comparaison des modèles ajustés

Modèle	Variance des résidus		Indépendance des résidus
	Période d'estimation	Période de validation	
(A)	302,998	270,195	OK
(B)	302,483	247,677	OK
(C)	304,258	271,385	OK
(D)	302,738	246,111	OK
(E)	301,222	246,25	OK

Le Tableau II.3.51 résume l'estimation des paramètres du modèle retenu SARIMA(1,1,1)(0,1,1)₁₂ et la signification statistique de ses termes. Les termes correspondant à une valeur de $P < 0.05$ sont statistiquement significativement différents de zéro à 95% de confiance.

Les paramètres du modèle sont stables, et la valeur du terme AR(1) est différente de Un (1). La valeur de P pour les termes MA(1) et SMA(1) est inférieure à 0.05 donc ces derniers sont significativement différents de zéro à 95% de confiance et le terme AR(1) peut être éliminé car statistiquement il est non significativement différent de zéro. La valeur de P pour la moyenne est très supérieure à 0.05 donc elle est statistiquement non significative elle a donc été éliminée, la constante aussi. L'estimation de l'écart-type du bruit blanc est de 17,49 et la moyenne de la variance des résidus est de 306,049 ce qui veut dire que le présent modèle explique 64% de la variance de la série modélisée qui est de 849.9.

Tableau II.3.51 Estimation des paramètres du modèle retenu SARIMA(1,1,1)(0,1,1)₁₂

Paramètre	Estimation	Ecart- type	t	P
AR(1)	0,109257	0,0743173	1,47015	0,142543
MA(1)	0,921139	0,0237703	39,3594	0,000000
SMA(1)	0,896051	0,0169778	52,7778	0,000000

Notation : SAR(n): Autorégressif saisonnier d'ordre n; SMA(n) : Moyenne mobile saisonnière d'ordre n.

La variance estimée du bruit blanc = 306,049 avec 308 DDL et l'écart type estimé du bruit blanc = 17,4942

Le Tableau II.3.52 résume la performance du modèle SARIMA(1,1,1)(0,1,1)₁₂ à ajuster les données considérées et ce par l'analyse des erreurs qu'il produit.

Tableau II.3.52 Statistiques des erreurs du modèle SARIMA(1,1,1)(0,1,1)₁₂

Statistique	Période d'estimation	Période de validation
EQM	301,222	246,25
EAM	9,9138	9,70998
EM	0,233444	-1,33486

Nous y trouvons :

- 1) L'erreur quadratique moyenne (EQM)
- 2) L'erreur absolue moyenne (EAM)
- 3) L'erreur moyenne (EM)

Chacune de ces statistiques est basée sur l'erreur de prédiction un mois à l'avance qui est la différence entre la valeur de la donnée au temps t et sa prédiction au temps $t - 1$. Les deux premières mesurent la magnitude des erreurs alors que la dernière mesure le biais. Ces statistiques d'erreurs sont calculées pour la période d'estimation et celle de validation. Pour le modèle considéré ces statistiques d'erreurs ne sont pas différentes d'une manière considérable pour les deux périodes d'estimation et de validation. Il s'ensuit que le modèle ajusté peut être considéré comme assez performant et donc bon à faire des prévisions des valeurs futures.

Le modèle retenu pour la prédiction est alors le SARIMA(1,1,1)(0,1,1)₁₂ sans constante qui peut être formulé comme suit:

$$(1 - 0,109257B)(1 - B)(1 - B^{12})X_t = (1 - 0,921139B)(1 - 0,896051B^{12})\varepsilon_t$$

ε_t est un bruit blanc de moyenne nulle et de variance 306,049.

II.3.2.5.4 Validation du modèle SARIMA(1,1,1)(0,1,1)₁₂

Il s'agit de vérifier que les résidus forment bien un bruit blanc de moyenne nulle. Les tests des autocorrélogrammes et celui du Portmanteau (Box-Pierce) sont utilisés pour tester l'indépendance des résidus et le test t de student est utilisé pour tester la nullité de leur moyenne.

- **Test des ACF et PACF**

L'auto-corrélogramme total et partiel des résidus du modèle ont été tracés tel que montré dans la Fig. II.3.32 et Fig. II.3.33.

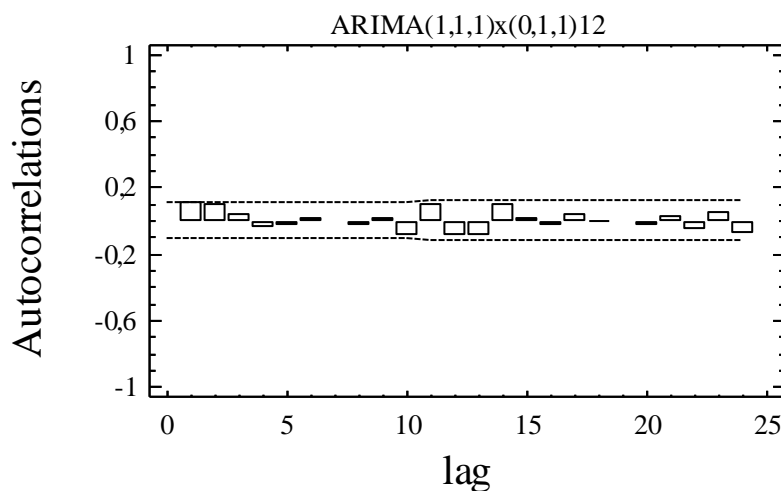


Fig. II.3.32 ACF des résidus du modèle SARIMA(1,1,1)(0,1,1)₁₂

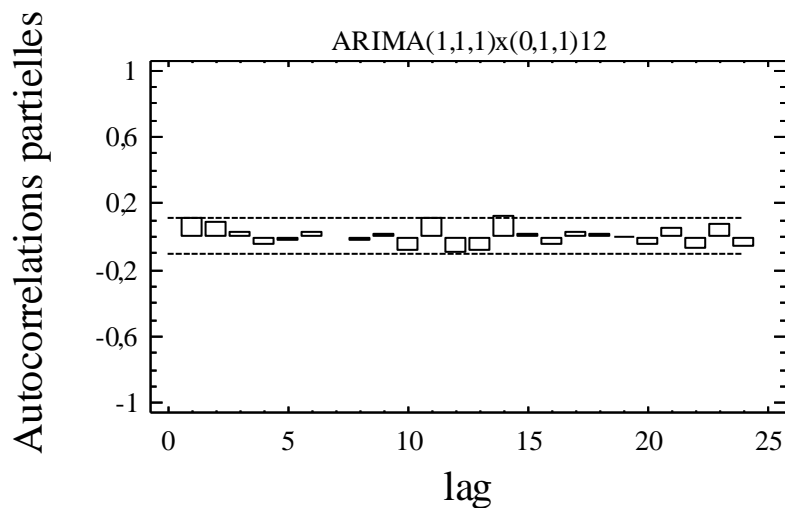


Fig. II.3.33 PACF des résidus du modèle SARIMA(1,1,1)(0,1,1)₁₂

Tous les coefficients sont inclus dans l'intervalle de confiance (IC) tracé à 95% de confiance ce qui implique leur indépendance. Leurs valeurs sont reportées dans le Tableau II.3.53 et Tableau II.3.54.

Tableau II.3.53 ACF des résidus du modèle SARIMA(1,1,1)(0,1,1)₁₂

Lag	Auto-corrélation	Ecart-type	limite inférieure IC 95%	limite supérieure IC 95%
1	0,104917	0,0567048	-0,11114	0,11114
2	0,0933735	0,0573256	-0,112356	0,112356
3	0,0366539	0,0578125	-0,113311	0,113311
4	-0,0477418	0,0578872	-0,113457	0,113457
5	-0,0262772	0,0580137	-0,113705	0,113705
6	0,00737216	0,0580519	-0,11378	0,11378
7	-0,015134	0,058055	-0,113786	0,113786
8	-0,0376854	0,0580676	-0,113811	0,113811
9	0,00944101	0,0581462	-0,113965	0,113965
10	-0,0990539	0,0581512	-0,113974	0,113974
11	0,0901919	0,0586912	-0,115033	0,115033
12	-0,0945002	0,0591352	-0,115903	0,115903
13	-0,0987037	0,0596188	-0,116851	0,116851
14	0,0912297	0,0601419	-0,117876	0,117876
15	0,00827679	0,0605852	-0,118745	0,118745
16	-0,0321294	0,0605889	-0,118752	0,118752
17	0,0342479	0,0606436	-0,11886	0,11886
18	0,00250965	0,0607058	-0,118981	0,118981
19	-0,0118706	0,0607061	-0,118982	0,118982
20	-0,0241878	0,0607136	-0,118997	0,118997
21	0,0224497	0,0607446	-0,119057	0,119057
22	-0,063617	0,0607712	-0,11911	0,11911
23	0,0518475	0,060985	-0,119529	0,119529
24	-0,0834962	0,0611266	-0,119806	0,119806

Tableau II.3.54 PACF des résidus du modèle SARIMA(1,1,1)(0,1,1)₁₂

Lag	Auto-corrélation partielle	Ecart-type	limite inférieure IC 95%	limite supérieure IC 95%
1	0,104917	0,0567048	-0,11114	0,11114
2	0,0832828	0,0567048	-0,11114	0,11114
3	0,0192801	0,0567048	-0,11114	0,11114
4	-0,0619641	0,0567048	-0,11114	0,11114
5	-0,0211872	0,0567048	-0,11114	0,11114
6	0,021317	0,0567048	-0,11114	0,11114
7	-0,0102945	0,0567048	-0,11114	0,11114
8	-0,0402651	0,0567048	-0,11114	0,11114
9	0,0157445	0,0567048	-0,11114	0,11114
10	-0,0948955	0,0567048	-0,11114	0,11114
11	0,11198	0,0567048	-0,11114	0,11114
12	-0,108176	0,0567048	-0,11114	0,11114
13	-0,0930201	0,0567048	-0,11114	0,11114
14	0,121059	0,0567048	-0,11114	0,11114
15	0,0136173	0,0567048	-0,11114	0,11114
16	-0,0565996	0,0567048	-0,11114	0,11114
17	0,0193303	0,0567048	-0,11114	0,11114
18	0,00768217	0,0567048	-0,11114	0,11114
19	0,0038542	0,0567048	-0,11114	0,11114
20	-0,0577332	0,0567048	-0,11114	0,11114
21	0,0525413	0,0567048	-0,11114	0,11114
22	-0,0866347	0,0567048	-0,11114	0,11114
23	0,0693089	0,0567048	-0,11114	0,11114
24	-0,0674547	0,0567048	-0,11114	0,11114

- **Test du Portmanteau**

Ce test est basé sur les 24 premières auto-corrélations des résidus, les résultats du test montrent une valeur de $P = 0,159092$ qui est supérieure ou égale à 0.10, donc on ne peut pas rejeter l'hypothèse nulle que les résidus sont indépendants avec un intervalle de confiance de 90% ou plus.

Test de Box-Pierce

Hypothèse nulle : résidus indépendants

Valeur calculée de la statistique $Q = 27,3651$

Valeur de $P = 0,159092$

- **Test de nullité de la moyenne**

La moyenne des résidus est de -0,270834, leur médiane est de -1,44594 et leur variance est de 291,35. La Figure II.3.34 montre la variabilité temporelle de ces derniers autour de leur moyenne. Les résultats du test t de Student montrent une valeur $P = 0,776075$ qui est supérieure à 0.05 ce qui implique que l'hypothèse nulle de nullité de la moyenne des résidus ne peut pas être rejetée à 95% de confiance.

Test t de Student

Hypothèse nulle : moyenne = 0,0

Alternative: non nulle.

Valeur calculée de la statistique $t = -0,284686$

Valeur de $P = 0,776075$

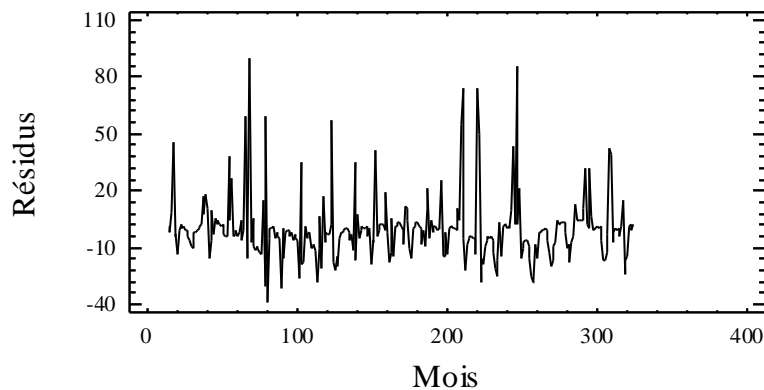


Fig. II.3.34 Résidus du modèle SARIMA(1,1,1)(0,1,1)₁₂

Les résultats de ces tests ainsi que les calculs des résidus montrent bien que les résidus du modèle SARIMA(1,1,1)(0,1,1)₁₂ forment bel et bien une séquence bruit blanc de moyenne nulle et de variance 291,35. Dans ce cas la variance expliquée par le modèle est de 65,7% de la variance totale.

II.3.2.5.5 Prédiction

En permettant au modèle retenu SARIMA(1,1,1)(0,1,1)₁₂ de faire de la prédiction en arrière nous pouvons obtenir les prédictions correspondant aussi bien à la période où les données sont disponibles qu'à celle où ces dernières ne le sont pas. Le Tableau II.3.55 présente les prédictions des 24 données futures pour la période où les données ne sont pas disponibles, avec les résidus correspondants. Les prédictions de ces 24 données futures sont présentées avec leurs intervalles de confiance à 95%. Ces limites de confiance indiquent où va être située la vraie valeur à une période déterminée dans le futur, dans l'hypothèse que le modèle ajusté est adéquat aux données en question. Tandis que les prédictions correspondant à la période où les données sont disponibles avec les résidus correspondants sont présentées en annexe A.II.3.5.

Les prédictions obtenues par le modèle, tel que présenté par la Fig. II.3.35, suivent en général l'allure des données observées qui sont représentées par des points carrés sur la dite figure. Pour la période où les données ne sont pas disponibles, les prédictions obtenues pour les 24 mois futurs suivant la fin de la série sont représentées avec leurs limites tracées à 95% de confiance.

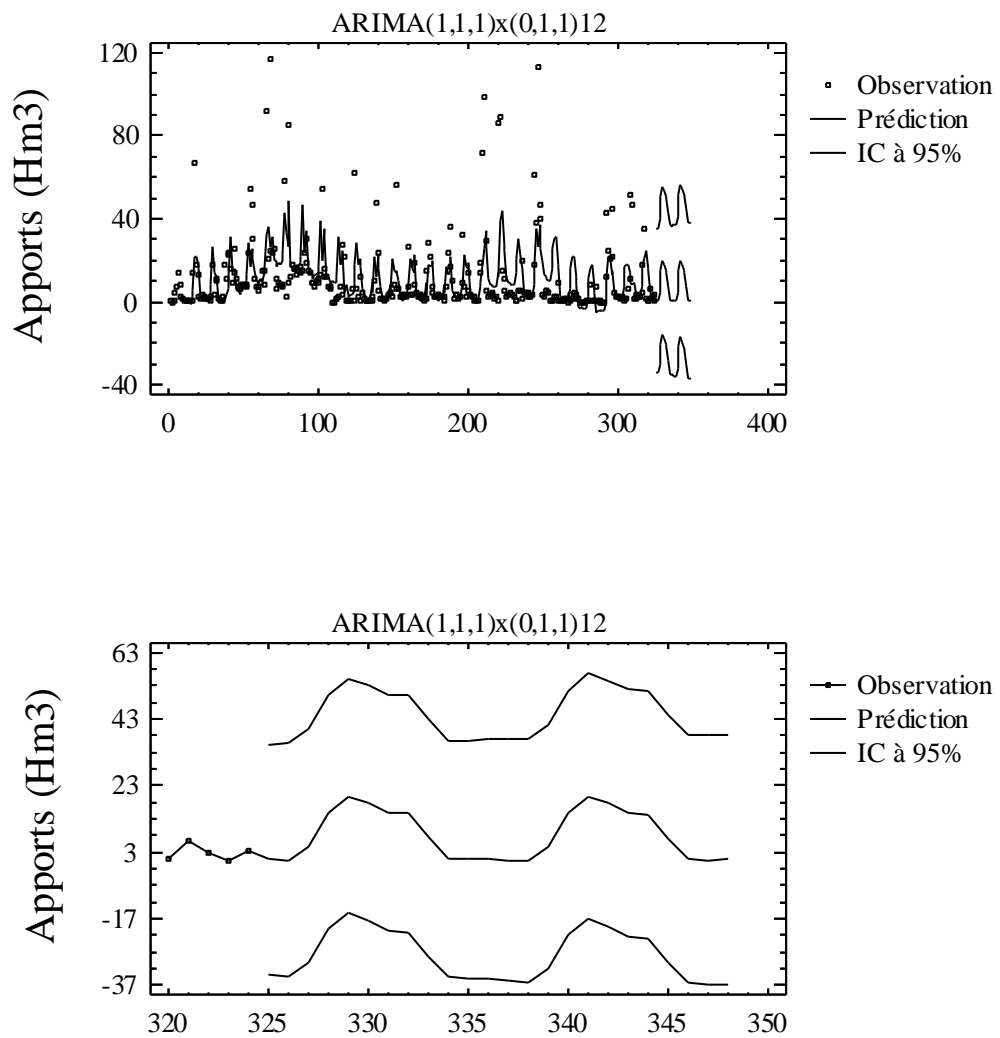


Fig. II.3.35 Prédictions des apports mensuels à CHEFFIA obtenues par le modèle SARIMA(1,1,1)(0,1,1)₁₂

Tableau II.3.55 Prévisions des apports mensuels à CHEFFIA pour les 24 mois futurs

Période	Prédiction	- IC 95%	+IC 95%
325	0,856141	-33,5673	35,2796
326	0,569336	-34,3694	35,508
327	4,89985	-30,1566	39,9563
328	14,9818	-20,1659	50,1295
329	19,8544	-15,3817	55,0905
330	17,7922	-17,5317	53,1162
331	14,8545	-20,557	50,266
332	14,52	-20,9789	50,0189
333	7,3319	-28,2542	42,918
334	1,01331	-34,6598	36,6864
335	0,763518	-34,9963	36,5234
336	0,983998	-34,8624	36,8304
337	0,47145	-35,8849	36,8278
338	0,397512	-36,0916	36,8867
339	4,75128	-31,8442	41,3468
340	14,8358	-21,8631	51,5347
341	19,7086	-17,093	56,5103
342	17,6465	-19,2576	54,5506
343	14,7088	-22,2975	51,7151
344	14,3743	-22,7339	51,4825
345	7,18619	-30,0237	44,396
346	0,867591	-36,4436	38,1788
347	0,617803	-36,7945	38,0301
348	0,838284	-36,6748	38,3514

II.3.2.6 KSOB

Les statistiques de la tendance centrale, de la variabilité et de la forme de la série chronologique des apports mensuels observés à KSOB durant les 324 mois allant de Septembre 1966 à Aout 1992 sont résumées dans le Tableau II.3.56 et la variabilité temporelle de cette série chronologique est illustrée par la Fig. II.3.36.

Tableau II.3.56 Statistiques des apports mensuels observés à KSOB (Sep1966 à Aout 1992)

Paramètre	Valeur
Nombre d'observations	324
Moyenne	2,74883
Variance	15,2115
Déviatoin Standard	3,90019
Minimum	0,0
Maximum	24,554
Std. skewness	19,5419
Std. kurtosis	30,2971

La moyenne de la série est de 2,74883alors que sa variance autour de cette moyenne est de 15,2115avec un écart-type de 3,90019. La skewness et la kurtosis standardisées indiquent un éloignement remarquable de la distribution normale.

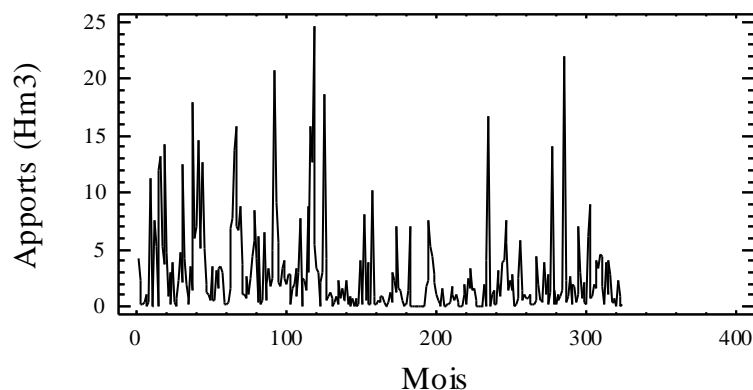


Fig. II.3.36 Apports mensuels observés à KSOB

II.3.2.6.1 Stationnarisation

Le détail du calcul des séries différenciées pour les 20 premiers mois est montré dans le Tableau II.3.57, tandis que les statistiques des séries différenciées pour l'ensemble des données sont représentées dans le Tableau II.3.58.

Tableau II.3.57 Calcul des séries différenciées (données des 20 premiers mois)

N°	Apport	DIFF(Apport,1)	SDIFF(Apports,12) = S	DIFF(S,1)
1	4,247			
2	2,22	-2,027		
3	0,127	-2,093		
4	0,283	0,156		
5	0,42	0,137		
6	0,999	0,579		
7	0,028	-0,971		
8	0,342	0,314		
9	11,289	10,947		
10	0,057	-11,232		
11	0,063	0,006		
12	7,607	7,544		
13	5,35	-2,257	1,103	
14	0,055	-5,295	-2,165	-3,268
15	12,01	11,955	11,883	14,048
16	13,169	1,159	12,886	1,003
17	5,213	-7,956	4,793	-8,093
18	3,673	-1,54	2,674	-2,119
19	14,21	10,537	14,182	11,508
20	4,801	-9,409	4,459	-9,723

Le Tableau II.3.58 résume les statistiques descriptives des séries différenciées. Dans le cas de la série étudiée, la stationnarité est obtenue en éliminant la saisonnalité ($S=12$) et ce effectuant les premières différences saisonnières ($D=1$). La variance des premières différences saisonnières $SDIFF(Apports,12)$ étant de 29.34 et la série est stationnaire pour $S=12$, $d=0$ et $D=1$.

Tableau II.3.58 Statistiques descriptives des séries différenciées pour les 324 mois.

Série	N	Moyenne	Variance	Ecart-type
Apports	324	2.74	15.21	3.9
DIFF(Apports,1)	323	-0.012	20.39	4.51
SDIFF(Apports,12)	312	-0.035	29.34	5.41
DIFF(SDIFF(Apports,12),1)	311	-0.003	39.15	6.25

DIFF(x,n) : différences régulières de x d'ordre n ; SDIFF(x,n,s) :différences saisonnières s de x d'ordre n.

II.3.2.6.2 Identification

Dans le cas de la station courante, les calculs de l'ACF et de la PACF des apports stationnaires sont reportés dans les Tableaux II.3.59 et 60 respectivement et leurs tracés sont illustrés par les Figures II.3.37 et 38.

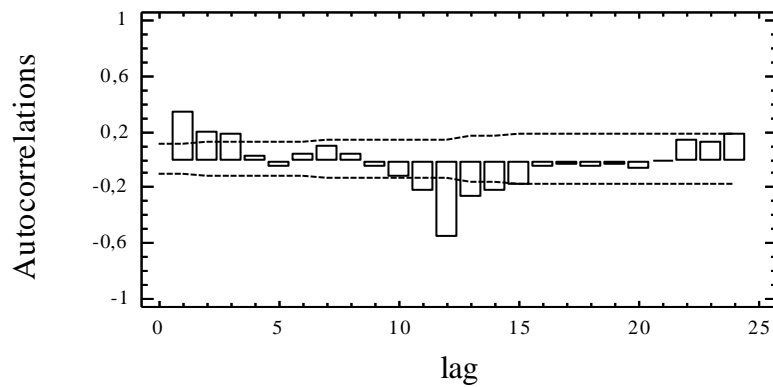


Fig. II.3.37 ACF des apports mensuels

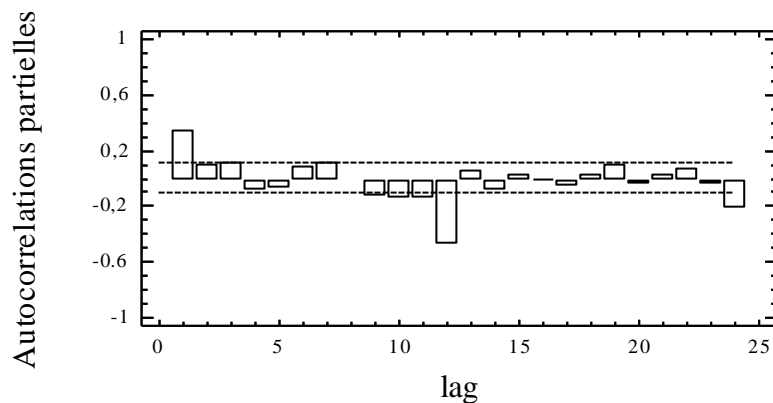


Fig. II.3.38 PACF des apports mensuels

Tableau II.3.59 ACF des apports mensuels stationnaires

Lag	Auto-corrélation	Ecart-type	-IC à 95%	IC à 95%
1	0,334912	0,0566139	-0,110961	0,110961
2	0,192735	0,062643	-0,122778	0,122778
3	0,184365	0,0645156	-0,126448	0,126448
4	0,0217399	0,0661827	-0,129716	0,129716
5	-0,0594291	0,0662056	-0,129761	0,129761
6	0,0337163	0,0663763	-0,130095	0,130095
7	0,0975812	0,0664312	-0,130203	0,130203
8	0,0311667	0,066889	-0,1311	0,1311
9	-0,0637738	0,0669356	-0,131192	0,131192
10	-0,139028	0,06713	-0,131573	0,131573
11	-0,233791	0,0680466	-0,133369	0,133369
12	-0,570869	0,0705742	-0,138323	0,138323
13	-0,277358	0,0840819	-0,164798	0,164798
14	-0,240173	0,0869649	-0,170448	0,170448
15	-0,194164	0,0890655	-0,174565	0,174565
16	-0,0656158	0,090412	-0,177205	0,177205
17	-0,0502704	0,0905645	-0,177503	0,177503
18	-0,0686173	0,0906538	-0,177679	0,177679
19	-0,0401395	0,0908202	-0,178005	0,178005
20	-0,0777588	0,090877	-0,178116	0,178116
21	-0,00188787	0,09109	-0,178533	0,178533
22	0,132308	0,0910901	-0,178534	0,178534
23	0,127666	0,091704	-0,179737	0,179737
24	0,179957	0,0922719	-0,18085	0,18085

Tableau II.3.60 PACF des apports mensuels stationnaires

Lag	Auto-corrélation	Ecart-type	-IC à 95%	IC à 95%
1	0,334912	0,0566139	-0,110961	0,110961
2	0,0907476	0,0566139	-0,110961	0,110961
3	0,10821	0,0566139	-0,110961	0,110961
4	-0,0918999	0,0566139	-0,110961	0,110961
5	-0,0845836	0,0566139	-0,110961	0,110961
6	0,0757392	0,0566139	-0,110961	0,110961
7	0,114774	0,0566139	-0,110961	0,110961
8	-0,0184185	0,0566139	-0,110961	0,110961
9	-0,136001	0,0566139	-0,110961	0,110961
10	-0,152736	0,0566139	-0,110961	0,110961
11	-0,147298	0,0566139	-0,110961	0,110961
12	-0,483624	0,0566139	-0,110961	0,110961
13	0,0462714	0,0566139	-0,110961	0,110961
14	-0,0986885	0,0566139	-0,110961	0,110961
15	0,0188302	0,0566139	-0,110961	0,110961
16	-0,0141871	0,0566139	-0,110961	0,110961
17	-0,0689533	0,0566139	-0,110961	0,110961
18	0,0217864	0,0566139	-0,110961	0,110961
19	0,0918408	0,0566139	-0,110961	0,110961
20	-0,0573752	0,0566139	-0,110961	0,110961
21	0,0193011	0,0566139	-0,110961	0,110961
22	0,0624539	0,0566139	-0,110961	0,110961
23	-0,0417521	0,0566139	-0,110961	0,110961
24	-0,22312	0,0566139	-0,110961	0,110961

L'aspect des auto-corrélogrammes de la série stationnaire considérée laisserait supposer les modèles suivants :

- (A) SARIMA(1,0,3)(1,1,1)₁₂ avec constante
- (B) SARIMA(0,0,3)(0,1,1)₁₂ avec constante
- (C) SARIMA(1,0,1)(0,1,1)₁₂ avec constante
- (D) SARIMA(1,0,2)(0,1,1)₁₂ avec constante
- (E) SARIMA(0,0,1)(0,1,1)₁₂ avec constante

II.3.2.6.3 Estimation des modèles ajustés

Pour l'estimation des paramètres de ces modèles, il faut rappeler que 324 données ont été utilisées dont les 100 dernières ont servi à la validation. Les paramètres des modèles ont alors été estimés sur une période de 224 mois.

L'analyse et la comparaison des modèles ajustés tel que présenté dans le Tableau II.3.61, montre que tous les modèles ajustés produisent des résidus complètement indépendants sauf le modèle (E). D'un autre côté, il faut noter que la moyenne de la variance des résidus produits par les modèles n'est pas considérablement différente entre la période d'estimation et celle de validation, ceci est valable pour tous les modèles ajustés. En outre, le modèle qui produit le minimum de cette variance est le modèle (B), les autres modèles présentent tous des termes qui sont statistiquement non significatifs à 95% de confiance et qui nous renvoient vers le modèle (B) ou (E). Comme le modèle (E) produit des résidus non indépendants, le choix est fixé sur le modèle (B) c'est lui qui est donc retenu pour la prédiction.

Tableau II.3.61 Comparaison des modèles ajustés

Modèle	Variance des résidus		Indépendance des résidus
	Période d'estimation	Période de validation	
(A)	14,2415	12,0634	OK
(B)	13,0116	12,9835	OK
(C)	13,5415	12,0312	OK
(D)	13,5757	12,0978	OK
(E)	14,247	11,9741	NON

Le Tableau II.3.62 résume l'estimation des paramètres du modèle retenu SARIMA(0,0,3)(0,1,1)₁₂ et la signification statistique de ses termes. Les termes correspondant à une valeur de $P < 0.05$ sont statistiquement significativement différents de zéro à 95% de confiance.

La valeur de P pour tous les termes MA(1), MA(2), MA(3) et SMA(1) est inférieure à 0.05 donc ces derniers sont significativement différents de zéro à 95% de confiance. La même chose peut se dire au sujet de la moyenne. L'estimation de l'écart-type du bruit blanc est de 3,68209 et la variance des résidus est de 13,5578 ce qui veut dire que le présent modèle explique 54% de la variance de la série modélisée qui est de 29.34.

Tableau II.3.62 Estimation des paramètres du modèle retenu SARIMA(0,0,3)(0,1,1)₁₂

Paramètre	Estimation	Ecart- type	t	P
MA(1)	-0,268866	0,0672876	-3,99577	0,000081
MA(2)	-0,210279	0,0677722	-3,10273	0,002096
MA(3)	-0,304158	0,0669498	-4,54308	0,000008
SMA(1)	0,962346	0,0157319	61,1715	0,000000
Moyenne	-0,213735	0,0679343	-3,1462	0,001816
Constante	-0,213735			

Notation : SAR(n): Autorégressif saisonnier d'ordre n; SMA(n) : Moyenne mobile saisonnière d'ordre n.

La variance estimée du bruit blanc = 13,5578 avec 307 DDL et l'écart type estimé du bruit blanc = 3,68209

Le Tableau II.3.63 résume la performance du modèle SARIMA(0,0,3)(0,1,1)₁₂ à ajuster les données considérées et ce par l'analyse des erreurs qu'il produit.

Tableau II.3.63 Statistiques des erreurs du modèle SARIMA(0,0,3)(0,1,1)₁₂

Statistique	Période d'estimation	Période de validation
EQM	13,0116	12,9835
EAM	2,41469	2,13038
EM	-0,316972	0,830743

Nous y trouvons :

- 1) L'erreur quadratique moyenne (EQM)
- 2) L'erreur absolue moyenne (EAM)
- 3) L'erreur moyenne (EM)

Chacune de ces statistiques est basée sur l'erreur de prédiction un mois à l'avance qui est la différence entre la valeur de la donnée au temps t et sa prédiction au temps $t - 1$. Les deux premières mesurent la magnitude des erreurs alors que la dernière mesure le biais. Ces statistiques d'erreurs sont calculées pour la période d'estimation et celle de validation. Pour le modèle considéré ces statistiques d'erreurs ne sont pas considérablement différentes entre la période d'estimation et celle de validation. Il s'ensuit que le modèle ajusté peut être considéré comme assez performant et donc bon à faire des prévisions des valeurs futures.

Le modèle retenu pour la prédiction est alors le SARIMA(0,0,3)(0,1,1)₁₂ avec constante qui peut être formulé comme suit:

$$(1 - B^{12})X_t = -0,213735 + (1 + 0,268866 B)(1 + 0,210279 B^2)(1 + 0,304158 B^3)(1 - 0,962346B^{12})\varepsilon_t$$

ε_t est une séquence bruit blanc de moyenne nulle et de variance 13,55.

II.3.2.6.4 Validation du modèle SARIMA(0,0,3)(0,1,1)₁₂

Il s'agit de vérifier que les résidus forment bien un bruit blanc de moyenne nulle. Les tests des autocorrélogrammes et celui du Portmanteau (Box-Pierce) sont utilisés pour tester l'indépendance des résidus et le test t de student est utilisé pour tester la nullité de leur moyenne.

- **Test des ACF et PACF**

L'autocorrélogramme total et partiel des résidus du modèle ont été tracés tel que montré dans la Fig. II.3.39 et Fig. II.3.40.

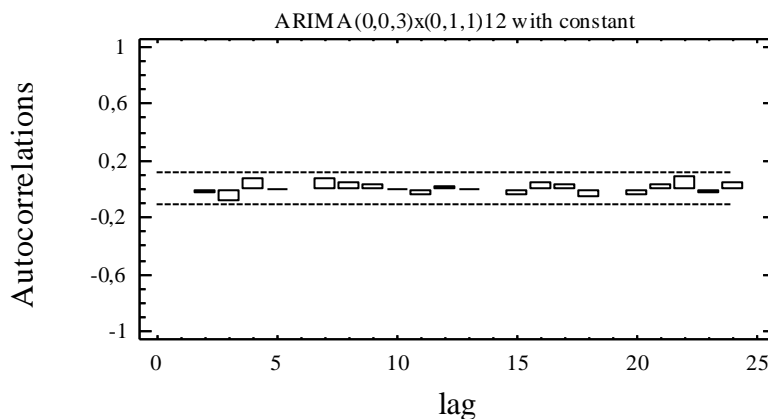


Fig. II.3.39 ACF des résidus du modèle SARIMA(0,0,3)(0,1,1)₁₂

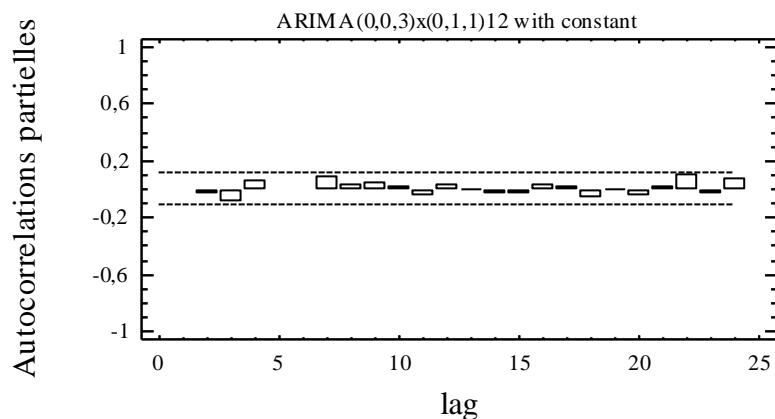


Fig. II.3.40 PACF des résidus du modèle SARIMA(0,0,3)(0,1,1)₁₂

Tous les coefficients sont inclus dans l'intervalle de confiance (IC) tracé à 95% de confiance ce qui implique leur indépendance. Leurs valeurs sont reportées dans le Tableau II.3.64 et Tableau II.3.65.

Tableau II.3.64 ACF des résidus du modèle SARIMA(0,0,3)(0,1,1)₁₂

Lag	Auto-corrélation	Ecart-type	limite inférieure IC 95%	limite supérieure IC 95%
1	-0,00870383	0,0566139	-0,110961	0,110961
2	-0,0347413	0,0566181	-0,11097	0,11097
3	-0,0965528	0,0566864	-0,111104	0,111104
4	0,0636048	0,0572111	-0,112132	0,112132
5	-0,00300351	0,0574373	-0,112575	0,112575
6	-0,0123537	0,0574378	-0,112576	0,112576
7	0,0664683	0,0574463	-0,112593	0,112593
8	0,0377265	0,0576923	-0,113075	0,113075
9	0,0342356	0,0577713	-0,11323	0,11323
10	-0,00149654	0,0578363	-0,113357	0,113357
11	-0,0505819	0,0578364	-0,113358	0,113358
12	0,0201063	0,057978	-0,113635	0,113635
13	0,00362253	0,0580004	-0,113679	0,113679
14	-0,0167329	0,0580011	-0,11368	0,11368
15	-0,0518164	0,0580166	-0,113711	0,113711
16	0,0360062	0,0581647	-0,114001	0,114001
17	0,0311659	0,0582361	-0,114141	0,114141
18	-0,0674208	0,0582895	-0,114246	0,114246
19	-0,00944004	0,058539	-0,114734	0,114734
20	-0,05664	0,0585438	-0,114744	0,114744
21	0,0286451	0,0587192	-0,115088	0,115088
22	0,0855473	0,058764	-0,115175	0,115175
23	-0,028004	0,0591618	-0,115955	0,115955
24	0,0483406	0,0592043	-0,116038	0,116038

Tableau II.3.65 PACF des résidus du modèle SARIMA(0,0,3)(0,1,1)₁₂

Lag	Auto-corrélation partielle	Ecart-type	limite inférieure IC 95%	limite supérieure IC 95%
1	-0,00870383	0,0566139	-0,110961	0,110961
2	-0,0348197	0,0566139	-0,110961	0,110961
3	-0,0972941	0,0566139	-0,110961	0,110961
4	0,0609762	0,0566139	-0,110961	0,110961
5	-0,00877862	0,0566139	-0,110961	0,110961
6	-0,0177523	0,0566139	-0,110961	0,110961
7	0,0790925	0,0566139	-0,110961	0,110961
8	0,0331734	0,0566139	-0,110961	0,110961
9	0,03806	0,0566139	-0,110961	0,110961
10	0,0181995	0,0566139	-0,110961	0,110961
11	-0,051645	0,0566139	-0,110961	0,110961
12	0,0234289	0,0566139	-0,110961	0,110961
13	-0,000441277	0,0566139	-0,110961	0,110961
14	-0,0318008	0,0566139	-0,110961	0,110961
15	-0,04626	0,0566139	-0,110961	0,110961
16	0,0249491	0,0566139	-0,110961	0,110961
17	0,0192492	0,0566139	-0,110961	0,110961
18	-0,0673923	0,0566139	-0,110961	0,110961
19	0,00373212	0,0566139	-0,110961	0,110961
20	-0,060698	0,0566139	-0,110961	0,110961
21	0,0132196	0,0566139	-0,110961	0,110961
22	0,101519	0,0566139	-0,110961	0,110961
23	-0,0371107	0,0566139	-0,110961	0,110961
24	0,0656715	0,0566139	-0,110961	0,110961

- **Test du Portmanteau**

Ce test est basé sur les 24 premières auto-corrélations des résidus, les résultats du test montrent une valeur de $P = 0,757138$ qui est supérieure ou égale à 0.10, donc on ne peut pas rejeter l'hypothèse nulle que les résidus sont indépendants avec un intervalle de confiance de 90% ou plus.

Test de Box-Pierce

Hypothèse nulle : résidus indépendants

Valeur calculée de la statistique $Q = 15,3315$

Valeur de $P = 0,757138$

- **Test de nullité de la moyenne**

La moyenne des résidus est de 0,050885, leur médiane est de -0,628771 et leur variance est de 12,83. La Figure II.3.41 montre la variabilité temporelle de ces derniers autour de leur moyenne. Les résultats du test t de Student montrent une valeur $P = 0,802053$ qui est supérieure à 0.05 ce qui implique que l'hypothèse nulle de nullité de la moyenne des résidus ne peut pas être rejetée à 95% de confiance.

Test t de Student

Hypothèse nulle : moyenne = 0,0

Alternative: non nulle.

Valeur calculée de la statistique $t = 0,250905$

Valeur de $P = 0,802053$

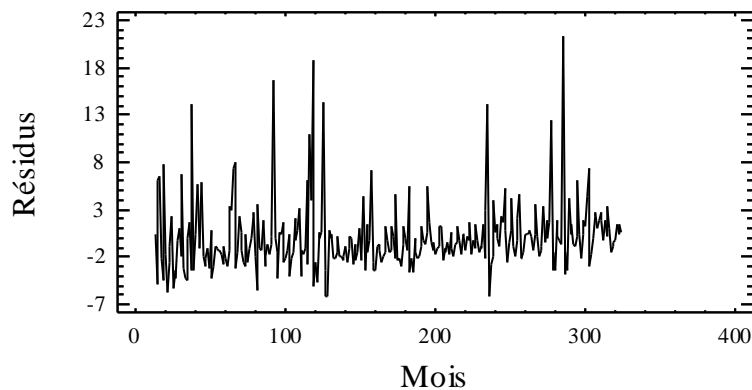


Fig. II.3.41 Résidus du modèle SARIMA(0,0,3)(0,1,1)₁₂

Les résultats de ces tests montrent bien que les résidus du modèle SARIMA(0,0,3)(0,1,1)₁₂ forment bel et bien une séquence bruit blanc de moyenne nulle et de variance 12,83. Dans ce cas la variance expliquée par le modèle est de 56,2%.

II.3.2.6.5 Prédiction

En permettant au modèle retenu SARIMA(0,0,3)(0,1,1)₁₂ de faire de la prédiction en arrière nous pouvons obtenir les prédictions correspondant aussi bien à la période où les données sont disponibles qu'à celle où ces dernières ne le sont pas. Le Tableau II.3.66 présente les prédictions des 24 données futures pour la période où les données ne sont pas disponibles, avec les résidus correspondants. Les prédictions de ces 24 données futures sont présentées avec leurs intervalles de confiance à 95%. Ces limites de confiance indiquent où va être située la vraie valeur à une période déterminée dans le futur, dans l'hypothèse que le modèle ajusté est adéquat aux données en question. Tandis que les prédictions correspondant à la période où les données sont disponibles avec les résidus correspondants sont présentées en annexe A.II.3.6.

Les prédictions obtenues par le modèle, tel que présenté par la Fig. II.3.42, suivent en moyenne l'allure générale des données observées qui sont représentées par des points carrés sur la dite figure. Pour la période où les données ne sont pas disponibles, les prédictions obtenues pour les 24 mois futurs suivant la fin de la série sont représentées avec leurs limites tracées à 95% de confiance.

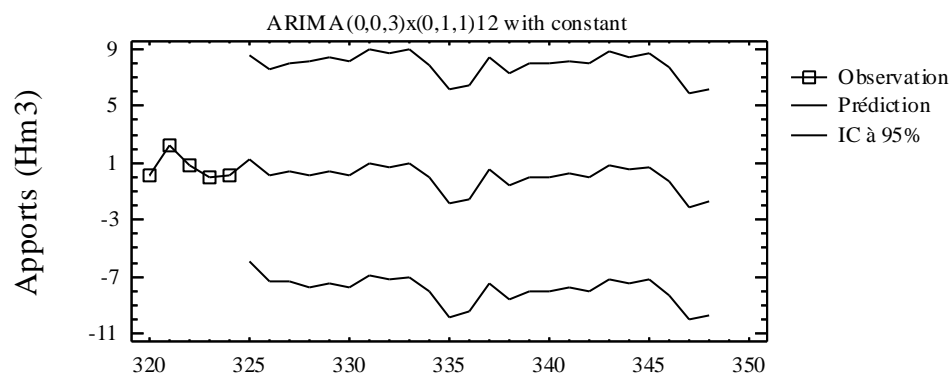
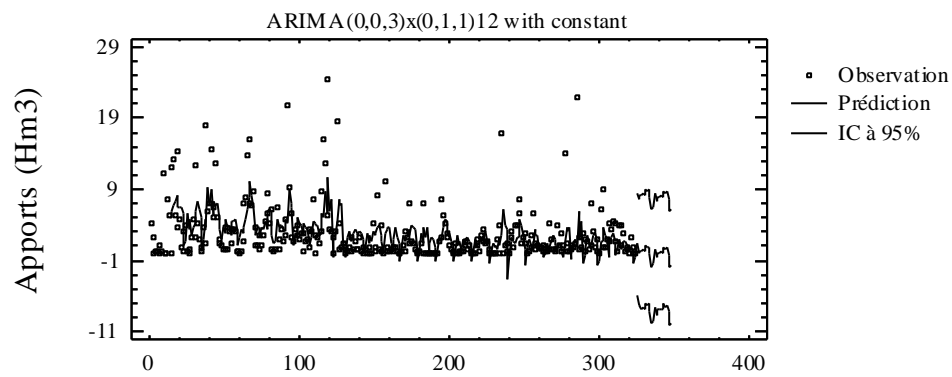


Fig. II.3.42 Prédictions des apports mensuels à KSOB
obtenues par le modèle SARIMA $(0,0,3)(0,1,1)_{12}$

Tableau II.3.66 Prévisions des apports mensuels à
KSOB pour les 24 mois futurs

Période	Prédiction	- IC 95%	+IC 95%
325	1,26831	-5,97703	8,51365
326	0,10869	-7,39396	7,61134
327	0,380811	-7,27497	8,03659
328	0,176252	-7,79039	8,14289
329	0,450467	-7,51617	8,41711
330	0,181638	-7,785	8,14828
331	1,02854	-6,9381	8,99518
332	0,729603	-7,23704	8,69624
333	0,953207	-7,01343	8,91985
334	-0,0706109	-8,03725	7,89603
335	-1,85275	-9,81939	6,11389
336	-1,51813	-9,48477	6,44851
337	0,487933	-7,48338	8,45924
338	-0,62524	-8,59689	7,34641
339	0,000527436	-7,97133	7,97238
340	-0,0374825	-8,00977	7,9348
341	0,236732	-7,73555	8,20902
342	-0,0320969	-8,00438	7,94019
343	0,814809	-7,15748	8,7871
344	0,515868	-7,45642	8,48815
345	0,739473	-7,23281	8,71176
346	-0,284345	-8,25663	7,68794
347	-2,06648	-10,0388	5,9058
348	-1,73186	-9,70415	6,24042

II.3.2.7 MEFROUCHE

Les statistiques de la tendance centrale, de la variabilité et de la forme de la série chronologique des apports mensuels observés à MEFROUCHE durant les 600 mois allant de Septembre 1943 à Aout 1992 sont résumées dans le Tableau II.3.67 et la variabilité temporelle de cette série chronologique est illustrée par la Fig. II.3.43.

Tableau II.3.67 Statistiques des apports mensuels observés à MEFROUCHE (Sep1943 à Aout 1992)

Paramètre	Valeur
Nombre d'observations	600
Moyenne	1,14692
Variance	3,5861
Déviatoin Standard	1,8937
Minimum	0,002
Maximum	18,863
Std. skewness	40,1639
Std. kurtosis	112,02

La moyenne de la série est de 1,14692 alors que sa variance autour de cette moyenne est de 3,5861 avec un écart-type de 1,8937. La skewness et la kurtosis standardisées indiquent un éloignement remarquable de la distribution normale.

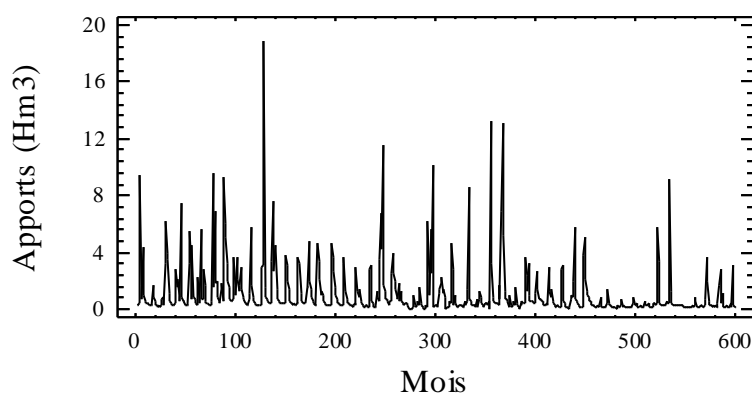


Fig. II.3.43 Apports mensuels observés à MEFROUCHE

II.3.2.7.1 Stationnarisation

Le détail du calcul des séries différenciées pour les 20 premiers mois est montré dans le Tableau II.3.68, tandis que les statistiques des séries différenciées pour l'ensemble des données sont représentées dans le Tableau II.3.69.

Tableau II.3.68 Calcul des séries différenciées (données des 20 premiers mois)

N°	Apport	DIFF(Apport,1)	SDIFF(Apports,12) = S	DIFF(S,1)
1	0,332			
2	0,358	0,026		
3	0,405	0,047		
4	9,466	9,061		
5	1,694	-7,772		
6	0,756	-0,938		
7	4,357	3,601		
8	1,018	-3,339		
9	0,571	-0,447		
10	0,408	-0,163		
11	0,424	0,016		
12	0,387	-0,037		
13	0,353	-0,034	0,021	
14	0,343	-0,01	-0,015	-0,036
15	0,332	-0,011	-0,073	-0,058
16	0,282	-0,05	-9,184	-9,111
17	1,643	1,361	-0,051	9,133
18	0,92	-0,723	0,164	0,215
19	0,515	-0,405	-3,842	-4,006
20	0,273	-0,242	-0,745	3,097

Dans le cas de la série étudiée, la stationnarité est obtenue en éliminant la saisonnalité ($S=12$) et ce en effectuant les premières différences saisonnières ($D=1$). La variance des premières différences saisonnières $SDIFF(Apports,12)$ est de 2.24 et la série est stationnaire pour $S=12$, $d=0$ et $D=1$.

Tableau II.3.69 Statistiques descriptives des séries différenciées pour les 600 mois.

Série	N	Moyenne	Variance	Ecart-type
Apports	600	1.14	3.58	1.89
DIFF(Apports,1)	599	-0.0002	4.86	2.2
SDIFF(Apports ,12)	588	-0.025	5.06	2.24
DIFF(SDIFF(Apports,12) ,1)	587	-0.0001	8.54	2.92

Notation : DIFF (x,n) : différences régulières de x d'ordre n ; SDIFF(x,n,s) :différences saisonnières s de x d'ordre n.

II.3.2.7.2 Identification

Dans le cas de la station courante, les calculs de l'ACF et de la PACF des apports stationnaires sont reportés dans les Tableaux II.3.70 et 71 respectivement et leurs tracés sont illustrés par les Figures II.3.44 et 45.

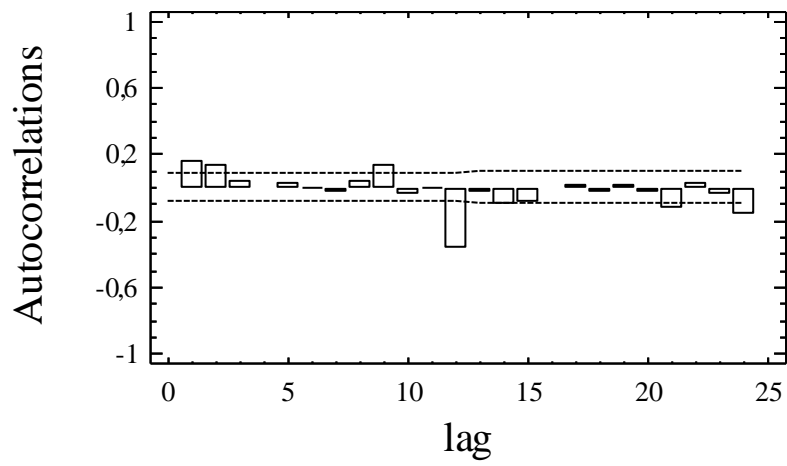


Fig. II.3.44 ACF des apports mensuels

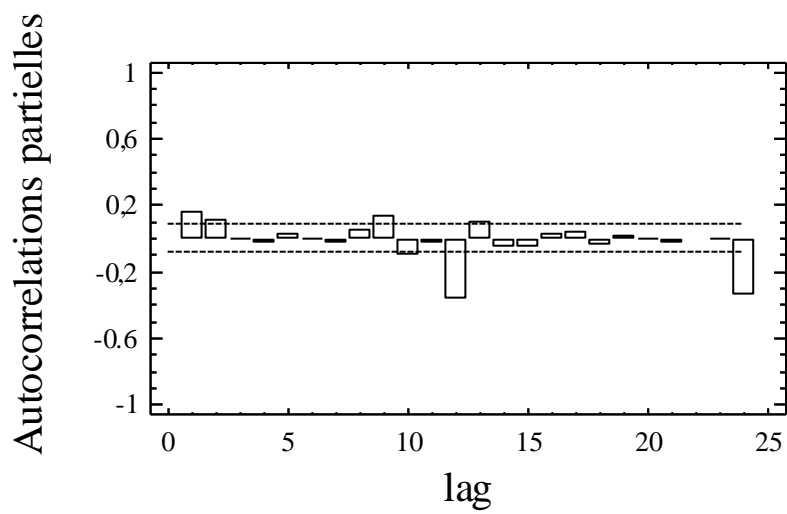


Fig. II.3.45 PACF des apports mensuels

Tableau II.3.70 ACF des apports mensuels stationnaires

Lag	Auto-corrélation	Ecart-type	-IC à 95%	IC à 95%
1	0,156957	0,0412393	-0,0808277	0,0808277
2	0,133372	0,042243	-0,082795	0,082795
3	0,0303368	0,0429532	-0,0841869	0,0841869
4	-0,00792336	0,0429896	-0,0842583	0,0842583
5	0,0199114	0,0429921	-0,0842631	0,0842631
6	0,0056605	0,0430078	-0,0842939	0,0842939
7	-0,029324	0,0430091	-0,0842964	0,0842964
8	0,0379633	0,043043	-0,084363	0,084363
9	0,132562	0,0431	-0,0844745	0,0844745
10	-0,0512515	0,0437879	-0,0858228	0,0858228
11	-0,00468067	0,0438898	-0,0860225	0,0860225
12	-0,370188	0,0438906	-0,0860242	0,0860242
13	-0,038042	0,0489132	-0,0958684	0,0958684
14	-0,109964	0,0489635	-0,095967	0,095967
15	-0,0969958	0,0493818	-0,0967867	0,0967867
16	-0,0136492	0,0497047	-0,0974196	0,0974196
17	0,012454	0,0497111	-0,0974321	0,0974321
18	-0,0250467	0,0497164	-0,0974425	0,0974425
19	0,0165533	0,0497379	-0,0974846	0,0974846
20	-0,0257792	0,0497472	-0,0975029	0,0975029
21	-0,130088	0,0497699	-0,0975475	0,0975475
22	0,0186727	0,0503449	-0,0986743	0,0986743
23	-0,051005	0,0503567	-0,0986974	0,0986974
24	-0,166853	0,0504444	-0,0988695	0,0988695

Tableau II.3.71 PACF des apports mensuels stationnaires

Lag	Auto-corrélation	Ecart-type	-IC à 95%	IC à 95%
1	0,156957	0,0412393	-0,0808277	0,0808277
2	0,111483	0,0412393	-0,0808277	0,0808277
3	-0,00598088	0,0412393	-0,0808277	0,0808277
4	-0,0272168	0,0412393	-0,0808277	0,0808277
5	0,0229416	0,0412393	-0,0808277	0,0808277
6	0,0045522	0,0412393	-0,0808277	0,0808277
7	-0,0366001	0,0412393	-0,0808277	0,0808277
8	0,0467868	0,0412393	-0,0808277	0,0808277
9	0,135291	0,0412393	-0,0808277	0,0808277
10	-0,103756	0,0412393	-0,0808277	0,0808277
11	-0,0209521	0,0412393	-0,0808277	0,0808277
12	-0,371132	0,0412393	-0,0808277	0,0808277
13	0,0933701	0,0412393	-0,0808277	0,0808277
14	-0,0542534	0,0412393	-0,0808277	0,0808277
15	-0,0595529	0,0412393	-0,0808277	0,0808277
16	0,0246847	0,0412393	-0,0808277	0,0808277
17	0,0375821	0,0412393	-0,0808277	0,0808277
18	-0,0532477	0,0412393	-0,0808277	0,0808277
19	0,00967864	0,0412393	-0,0808277	0,0808277
20	-0,00336715	0,0412393	-0,0808277	0,0808277
21	-0,031861	0,0412393	-0,0808277	0,0808277
22	-0,00775943	0,0412393	-0,0808277	0,0808277
23	-0,000085676	0,0412393	-0,0808277	0,0808277
24	-0,351734	0,0412393	-0,0808277	0,0808277

L'aspect des auto-corrélogrammes de la série stationnaire considérée laisserait supposer les modèles suivants :

- (A) SARIMA(1,0,1)(1,1,1)₁₂ avec constante
- (B) SARIMA(0,0,1)(0,1,1)₁₂ sans constante
- (C) SARIMA(0,0,2)(0,1,1)₁₂ avec constante
- (D) SARIMA(2,0,0)(0,1,2)₁₂ avec constante
- (E) SARIMA(1,0,0)(0,1,1)₁₂ avec constante

II.3.2.7.3 Estimation des modèles ajustés

Pour l'estimation des paramètres de ces modèles, il faut rappeler que 600 données ont été utilisées dont les 200 dernières ont servi à la validation. Les paramètres des modèles ont alors été estimés sur une période de 400 mois.

L'analyse et la comparaison des modèles ajustés tel que présenté dans le Tableau II.3.72, montre que tous les modèles ajustés produisent des résidus complètement indépendants sauf le modèle (A). D'un autre côté, il faut noter que la moyenne de la variance des résidus entre la période d'estimation et celle de validation ne diffère pas considérablement. En outre, le modèle qui produit le minimum de cette variance est le modèle (D), mais ses termes AR(2) et SMA(2) sont non significativement différents de zéro ce qui induit leur réduction pour tomber sur le modèle (E). Ce dernier modèle présente dans ce cas le minimum de variance des résidus comparé au modèle (B) et au (C) qui, quant à lui présente un terme MA(2) à réduire. C'est donc le modèle (E) qui est retenu pour la prédiction.

Tableau II.3.72 Comparaison des modèles ajustés

Modèle	Variance des résidus		Indépendance des résidus
	Période d'estimation	Période de validation	
(A)	3,59766	1,14234	NON
(B)	3,60756	1,19119	OK
(C)	3,59033	1,18538	OK
(D)	3,58655	1,1222	OK
(E)	3,59025	1,1827	OK

Le Tableau II.3.73 résume l'estimation des paramètres du modèle retenu SARIMA(1,0,0)(0,1,1)₁₂ et la signification statistique de ses termes. Les termes correspondant à une valeur de $P < 0.05$ sont statistiquement significativement différents de zéro à 95% de confiance.

Les paramètres du modèle sont stables, et la valeur du terme AR(1) est différente de Un (1). La valeur de P pour les termes AR(1) et SMA(1) est inférieure à 0.05 donc ces derniers sont significativement différents de zéro à 95% de. La valeur de P pour la moyenne est très supérieure à 0.05 donc elle est statistiquement non significative elle peut donc être éliminée, la constante aussi car trop petite. L'estimation de l'écart-type du bruit blanc est de 1,94595 et la variance des résidus est de 3,78671 ce qui veut dire que le présent modèle explique seulement 25% de la variance de la série modélisée qui est de 5.06.

Tableau II.3.73 Estimation des paramètres du modèle retenu.
SARIMA(1,0,0)(0,1,1)₁₂

Paramètre	Estimation	Ecart- type	t	P
AR(1)	0,178697	0,0501262	3,56494	0,000394
SMA(1)	0,961779	0,00709427	135,571	0,000000
Moyenne	-0,0151489	0,0121813	-1,24361	0,214140
Constante	-0,0124418			

Notation : SAR(n): Autorégressif saisonnier d'ordre n; SMA(n) : Moyenne mobile saisonnière d'ordre n.

La variance estimée du bruit blanc = 3,78671 avec 585 DDL et l'écart type estimé du bruit blanc = 1,94595

Le Tableau II.3.74 résume la performance du modèle SARIMA(1,0,0)(0,1,1)₁₂ à ajuster les

données considérées et ce par l'analyse des erreurs qu'il produit.

Tableau II.3.74 Statistiques des erreurs du modèle
SARIMA(1,0,0)(0,1,1)₁₂

Statistique	Période d'estimation	Période de validation
EQM	3,59025	1,1827
EAM	0,990148	0,663528
EM	0,0515807	-0,234445

Nous y trouvons :

- 4) L'erreur quadratique moyenne (EQM)
- 5) L'erreur absolue moyenne (EAM)
- 6) L'erreur moyenne (EM)

Chacune de ces statistiques est basée sur l'erreur de prédiction un mois à l'avance qui est la différence entre la valeur de la donnée au temps t et sa prédiction au temps $t - 1$. Les deux premières mesurent la magnitude des erreurs alors que la dernière mesure le biais. Ces statistiques d'erreurs sont calculées pour la période d'estimation et celle de validation. Pour le modèle considéré ces statistiques d'erreurs ne sont pas trop différentes entre la période d'estimation et celle de validation sauf pour la EQM. Il s'ensuit que le modèle ajusté peut être employé à la prévision mais il faudra faire attention car la variance des valeurs futures peut être sous-estimée.

Le modèle retenu pour la prédiction est alors le SARIMA(1,0,0)(0,1,1)₁₂ sans constante qui peut être formulé comme suit:

$$(1 - 0,178697B)(1 - B^{12})X_t = (1 - 0,961779B^{12})\varepsilon_t$$

ε_t est un bruit blanc de moyenne nulle et de variance égale à 3.79.

II.3.2.7.4 Validation du modèle SARIMA(1,0,0)(0,1,1)₁₂

Il s'agit de vérifier que les résidus forment bien un bruit blanc de moyenne nulle. Les tests des autocorrélogrammes et celui du Portmanteau (Box-Pierce) sont utilisés pour tester l'indépendance des résidus et le test t de student est utilisé pour tester la nullité de leur moyenne.

- **Test des ACF et PACF**

L'auto-corrélogramme total et partiel des résidus du modèle ont été tracés tel que montré dans la Fig. II.3.46 et Fig. II.3.47.

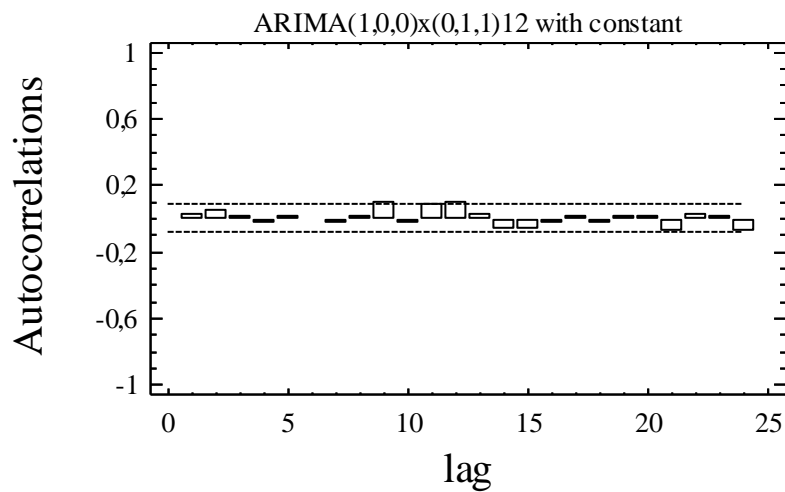


Fig. II.3.46 ACF des résidus du modèle SARIMA(1,0,0)(0,1,1)₁₂

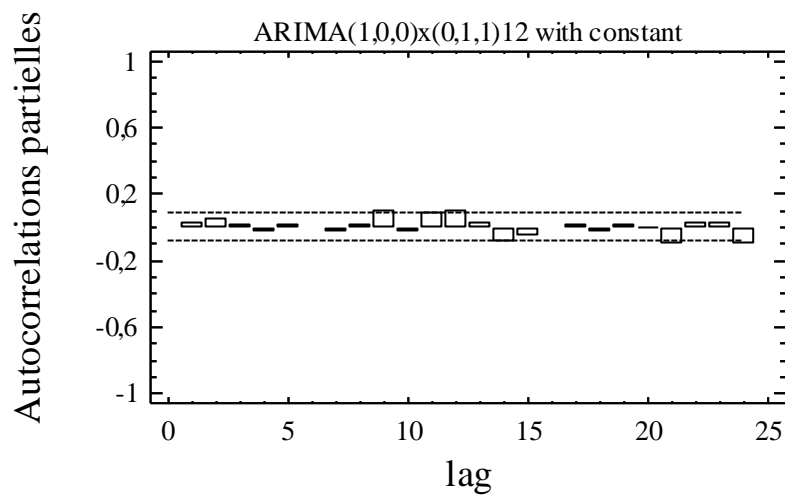


Fig. II.3.47 PACF des résidus du modèle SARIMA(1,0,0)(0,1,1)₁₂

Tous les coefficients sont inclus dans l'intervalle de confiance (IC) tracé à 95% de confiance ce qui implique leur indépendance. Leurs valeurs sont reportées dans le Tableau II.3.75 et Tableau II.3.76.

Tableau II.3.75 ACF des résidus du modèle SARIMA(1,0,0)(0,1,1)₁₂

Lag	Auto-corrélation	Ecart-type	limite inférieure IC 95%	limite supérieure IC 95%
1	0,0282139	0,0412393	-0,0808277	0,0808277
2	0,0534351	0,0412721	-0,080892	0,080892
3	0,00895292	0,0413896	-0,0811223	0,0811223
4	-0,0204109	0,0413929	-0,0811288	0,0811288
5	0,0130194	0,04141	-0,0811623	0,0811623
6	-0,0107224	0,041417	-0,0811759	0,0811759
7	-0,0268821	0,0414217	-0,0811852	0,0811852
8	0,0142861	0,0414514	-0,0812433	0,0812433
9	0,090825	0,0414597	-0,0812597	0,0812597
10	-0,032166	0,0417967	-0,0819203	0,0819203
11	0,088888	0,0418388	-0,0820027	0,0820027
12	0,0942477	0,0421588	-0,0826298	0,0826298
13	0,0285655	0,0425156	-0,0833292	0,0833292
14	-0,0703404	0,0425482	-0,0833931	0,0833931
15	-0,0716419	0,0427455	-0,0837798	0,0837798
16	-0,0242135	0,0429492	-0,0841791	0,0841791
17	0,0122986	0,0429724	-0,0842246	0,0842246
18	-0,0238513	0,0429784	-0,0842363	0,0842363
19	0,0062003	0,0430009	-0,0842804	0,0842804
20	0,0124244	0,0430025	-0,0842834	0,0842834
21	-0,0844684	0,0430086	-0,0842954	0,0842954
22	0,0294242	0,0432898	-0,0848466	0,0848466
23	0,0136217	0,0433238	-0,0849132	0,0849132
24	-0,0888776	0,0433311	-0,0849275	0,0849275

Tableau II.3.76 PACF des résidus du modèle SARIMA(1,0,0)(0,1,1)₁₂

Lag	Auto-corrélation partielle	Ecart-type	limite inférieure IC 95%	limite supérieure IC 95%
1	0,0282139	0,0412393	-0,0808277	0,0808277
2	0,052681	0,0412393	-0,0808277	0,0808277
3	0,00606003	0,0412393	-0,0808277	0,0808277
4	-0,0237101	0,0412393	-0,0808277	0,0808277
5	0,0134477	0,0412393	-0,0808277	0,0808277
6	-0,00916333	0,0412393	-0,0808277	0,0808277
7	-0,0275499	0,0412393	-0,0808277	0,0808277
8	0,0162099	0,0412393	-0,0808277	0,0808277
9	0,094184	0,0412393	-0,0808277	0,0808277
10	-0,0394844	0,0412393	-0,0808277	0,0808277
11	0,0807313	0,0412393	-0,0808277	0,0808277
12	0,095309	0,0412393	-0,0808277	0,0808277
13	0,0182557	0,0412393	-0,0808277	0,0808277
14	-0,0900415	0,0412393	-0,0808277	0,0808277
15	-0,0653659	0,0412393	-0,0808277	0,0808277
16	-0,00688571	0,0412393	-0,0808277	0,0808277
17	0,0164607	0,0412393	-0,0808277	0,0808277
18	-0,0279115	0,0412393	-0,0808277	0,0808277
19	0,0175629	0,0412393	-0,0808277	0,0808277
20	-0,00194408	0,0412393	-0,0808277	0,0808277
21	-0,106496	0,0412393	-0,0808277	0,0808277
22	0,0257864	0,0412393	-0,0808277	0,0808277
23	0,0294638	0,0412393	-0,0808277	0,0808277
24	-0,104098	0,0412393	-0,0808277	0,0808277

- **Test du Portmanteau**

Ce test est basé sur les 24 premières auto-corrélations des résidus, les résultats du test montrent une valeur de $P = 0,0366905$ qui est supérieure ou égale à 0.01, donc on ne peut pas rejeter l'hypothèse nulle que les résidus sont indépendants avec un intervalle de confiance de 99 %.

Test de Box-Pierce

Hypothèse nulle : résidus indépendants

Valeur calculée de la statistique $Q = 35,2258$

Valeur de $P = 0,0366905$.

- **Test de nullité de la moyenne**

La moyenne des résidus est de -0,11402 leur médiane est de -0,198453 et leur variance est de 2,80. La Figure II.3.48 montre la variabilité temporelle de ces derniers autour de leur moyenne. Les résultats du test t de Student montrent une valeur $P = 0,0991896$ qui est supérieure à 0.05 ce qui implique que l'hypothèse nulle de nullité de la moyenne des résidus ne peut pas être rejetée à 95% de confiance.

Test t de Student

Hypothèse nulle : moyenne = 0,0

Alternative: non nulle.

Valeur calculée de la statistique $t = -1,65141$

Valeur de $P = 0,0991896$.

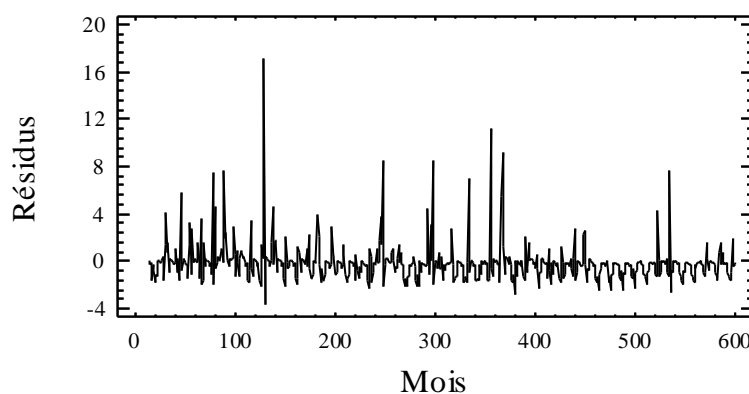


Fig. II.3.48 Résidus du modèle SARIMA(1,0,0)(0,1,1)₁₂

Les résultats de ces tests ainsi que les calculs des résidus montrent bien que les résidus du modèle SARIMA(1,0,0)(0,1,1)₁₂ forment bel et bien une séquence bruit blanc de moyenne nulle et de variance 2,8. Dans ce cas on peut estimer à 44,6 % la variation expliquée par le modèle.

II.3.2.7.5 Prédiction

En permettant au modèle retenu SARIMA(1,0,0)(0,1,1)₁₂ de faire de la prédiction en arrière nous pouvons obtenir les prédictions correspondant aussi bien à la période où les données sont disponibles qu'à celle où ces dernières ne le sont pas. Le Tableau II.3.77 présente les prédictions des 24 données futures pour la période où les données ne sont pas disponibles, avec les résidus correspondants. Les prédictions de ces 24 données futures sont présentées avec leurs intervalles de confiance à 95%. Ces limites de confiance indiquent où va être située la vraie valeur à une période déterminée dans le future, dans l'hypothèse que le modèle ajusté est adéquat aux données en question. Tandis que les prédictions correspondant à la période où les données sont disponibles avec les résidus correspondants sont présentées en annexe A.II.3.7.

Les prédictions obtenues par le modèle, tel que présenté par la Fig. II.3.49, suivent en moyenne l'allure générale des données observées qui sont représentées par des points carrés sur la dite figure. Pour la période où les données ne sont pas disponibles, les prédictions obtenues pour les 24 mois futurs suivant la fin de la série sont représentées avec leurs limites tracées à 95% de confiance.

Tableau II.3.77 Prévisions des apports mensuels à
MEFROUCHE pour les 24 mois futurs

Période	Prédiction	- IC 95%	+IC 95%
601	0,0339589	-3,78794	3,85586
602	0,181336	-3,70111	4,06378
603	0,126807	-3,75755	4,01117
604	0,849874	-3,03455	4,7343
605	1,2304	-2,65402	5,11483
606	1,5274	-2,35702	5,41182
607	1,94464	-1,93978	5,82907
608	1,08418	-2,80024	4,96861
609	1,09326	-2,79117	4,97768
610	0,100971	-3,78345	3,9854
611	0,0831249	-3,8013	3,96755
612	-0,0279871	-3,91241	3,85644
613	-0,0117182	-3,89889	3,87545
614	0,160732	-3,72652	4,04799
615	0,110683	-3,77658	3,99794
616	0,834551	-3,05271	4,72181
617	1,21522	-2,67204	5,10248
618	1,51225	-2,37501	5,39951
619	1,9295	-1,95776	5,81676
620	1,06903	-2,81823	4,95629
621	1,07811	-2,80915	4,96537
622	0,0858224	-3,80144	3,97308
623	0,067976	-3,81928	3,95524
624	-0,0431359	-3,9304	3,84412

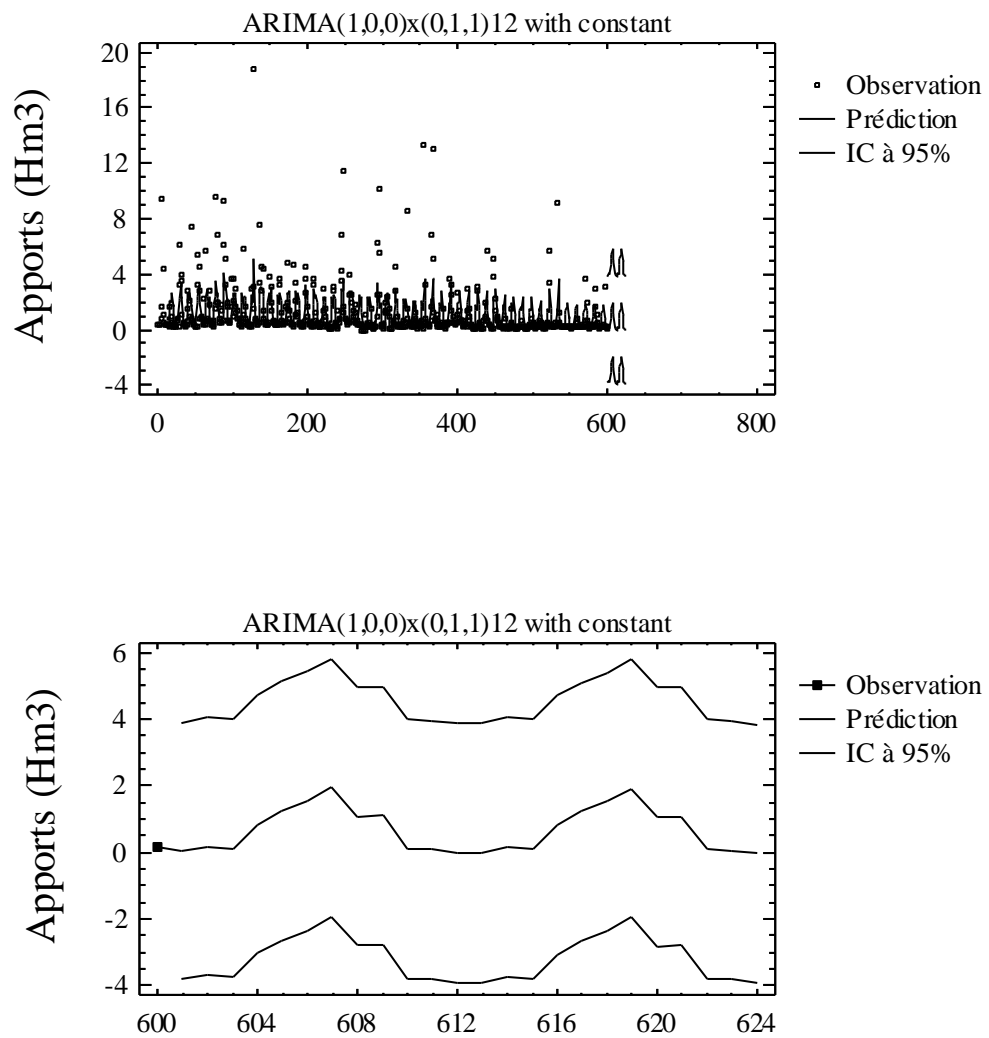


Fig. II.3.49 Prédictions des apports mensuels à MEFROUCHE obtenues par le modèle SARIMA(1,0,0)(0,1,1)₁₂

II.3.2.8 MIREBECK

Les statistiques de la tendance centrale, de la variabilité et de la forme de la série chronologique des apports mensuels observés à MIREBECK durant les 324 mois allant de Septembre 1968 à Aout 1993 sont résumées dans le Tableau II.3.78 et la variabilité temporelle de cette série chronologique est illustrée par la Fig. II.3.50.

Tableau II.3.78 Statistiques des apports mensuels observés à MEFROUCHE (Sep1968 à Aout 1993)

Paramètre	Valeur
Nombre d'observations	324
Moyenne	27,8447
Variance	2706,16
Déviatoin Standard	52,0207
Minimum	0,0032
Maximum	509,15
Std. skewness	32,1123
Std. kurtosis	101,189

La moyenne de la série est de 27,8447 alors que sa variance autour de cette moyenne est de 2706,16 avec un écart-type de 7,56469. La skewness et la kurtosis standardisées indiquent un éloignement remarquable de la distribution normale.

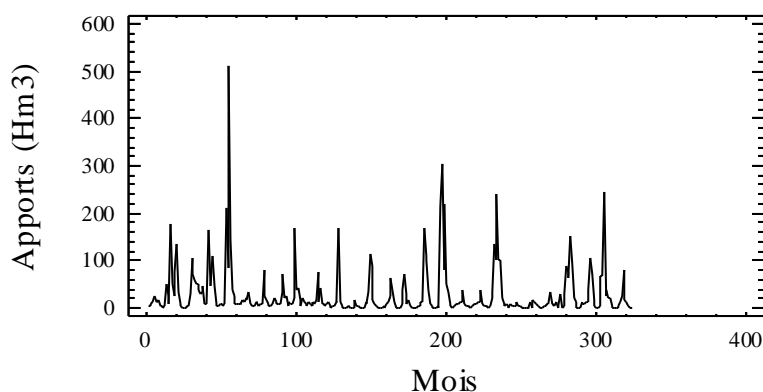


Fig. II.3.50 Apports mensuels observés à MIREBECK

II.3.2.8.1 Stationnarisation

Le détail du calcul des séries différenciées pour les 20 premiers mois est montré dans le Tableau II.3.79, tandis que les statistiques des séries différenciées pour l'ensemble des données sont représentées dans le Tableau II.3.80.

Tableau II.3.79 Calcul des séries différenciées (données des 20 premiers mois)

N°	Apport	DIFF(Apport,1)	SDIFF(Apports,12) = S	DIFF(S,1)
	5,85			
1	4,21	-1,64		
2	10,08	5,87		
3	13,32	3,24		
4	25,06	11,74		
5	11,47	-13,59		
6	14,89	3,42		
7	16,69	1,8		
8	5	-11,69		
9	4,19	-0,81		
10	2,01	-2,18		
11	4,38	2,37		
12	50,46	46,08	44,61	
13	27,1	-23,36	22,89	-21,72
14	9,29	-17,81	-0,79	-23,68
15	177,55	168,26	164,23	165,02
16	49,27	-128,28	24,21	-140,02
17	25,37	-23,9	13,9	-10,31
18	75,28	49,91	60,39	46,49
19	133,38	58,1	116,69	56,3
20				

Dans le cas de la série étudiée, la stationnarité est obtenue en éliminant la saisonnalité ($S=12$) et ce effectuant les premières différences saisonnières ($D=1$). La variance des premières différences saisonnières $SDIFF(Apports,12)$ est de 4802.36 et la série est stationnaire pour $S=12$, $d=0$ et $D=1$.

Tableau II.3.80 Statistiques descriptives des séries différenciées pour les 324 mois.

Série	N	Moyenne	Variance	Ecart-type
Apports	324	27.84	2706.16	52.02
DIFF(Apports,1)	323	-0.015	3143.76	56.06
SDIFF(Apports ,12)	312	0.153	4802.36	69.29
DIFF(SDIFF(Apports,12) ,1)	311	-0.145	6188.47	78.66

Notation : DIFF (x,n) : différences régulières de x d'ordre n ; SDIFF(x,n,s) :différences saisonnières s de x d'ordre n.

II.3.2.8.2 Identification

Dans le cas de la station courante, les calculs de l'ACF et de la PACF des apports stationnaires sont reportés dans les Tableaux II.3.81 et 82 respectivement et leurs tracés sont illustrés par les Figures II.3.51 et 52.

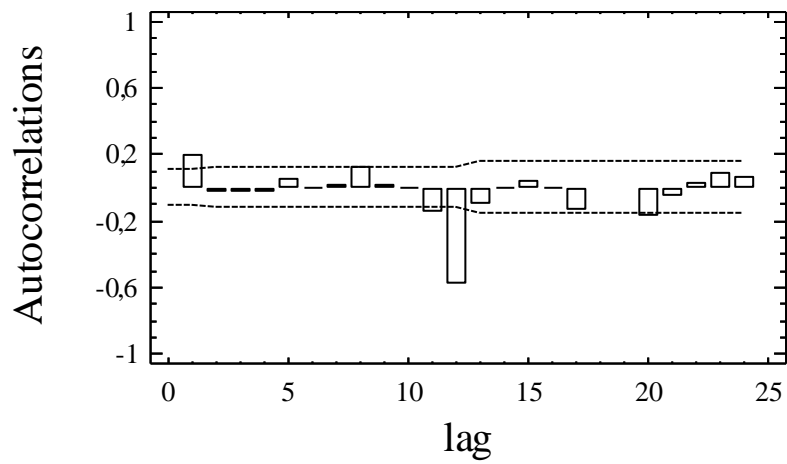
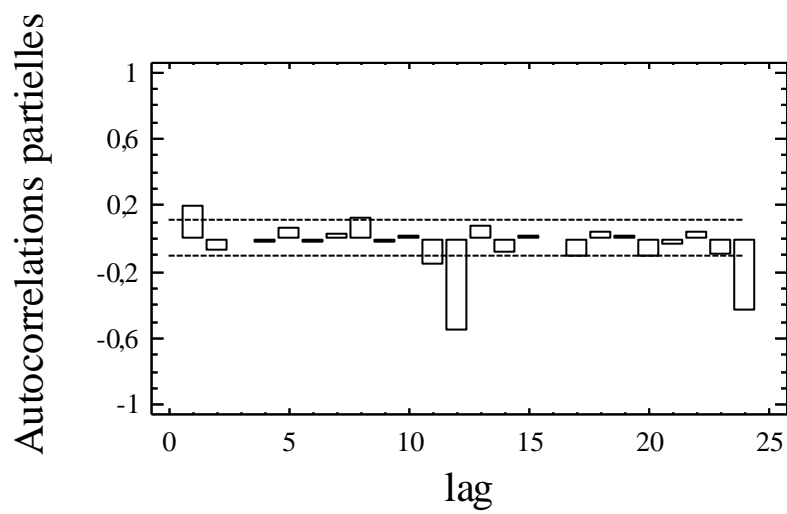
**Fig. II.3.51** ACF des apports mensuels**Fig. II.3.52** PACF des apports mensuels

Tableau II.3.81 ACF des apports mensuels stationnaires

Lag	Auto-corrélation	Ecart-type	-IC à 95%	IC à 95%
1	0,191467	0,0566139	-0,110961	0,110961
2	-0,0391037	0,0586526	-0,114957	0,114957
3	-0,0388445	0,0587361	-0,115121	0,115121
4	-0,0337568	0,0588184	-0,115282	0,115282
5	0,0536117	0,0588804	-0,115404	0,115404
6	0,00245332	0,0590367	-0,11571	0,11571
7	0,0108421	0,059037	-0,115711	0,115711
8	0,116725	0,0590434	-0,115723	0,115723
9	0,015407	0,0597784	-0,117164	0,117164
10	-0,00365497	0,0597911	-0,117189	0,117189
11	-0,161053	0,0597919	-0,11719	0,11719
12	-0,58896	0,0611665	-0,119884	0,119884
13	-0,10405	0,0772327	-0,151374	0,151374
14	-0,00585992	0,0776807	-0,152252	0,152252
15	0,0311922	0,0776821	-0,152254	0,152254
16	-0,00104375	0,0777222	-0,152333	0,152333
17	-0,142398	0,0777223	-0,152333	0,152333
18	-0,0150112	0,078554	-0,153963	0,153963
19	-0,0147504	0,0785632	-0,153981	0,153981
20	-0,182192	0,0785721	-0,153999	0,153999
21	-0,0648897	0,0799147	-0,15663	0,15663
22	0,0250997	0,0800834	-0,156961	0,156961
23	0,0855421	0,0801086	-0,15701	0,15701
24	0,0588272	0,0804008	-0,157583	0,157583

Tableau II.3.82 PACF des apports mensuels stationnaires

Lag	Auto-corrélation	Ecart-type	-IC à 95%	IC à 95%
1	0,191467	0,0566139	-0,110961	0,110961
2	-0,0786463	0,0566139	-0,110961	0,110961
3	-0,0164099	0,0566139	-0,110961	0,110961
4	-0,0267316	0,0566139	-0,110961	0,110961
5	0,0650479	0,0566139	-0,110961	0,110961
6	-0,0267259	0,0566139	-0,110961	0,110961
7	0,0215714	0,0566139	-0,110961	0,110961
8	0,116653	0,0566139	-0,110961	0,110961
9	-0,0284306	0,0566139	-0,110961	0,110961
10	0,0081238	0,0566139	-0,110961	0,110961
11	-0,164458	0,0566139	-0,110961	0,110961
12	-0,565419	0,0566139	-0,110961	0,110961
13	0,0778687	0,0566139	-0,110961	0,110961
14	-0,0964423	0,0566139	-0,110961	0,110961
15	0,0123361	0,0566139	-0,110961	0,110961
16	-0,0165891	0,0566139	-0,110961	0,110961
17	-0,117118	0,0566139	-0,110961	0,110961
18	0,0346933	0,0566139	-0,110961	0,110961
19	0,0166043	0,0566139	-0,110961	0,110961
20	-0,115447	0,0566139	-0,110961	0,110961
21	-0,0439087	0,0566139	-0,110961	0,110961
22	0,0337879	0,0566139	-0,110961	0,110961
23	-0,10247	0,0566139	-0,110961	0,110961
24	-0,438275	0,0566139	-0,110961	0,110961

L'aspect des auto-corrélogrammes de la série stationnaire considérée laisserait supposer les modèles suivants :

- (A) SARIMA(2,0,0)(0,1,1)₁₂ avec constante
- (B) SARIMA(1,0,1)(0,1,1)₁₂ sans constante
- (C) SARIMA(1,0,0)(0,1,1)₁₂ avec constante
- (D) SARIMA(0,0,2)(0,1,1)₁₂ avec constante
- (E) SARIMA(0,1,1)(1,1,1)₁₂ avec constante

II.3.2.8.3 Estimation des modèles ajustés

Pour l'estimation des paramètres de ces modèles, il faut rappeler que 324 données ont été utilisées dont les 100 dernières ont servi à la validation. Les paramètres des modèles ont alors été estimés sur une période de 224 mois.

L'analyse et la comparaison des modèles ajustés tel que présenté dans le Tableau II.3.83, montre que seulement les modèles ajustés (A) et (B) produisent des résidus complètement indépendants. Le modèle (B) présente un terme MA(1) non significativement différent de zéro. Le modèle (A) produit une variance des résidus minimale et tous ses termes sont statistiquement différents de zéro, c'est lui qui est donc retenu pour la prédiction. D'un autre côté, il faut noter que la variance des résidus produits par les modèles (A) et (B) diffère entre la période d'estimation et celle de la validation.

Tableau II.3.83 Comparaison des modèles ajustés

Modèle	Variance des résidus		Indépendance des résidus
	Période d'estimation	Période de validation	
(A)	2428,46	1240,27	OK
(B)	2449,85	1272,83	OK
(C)	2497,95	1333,48	NON
(D)	2550,44	1466,37	NON
(E)	2576,92	1272,07	NON

Le Tableau II.3.84 résume l'estimation des paramètres du modèle retenu SARIMA(2,0,0)(0,1,1)₁₂ et la signification statistique de ses termes. Les termes correspondant à une valeur de $P < 0.05$ sont statistiquement significativement différents de zéro à 95% de confiance.

Les paramètres du modèle sont stables, et la valeur des termes AR(1), AR(2) est différente de Un (1). La valeur de P pour les termes AR(1), AR(2) et SMA(1) est inférieure à 0.05 donc ces derniers sont significativement différents de zéro à 95% de confiance. La valeur de P pour la moyenne est supérieure à 0.05 donc elle est statistiquement non significative elle peut donc être éliminée. L'estimation de l'écart-type du bruit blanc est de 49,0995 et l'estimation de la variance des résidus est de 2410,76 ce qui veut dire que le présent modèle explique près de 50% de la variance de la série modélisée qui est de 4802.36.

Tableau II.3.84 Estimation des paramètres du modèle retenu SARIMA(2,0,0)(0,1,1)₁₂

Paramètre	Estimation	Ecart- type	t	P
AR(1)	0,269507	0,0683947	3,94047	0,000101
AR(2)	0,181845	0,0682527	2,66429	0,008121
SMA(1)	0,967746	0,0181164	53,4182	0,000000
Moyenne	-1,22301	0,938653	-1,30295	0,193567
Constante	-0,671004			

Notation : SAR(n): Autorégressif saisonnier d'ordre n; SMA(n) : Moyenne mobile saisonnière d'ordre n.

La variance estimée du bruit blanc = 2410,76 avec 308 DDL et l'écart type estimé du bruit blanc = 49,0995

Le Tableau II.3.85 résume la performance du modèle SARIMA(2,0,0)(0,1,1)₁₂ à ajuster les données considérées et ce par l'analyse des erreurs qu'il produit.

Tableau II.3.85 Statistiques des erreurs du modèle SARIMA(2,0,0)(0,1,1)₁₂

Statistique	Période d'estimation	Période de validation
EQM	2428,46	1240,27
EAM	24,3933	19,3621
EM	-0,555994	6,0065

Nous y trouvons :

- 1) L'erreur quadratique moyenne (EQM)
- 2) L'erreur absolue moyenne (EAM)
- 3) L'erreur moyenne (EM)

Chacune de ces statistiques est basée sur l'erreur de prédiction un mois à l'avance qui est la différence entre la valeur de la donnée au temps t et sa prédiction au temps $t - 1$. Les deux premières mesurent la magnitude des erreurs alors que la dernière mesure le biais. Ces statistiques d'erreurs sont calculées pour la période d'estimation et celle de validation et les différences observées entre les deux périodes pour la EQM nous poussent à prendre les résultats du modèle retenu avec prudence quant à la variance car celle-ci peut être sous-estimée.

Le modèle retenu pour la prédiction est alors le SARIMA(2,0,0)(0,1,1)₁₂ sans constante qui peut être formulé comme suit:

$$(1 - 0,269507 B - 0,181845 B^2)(1 - B^{12})X_t = (1 - 0,967746 B^{12})\varepsilon_t - 0,671004$$

ε_t est une séquence bruit blanc avec une moyenne nulle et une variance estimée à 2410,76.

II.3.2.8.4 Validation du modèle SARIMA(2,0,0)(0,1,1)₁₂

Il s'agit de vérifier que les résidus forment bien un bruit blanc de moyenne nulle. Les tests des autocorrélogrammes et celui du Portmanteau (Box-Pierce) sont utilisés pour tester l'indépendance des résidus et le test t de student est utilisé pour tester la nullité de leur moyenne.

- **Test des ACF et PACF**

L'auto-corrélogramme total et partiel des résidus du modèle ont été tracés tel que montré dans la Fig. II.3.53 et Fig. II.3.54.

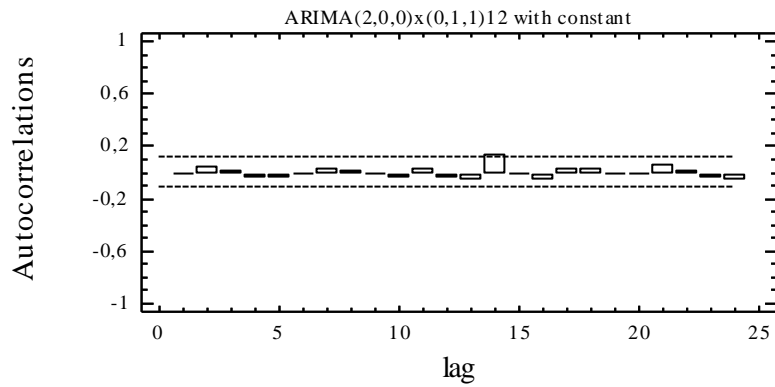


Fig. II.3.53 ACF des résidus du modèle SARIMA(2,0,0)(0,1,1)₁₂

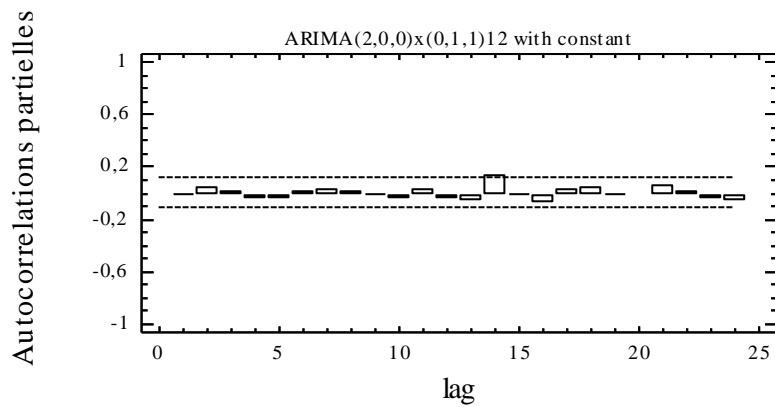


Fig. II.3.54 PACF des résidus du modèle SARIMA(2,0,0)(0,1,1)₁₂

Tous les coefficients sont inclus dans l'intervalle de confiance (IC) tracé à 95% de confiance ce qui implique leur indépendance. Leurs valeurs sont reportées dans les Tableaux II.3.86 et 87.

Tableau II.3.86 ACF des résidus du modèle SARIMA(2,0,0)(0,1,1)₁₂

Lag	Auto-corrélation	Ecart-type	limite inférieure IC 95%	limite supérieure IC 95%
1	-0,00268804	0,0566139	-0,110961	0,110961
2	0,0348418	0,0566143	-0,110962	0,110962
3	0,00505769	0,0566829	-0,111097	0,111097
4	-0,0387991	0,0566844	-0,1111	0,1111
5	-0,037782	0,0567694	-0,111266	0,111266
6	-0,00218789	0,05685	-0,111424	0,111424
7	0,0176998	0,0568503	-0,111425	0,111425
8	0,00864901	0,0568679	-0,111459	0,111459
9	-0,00711101	0,0568721	-0,111468	0,111468
10	-0,0518017	0,056875	-0,111473	0,111473
11	0,015394	0,057026	-0,111769	0,111769
12	-0,0413756	0,0570393	-0,111795	0,111795
13	-0,0681498	0,0571354	-0,111984	0,111984
14	0,137341	0,0573954	-0,112493	0,112493
15	-0,0132583	0,0584392	-0,114539	0,114539
16	-0,0628816	0,0584489	-0,114558	0,114558
17	0,0292049	0,0586653	-0,114982	0,114982
18	0,0304033	0,0587119	-0,115073	0,115073
19	-0,00679004	0,0587623	-0,115172	0,115172
20	-0,0110336	0,0587648	-0,115177	0,115177
21	0,0566929	0,0587715	-0,11519	0,11519
22	0,0148614	0,0589465	-0,115533	0,115533
23	-0,0490001	0,0589585	-0,115557	0,115557
24	-0,0740262	0,0590889	-0,115812	0,115812

Tableau II.3.87 PACF des résidus du modèle SARIMA(2,0,0)(0,1,1)₁₂

Lag	Auto-corrélation partielle	Ecart-type	limite inférieure IC 95%	limite supérieure IC 95%
1	-0,00268804	0,0566139	-0,110961	0,110961
2	0,0348348	0,0566139	-0,110961	0,110961
3	0,00524813	0,0566139	-0,110961	0,110961
4	-0,0400351	0,0566139	-0,110961	0,110961
5	-0,0384611	0,0566139	-0,110961	0,110961
6	0,00033419	0,0566139	-0,110961	0,110961
7	0,020956	0,0566139	-0,110961	0,110961
8	0,00779584	0,0566139	-0,110961	0,110961
9	-0,0115174	0,0566139	-0,110961	0,110961
10	-0,0545525	0,0566139	-0,110961	0,110961
11	0,0170823	0,0566139	-0,110961	0,110961
12	-0,035246	0,0566139	-0,110961	0,110961
13	-0,0695098	0,0566139	-0,110961	0,110961
14	0,135923	0,0566139	-0,110961	0,110961
15	-0,0109932	0,0566139	-0,110961	0,110961
16	-0,0768998	0,0566139	-0,110961	0,110961
17	0,0246935	0,0566139	-0,110961	0,110961
18	0,042938	0,0566139	-0,110961	0,110961
19	-0,000250732	0,0566139	-0,110961	0,110961
20	-0,0199478	0,0566139	-0,110961	0,110961
21	0,0534582	0,0566139	-0,110961	0,110961
22	0,0136622	0,0566139	-0,110961	0,110961
23	-0,0560049	0,0566139	-0,110961	0,110961
24	-0,0642794	0,0566139	-0,110961	0,110961

- **Test du Portmanteau**

Ce test est basé sur les 24 premières auto-corrélations des résidus, les résultats du test montrent une valeur de $P = 0,789101$ qui est supérieure ou égale à 0.10, donc on ne peut pas rejeter l'hypothèse nulle que les résidus sont indépendants avec un intervalle de confiance de 90% ou plus.

Test de Box-Pierce

Hypothèse nulle : résidus indépendants

Valeur calculée de la statistique $Q = 15,6478$

Valeur de $P = 0,789101$

- **Test de nullité de la moyenne**

La moyenne des résidus est de 1,54737, leur médiane est de -2,5203 alors que leur variance est de 2028,58. La Figure II.3.55 montre la variabilité temporelle de ces derniers autour de leur moyenne. Les résultats du test t de Student montrent une valeur $P = 0,543604$ qui est supérieure à 0.05 ce qui implique que l'hypothèse nulle de nullité de la moyenne des résidus ne peut pas être rejetée à 95% de confiance.

Test t de Student

Hypothèse nulle : moyenne = 0,0

Alternative: non nulle.

Valeur calculée de la statistique $t = 0,608042$

Valeur de $P = 0,543604$

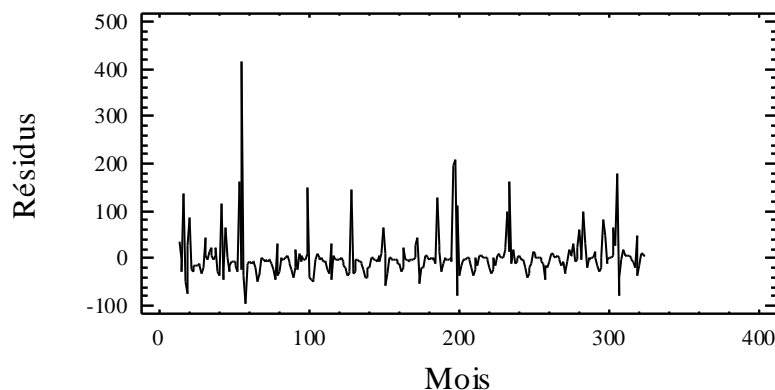


Fig. II.3.55 Résidus du modèle SARIMA(2,0,0)(0,1,1)₁₂

Les résultats de ces tests ainsi que les calculs des résidus montrent bien que les résidus du modèle SARIMA(2,0,0)(0,1,1)₁₂ forment bel et bien une séquence bruit blanc de moyenne nulle et de variance 2028,58. Dans ce cas on peut dire que la variance expliquée par le modèle est de 58 % de la variance totale.

II.3.2.8.5 Prédiction

En permettant au modèle retenu SARIMA(2,0,0)(0,1,1)₁₂ de faire de la prédiction en arrière nous pouvons obtenir les prédictions correspondant aussi bien à la période où les données sont disponibles qu'à celle où ces dernières ne le sont pas. Le Tableau II.3.88 présente les prédictions des 24 données futures pour la période où les données ne sont pas disponibles, avec les résidus correspondants. Les prédictions de ces 24 données futures sont présentées avec leurs intervalles de confiance à 95%. Ces limites de confiance indiquent où va être située la vraie valeur à une période déterminée dans le future, dans l'hypothèse que le modèle ajusté est adéquat aux données en question. Tandis que les prédictions correspondant à la période où les données sont disponibles avec les résidus correspondants sont représentées en annexe A.II.3.8.

Les prédictions obtenues par le modèle, tel que présenté par la Fig. II.3.56, suivent en moyenne l'allure générale des données observées qui sont représentées par des points carrés sur la dite figure. Pour la période où les données ne sont pas disponibles, les prédictions obtenues pour les 24 mois futurs suivant la fin de la série sont représentées avec leurs limites tracées à 95% de confiance.

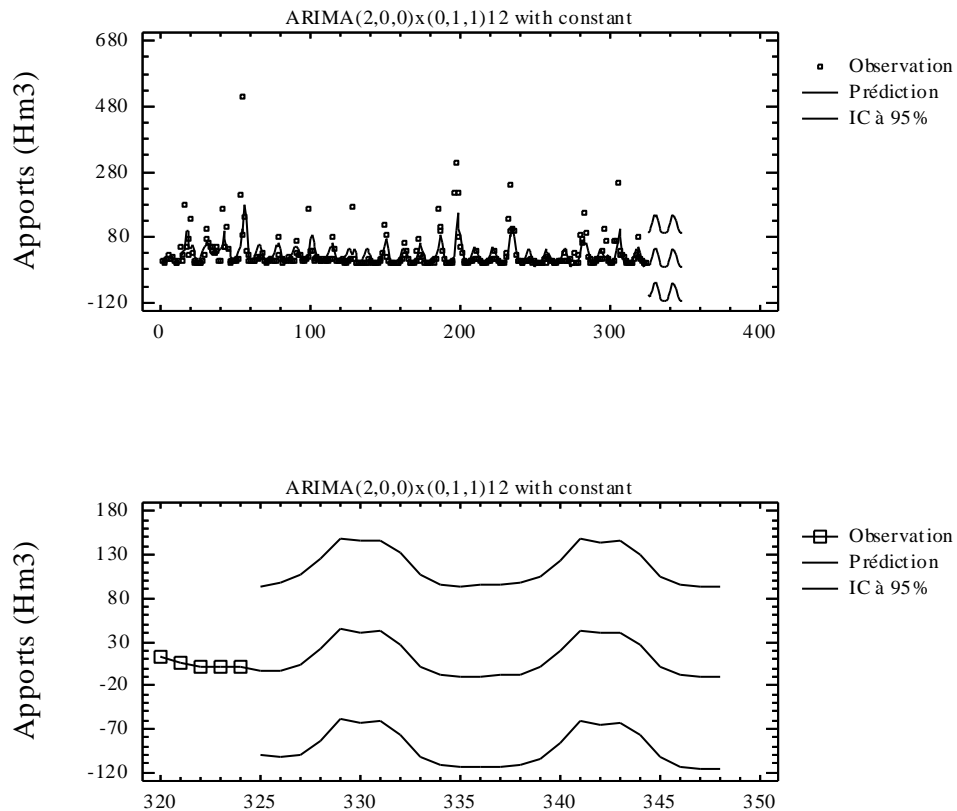


Fig. II.3.56 Prédictions des apports mensuels à MIREBECK obtenues par le modèle SARIMA(2,0,0)(0,1,1)₁₂

Tableau II.3.88 Prévisions des apports mensuels à
MIREBECK pour les 24 mois futurs

Période	Prédiction	- IC 95%	+IC 95%
325	-3,1611	-99,7742	93,452
326	-2,97875	-103,039	97,0816
327	3,3782	-99,6584	106,415
328	20,8685	-82,7925	124,53
329	45,4385	-58,4959	149,373
330	41,1007	-62,9143	145,116
331	42,254	-61,7905	146,298
332	27,645	-76,409	131,699
333	2,3146	-101,743	106,372
334	-7,45677	-111,515	96,6017
335	-9,89281	-113,952	94,166
336	-9,55563	-113,615	94,5034
337	-8,4627	-112,571	95,6458
338	-6,94112	-111,053	97,1712
339	0,675244	-103,44	104,791
340	18,7485	-85,3678	122,865
341	43,7046	-60,4121	147,821
342	39,5769	-64,5398	143,694
343	40,857	-63,2598	144,974
344	26,3204	-77,7964	130,437
345	1,03257	-103,084	105,149
346	-8,71416	-112,831	95,4026
347	-11,1358	-115,253	92,981
348	-10,7903	-114,907	93,3265

II.3.2.9 PIERRE DU CHAT

Les statistiques de la tendance centrale, de la variabilité et de la forme de la série chronologique des apports mensuels observés à PIERRE DU CHAT durant les 504 mois allant de Septembre 1952 à Aout 1993 sont résumées dans le Tableau II.3.89 et la variabilité temporelle de cette série chronologique est illustrée par la Fig. II.3.57.

Tableau II.3.89 Statistiques des apports mensuels observés à PIERRE DU CHAT (Sep1952 à Aout 1993)

Paramètre	Valeur
Nombre d'observations	504
Moyenne	15,6913
Variance	633,538
Déviatoin Standard	25,1702
Minimum	0,0007
Maximum	222,64
Std. skewness	41,1911
Std. kurtosis	126,899

La moyenne de la série est de 15,6913 alors que sa variance autour de cette moyenne est de 633,538 avec un écart-type de 25,1702. La skewness et la kurtosis standardisées indiquent un éloignement remarquable de la distribution normale.

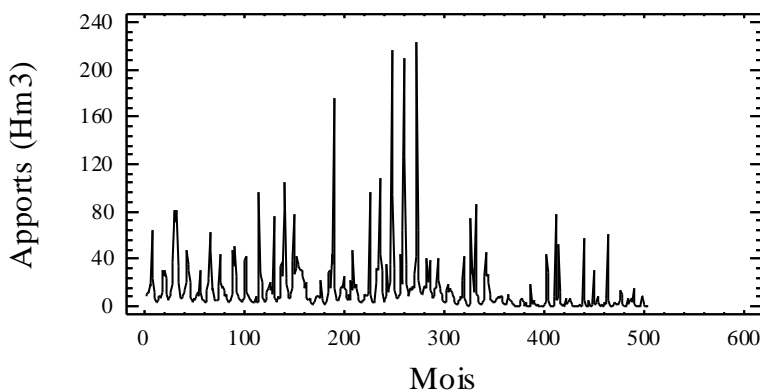


Fig. II.3.57 Apports mensuels observés à PIERRE DU CHAT.

II.3.2.9.1 Stationnarisation

Le détail du calcul des séries différenciées pour les 20 premiers mois est montré dans le Tableau II.3.90, tandis que les statistiques des séries différenciées pour l'ensemble des données sont représentées dans le Tableau II.3.91.

Tableau II.3.90 Calcul des séries différenciées (données des 20 premiers mois)

	Apport	DIFF(Apport,1)	SDIFF(Apports,12) = S	DIFF(S,1)
1	8,1			
2	10,15	2,05		
3	11,18	1,03		
4	12	0,82		
5	19,18	7,18		
6	17,12	-2,06		
7	64,15	47,03		
8	24,36	-39,79		
9	11,25	-13,11		
10	6,22	-5,03		
11	3,68	-2,54		
12	3,49	-0,19		
13	5,8	2,31	-2,3	
14	9,04	3,24	-1,11	1,19
15	8,51	-0,53	-2,67	-1,56
16	6,94	-1,57	-5,06	-2,39
17	11,3	4,36	-7,88	-2,82
18	29,99	18,69	12,87	20,75
19	25,71	-4,28	-38,44	-51,31
20	30,43	4,72	6,07	44,51

Dans le cas de la série étudiée, la stationnarité est obtenue en éliminant la saisonnalité ($S=12$) et ce en effectuant les premières différences saisonnières ($D=1$). La variance des premières différences saisonnières $SDIFF(Apports,12)$ est de 709 et la série est stationnaire pour $S=12$, $d=0$ et $D=1$.

Tableau II.3.91 Statistiques descriptives des séries différenciées pour les 504 mois.

Série	N	Moyenne	Variance	Ecart-type
Apports	504	15.69	633.53	25.17
DIFF(Apports,1)	503	-0.016	685.7	26.18
SDIFF(Apports,12)	492	-0.34	709	26.62
DIFF(SDIFF(Apports,12),1)	491	0.004	1119.79	33.46

Notation : DIFF (x,n) : différences régulières de x d'ordre n ; SDIFF(x,n,s) :différences saisonnières s de x d'ordre n.

II.3.2.9.2 Identification

Dans le cas de la station courante, les calculs de l'ACF et de la PACF des apports stationnaires sont reportés dans les Tableaux II.3.92 et 93 respectivement et leurs tracés sont illustrés par les Figures II.3.58 et 59.

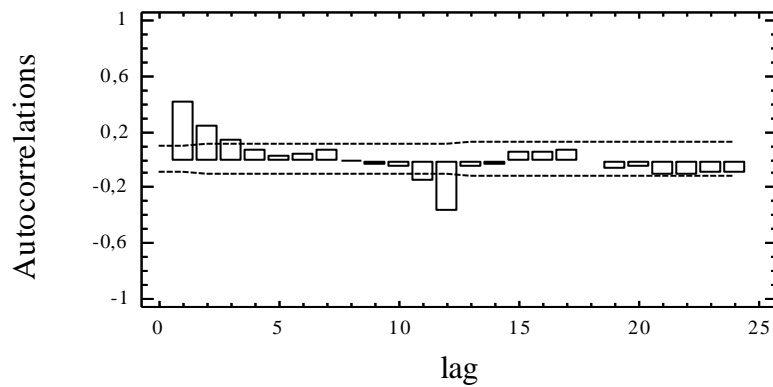


Fig. II.3.58 ACF des apports mensuels

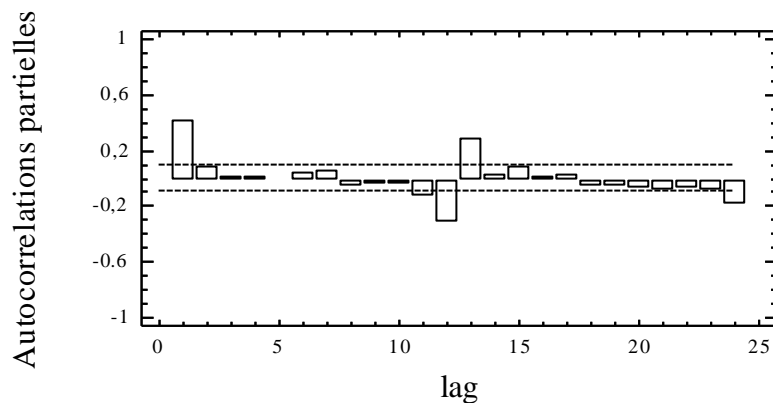


Fig. II.3.59 PACF des apports mensuels

Tableau II.3.92 ACF des apports mensuels stationnaires

Lag	Auto-corrélation	Ecart-type	-IC à 95%	IC à 95%
1	0,406681	0,0450835	-0,0883622	0,0883622
2	0,235517	0,052008	-0,101934	0,101934
3	0,130928	0,0541324	-0,106098	0,106098
4	0,0720005	0,0547722	-0,107352	0,107352
5	0,0195643	0,0549643	-0,107728	0,107728
6	0,0390435	0,0549784	-0,107756	0,107756
7	0,061938	0,0550347	-0,107866	0,107866
8	-0,00885845	0,0551762	-0,108144	0,108144
9	-0,0318023	0,0551791	-0,108149	0,108149
10	-0,0715954	0,0552164	-0,108222	0,108222
11	-0,166057	0,0554047	-0,108592	0,108592
12	-0,389365	0,0564073	-0,110556	0,110556
13	-0,070481	0,0616284	-0,12079	0,12079
14	-0,0346316	0,061792	-0,12111	0,12111
15	0,0476808	0,0618314	-0,121188	0,121188
16	0,0555848	0,0619061	-0,121334	0,121334
17	0,066488	0,0620075	-0,121533	0,121533
18	-0,0207471	0,0621522	-0,121816	0,121816
19	-0,0800524	0,0621663	-0,121844	0,121844
20	-0,0712906	0,0623755	-0,122254	0,122254
21	-0,117387	0,0625409	-0,122578	0,122578
22	-0,124381	0,0629871	-0,123453	0,123453
23	-0,102591	0,0634844	-0,124427	0,124427
24	-0,106639	0,0638204	-0,125086	0,125086

Tableau II.3.93 PACF des apports mensuels stationnaires

Lag	Auto-corrélation	Ecart-type	-IC à 95%	IC à 95%
1	0,406681	0,0450835	-0,0883622	0,0883622
2	0,0840242	0,0450835	-0,0883622	0,0883622
3	0,0108899	0,0450835	-0,0883622	0,0883622
4	0,000103708	0,0450835	-0,0883622	0,0883622
5	-0,0244471	0,0450835	-0,0883622	0,0883622
6	0,0387609	0,0450835	-0,0883622	0,0883622
7	0,0453929	0,0450835	-0,0883622	0,0883622
8	-0,065064	0,0450835	-0,0883622	0,0883622
9	-0,0299132	0,0450835	-0,0883622	0,0883622
10	-0,0554936	0,0450835	-0,0883622	0,0883622
11	-0,134486	0,0450835	-0,0883622	0,0883622
12	-0,327972	0,0450835	-0,0883622	0,0883622
13	0,275822	0,0450835	-0,0883622	0,0883622
14	0,0217289	0,0450835	-0,0883622	0,0883622
15	0,086372	0,0450835	-0,0883622	0,0883622
16	0,0115893	0,0450835	-0,0883622	0,0883622
17	0,0195303	0,0450835	-0,0883622	0,0883622
18	-0,0698995	0,0450835	-0,0883622	0,0883622
19	-0,059952	0,0450835	-0,0883622	0,0883622
20	-0,0757089	0,0450835	-0,0883622	0,0883622
21	-0,0960173	0,0450835	-0,0883622	0,0883622
22	-0,0863125	0,0450835	-0,0883622	0,0883622
23	-0,0937684	0,0450835	-0,0883622	0,0883622
24	-0,194774	0,0450835	-0,0883622	0,0883622

L'aspect des auto-corrélogrammes de la série stationnaire considérée laisserait supposer les modèles suivants :

- (A) SARIMA(1,0,1)(1,1,1)₁₂
- (B) SARIMA(1,0,0)(1,1,1)₁₂
- (C) SARIMA(0,0,1)(1,1,1)₁₂ avec constante
- (D) SARIMA(2,0,0)(1,1,1)₁₂ avec constante
- (E) SARIMA(2,0,3)(2,1,1)₁₂ avec constante

II.3.2.9.3 Estimation des modèles ajustés

Pour l'estimation des paramètres de ces modèles, il faut rappeler que 504 données ont été utilisées dont les 200 dernières ont servi à la validation. Les paramètres des modèles ont alors été estimés sur une période de 304 mois.

L'analyse et la comparaison des modèles ajustés tel que présenté dans le Tableau II.3.94, montre que tous les modèles ajustés produisent des résidus non complètement indépendants sauf le modèle (B). D'un autre côté, il faut noter que tous les modèles présentent des termes non significativement différents de zéro qui appellent à la réduction de l'ordre du modèle, à l'exception du modèle (B), qui en outre produit une variance des résidus minimale. Il faut noter également que la moyenne de la variance des résidus produits par tous les modèles diffère entre la période d'estimation que celle de la validation.

Tableau II.3.94 Comparaison des modèles ajustés

Modèle	Variance des résidus		Indépendance des résidus
	Période d'estimation	Période de validation	
(A)	617,728	203,943	NON
(B)	615,175	200,868	OK
(C)	623,53	244,933	NON
(D)	619,32	226,716	NON
(E)	624,695	255,409	NON

Le Tableau II.3.95 résume l'estimation des paramètres du modèle retenu SARIMA(1,0,0)(1,1,1)₁₂ et la signification statistique de ses termes. Les termes correspondant à une valeur de $P < 0.05$ sont statistiquement significativement différents de zéro à 95% de confiance.

Les paramètres du modèle sont stables, et la valeur du terme AR(1) est différente de Un (1). La valeur de P pour les termes AR(1), SAR(1) et SMA(1) est inférieure à 0.05 donc ces derniers sont significativement différents de zéro à 95% de confiance. La valeur de P pour la moyenne est très supérieure à 0.05 donc elle est statistiquement non significative elle a donc été éliminée, la constante aussi car trop faible. L'estimation de l'écart-type du bruit blanc est de 24,8541 et la variance des résidus est de 617,727 ce qui veut dire que le présent modèle explique près de 13% de la variance de la série modélisée qui est de 709.

Tableau II.3.95 Estimation des paramètres du modèle retenu SARIMA(1,0,0)(1,1,1)₁₂

Paramètre	Estimation	Ecart- type	t	P
AR(1)	0,297104	0,0562594	5,28097	0,000000
SAR(1)	0,221882	0,0632896	3,50582	0,000497
SMA(1)	0,957414	0,0189392	50,552	0,000000

Notation : SAR(n): Autorégressif saisonnier d'ordre n; SMA(n) : Moyenne mobile saisonnière d'ordre n.

La variance estimée du bruit blanc = 617,727 avec 489 DDL et l'écart type estimé du bruit blanc = 24,8541

Le Tableau II.3.96 résume la performance du modèle SARIMA(1,0,0)(1,1,1)₁₂ à ajuster les données considérées et ce par l'analyse des erreurs qu'il produit.

Tableau II.3.96 Statistiques des erreurs du modèle SARIMA(1,0,0)(1,1,1)₁₂

Statistique	Période d'estimation	Période de validation
EQM	615,175	200,868
EAM	12,1837	9,54552
EM	2,46325	-5,49183

Nous y trouvons :

- 1) L'erreur quadratique moyenne (EQM)
- 2) L'erreur absolue moyenne (EAM)
- 3) L'erreur moyenne (EM)

Chacune de ces statistiques est basée sur l'erreur de prédiction un mois à l'avance qui est la différence entre la valeur de la donnée au temps t et sa prédiction au temps $t - 1$. Les deux premières mesurent la magnitude des erreurs alors que la dernière mesure le biais. Ces statistiques d'erreurs sont calculées pour la période d'estimation et celle de validation. Pour le modèle considéré ces statistiques d'erreurs sont pour la EQM en particulier, différentes d'une manière considérable et ce entre les deux périodes d'estimation et de validation. Il s'ensuit que le modèle ajusté peut être pris avec prudence quant aux prévisions des valeurs futures qui peuvent avoir une variance sous-estimée.

Le modèle retenu pour la prédiction est alors le SARIMA(1,0,0)(1,1,1)₁₂ sans constante qui peut être formulé comme suit:

$$(1 - 0,297104B)(1 - 0,221882B^{12})(1 - B^{12})X_t = (1 - 0,957414B^{12})\varepsilon_t$$

ε_t est un bruit blanc de moyenne nulle et de variance estimée 617,727.

II.3.2.9.4 Validation du modèle SARIMA(1,0,0)(1,1,1)₁₂

Il s'agit de vérifier que les résidus forment bien un bruit blanc de moyenne nulle. Les tests des autocorrélogrammes et celui du Portmanteau (Box-Pierce) sont utilisés pour tester l'indépendance des résidus et le test t de student est utilisé pour tester la nullité de leur moyenne.

- **Test des ACF et PACF**

L'auto-corrélogramme total et partiel des résidus du modèle ont été tracés tel que montré dans la Fig. II.3.60 et Fig. II.3.61.

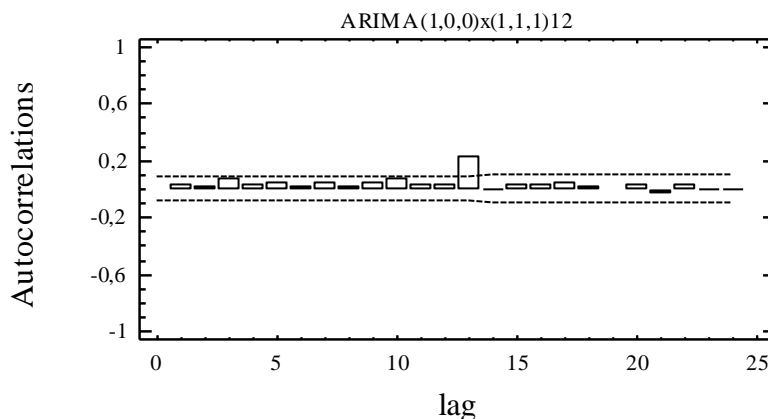


Fig. II.3.60 ACF des résidus du modèle SARIMA(1,0,0)(1,1,1)₁₂

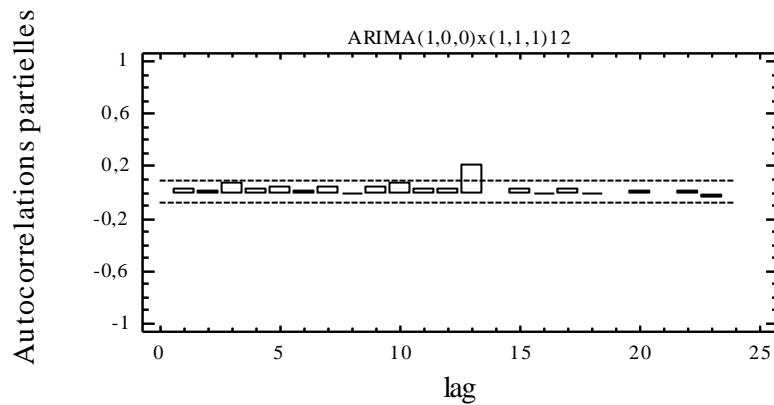


Fig. II.3.61 PACF des résidus du modèle SARIMA(1,0,0)(1,1,1)₁₂

Tous les coefficients sont pratiquement inclus dans l'intervalle de confiance (IC) tracé à 95% de confiance ce qui implique leur indépendance. Leurs valeurs sont reportées dans le Tableau II.3.97 et Tableau II.3.98.

Tableau II.3.97 ACF des résidus du modèle SARIMA(1,0,0)(1,1,1)₁₂

Lag	Auto-corrélation	Ecart-type	limite inférieure IC 95%	limite supérieure IC 95%
1	0,0298536	0,0450835	-0,0883622	0,0883622
2	0,0150372	0,0451236	-0,0884409	0,0884409
3	0,0719027	0,0451338	-0,0884608	0,0884608
4	0,033286	0,0453661	-0,088916	0,088916
5	0,0440138	0,0454157	-0,0890132	0,0890132
6	0,0164289	0,0455023	-0,089183	0,089183
7	0,0432686	0,0455143	-0,0892066	0,0892066
8	0,00790529	0,0455979	-0,0893703	0,0893703
9	0,0419865	0,0456006	-0,0893758	0,0893758
10	0,0749941	0,0456792	-0,0895297	0,0895297
11	0,0296709	0,0459287	-0,0900188	0,0900188
12	0,0297354	0,0459677	-0,0900951	0,0900951
13	0,223988	0,0460067	-0,0901717	0,0901717
14	0,00365533	0,0481723	-0,0944161	0,0944161
15	0,032879	0,0481728	-0,0944172	0,0944172
16	0,0296399	0,0482184	-0,0945065	0,0945065
17	0,0412145	0,0482554	-0,0945791	0,0945791
18	0,0141119	0,0483269	-0,0947192	0,0947192
19	-0,00707193	0,0483353	-0,0947356	0,0947356
20	0,0346488	0,0483374	-0,0947398	0,0947398
21	-0,0221134	0,0483879	-0,0948386	0,0948386
22	0,0228997	0,0484084	-0,0948789	0,0948789
23	-0,00526753	0,0484304	-0,094922	0,094922
24	-0,00444406	0,0484316	-0,0949243	0,0949243

Tableau II.3.98 PACF des résidus du modèle SARIMA(1,0,0)(1,1,1)₁₂

Lag	Auto-corrélation partielle	Ecart-type	limite inférieure IC 95%	limite supérieure IC 95%
1	0,0298536	0,0450835	-0,0883622	0,0883622
2	0,0141585	0,0450835	-0,0883622	0,0883622
3	0,0711151	0,0450835	-0,0883622	0,0883622
4	0,0291166	0,0450835	-0,0883622	0,0883622
5	0,0406522	0,0450835	-0,0883622	0,0883622
6	0,00844416	0,0450835	-0,0883622	0,0883622
7	0,0376332	0,0450835	-0,0883622	0,0883622
8	-0,00124992	0,0450835	-0,0883622	0,0883622
9	0,0372144	0,0450835	-0,0883622	0,0883622
10	0,0656761	0,0450835	-0,0883622	0,0883622
11	0,0224163	0,0450835	-0,0883622	0,0883622
12	0,0190002	0,0450835	-0,0883622	0,0883622
13	0,213552	0,0450835	-0,0883622	0,0883622
14	-0,0174881	0,0450835	-0,0883622	0,0883622
15	0,0227495	0,0450835	-0,0883622	0,0883622
16	-0,00444258	0,0450835	-0,0883622	0,0883622
17	0,0264184	0,0450835	-0,0883622	0,0883622
18	-0,0107356	0,0450835	-0,0883622	0,0883622
19	-0,0178959	0,0450835	-0,0883622	0,0883622
20	0,00823912	0,0450835	-0,0883622	0,0883622
21	-0,0300343	0,0450835	-0,0883622	0,0883622
22	0,00249451	0,0450835	-0,0883622	0,0883622
23	-0,0387558	0,0450835	-0,0883622	0,0883622
24	-0,0155761	0,0450835	-0,0883622	0,0883622

- **Test du Portmanteau**

Ce test est basé sur les 24 premières auto-corrélations des résidus, les résultats du test montrent une valeur de $P = 0,0132248$ qui est supérieure ou égale à 0.01, donc on ne peut pas rejeter l'hypothèse nulle que les résidus sont indépendants avec un intervalle de confiance de 99%.

Test de Box-Pierce

Hypothèse nulle : résidus indépendants

Valeur calculée de la statistique $Q = 37,9045$

Valeur de $P = 0,0132248$

- **Test de nullité de la moyenne**

La moyenne des résidus est de -0,770519, leur médiane est de -2,28655 et leur variance est de 443,3. La Figure II.3.62 montre la variabilité temporelle de ces derniers autour de leur moyenne. Les résultats du test t de Student montrent une valeur $P = 0,417341$ qui est supérieure à 0.05 ce qui implique que l'hypothèse nulle de nullité de la moyenne des résidus ne peut pas être rejetée à 95% de confiance.

Test t de Student

Hypothèse nulle : moyenne = 0,0

Alternative: non nulle.

Valeur calculée de la statistique $t = -0,811728$

Valeur de $P = 0,417341$

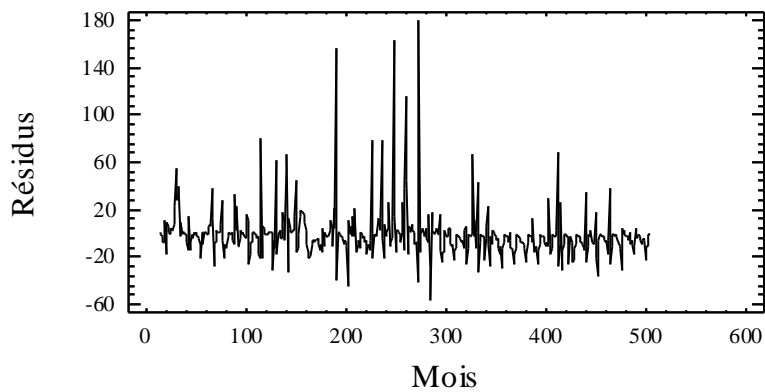


Fig. II.3.62 Résidus du modèle SARIMA(1,0,0)(1,1,1)₁₂

Les résultats de ces tests montrent bien que les résidus du modèle SARIMA(1,0,0)(1,1,1)₁₂ forment bel et bien une séquence bruit blanc de moyenne nulle et de variance 443,3. Dans ce cas, la variance expliquée par le modèle est de seulement 37,4 % de la variance totale.

II.3.2.9.5 Prédiction

En permettant au modèle retenu SARIMA(1,0,0)(1,1,1)₁₂ de faire de la prédiction en arrière nous pouvons obtenir les prédictions correspondant aussi bien à la période où les données sont disponibles qu'à celle où ces dernières ne le sont pas. Le Tableau II.3.99 présente les prédictions des 24 données futures pour la période où les données ne sont pas disponibles, avec les résidus correspondants. Les prédictions de ces 24 données futures sont présentées avec leurs intervalles de confiance à 95%. Ces limites de confiance indiquent où va être située la vraie valeur à une période déterminée dans le futur, dans l'hypothèse que le modèle ajusté est adéquat aux données en question. Tandis que les prédictions correspondant à la période où les données sont disponibles avec les résidus correspondants sont représentées en annexe A.II.3.9.

Les prédictions obtenues par le modèle, tel que présenté par la Fig. II.3.63, suivent en moyenne l'allure générale des données observées qui sont représentées par des points carrés sur la dite figure. Pour la période où les données ne sont pas disponibles, les prédictions obtenues pour les 24 mois futurs suivant la fin de la série sont représentées avec leurs limites tracées à 95% de confiance.

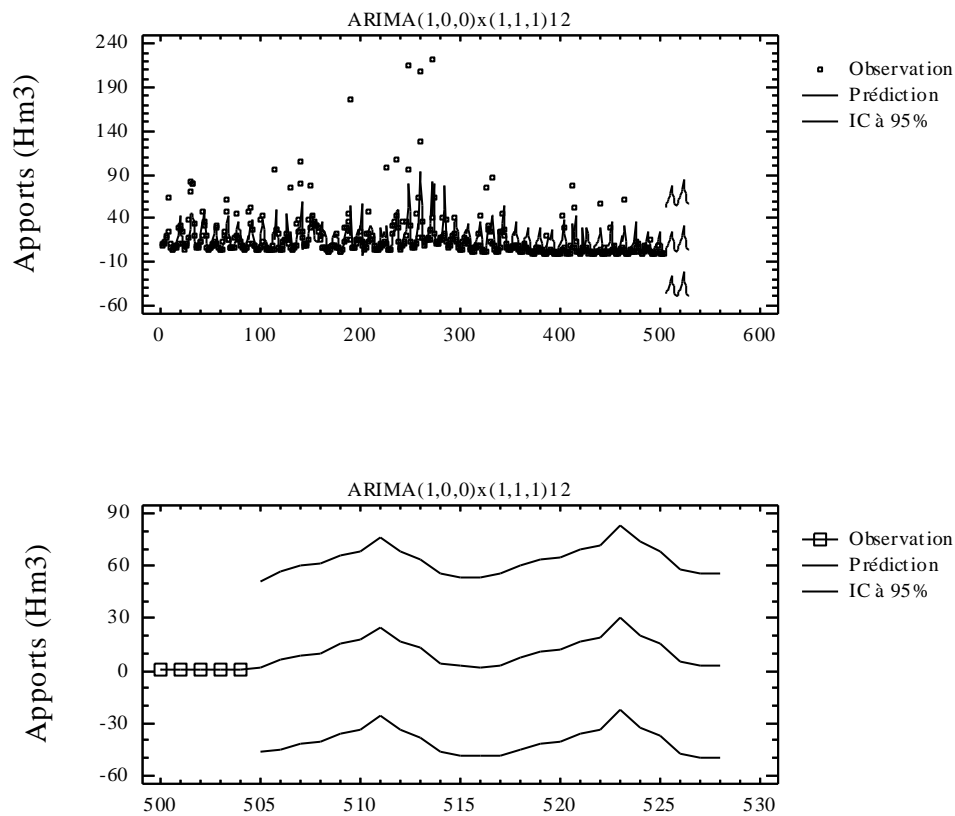


Fig. II.3.63 Prédictions des apports mensuels à PIERRE DU CHAT obtenues par le modèle SARIMA(1,0,0)(1,1,1)₁₂

Tableau II.3.99 Prévisions des apports mensuels à PIERRE DU CHAT pour les 24 mois futurs

Période	Prédiction	- IC 95%	+IC 95%
505,0	2,32326	-46,5109	51,1574
506,0	6,02043	-44,9234	56,9643
507,0	8,75502	-42,3709	59,8809
508,0	9,88387	-41,2581	61,0258
509,0	15,2687	-35,8746	66,4121
510,0	17,3426	-33,8009	68,4861
511,0	25,3302	-25,8133	76,4737
512,0	16,9571	-34,1864	68,1006
513,0	13,0211	-38,1224	64,1646
514,0	4,43813	-46,7054	55,5816
515,0	2,65857	-48,4849	53,8021
516,0	2,30531	-48,8382	53,4488
517,0	3,45617	-49,2928	56,2052
518,0	7,48632	-45,4021	60,3747
519,0	10,6244	-42,2763	63,525
520,0	11,8976	-41,0042	64,7993
521,0	16,7904	-36,1114	69,6923
522,0	19,2249	-33,6769	72,1268
523,0	30,4728	-22,429	83,3747
524,0	20,6975	-32,2043	73,5994
525,0	15,8881	-37,0137	68,79
526,0	5,40069	-47,5012	58,3026
527,0	3,22628	-49,6756	56,1281
528,0	2,79463	-50,1072	55,6965

II.3.2.10 REMCHI

Les statistiques de la tendance centrale, de la variabilité et de la forme de la série chronologique des apports mensuels observés à REMCHI durant les 516 mois allant de Septembre 1948 à Aout 1993 sont résumées dans le Tableau II.3.100 et la variabilité temporelle de cette série chronologique est illustrée par la Fig. II.3.64.

Tableau II.3.100 Statistiques des apports mensuels observés à REMCHI(Sep1948 à Aout 1993)

Paramètre	Valeur
Nombre d'observations	516
Moyenne	8,01665
Variance	239,136
Déviation Standard	15,464
Minimum	0,0003
Maximum	177,04
Std. skewness	56,2314
Std. kurtosis	243,002

La moyenne de la série est de 8,01665 alors que sa variance autour de cette moyenne est de 239,136 avec un écart-type de 15,464. La skewness et la kurtosis standardisées indiquent un éloignement remarquable de la distribution normale.

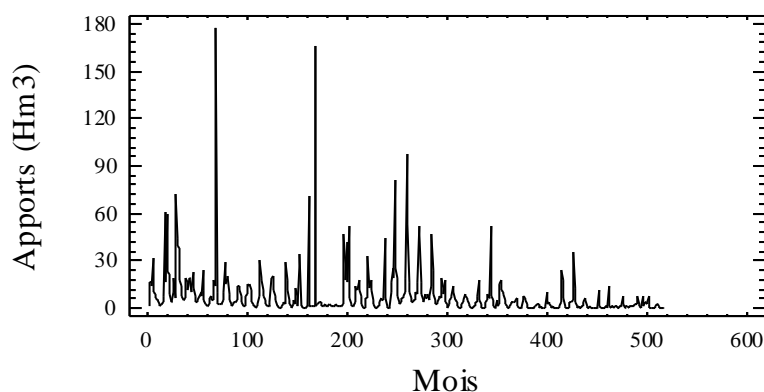


Fig. II.3.64 Apports mensuels observés à REMCHI

II.3.2.10.1 Stationnarisation

Le détail du calcul des séries différenciées pour les 20 premiers mois est montré dans le Tableau II.3.101, tandis que les statistiques des séries différenciées pour l'ensemble des données sont représentées dans le Tableau II.3.102.

Tableau II.3.101 Calcul des séries différenciées (données des 20 premiers mois)

N°	Apport	DIFF(Apport,1)	SDIFF(Apports,12) = S	DIFF(S,1)
1	1,04			
2	16,89	15,85		
3	16,93	0,04		
4	13,46	-3,47		
5	31,29	17,83		
6	10,36	-20,93		
7	8,6	-1,76		
8	6,2	-2,4		
9	5,1	-1,1		
10	4,74	-0,36		
11	2,56	-2,18		
12	1,18	-1,38		
13	2,2	1,02	1,16	
14	3,02	0,82	-13,87	-15,03
15	4,57	1,55	-12,36	1,51
16	7,35	2,78	-6,11	6,25
17	60,7	53,35	29,41	35,52
18	16,82	-43,88	6,46	-22,95
19	59,51	42,69	50,91	44,45
20	22,73	-36,78	16,53	-34,38

Dans le cas de la série étudiée, la stationnarité est obtenue en éliminant la saisonnalité ($S=12$) et ce effectuant les premières différences saisonnières ($D=1$). La variance des premières différences saisonnières $SDIFF(Apports,12)$ est de 351.96 et la série est stationnaire pour $S=12$, $d=0$ et $D=1$.

Tableau II.3.102 Statistiques descriptives des séries différenciées pour les 516 mois.

Série	N	Moyenne	Variance	Ecart-type
Apports	516	8.01	239.13	15.46
DIFF(Apports,1)	515	-0.002	352.08	18.76
SDIFF(Apports ,12)	504	-0.217	351.96	18.76
DIFF(SDIFF(Apports,12) ,1)	503	-0.002	648.51	25.46

Notation : DIFF (x,n) : différences régulières de x d'ordre n ; SDIFF(x,n,s) :différences saisonnières s de x d'ordre n.

II.3.2.10.2 Identification

Dans le cas de la station courante, les calculs de l'ACF et de la PACF des apports stationnaires sont reportés dans les Tableaux II.3.103 et 104 respectivement et leurs tracés sont illustrés par les Figures II.3.65 et 66.

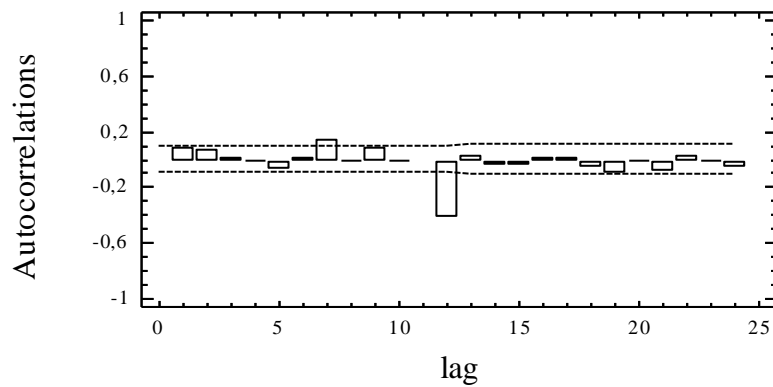


Fig. II.3.65 ACF des apports mensuels

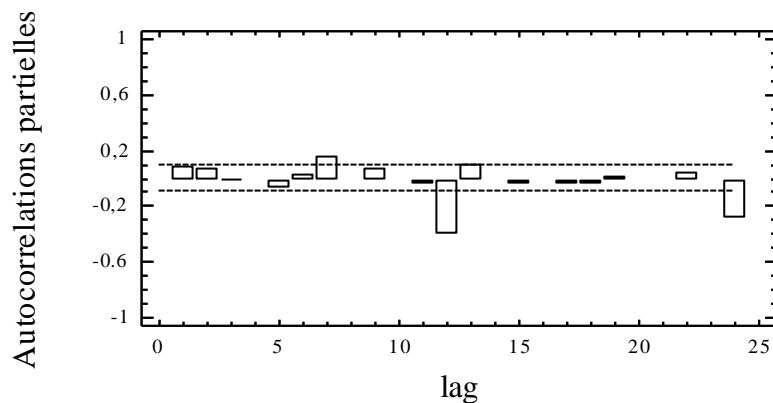


Fig. II.3.66 PACF des apports mensuels

Tableau II.3.103 ACF des apports mensuels stationnaires

Lag	Auto-corrélation	Ecart-type	-IC à 95%	IC à 95%
1	0,0805515	0,0445435	-0,0873039	0,0873039
2	0,0670336	0,0448316	-0,0878686	0,0878686
3	0,00514241	0,0450301	-0,0882575	0,0882575
4	-0,014058	0,0450312	-0,0882598	0,0882598
5	-0,0763252	0,0450399	-0,0882768	0,0882768
6	0,00922613	0,0452958	-0,0887784	0,0887784
7	0,136763	0,0452996	-0,0887857	0,0887857
8	-0,00436177	0,0461115	-0,0903771	0,0903771
9	0,0794266	0,0461123	-0,0903787	0,0903787
10	-0,00351647	0,046383	-0,0909092	0,0909092
11	-0,0284356	0,0463835	-0,0909102	0,0909102
12	-0,420554	0,0464181	-0,090978	0,090978
13	0,0146923	0,0534461	-0,104753	0,104753
14	-0,0502308	0,0534542	-0,104768	0,104768
15	-0,0426733	0,0535477	-0,104952	0,104952
16	0,005895	0,0536152	-0,105084	0,105084
17	0,0047993	0,0536164	-0,105086	0,105086
18	-0,0580303	0,0536173	-0,105088	0,105088
19	-0,104626	0,0537418	-0,105332	0,105332
20	-0,0015059	0,0541444	-0,106121	0,106121
21	-0,0921029	0,0541445	-0,106121	0,106121
22	0,0160828	0,0544545	-0,106729	0,106729
23	-0,00436248	0,0544639	-0,106747	0,106747
24	-0,065823	0,0544646	-0,106749	0,106749

Tableau II.3.104 PACF des apports mensuels stationnaires

Lag	Auto-corrélation	Ecart-type	-IC à 95%	IC à 95%
1	0,0805515	0,0445435	-0,0873039	0,0873039
2	0,0609405	0,0445435	-0,0873039	0,0873039
3	-0,00488677	0,0445435	-0,0873039	0,0873039
4	-0,0183519	0,0445435	-0,0873039	0,0873039
5	-0,074654	0,0445435	-0,0873039	0,0873039
6	0,0230124	0,0445435	-0,0873039	0,0873039
7	0,146611	0,0445435	-0,0873039	0,0873039
8	-0,0281833	0,0445435	-0,0873039	0,0873039
9	0,0619284	0,0445435	-0,0873039	0,0873039
10	-0,0200107	0,0445435	-0,0873039	0,0873039
11	-0,0319655	0,0445435	-0,0873039	0,0873039
12	-0,410528	0,0445435	-0,0873039	0,0873039
13	0,09283	0,0445435	-0,0873039	0,0873039
14	-0,0164535	0,0445435	-0,0873039	0,0873039
15	-0,0329181	0,0445435	-0,0873039	0,0873039
16	-0,0238871	0,0445435	-0,0873039	0,0873039
17	-0,0325108	0,0445435	-0,0873039	0,0873039
18	-0,0544412	0,0445435	-0,0873039	0,0873039
19	0,00525118	0,0445435	-0,0873039	0,0873039
20	-0,0170239	0,0445435	-0,0873039	0,0873039
21	-0,017832	0,0445435	-0,0873039	0,0873039
22	0,0327651	0,0445435	-0,0873039	0,0873039
23	-0,0223335	0,0445435	-0,0873039	0,0873039
24	-0,295964	0,0445435	-0,0873039	0,0873039

L'aspect des auto-corrélogrammes de la série stationnaire considérée laisserait supposer les modèles suivants :

- (A) SARIMA(0,0,1)(1,1,1)₁₂ avec constante
- (B) SARIMA(0,0,1)(0,1,1)₁₂ avec constante
- (C) SARIMA(0,0,1)(0,1,2)₁₂ avec constante
- (D) SARIMA(0,0,0)(1,1,1)₁₂ avec constante
- (E) SARIMA(1,0,1)(0,1,1)₁₂ avec constante

II.3.2.10.3 Estimation des modèles ajustés

Pour l'estimation des paramètres de ces modèles, il faut rappeler que 516 données ont été utilisées dont les 200 dernières ont servi à la validation. Les paramètres des modèles ont alors été estimés sur une période de 316 mois.

L'analyse et la comparaison des modèles ajustés tel que présenté dans le Tableau II.3.105, montre que tous les modèles ajustés produisent des résidus complètement indépendants sauf le modèle (A). D'un autre côté, il faut noter que la moyenne de la variance des résidus produits pour tous les modèles ne diffère pas considérablement entre la période d'estimation que celle de la validation. En outre, le modèle qui produit le minimum de cette variance est le modèle (E), mais ses termes AR(1) et MA(1) sont à réduire car ils sont statistiquement non significatifs. Le modèle (C) qui présente alors le minimum de cette variance doit revoir ses termes MA(1) et SMA(2) à la baisse car ils sont également statistiquement non significatifs. Le modèle (D) doit aussi voir son terme SAR(1) à la baisse pour la même raison, ce qui nous renvoie au modèle (B) qui présente dans ce cas le minimum de la variance des résidus, c'est lui qui est donc retenu pour la prédiction.

Tableau II.3.105 Comparaison des modèles ajustés

Modèle	Variance des résidus		Indépendance des résidus
	Période d'estimation	Période de validation	
(A)	312,957	34,579	NON
(B)	314,079	39,6037	OK
(C)	312,866	35,1774	OK
(D)	316,144	37,4275	OK
(E)	312,734	37,9258	OK

Le Tableau II.3.106 résume l'estimation des paramètres du modèle retenu SARIMA(0,0,1)(0,1,1)₁₂ et la signification statistique de ses termes. Les termes correspondant à une valeur de $P < 0.05$ sont statistiquement significativement différents de zéro à 95% de confiance.

Les paramètres du modèle sont inversibles, et la valeur du terme MA(1) est différente de 'un'. La valeur de P pour les termes MA(1) et SMA(1) est inférieure à 0.05 donc ces derniers sont significativement différents de zéro à 95% de confiance. La valeur de P pour la moyenne est supérieure à 0.05 donc elle est statistiquement non significative et peut être éliminée. L'estimation de l'écart-type du bruit blanc est de 17,7732 et la variance des résidus est de 176.84 ce qui veut dire que le présent modèle explique 50% de la variance de la série modélisée qui est de 351.96.

Tableau II.3.106 Estimation des paramètres du modèle retenu SARIMA(0,0,1)(0,1,1)₁₂

Paramètre	Estimation	Ecart- type	t	P
MA(1)	-0,114438	0,057445	-1,99212	0,046900
SMA(1)	0,966848	0,0117654	82,1771	0,000000
Moyenne	-0,176678	0,137671	-1,28334	0,199967
Constante	-0,176678			

SAR(n): Autorégressif saisonnier d'ordre n; SMA(n) : Moyenne mobile saisonnière d'ordre n.

La variance estimée du bruit blanc = 315,888 avec 501 DDL et l'écart type estimé du bruit blanc = 17,7732

Le Tableau II.3.107 résume la performance du modèle SARIMA(0,0,1)(0,1,1)₁₂ à ajuster les données considérées et ce par l'analyse des erreurs qu'il produit.

Tableau II.3.107 Statistiques des erreurs du modèle SARIMA(0,0,1)(0,1,1)₁₂

Statistique	Période d'estimation	Période de validation
EQM	314,079	39,6037
EAM	7,71663	4,28019
EM	1,39862	-2,48402

Nous y trouvons :

- 1) L'erreur quadratique moyenne (EQM)
- 2) L'erreur absolue moyenne (EAM)
- 3) L'erreur moyenne (EM)

Chacune de ces statistiques est basée sur l'erreur de prédiction un mois à l'avance qui est la différence entre la valeur de la donnée au temps t et sa prédiction au temps $t - 1$. Les deux premières mesurent la magnitude des erreurs alors que la dernière mesure le biais. Ces statistiques d'erreurs sont calculées pour la période d'estimation et celle de validation. Pour le modèle considéré ces statistiques d'erreurs sont considérablement différentes entre les deux périodes d'estimation et de validation en particulier pour la EQM. Il s'ensuit que le modèle ajusté peut être utilisé avec prudence pour faire la prévision des valeurs futures qui peuvent avoir une sous-estimation de la variance.

Le modèle retenu pour la prédiction est alors le SARIMA(0,0,1)(0,1,1)₁₂ sans constante qui peut être formulé comme suit:

$$(1 - B^{12})X_t = (1 + 0,114438B)(1 - 0,966848B^{12})\varepsilon_t - 0,176678$$

ε_t est un bruit blanc de moyenne nulle et de variance estimée 176.84.

II.3.2.10.4 Validation du modèle SARIMA(0,0,1)(0,1,1)₁₂

Il s'agit de vérifier que les résidus forment bien un bruit blanc de moyenne nulle. Les tests des autocorrélogrammes et celui du Portmanteau (Box-Pierce) sont utilisés pour tester l'indépendance des résidus et le test t de student est utilisé pour tester la nullité de leur moyenne.

- **Test des ACF et PACF**

L'auto-corrélogramme total et partiel des résidus du modèle ont été tracés tel que montré dans la Fig. II.3.67 et Fig. II.3.68.

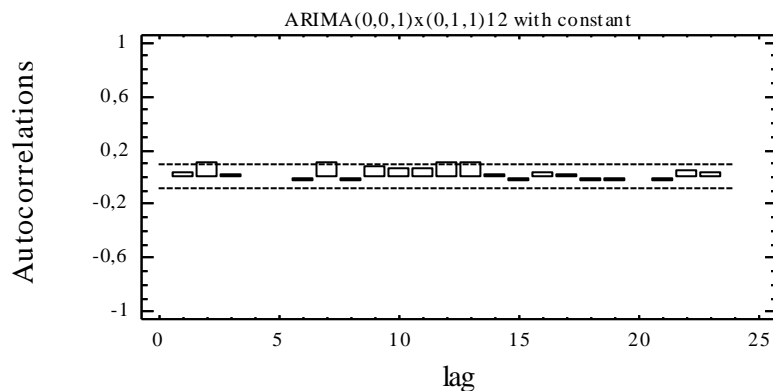


Fig. II.3.67 ACF des résidus du modèle SARIMA(0,0,1)(0,1,1)₁₂

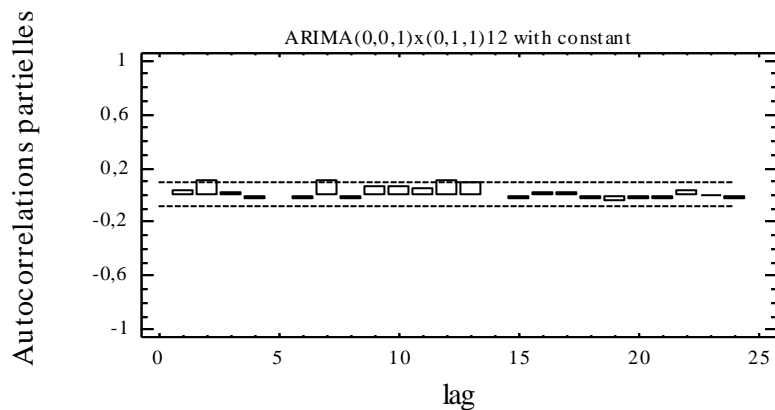


Fig. II.3.68 PACF des résidus du modèle SARIMA(0,0,1)(0,1,1)₁₂

Tous les coefficients sont pratiquement inclus dans l'intervalle de confiance (IC) tracé à 95% de confiance ce qui implique leur indépendance. Leurs valeurs sont reportées dans le Tableau II.3.108 et Tableau II.3.109.

Tableau II.3.108 ACF des résidus du modèle SARIMA(0,0,1)(0,1,1)₁₂

Lag	Auto-corrélation	Ecart-type	limite inférieure IC 95%	limite supérieure IC 95%
1	0,0337794	0,0445435	-0,0873039	0,0873039
2	0,0999848	0,0445943	-0,0874035	0,0874035
3	0,0161837	0,0450369	-0,0882709	0,0882709
4	-0,0156345	0,0450485	-0,0882936	0,0882936
5	-0,0139125	0,0450592	-0,0883146	0,0883146
6	-0,0289667	0,0450678	-0,0883314	0,0883314
7	0,0967541	0,0451047	-0,0884037	0,0884037
8	-0,0446368	0,0455146	-0,0892072	0,0892072
9	0,0687246	0,0456014	-0,0893773	0,0893773
10	0,0559787	0,0458064	-0,0897791	0,0897791
11	0,0565118	0,045942	-0,0900448	0,0900448
12	0,111091	0,0460797	-0,0903147	0,0903147
13	0,0970829	0,0466081	-0,0913503	0,0913503
14	0,017232	0,0470076	-0,0921333	0,0921333
15	-0,0298074	0,0470201	-0,0921579	0,0921579
16	0,0249001	0,0470576	-0,0922313	0,0922313
17	0,00771276	0,0470837	-0,0922826	0,0922826
18	-0,0281621	0,0470862	-0,0922875	0,0922875
19	-0,0343486	0,0471196	-0,092353	0,092353
20	-0,0161933	0,0471693	-0,0924503	0,0924503
21	-0,0255179	0,0471803	-0,0924719	0,0924719
22	0,0415721	0,0472077	-0,0925255	0,0925255
23	0,0294382	0,0472803	-0,0926678	0,0926678
24	-0,0159987	0,0473166	-0,0927391	0,0927391

Tableau II.3.109 PACF des résidus du modèle SARIMA(0,0,1)(0,1,1)₁₂

Lag	Auto-corrélation partielle	Ecart-type	limite inférieure IC 95%	limite supérieure IC 95%
1	0,0337794	0,0445435	-0,0873039	0,0873039
2	0,0989566	0,0445435	-0,0873039	0,0873039
3	0,00990602	0,0445435	-0,0873039	0,0873039
4	-0,026603	0,0445435	-0,0873039	0,0873039
5	-0,015438	0,0445435	-0,0873039	0,0873039
6	-0,0243265	0,0445435	-0,0873039	0,0873039
7	0,103162	0,0445435	-0,0873039	0,0873039
8	-0,0463116	0,0445435	-0,0873039	0,0873039
9	0,0530209	0,0445435	-0,0873039	0,0873039
10	0,0577617	0,0445435	-0,0873039	0,0873039
11	0,0463379	0,0445435	-0,0873039	0,0873039
12	0,0990994	0,0445435	-0,0873039	0,0873039
13	0,0887477	0,0445435	-0,0873039	0,0873039
14	-0,0164895	0,0445435	-0,0873039	0,0873039
15	-0,0339696	0,0445435	-0,0873039	0,0873039
16	0,0209293	0,0445435	-0,0873039	0,0873039
17	0,0194703	0,0445435	-0,0873039	0,0873039
18	-0,0318576	0,0445435	-0,0873039	0,0873039
19	-0,0572149	0,0445435	-0,0873039	0,0873039
20	-0,0278431	0,0445435	-0,0873039	0,0873039
21	-0,0255711	0,0445435	-0,0873039	0,0873039
22	0,0328468	0,0445435	-0,0873039	0,0873039
23	0,00454942	0,0445435	-0,0873039	0,0873039
24	-0,0468745	0,0445435	-0,0873039	0,0873039

- **Test du Portmanteau**

Ce test est basé sur les 24 premières auto-corrélations des résidus, les résultats du test montrent une valeur de $P = 0,0695338$ qui est supérieure ou égale à 0.01, donc on ne peut pas rejeter l'hypothèse nulle que les résidus sont indépendants avec un intervalle de confiance de 99%.

Test de Box-Pierce

Hypothèse nulle : résidus indépendants

Valeur calculée de la statistique $Q = 32,4825$

Valeur de $P = 0,0695338$

- **Test de nullité de la moyenne**

La moyenne des résidus est de -0,14211, leur médiane est de -1,68954 et la variance est de 203,675. La Figure II.3.69 montre la variabilité temporelle de ces derniers autour de leur moyenne. Les résultats du test t de Student montrent une valeur $P = 0,823199$ qui est supérieure à 0.05 ce qui implique que l'hypothèse nulle de nullité de la moyenne des résidus ne peut pas être rejetée à 95% de confiance.

Test t de Student

Hypothèse nulle : moyenne = 0,0

Alternative: non nulle.

Valeur calculée de la statistique $t = -0,223549$

Valeur de $P = 0,823199$

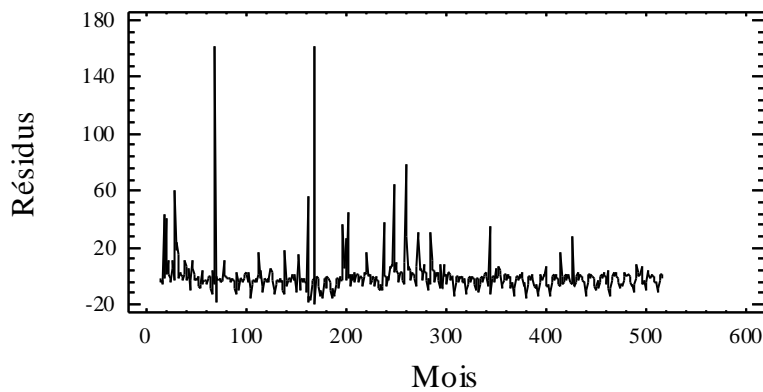


Fig. II.3.69 Résidus du modèle SARIMA(0,0,1)(0,1,1)₁₂

Les résultats de ces tests montrent bien que les résidus du modèle SARIMA(0,0,1)(0,1,1)₁₂ forment bel et bien une séquence bruit blanc de moyenne nulle et de variance 203,675.

II.3.2.10.5 Prédiction

En permettant au modèle retenu SARIMA(0,0,1)(0,1,1)₁₂ de faire de la prédiction en arrière nous pouvons obtenir les prédictions correspondant aussi bien à la période où les données sont disponibles qu'à celle où ces dernières ne le sont pas. Le Tableau II.3.110 présente les prédictions des 24 données futures pour la période où les données ne sont pas disponibles, avec les résidus correspondants. Les prédictions de ces 24 données futures sont présentées avec leurs intervalles de confiance à 95%. Ces limites de confiance indiquent où va être située la vraie valeur à une période déterminée dans le future, dans l'hypothèse que le modèle ajusté est adéquat aux données en question. Tandis que les prédictions correspondant à la période où les données sont disponibles avec les résidus correspondants sont présentées en annexe A.II.3.10.

Les prédictions obtenues par le modèle, tel que présenté par la Fig. II.3.70, suivent en général l'allure des données observées qui sont représentées par des points carrés sur la dite figure. Pour la période où les données ne sont pas disponibles, les prédictions obtenues pour les 24 mois futurs suivant la fin de la série sont représentées avec leurs limites tracées à 95% de confiance.

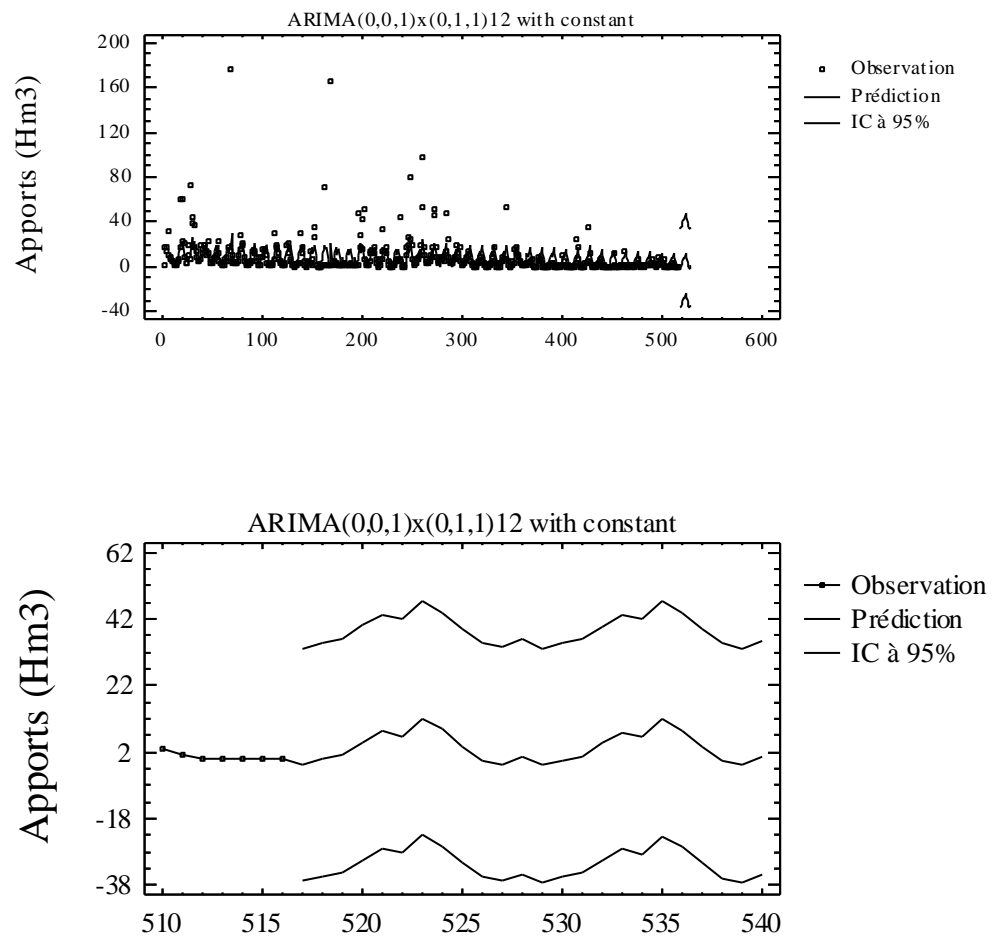


Fig. II.3.70 Prédictions des apports mensuels à REMCHI obtenues par le modèle SARIMA(0,0,1)(0,1,1)₁₂

Tableau II.3.110 Prévisions des apports mensuels à REMCHI pour les 24 mois futurs

Période	Prédiction	- IC 95%	+IC 95%
517,0	-1,82661	-36,7459	33,0927
518,0	-0,276403	-35,4236	34,8708
519,0	1,06172	-34,0855	36,2089
520,0	4,7815	-30,3657	39,9287
521,0	8,23857	-26,9087	43,3858
522,0	6,7404	-28,4068	41,8876
523,0	12,1129	-23,0343	47,2602
524,0	8,71222	-26,435	43,8594
525,0	3,89791	-31,2493	39,0451
526,0	-0,405124	-35,5524	34,7421
527,0	-1,71833	-36,8656	33,4289
528,0	0,627778	-34,5194	35,775
529,0	-1,89176	-37,058	33,2745
530,0	-0,453081	-35,6196	34,7135
531,0	0,88504	-34,2815	36,0516
532,0	4,60483	-30,5617	39,7714
533,0	8,06189	-27,1046	43,2284
534,0	6,56372	-28,6028	41,7303
535,0	11,9363	-23,2303	47,1028
536,0	8,53554	-26,631	43,7021
537,0	3,72123	-31,4453	38,8878
538,0	-0,581802	-35,7483	34,5847
539,0	-1,89501	-37,0615	33,2715
540,0	0,451099	-34,7154	35,6176

II.2.3 RECAPITULATIF ET CONCLUSIONS

L'application de la méthode de Box-Jenkins à la modélisation stochastique des apports liquides mensuels observés aux 10 stations hydrométriques au Nord d'Algérie, est récapitulée dans les Tableaux II.3.111 et 112.

Les résultats de cette modélisation montrent pour chaque série chronologique, le modèle SARIMA obtenu ainsi que le pourcentage de la variance qu'il explique, en plus des statistiques des erreurs de prédiction obtenues pendant la période d'estimation et la période de validation.

Il est clair que la partie structurée des séries chronologiques mensuelles est plus forte que celle des séries chronologiques annuelles. Ceci est dû à l'existence de la composante saisonnière qui crée une sorte de dépendance au sein des données, et qui fait que les modèles SARIMA obtenus prennent mieux en charge l'explication de la variabilité des séries considérées, avec un taux d'explication moyen de 55% de la variance totale qui est presque le double du taux d'explication des modèles annuels. Ces modèles s'adaptent mieux aux données, car les statistiques des erreurs sont plus faibles en moyenne pendant la période de validation comparativement à la période d'estimation et les prédictions qu'ils produisent sont satisfaisantes.

Tableau II.3.111 Récapitulatif des principaux résultats obtenus à l'échelle mensuelle.

N°	Station	N	DIFF	σ^2	Modèle	σ_ε^2	σ^2	TEM	TEV
				(totale)		(calculée)	(expliquée en %)		
01	Ain berda	372	d=0, D=1 S=12	10	SARIMA (1, 0, 0) x (0, 1, 1) ₁₂	4,17	58,3	OK	NON
02	Béni bahdel	828	d=1, D=1 S=12	103,73	SARIMA (1, 1, 1) x (2, 1, 1) ₁₂	39,37	62	OK	NON
03	Boucheougouf	408	d=0, D=1 S=12	339,98	SARIMA (1, 0, 0) x (0, 1, 1) ₁₂	158,28	53,4	NON	OK
04	Bouhnifia	708	d=1, D=1 S=12	443,5	SARIMA (1, 1, 1) x (0, 1, 1) ₁₂	128,29	71	OK	NON
05	Cheffia	324	d=1, D=1 S=12	849,9	SARIMA (1, 1, 1) x (0, 1, 1) ₁₂	291,35	65,7	OK	OK
06	Ksob	324	d=0, D=1 S=12	29,34	SARIMA (0, 0, 3) x (0, 1, 1) ₁₂	12,83	56,2	OK	OK
07	Mefrouche	600	d=0, D=1 S=12	5,06	SARIMA (1, 0, 0) x (0, 1, 1) ₁₂	2,80	44,6	OK	OK
08	Mirebeck	324	d=0, D=1 S=12	4802,36	SARIMA (2, 0, 0) x (0, 1, 1) ₁₂	2020,58	58	OK	NON
09	Pierre du chat	504	d=0, D=1 S=12	709	SARIMA (1, 0, 0) x (1, 1, 1) ₁₂	443,3	37,4	OK	NON
10	Remchi	516	d=0, D=1 S=12	351,96	SARIMA (0, 0, 1) x (0, 1, 1) ₁₂	203,67	42	OK	NON
Moyenne				764,48		330,48	54,86		

Notation : N : nombre de données ; DIFF : différenciation ; σ^2 (Totale) : variance totale après la différenciation ; σ_ε^2 (calculée) : variance calculée des résidus ; σ^2 (expliquée en %) : variance expliquée en pourcentage ; TEM : test d'égalité de la moyenne ; TEV : test d'égalité de la variance.

Tableau II.3.112 Statistiques des erreurs des modèles ajustés à l'échelle mensuelle.

N°	Station	EQM		EAM		EM	
		Période d'estimation	Période de validation	Période d'estimation	Période de validation	Période d'estimation	Période de validation
01	Ain berda	4,74	2,79	1,05	0,96	0,009	0,17
02	Beni bahdel	47,34	15,85	3,30	2,46	0,13	-0,39
03	Boucheougouf	177,65	103,25	7,16	6,00	-0,05	-0,97
04	Bouhnifia	155,90	64,03	5,57	4,13	1,20	-0,10
05	Cheffia	301,22	246,25	9,91	9,70	0,23	-1,33
06	Ksob	13,01	12,98	2,41	2,13	-0,31	0,83
07	Mefrouche	3,59	1,18	0,99	0,66	0,05	-0,23
08	Mirebeck	2428,46	1240,27	24,39	19,36	-0,55	6,00
09	Pierre du chat	615,17	200,86	12,18	9,54	2,46	-5,49
10	Remchi	314,07	39,60	7,71	4,28	1,39	-2,48
Moyenne		406,11	192,70	7,46	5,92	0,45	-0,39

Notation : EQM : erreur quadratique moyenne ; EAM : erreur absolue moyenne ; EM : erreur moyenne.

Les principaux résultats peuvent être résumés comme suit:

7. Pour cette catégorie de données, l'existence de la variabilité saisonnière compromet largement la normalité des séries considérées. En effet, il se trouve que pour la totalité des stations les coefficients de la skewness et la kurtosis standardisées sont en dehors de l'intervalle ± 2 , indiquant les limites pour une distribution Normale. Dans le but de se ramener à la Normalité et satisfaire l'une des hypothèses de base de la méthode de Box-Jenkins, les transformations de type logarithmique ou racine carré se sont avérées suffisantes parmi toute la panoplie de transformations proposées par le logiciel de travail SGWIN+.
8. Les stations hydrométriques étudiées ont toutes été sujettes à une instabilité dans la variance ou la moyenne ce qui confirme le caractère non stationnaire des séries chronologiques des apports mensuels et nécessite des transformations mathématiques de ces dernières. A cet effet, les transformations visant la stationnarité des séries chronologiques étudiées ont été pour la totalité des différenciations saisonnières d'ordre 1 ($D=1$). Cette transformation permet d'éliminer la variation saisonnière qui est de l'ordre de 12 en permettant une stabilité dans la variance. En plus de cette différenciation saisonnière, une autre différenciation mais de type régulier d'ordre 1 ($d=1$) a été appliquée aux données des stations BENI BAHDEL, BOUHNIFIA et CHEFFIA afin de stabiliser le niveau de leur moyenne.
9. L'application de la méthode de Box-Jenkins aux données mensuelles met en évidence l'importance de la contribution saisonnière dans la variabilité des séries de données considérées. En effet, pour la totalité des stations la partie structurée en moyenne représente un peu près de 57 % de la variation totale ce qui présente un peu plus du double de la variabilité moyenne expliquée par les modèles annuels.
10. La modélisation des apports mensuels par la méthode de Box-Jenkins a conduit à la détermination de différents modèles saisonniers SARIMA qui sont simples et également parcimonieux. Les termes AR dans ces modèles sont en général d'ordre 1 sauf pour REMCHI, et KSOB où ce terme est réduit à zéro ou encore pour MIREBECK où ce terme est de l'ordre de 2. Ce résultat est conforme au sens physique car l'apport du mois présent dans n'importe quel cours d'eau est étroitement lié à l'apport du mois passé, cette relation peut être éventuellement étendue au mois avant le mois passé par le biais des écoulements hypodermiques. Cela dit, les termes MA et SMA sont en général de l'ordre de 1, ce qui a également un sens physique puisque cela peut être attribué à la contribution aléatoire de la précipitation par exemple. Parmi les modèles obtenus le SARIMA(1,0,0)(0,1,1)₁₂ revient 3 fois sur 10, le SARIMA(1,1,1)(0,1,1)₁₂ revient 2 fois sur 10, ceci ne peut pas nous conduire à une conclusion quant au modèle qui peut être adopté pour les apports mensuels dans la partie Nord d'Algérie, pour cela l'extension de l'étude à toutes les stations hydrométriques disponibles est fortement recommandée.
11. Pour l'estimation des modèles SARIMA, les séries chronologiques ont été divisées en deux périodes : la période sur laquelle l'estimation des paramètres du modèle est réalisée qui représente en général les 2/3 des données et la période sur laquelle la validation de ce modèle est effectuée et qui représente le 1/3. Une fois le modèle estimé et validé, il a été employé à la prédiction en avant comme en arrière. Les erreurs de prédiction obtenues

alors sur les deux périodes forment un des critères marquant la performance du modèle en question. En effet, il est clair d'après le Tableau II.3.88 que la moyenne de l'erreur quadratique moyenne (EQM) des 10 modèles ajustés est visiblement meilleure sur la période de validation comparativement à la période d'estimation. La même chose peut se dire de la moyenne de l'erreur absolue moyenne (EAM) ou encore de celle de l'erreur moyenne (EM). Ceci est sans doute la traduction de la validité de ces modèles dans la formulation du processus étudié et une indication sur la performance de ces derniers quant à la prévision.

12. Pour les 10 stations étudiées, il faut noter que les 10 modèles correspondants produisent tous des résidus qui forment un bruit blanc gaussien. Il faudra noter cependant, que lorsqu'il s'agit de la stationnarité de ces derniers les deux points suivants sont à soulever : d'une part, remarquons que le test d'égalité de la moyenne sur la période d'estimation et de validation est vérifié pour 9 stations parmi 10 de la présente étude ; ce qui renforce la stationnarité de la moyenne et consolide la performance des modèles ajustés. D'une autre part, nous constatons que le test d'homoscédasticité n'est vérifié que pour 4 stations. Ce défaut de stationnarité dans la variance est obtenu inévitablement avec le meilleur modèle selon la méthodologie de Box-Jenkins, il reflète la nature « un peu naturelle » des résidus et peut être dû à la complexité de la variation saisonnière qui n'est pas bien prise en charge par les techniques de dessaisonnalisation de Box-Jenkins (**Koutsoyiannis, 2001**).

CONCLUSION II

La problématique dans cette seconde partie était d'abord de trouver un modèle pratique pour la prédiction des apports de rivières au Nord d'Algérie, et ensuite de l'estimer. Afin de répondre à ces préoccupations, l'analyse des séries chronologiques que nous avons effectuée jusque là, a conduit pour chacune des 10 séries considérées, à la construction d'un modèle pour la variable hydrologique des apports, aussi bien au pas annuel que mensuel. Ce modèle pratique approche le plus possible le processus générateur qui est inconnu et de nature stochastique, et les prédictions qu'il procure sont acceptables de manière générale.

Le chapitre II.1 présente une orientation vers un type particulier de modèles paramétriques : les modèles ARMA. A cet effet, la méthodologie de Box-Jenkins pour la modélisation stochastique des séries chronologiques y est exposée avec les étapes principales correspondantes, au même titre que les outils mathématiques indispensables à l'accomplissement de chacune de ces dernières. Un accent particulier a été mis sur les hypothèses de normalité et de stationnarité sur lesquelles reposent les fondements théoriques de la méthode, et la structure de dépendance interne des données qui est à la base de cette dernière a été mise en valeur ainsi que le caractère itératif de la procédure. Enfin, nous avons exposé comment le modèle ARMA résultant peut être mis en œuvre pour faire de la prédiction tout en indiquant les intervalles de confiance de ces dernières.

Dans les chapitres II.2 et II.3, nous avons exposé en détail et pour chacune des 10 stations, comment la méthode de Box-Jenkins a été mise en œuvre dans le but de trouver le modèle le plus adéquat aux données observées. Le chapitre II.2 a été consacré à l'application de la dite méthode aux données annuelles qui sont généralement stationnaires, normalement distribuées avec une faible auto-corrélation dans leur structure de dépendance interne, tandis que le chapitre II.3 concernait les données mensuelles qui sont, au contraire caractérisées par une instabilité saisonnière au niveau de la variance, par une distribution non normale et par une forte auto-corrélation régulière et saisonnière dans la structure de dépendance interne. Les résultats obtenus dans ces deux chapitres ont conduit à la détermination de modèles pratiques, simples et parcimonieux qui représentent le plus possible le processus générateur de la variable hydrologique en question et reflètent l'ampleur de la variabilité aléatoire de celle-ci et les prédictions obtenues sont acceptables.

L'usage de ces classes de modèles pour des fins hydrologiques telles que la prédiction est devenu classique, mais il faut dire qu'ils sont toujours très populaires et les résultats qu'ils procurent sont toujours satisfaisants.

Cependant, ces types de modèles sont loin d'être l'instrument mathématique idéal pour les processus hydrologiques et ce pour plusieurs raisons : (1) l'une de ces raisons principales est que ces modèles reposent largement sur l'hypothèse de Normalité alors qu'il est connu que les processus hydrologiques, particulièrement à un pas inférieur à l'annuel, s'éloignent de la distribution Normale et tendent vers des distributions de type puissance. L'autre hypothèse d'appui de ces modèles est la stationnarité alors que le comportement saisonnier exhibé par les processus à une échelle inférieure à l'année est complexe (type de distribution et structure de dépendance qui changent au sein de l'année) et ne peut pas être pris en charge par les techniques de dessaisonnalisation de Box-Jenkins (**Koutsoyannis, 2006**). (2) Les modèles Box-Jenkins qui se basent essentiellement sur la structure de dépendance des séries chronologiques ont tous une structure de dépendance de court terme, c'est-à-dire une fonction d'auto-corrélation qui diminue exponentiellement avec le pas de temps, alors qu'à l'opposé les processus hydrologiques montrent une dépendance de long terme, c'est-à-dire une décroissance en puissance connue par le phénomène de Hurst (**Koutsoyannis, 2005b**). (3) A l'exception des modèles simplifiés AR(1) ou ARMA(1,1) les modèles ne sont pas parcimonieux

car ils font intervenir plusieurs paramètres estimés à partir des données. En plus, les échantillons statistiques tels que les séries chronologiques ne permettent pas l'estimation fiable de plusieurs paramètres. Ceci est particulièrement le cas des processus hydrologiques avec une dépendance de long terme puisque ce comportement induit des incertitudes importantes par rapport aux processus avec une dépendance de court terme (**Koutsoyiannis, 2005a et b**).⁽⁴⁾ De nouveau, à l'exception des modèles AR(1) et ARMA(1,1), d'autres processus de la même famille ne possèdent pas de sens physique et sont en plus utilisés dans un concept de boîte noire (**Koutsoyiannis et al., 2008**).

La méthode de Box-Jenkins est toujours valable si l'on est à la recherche d'un modèle avec une prise en charge de la nature stochastique d'un processus naturel quelconque et les résultats qu'elle produit sont toujours satisfaisants en prévision (**Boukharouba et al., 2002 ; Boukharouba & Kettab, 2003b**) et en simulation (**Boukharouba & Kettab, 2003a, 2005b, 2006**). Cependant, les contraintes imposées par la méthode sur le processus naturel, telles que la stationnarité ou encore plus la Normalité font de ces modèles loin d'être l'instrument mathématique idéal pour ce type de problèmes. Ceci fait appel à d'autres outils mathématiques qui répondent mieux à la nature dynamique et au comportement saisonnier des processus hydrologiques. C'est l'une des raisons qui justifient le recours au Filtre de Kalman qui sera abordé dans la troisième partie de ce travail.

[Tapez un texte]

PARTIE III

Le filtre de Kalman pour la prédiction multi-site des apports liquides annuels au nord d'Algérie

PREAMBULE III

La prédiction peut être définie comme le comportement futur probable d'un processus et la confiance qui lui est assignée dépend de la nature du processus étudié et de la qualité de l'adéquation du modèle adopté (**Brown and Hwang, 1997**). Dans ce contexte, il est important de distinguer entre les deux modes avec lesquels un modèle de prédiction peut fonctionner. En mode de simulation, on peut utiliser les sorties du modèle comme nouvelles entrées dans le calcul de nouvelles sorties mais, il est exclu d'utiliser la sortie observée comme entrée du modèle. Par contre, en mode adaptatif, la sortie courante du modèle peut être basée sur les sorties observées antérieures du système. Dans la prédiction en temps réel, il est nécessaire de travailler dans le mode adaptatif, ce qui implique l'utilisation d'un modèle ayant une structure rétroactive qui fait que la sortie du modèle au temps présent est reliée aux sorties observées précédentes, c'est le cas du filtre de Kalman (FK).

Le FK est une version moderne et probabiliste du concept des moindres carrés qui se base sur la description du système en se basant sur la notion d'état, il peut être défini d'une manière simple comme étant un algorithme optimal et récursif de traitement de données. Son principe est simple, il consiste à combiner deux estimations indépendantes pour former une estimation pondérée ou prédiction. La première estimation peut être une prédiction à-priori ou une estimation basée sur une information à-priori, la seconde, une prédiction basée sur la nouvelle information (nouvelle mesure). L'objectif du FK est justement de combiner ces deux pièces d'information afin d'obtenir une estimation améliorée et son grand avantage est de procurer avec précision la covariance des erreurs de prédiction qui constitue en soi un critère d'optimalité. Un autre aspect d'optimalité du filtre de Kalman c'est qu'il incorpore toute l'information disponible sur le système et ses perturbations et peut être amorcé avec le minimum d'information disponible.

L'optimalité des résultats obtenus depuis sa mise en application dans les années soixante du siècle dernier jusqu'à ce jour, fait du filtre de Kalman l'outil favori des ingénieurs et chercheurs travaillant dans les domaines où la prédiction ou l'estimation linéaire sont requises.

En effet, depuis son introduction par **Kalman (1960)** et ses applications initiales par Kalman et Bucy une année après (**Kalman et Bucy, 1961**), le filtre de Kalman n'a pas cessé d'intriguer les chercheurs. Plusieurs recherches théoriques ont été stimulées par les problèmes rencontrés lors de l'application de ce filtre à des problèmes pratiques entre autres (**Jazwinski, 1969**), et depuis on a constaté l'explosion de son application à des problèmes divers dans plusieurs domaines d'intérêt. L'une de ses applications particulières est la navigation autonome et assistée dans la dynamique des navires et des avions (**Menaham, 1983**), (**Boone, 2000**) et récemment dans l'automatisme (**Welsh et al., 2001**; **Welch et Bishop, 2003**). Dans les sciences de la terre et de l'environnement nous pouvons citer à titre d'exemple (**Alpuim et Barbosa, 1999**; **Bertino et al 2002**; **Brammer et al. 1983**; **Guy, 1983**; **Schreider et al. 2001**; **Sorenson, 1970**;) parmi tant d'autres, en économie (**Gourrierroux et Montfort, 1990**; **Granger, 1986**), en procédés industriels (**Bialkovski, 1983**) et en systèmes énergétiques (**Wallace et Clarke, 1983**).

Ce puissant outil a également été appliqué à une variété de problèmes hydrologiques (**Hino, 1973**) ou à des problèmes en relation avec l'eau (**Todini et Boillot, 1975**) car ces derniers sont souvent sujets à des influences aléatoires d'origines inconnues (**Bras et Rodriguez, 1993**). Son avantage est la prise en compte de la variabilité temporelle du modèle, des paramètres et des variances des bruits, ainsi que la procuration de la covariance des erreurs de prédiction d'une manière précise. L'une des applications intéressantes du filtre de Kalman est l'assimilation des

données (**Bertino et al., 2002; Hartnack et Madsen, 2001; McLaughlin, 2002; Troch et al., 2003**). Un autre domaine particulier où le filtre de Kalman a également été l'objet d'une large recherche et application est l'estimation et la prédiction hydrologique (**Schreider et al., 2001; Husain, 1985; Georgakakos et Smith, 1990**).

En ce qui concerne l'application de la théorie du filtrage optimal en hydrologie, les contributions les plus significatives peuvent être retracées chronologiquement par les stations suivantes :

Jazwinski (1970) dans son livre, présenta un traitement unifié de la théorie de filtrage pour les systèmes linéaires et non linéaires, en insistant suffisamment sur son application afin de rendre possible son utilisation par le lecteur. Des résultats du filtrage non linéaire sont dérivés en premier, ensuite ils ont été spécialisés aux systèmes linéaires. Le traitement du filtrage linéaire inclut la stabilité du filtre ainsi que la sensibilité des erreurs du modèle. La performance de ces filtres non linéaires a été analysée d'une manière très critique.

D'un autre côté, **Gelb (1974)** présenta le premier livre sur l'estimation optimale par le filtre de Kalman qui plaça son importance majeure dans ses applications pratiques. Ceci dit, les concepts théoriques et mathématiques du filtre de Kalman sont introduits et suffisamment développés ce qui attribut au livre la particularité d'une source assez complète pour l'enseignement du lecteur sans qu'il y ait besoin de connaissance des principes de base du domaine.

Chiu (1978) présenta un livre qui regroupe des lectures et des articles présentés à Chapman conférence sur les applications théoriques et techniques du filtre de Kalman en hydrologie, en hydraulique et aux ressources en eau, de l'American Geophysical Union. L'objectif de la conférence était de donner au filtre de Kalman un plus grand angle d'ouverture pour les scientifiques de l'eau en donnant une introduction au fondamental du Filtre de Kalman théorie et technique, d'identifier et illustrer les cas utilisant le filtre de Kalman en hydrologie ; hydraulique et ressources en eau et de déterminer les directions d'études futures et finalement d'investir les domaines ou les applications du FK sont plus effectives. Cet ouvrage couvre le filtre de Kalman et d'autres estimateurs, avec des applications du FK à un spectre extensif de sujets en hydrologie, hydraulique et les ressources en eau, tel que le design de l'expérimental, ou les systèmes moniteurs, études de systèmes pluie –débit, la modélisation et prévision des débits de rivières, l'hydraulique des écoulements et d'autres processus de transport dans les cours d'eau et dans les rivières, les études de qualité d'eau, les problèmes d'eau souterraines, et d'autres domaines de ressources en eau et de géophysique.

La théorie mathématique du FK et ses implications ne sont pas bien compris même parmi la multitude d'ingénieurs et de mathématiciens appliqués. En fait, la plus part des praticiens sont juste informés des algorithmes de FK et non pas pourquoi ils fonctionnent si bien.

Chui et Chen (1987) ont apporté des réponses à quelques unes de ces questions en présentant littéralement une discussion détaillée de sa théorie mathématique et ses applications variées à des problèmes élémentaires en temps réel. Une dérivation très élémentaire des équations du filtrage a d'abord été présentée pour comprendre l'optimalité du FK. Le filtrage des systèmes non linéaires avec une application à l'identification adaptative des systèmes est également discuté dans le même livre. En plus, la limitation de la théorie de FK dans l'espace-état et les schémas de calcul efficaces tels que les algorithmes séquentiels et de racine carrée sont inclus pour des objectifs d'application en temps réel.

La planification effective, le dimensionnement et l'exploitation de n'importe quel système de ressource en eau, dépend de volume d'eau disponible qui peut être déterminé par l'étude des caractéristiques statistiques de plusieurs séries hydrologiques telles que les précipitations, les débits, les niveaux d'eau souterraines etc., c'est pour cela que le concept de modèles multi-variés devient très demandé. Les modèles multi-variés mettent en jeu de grandes matrices dont l'inversion crée des difficultés mathématiques et computationnelles non négligeables. **Şen (1980)** présente un modèle de désagrégation multi-site récursif qui conduit à une prédiction optimum d'événements saisonniers à partir de séquences annuelles observées à ces sites. Dans son modèle, une technique de partition a été employée avec la théorie du filtrage linéaire de Kalman et Bucy ce qui inclut des difficultés computationnelles en particulier dans l'application à des systèmes de grandes dimensions.

Şen(1991) a combiné le FK avec les séries Walsh orthogonales qui sont proposées comme un modèle effectif au calcul des périodicités dans les séries hydrologiques observées. Cette combinaison a conduit à une procédure de prédiction en temps réel des variables d'état qui sont de variables hydrologiques mensuelles. Une application générale est effectuée pour des volumes de débits et de pluie mensuelle. La méthode a été appliquée à des données mensuelles de débits de rivières observées en Turquie et aux USA, ainsi qu'aux précipitations mensuelles d'Arabie Saoudite comme représentative des zones extrêmement arides. La comparaison avec les résultats déjà disponibles, indique que la combinaison des fonctions Walsh avec le FK conduit à de meilleures prédictions adaptatives que les séries de Fourier. En termes de temps et mémoire de calcul, les fonctions Walsh est l'approche la plus économique parce que leurs propriétés de linéarité, orthogonalité et symétrie résultent dans les opérations de base d'addition et soustraction seulement.

Ahsan and Connor (1994) ont revu quelques unes des applications de la technique du FK dans la prévision des apports de rivières. Il a été argumenté que le minimum de l'erreur moyenne quadratique de la prévision telle qu'obtenue par le FK est identique avec ceux obtenues en utilisant la méthode conventionnelle de Box et Jenkins pour la prévision des séries temporelles d'apports de rivières, quand le modèle est un ARMA dont les données correspondantes sont libres de toute erreur de mesure. Cependant, dans l'hypothèse de présence de bruits de mesure dans la série chronologique des apports de rivières, l'utilisation du FK marque sa supériorité.

Le FK est l'approche la plus générale à l'estimation statistique et la prédiction. Il a été montré par **Harrison et Stevens (1975a,b)** que toutes les méthodes de prédiction sont des cas spéciaux du FK. Ce Filtre permet des changements de modèle, de paramètres et de variances. La difficulté avec le FK est que plusieurs questions techniques n'ont pas encore été répondues d'une manière satisfaisante. L'approche elle-même ayant évolué au sein des applications dans l'engineering. Par conséquent, plusieurs statisticiens et chercheurs opérationnels connaissent peu sur lui, ou le trouvent difficile à comprendre par ce qu'il est souvent formulé dans l'espace d'état. En plus, plusieurs difficultés pratiques existent encore telles que les estimations initiales des paramètres, variances, covariances ainsi que la matrice de transition.

Latif (1999), a développé une technique de FK et son application à des précipitations en Turquie. Son software reste applicable à n'importe qu'elle série de mesure de précipitations dans n'importe qu'elle région du monde.

Parmi les publications les plus récentes référant à l'application du filtre de Kalman nous pouvons citer à titre d'exemple **Şen et al. (2004)**, **Moradkhani et al.(2005)**, **Weerts et El Serafy (2006)** et **Ouachani et al. (2007)**.

Concernant la modélisation et la prédiction des processus d'écoulements souterrains ou de surface, **Bergman et Delleur (1985)** ont utilisé le filtre de Kalman pour la prédiction des

écoulements journaliers, **Ngan et Russel (1986)** ont étudié le filtre de Kalman combiné avec les modèles ARMAX pour les écoulements journaliers, **Jimenez et al. (1989)** ont essayé moyennant le filtre de Kalman d'identifier des modèles PARMA (*periodic autogressive-moving average*) pour des écoulements mensuels, **Awwad et Valdés (1992)** ont appliqué le filtre de Kalman afin d'identifier des modèles ARMAX multi-site pour les écoulement journaliers, et **Awwad et al (1994)** ont utilisé le filtre de Kalman et les modèles ARMAX dans la prévision multi-site des écoulements avec plusieurs intervalles à l'avance et à un pas de temps de 6 heures.

En Algérie et jusqu'à ce jour, la recherche bibliographique a montré que les modèles dynamiques basés sur la description d'état n'ont jamais eu une véritable chance d'être appliqués à la prédiction hydrologique et ce malgré le succès remarquable observé depuis l'introduction de ces derniers dans les années soixante. Particulièrement dans le domaine de l'eau, cette technique a toujours été évitée dans les centres de recherche et de prise de décision, exception faite de quelques contributions modestes dans ce domaine (**Boukharouba, 1996**), (**Boukharouba et Mokrane , 1997**), (**Boukharouba et Kettab, 2000**), et (**Boukharouba et Kettab, 2006**) mais qui sont restées inexploitées malgré la qualité des résultats obtenus. L'usage du filtre de Kalman pour des fins hydrologiques telles que la prédiction ou la simulation peut être d'un grand apport pour les décideurs du secteur de l'eau ainsi que pour les autres opérateurs et partenaires économiques, industriels et environnementaux. L'une des raisons majeures qui n'encouragent pas le recours au FK est justement cette notion d'état et la représentation du système dans son espace. Une autre difficulté réside dans la détermination des conditions initiales entre autres les matrices de covariance des erreurs du système et de la mesure.

Cette troisième partie de la thèse présente une contribution au comblement de cette lacune dans le domaine de la prédiction hydrologique en Algérie, son objectif est justement l'étude de l'applicabilité de la technique du FK discret à la modélisation et la prédiction multi-site des apports annuels des rivières au Nord d'Algérie, ainsi que l'amélioration des erreurs de prédiction. Le modèle recherché serait un moyen de prédiction en ligne qui non seulement procure des prédictions multi-sites optimales mais aussi des prédictions qui tiennent en compte la nature dynamique du système hydrologique concerné. Le premier chapitre sera dédié au filtre de Kalman proprement dit, où l'accent sera mis sur le modèle mathématique en continu et en discret, sur l'algorithme de calcul et enfin sur les conditions de son application. Le second chapitre en sera une application pratique. C'est un travail de modélisation stochastique pour la prédiction multi-site ayant pour objet les séries chronologiques des apports liquides annuels observés à dix stations hydrométriques au Nord d'Algérie de 1968 à 1992.

Chapitre III.1

LA METHODE DU FILTRE DE KALMAN

III.1.1 INTRODUCTION

D'une manière générale, la méthode de Kalman est une version moderne et probabiliste de la méthode des moindres carrés, qui concerne essentiellement les systèmes linéaires (**Kailath, 1977**) ce qui nous conduit à faire quelques rappels à leur sujet au paragraphe suivant. C'est une méthode d'estimation qui introduit la notion d'état, dans la représentation du système physique étudié et se base sur la notion de filtrage optimal (**Lewis, 1986**), d'où son appellation : Filtre de Kalman (FK). C'est un outil mathématique assez performant qui est entrain de jouer un rôle de plus en plus important dans toutes les filières des sciences de l'ingénieur. En effet, Le FK permettant de donner une estimation de l'état de système à partir d'une information a priori sur l'évolution de cet état et de mesures réelles, il est utilisé à des fins de prédiction à court terme et des fins d'identification paramétrique (**Mandelbrot & Wallis, 1968**), (**Wood & O'Connell, 1985**).

Son grand avantage, pour nous en tant qu'ingénieurs, c'est qu'on n'est pas obligé d'être un génie en mathématiques pour pouvoir comprendre et utiliser ce Filtre d'une manière efficace.

Cependant, avant de rentrer dans les détails propres du filtre de Kalman, les premières questions qui surgissent sont d'abord (1) qu'est-ce qu'un FK ? (2) Qu'est ce qui justifie le recours à un tel outil, dans le cadre de cette thèse ?

Pour répondre à la première question, le FK peut être défini d'une manière simple comme étant un algorithme optimal et récursif de traitement de données (**Maybeck, 1979**).

Dans cette définition, le terme « optimal » peut être approché de différentes façons dépendamment des critères choisis pour évaluer la performance du filtre. Un des aspects de cette optimalité est que le FK incorpore toute l'information qui peut lui être procurée. Il traite la totalité des mesures disponibles, indépendamment de leur précision, pour estimer les valeurs courantes des variables d'intérêt. Ceci est fait en utilisant (1) la connaissance des modèles dynamiques du système et des mesures (2) la description statistique des bruits du système, des erreurs de mesure ainsi que les incertitudes dans les modèles dynamiques (3) n'importe qu'elle information disponible sur les conditions initiales des variables en question. Cela étant, l'aspect principal de cette optimalité est que le FK combine toute cette information disponible pour produire une estimation des variables en question, de telle sorte que l'erreur d'estimation est minimale.

Le terme « récursif » dans la description précédente, veut dire que le FK n'exige pas de garder ou stocker en mémoire l'intégralité des données précédentes, pour être prises en considération dans les calculs, à chaque fois qu'une nouvelle mesure lui est fournie. Ceci est une propriété très importante qui rend l'implémentation du filtre très pratique.

Cela dit, le FK peut alors être considéré comme tout juste un algorithme de traitement de données dans un processeur central. Comme tel, il incorpore des échantillons de mesures discrètes.

Conceptuellement, ce que n'importe quel filtre essaye d'obtenir c'est une estimation « optimale » des variables en question à partir des données procurées par un environnement bruité. Il s'agit d'extraire une information utile à partir d'une information bruitée. Le terme « optimale » ici veut dire minimum d'erreur. Pour aboutir à cet objectif, il existe plusieurs approches entre autres, l'approche de Bayes (**Harrison & Stevens, 1975b**). Si on adopte le point de vue de Bayes, alors on souhaiterait un filtre qui propage la densité de probabilité conditionnelle des variables en question, conditionnée par la connaissance des données actuelles provenant des systèmes de mesures.

La probabilité de la valeur d'une quantité scalaire x au temps i , $x(i)$, conditionnée par la connaissance des mesures z jusqu'au temps i : ($z(1) = z_1, z(2) = z_2, \dots, z(i) = z_i$) est tracée en fonction des $x(i)$ possibles, pour donner le tracé de la fonction de densité de probabilité conditionnelle notée : $f_{x(i)/z(1),z(2),z(i)}(x/z_1, z_2, \dots, z_i)$. Cette fonction contient toute l'information disponible sur $x(i)$, elle indique pour la valeur donnée de toutes les mesures jusqu'à l'instant i , ce que serait la probabilité de $x(i)$. Elle est appelée « conditionnelle » car sa forme et sa localisation sur l'axe des x dépendent des valeurs des mesures prises. Sa forme révèle sur le taux de certitude que l'on a sur la connaissance de la valeur de x : Si le tracé de la fonction de densité de probabilité conditionnelle est proche d'un « pic » cela signifie que la plupart de la probabilité est concentrée autour d'une plage de valeurs de x , on est alors plus surs de la valeur de x . Si le tracé est plus disparate et plus plat, cela signifie que l'on est moins surs de la valeur de x .

Une fois une telle densité de probabilité conditionnelle propagée et tracée, l'estimation « optimale » peut alors être définie. Les choix possibles peuvent inclure :

- (1) La moyenne : la valeur de x correspondant au centre de masse de la probabilité conditionnelle.
- (2) Le mode : la valeur de x correspondant au maximum de probabilité conditionnelle.
- (3) La médiane : la valeur de x telle que 50% de probabilité soit à sa gauche et 50% de probabilité soit à sa droite.

Le FK performe cette propagation de la fonction de densité de probabilité conditionnelle, pour les problèmes dans lesquels le système peut être décrit par un modèle linéaire et dans lesquels les bruits du système et de la mesure sont blancs et gaussiens (**Brown & Hwang, 1997**). Dans ces conditions, la moyenne, le mode et la médiane coïncident, de telle sorte qu'il existe en fait, une estimation « meilleure » unique de la valeur de x . En d'autres termes, dans ces trois conditions, le FK est le meilleur filtre parmi toutes les formes concevables, dans le sens où l'estimation qu'il fournit est une estimation à variance minimum (**Welch & Bishop, 2003**). Ceci est la raison principale qui justifie notre recours, dans le cadre de cette thèse, à un tel outil mathématique parmi tant d'autres.

III.1.2 LES SYSTEMES LINEAIRES

III.1.2.1 Variables et équations d'état

Nous considérons un système dont l'état est défini par un ensemble (fini) de variables d'état $X_1(t), X_2(t), \dots, X_n(t)$ obéissant à un système d'équations de la forme suivante :

$$\begin{cases} \frac{dX_1}{dt} = A_{11}X_1 + A_{12}X_2 + \dots + A_{1n}X_n + B_1 \\ \frac{dX_2}{dt} = A_{21}X_1 + A_{22}X_2 + \dots + A_{2n}X_n + B_2 \\ \frac{dX_n}{dt} = A_{n1}X_1 + A_{n2}X_2 + \dots + A_{nn}X_n + B_n \end{cases} \quad (\text{III. 1.1})$$

Nous admettons que la variable indépendante t représente le temps, bien qu'il n'y ait là aucune limitation essentielle. Les coefficients A_{ij} peuvent être constants, ou encore être fonction régulière de t .

Les seconds membres du système d'équations (III. 1.1) sont des formes linéaires des variables d'état X_i : on dit que ce système est linéaire.

Les variables d'état X_i sont fréquemment assimilées aux composantes d'un vecteur à n dimensions, appelé vecteur d'état.

Les coefficients A_{ij} peuvent être constants ; le système est alors dit à coefficients constants. Ces coefficients peuvent aussi être fonction de la variable indépendante t ; le système est alors dit « à coefficients variables ».

Lorsque les B_i sont nuls, le système est dit homogène (il est en effet de degré 1 par rapport à l'ensemble des variables d'état).

III.1.2.2 Forme matricielle des équations d'état

Il est très commode d'écrire (III. 1.1) sous forme matricielle (**Sorenson,1985**). Dans ce but, on considère :

- Une matrice colonne définissant le vecteur d'état

$$X(t) = \begin{bmatrix} X_1(t) \\ X_2(t) \\ \vdots \\ X_n(t) \end{bmatrix} \quad (\text{III. 1.2})$$

- Une matrice carrée définissant les coefficients du système :

$$A(t) = |A_{ij}(t)| \quad (\text{III. 1.3})$$

- Une matrice colonne définissant l'excitation ou la commande du système :

$$B(t) = \begin{bmatrix} B_1(t) \\ B_2(t) \\ \vdots \\ B_n(t) \end{bmatrix} \quad (\text{III. 1.4})$$

Dans ces conditions, le système (III. 1.1) s'écrit simplement

$$\dot{X}(t) = \frac{dX(t)}{dt} = A(t)X(t) + B(t) \quad (\text{III. 1.5})$$

III.1.2.3 Résolution des équations d'état (matrice de transition)

Etant donné une équation homogène :

$$\dot{X}(t) = A(t)X(t) \quad (\text{III. 1.6})$$

assortie de conditions initiales

$$X(t_0) = [X(t)]_{t=0}$$

Il est en général possible de mettre sa solution sous la forme :

$$X(t) = \varphi(tt_0)X(t_0) \quad (\text{III. 1.7})$$

$\varphi(tt_0)$, appelée matrice de transition, peut être considérée comme un opérateur linéaire transformant $X(t_0)$ en $X(t)$. On peut vérifier que cette matrice carrée est solution de l'équation différentielle matricielle suivante :

$$\dot{\varphi}(t) = A(t)\varphi(t) \quad (\text{III. 1.8})$$

Par sa définition même, l'opérateur $\varphi(tt_0)$, applicable à tous les instants t_0 et t , doit obéir aux relations :

$$\varphi(tt) = I \quad (\text{matrice unité}) \quad (\text{III. 1.9})$$

$$\varphi(t_3t_2)\varphi(t_2t_1) = \varphi(t_3t_1) \quad (\text{III. 1.10})$$

$$\varphi(t_2t_1) = \varphi^{-1}(t_2t_1) \quad (\text{III. 1.11})$$

La matrice de transition φ , définie à propos de la résolution du système homogène (III. 1.6), permet de résoudre le système complet (III. 1.5) ; en effet on démontre que la solution de (III. 1.5) peut être mise sous la forme suivante :

$$X(t) = \varphi(tt_0)X(t_0) + \int_{t_0}^t \varphi(t\tau)B(\tau)d\tau \quad (\text{III. 1.12})$$

- $X(t_0)$ est la valeur initiale de $X(t)$;
- $\varphi(tt_0)$ est la matrice de transition (solution de (III. 1.8)) ;
- τ est une variable d'intégration.

Le premier terme de (III.1.12), égal à l'expression (III.1.7), décrit l'influence de l'état initial du système étudié. Le deuxième terme de (III.1.12) décrit l'influence de « l'excitation » $B(t)$ sur le système.

Jusqu'à présent, nous avons considéré des variables d'état variant d'une manière continue, ce qui est d'ailleurs, en pratique, le cas d'un grand nombre de problèmes. Dans ce qui suit, nous allons nous intéresser aux suites de nombres obtenus par échantillonnage de ces variables d'état, à des instants successifs $(t_0, t_1, t_2, \dots, t_k, \dots, t_N)$.

Dans le cas d'une équation homogène, le passage de $X(t_{k-1})$ à $X(t_k)$ se fait exactement, par application de (III. 1.7) ; on a simplement :

$$X(t_k) = \varphi_{k\ k-1}X(t_{k-1}) \quad (\text{III. 1.13})$$

Avec

$$\varphi_{k\ k-1} = \varphi(t_k t_{k-1}) \quad (\text{matrice de transition})$$

Dans le cas d'une équation complète (III. 1.5), nous avons :

$$X(t_k) = \varphi_{k\ k-1}X(t_{k-1}) + F_{k-1} \quad (\text{III. 1.14})$$

Avec

$$F_{k-1} = \int_{t_{k-1}}^{t_k} \varphi(t_k \tau)B(\tau)d\tau \quad (\text{III. 1.15})$$

L'équation (III.1.14) décrit l'évolution des grandeurs d'état échantillonnées du système étudié ; elle constitue la forme digitalisée de l'équation (III.1.12). Lorsque l'intervalle de temps $(t_{k-1}t_k)$ est suffisamment petit pour qu'on puisse assimiler $\varphi_{k\ k-1}$ à une matrice unité, et $B(\tau)$ à une constante (B_{k-1}), l'expression de F_{k-1} se réduit à

$$F_{k-1} \simeq B_{k-1}(t_k - t_{k-1}) \quad (\text{III. 1.16})$$

III.1.3 CONDITIONS D'APPLICATION DU FILTRE DE KALMAN

L'application du filtre de Kalman à l'étude d'un système physique suppose que l'on ait réussi à représenter ce dernier par un système d'équations linéaires (S) (**Gelb, 1974**). Par ailleurs, le filtre lui-même est également constitué d'un système d'équations linéaires (F). Les systèmes (S) et (F) peuvent présenter deux formes :

- continues, il s'agit alors d'équations différentielles ;
- digitales, il s'agit alors d'équations aux différences.

La théorie du filtre est moins difficile à établir sous la forme digitale, ceci nous incite évidemment à ne considérer que la forme digitale du filtre de Kalman ; cependant, il se trouve que, dans de nombreux cas, le modèle mathématique représentant le système linéaire à étudier est constitué d'un système différentiel, auquel devrait donc être normalement appliqué le filtre continu de Kalman ; la solution habituellement adoptée consiste à digitaliser le modèle mathématique du système, c'est-à-dire à le remplacer par un système d'équations aux différences équivalent, auquel peut commodément être appliqué le filtre de Kalman digital.

Ainsi, l'objet des paragraphes suivants va être tout d'abord de préciser la forme du modèle mathématique décrivant le système physique étudié, puis de donner le moyen d'en effectuer la digitalisation. Une autre alternative est de développer directement les équations du filtre de Kalman dans le cas discret, comme nous allons faire dans le paragraphe 4 du présent chapitre (**Grewal & Andrews, 2001**).

III.1.3.1 Modèle mathématique adopté

L'état du système est représenté par un vecteur d'état ($n \times 1$), $X(t)$, obéissant à une équation de la forme suivante :

$$\frac{dX(t)}{dt} = A(t)X(t) + F(t) + W(t) \quad (\text{III. 1.17})$$

matrice représentant la dynamique du système
vecteur d'état
Commande (excitation)
bruit du système (bruit blanc)

D'après (III. 1.17), la solution de cette équation est :

$$X(t) = \varphi(t, t_0) X(t_0) + \int_{t_0}^t \varphi(t, \tau) F(\tau) d\tau + \int_{t_0}^t \varphi(t, \tau) W(\tau) d\tau \quad (\text{III. 1.18})$$

Matrice de Transition
vecteur d'état Initial
terme commandé
terme aléatoire

Par ailleurs, l'application du filtre de Kalman suppose que le système (S) est soumis à m mesures portant sur certaines combinaisons linéaires des variables d'état (en particulier, les mesures

peuvent porter sur les variables d'état elles-mêmes). Ces mesures sont les composantes d'un vecteur $Z(t)$ obéissant à la relation suivante

$$Z(t) = H(t) X(t) + V(t) \quad (\text{III. 1.19})$$

Vecteur De mesure (m x 1) matrice de mesure (m x n) vecteur d'état (n x 1) bruit de mesure (bruit blanc) (m x 1)

Le nombre des mesures m peut très bien ne pas être égal au nombre de variables d'état n ; la matrice H est alors rectangulaire.

III.1.3.2 Digitalisation du modèle mathématique

- Digitalisation de l'équation de transition

Nous posons

$$X(t_k) = X_k \quad (\text{III. 1.20})$$

$$\varphi(t_k, t_{k-1}) = \varphi_{k, k-1} \quad (\text{III. 1.21})$$

Il en résulte que l'équation de transition (III.1.18) peut être mise sous la forme récurrente suivante :

$$X_k = \varphi_{k, k-1} X_{k-1} + F_{k-1} + W_{k-1} \quad (\text{III. 1.22})$$

Les termes $\varphi_{k, k-1} X_{k-1}$ et F_{k-1} sont obtenus par application des formules établies au paragraphe (III.1.2.3), le terme aléatoire W_{k-1} d'après (III. 1.15), admet pour expression

$$W_{k-1} = \int_{t_{k-1}}^{t_k} \varphi(t_k, \tau) W(\tau) d\tau \quad (\text{III. 1.23})$$

Ce qui implique

$$\begin{aligned}
 EW_{k-1} &= E \int_{t_{k-1}}^{t_k} \varphi(t_k, \tau) W(\tau) d\tau = \int_{t_{k-1}}^{t_k} \varphi(t_k, \tau) [EW(\tau) d\tau] d\tau \\
 &= \text{zéro}
 \end{aligned} \quad (\text{III. 1.24})$$

La valeur moyenne du terme W_{k-1} est donc nulle.

- **Digitalisation de l'équation de mesure (III. 1.19)**

Nous posons

$$H(t_k) = H_k \quad (\text{III. 1.25})$$

$$Z(t_k) = Z_k \quad (\text{III. 1.26})$$

Dans ces conditions, l'équation (III. 1.19) devient

$$Z_k = H_k X_k + V_k \quad (\text{III. 1.27})$$

Les V_k formant une suite de vecteurs aléatoires respectant

$$E V_k = 0 \quad (\text{valeurs moyennes nulles}) \quad (\text{III. 1.28})$$

$$E V_k V_j^T = R_k \delta_{kj} \quad (\text{les } V_k \text{ successifs sont nulles}) \quad (\text{III. 1.29})$$

Ces relations traduisent, sous forme digitale, les propriétés d'un bruit pseudo-blanc. La valeur R_k doit être déduite des densités spectrales S_{ii} et des densités spectrales mutuelles S_{ij} des composantes $v_i(t)$ de $V(t)$.

Nous avons :

$$R_k = \frac{1}{t_k - t_{k-1}} \begin{vmatrix} S_{11} & S_{12} & \dots & S_{1m} \\ S_{21} & S_{22} & \dots & S_{2m} \\ & & \dots & \\ S_{m1} & S_{m2} & \dots & S_{mm} \end{vmatrix} \quad (\text{III. 1.30})$$


Cependant, en pratique, il arrive fréquemment que les propriétés statiques des erreurs V_k soient simplement définies par une matrice de covariance, que l'on égale à R_k . Par exemple, si les erreurs $(v_{k1}, v_{k2}, \dots, v_{km})$, à l'instant t_k , sont indépendantes et valent « à ε » $\varepsilon_{k1}\varepsilon_{k2} \dots \varepsilon_{km}$ on a :

$$R_k = \begin{vmatrix} \varepsilon_{k1}^2 & 0 & \dots & 0 \\ 0 & \varepsilon_{k2}^2 & & 0 \\ & & \dots & \\ 0 & 0 & & \varepsilon_{kn}^2 \end{vmatrix} \quad (\text{III. 1.31})$$

III.1.3.3 Les entrées et sorties du filtre de Kalman

Le but du filtre de Kalman est d'obtenir une estimation du vecteur d'état $X_p = X(t_p)$, compte tenu des informations disponibles à un instant t_q (suivant les cas, on peut avoir $t_q < t_k$, $t_q = t_k$, $t_q > t_k$). Cette estimation est notée de la manière suivante :

$$[X(t_p)] = \hat{X}_{p/q}$$


 Estimé en t_p sachant
les informations disponibles jusqu'à t_q

Le filtre fournit de plus les matrices de covariance des erreurs d'estimation

$$E(X_p - \hat{X}_{p/q})(X_p - \hat{X}_{p/q})^T = P_{p/q} \quad (\text{III. 1.32})$$

Les grandeurs sont calculées pas à pas en fonction des données suivantes :

- Matrice de transition $\varphi_{k \ k-1}$
- Matrice de mesure H_k
- Estimation initiale du vecteur d'état, notée $\hat{X}_{0/0}$
- Covariance de l'erreur d'estimation initiale $(X_0 - \hat{X}_{0/0})$, notée $P_{0/0}$
- Propriétés statistiques des bruits :
 - Valeurs moyennes nulles

$$E V_k = 0 \quad (\text{III. 1.33})$$

$$E W_k = 0 \quad (\text{III. 1.34})$$

- Matrices de covariance

$$E W_j W_k^T = 0 \quad \text{si } (j \neq k) \quad (\text{III. 1.35})$$

$$E W_k W_k^T = Q_k \quad (\text{III. 1.36})$$

$$E V_j V_k^T = 0 \quad \text{si } (j \neq k) \quad (\text{III. 1.37})$$

$$E V_k V_k^T = R_k \quad (\text{III. 1.38})$$

- Indépendance des bruits du système et de mesure

$$\Rightarrow E W_k V_j^T = 0 \quad (\text{III. 1.39})$$

- Indépendance du bruit de mesure et de l'erreur initiale d'estimation

$$\Rightarrow E V_k (X_0 - \hat{X}_{0/0})^T = 0 \quad (\text{III. 1.40})$$

Note

Il est possible de définir un filtre plus général pour lequel (39) est remplacée par

$$E W_k V_j^T = S_k \delta_{ki} \quad (\text{III. 1.39 bis})$$

III.1.4 MECANISME D'ESTIMATION DU FILTRE DE KALMAN

Afin d'aboutir au mécanisme d'estimation du Filtre de Kalman proprement dit, dont la structure est composée de deux termes, l'un déterministe et l'autre aléatoire, essayons d'étudier l'estimation de ces deux types de processus, d'abord chacun tout seul, puis les deux combinés ensemble (**Jacobs,1993**).

III.1.4.1 Prédiction d'un signal analytique

Considérons tout d'abord un système idéal qui ne serait soumis à aucun bruit (d'où $W_k = 0$) et dont le vecteur d'état serait de mesures parfaites, également sans bruit (d'où $V_k = 0$). L'évolution de ce système serait intégralement fixée, dans ce cas idéal, par son équation de transition (III. 1.12). La prévision de cette évolution consisterait simplement à appliquer cette équation en utilisant un jeu de mesures comme conditions initiales. Les astronomes, par exemple, se trouvent dans des conditions suffisamment proches de ce cas idéal pour être capables de prévoir assez exactement les positions des planètes.

III.1.4.2 Prévision d'un signal aléatoire

Considérons maintenant le cas d'un signal Y aléatoire, de valeur moyenne $E(Y)$ nulle, de variance $E(Y^2)$ déterminée. Comment prédire un tel signal ? Il existe une réponse simple : la meilleure estimation (la moins mauvaise) est simplement :

$$\hat{Y} = \text{zéro} \quad (\text{III. 1.41})$$

Ce résultat, quelque peu décevant, est conforme au bon sens ; en effet, d'après ce que l'on sait de Y (soit : $E(Y) = 0$ et $E(Y^2)$ fixé), on n'a vraiment aucune raison d'en prédire une valeur positive ou une valeur négative. De plus, l'estimation (41) peut être justifiée en notant qu'elle conduit à une variance d'erreur minimale. On a en effet :

$$\begin{aligned} \text{Variance de } (Y - \hat{Y}) &= E(Y - \hat{Y})^2 = EY^2 - 2\hat{Y}EY + \hat{Y}^2 \\ &= EY^2 + \hat{Y}^2 \end{aligned} \quad (\text{III. 1.42})$$

La variance EY^2 étant fixée, cette quantité est effectivement minimale quand $\hat{Y} = 0$.

III.1.4.3 Estimation d'un signal mixte

Nous en arrivons au vecteur d'état d'un système linéaire régi par des équations de la forme (III. 1.22, 27), soumis à un bruit W_k et objet de mesures Z_k .

Supposons que nous disposions d'une estimation $\hat{X}_{k-1/k-1}$ de X , relative à l'instant t_{k-1} . Que pouvons-nous en déduire relativement à l'instant t_k ?

Nous disposons de deux informations :

- 1) Une estimation de $X(t_k)$, par application de l'équation de transition (III. 1.14) à l'estimation $\hat{X}_{k-1/k-1}$ nous obtenons ainsi :

$$\hat{X}_{k/k-1} = \varphi_{k/k-1} \hat{X}_{k-1/k-1} + F_{k-1} \quad (\text{III. 1.43})$$

Cette formule, peut être déduite de (III. 1.22) et de (III. 1.41), si l'on admet que l'opération estimation est linéaire.

- 2) Une estimation des résultats de mesure que l'on doit obtenir à l'instant t_k :

$$\hat{Z}_{k/k-1} = H_k \hat{X}_{k/k-1} \quad (\text{III. 1.44})$$

Cette formule traduit qu'une estimation correcte $\hat{X}_{k/k-1}$ doit conduire à une estimation correcte $\hat{Z}_{k/k-1}$. Le problème se pose maintenant d'améliorer l'estimation $\hat{Z}_{k/k-1}$ compte tenu du résultat de mesure Z_k . Par souci de simplicité, nous adoptons une correction linéaire de la forme suivante :

$$\hat{X}_{k/k} = \hat{X}_{k/k-1} + \overbrace{K_k [Z_k - H_k \hat{X}_{k/k-1}]}^{\text{terme correctif}} \quad (\text{III. 1.45})$$

La matrice K_k permettant le calcul du terme correctif en fonction de l'écart $(Z_k - H_k \hat{X}_{k/k-1})$ est appelée matrice de correction.

Remplaçons $\hat{X}_{k/k-1}$ par son expression (III. 1.43) dans la relation (III. 1.45) ; nous obtenons :

$$\hat{X}_{k/k} = (I - K_k H_k)(\varphi_{k/k-1} \hat{X}_{k-1/k-1} + F_{k-1}) + K_k Z_k \quad (\text{III. 1.46})$$

Cette relation est applicable aux estimations successives $\hat{X}_{1/1}$, $\hat{X}_{2/2}$... $\hat{X}_{k/k}$; l'hypothèse de linéarité traduite par (III. 1.45) impose donc les conséquences suivantes :

- Les estimations successives sont des formes linéaires des conditions initiales $\hat{X}_{0/0}$ et des résultats de mesure $Z_1 Z_2 \dots Z_k$.

- Pour le calcul de $\hat{X}_{k/k}$, l'estimation précédente $\hat{X}_{k/k-1}$ et la mesure nouvelle Z_k constituent un résumé suffisant de X_0 et de $Z_1 Z_2 \dots Z_k$. il en résulte une économie de mémoire, dans le cas d'un calcul automatique.

La forme (III. 1.45) de la relation de correction étant admise, il reste à expliciter les matrices K_k en fonction des propriétés du système (traduites par les matrices de transition $\phi_{k+1/k}$), des propriétés des bruits (traduites par les matrices de covariances Q_k et R_k), et enfin de la définition des mesures (traduite par la matrice de mesure H_k). Ce travail est fait dans le paragraphe suivant.

III.1.4.4 Développement des formules récurrentes du FK

Le principe du filtre de Kalman est simple ; il utilise une prédiction qui s'appuie sur le modèle déterministe et un recalage qui s'appuie sur l'innovation (différence entre la mesure et la sortie prédite). Dans ce qui suit, nous allons nous intéresser au filtre de Kalman dans le cas discret, le terme de l'excitation F dans l'équation (III. 1.27) sera omis car nous supposons que la suite du développement est faite pour des séries chronologiques. Dans ce cas, les formules utilisées sont de type récurrent.

Les équations fondamentales du FK peuvent alors être exprimées comme suit :

$$X_k = \phi_{k/k-1} X_{k-1} + W_{k-1} \quad (\text{III. 1.47})$$

$$Z_k = H_k X_k + V_k \quad (\text{III. 1.48})$$

L'équation (III. 1.47) exprime le modèle dynamique de l'état. Elle permet la transition de l'état X_{k-1} au temps t_{k-1} à l'état X_k au temps t_k par le biais de la matrice de transition $\phi_{k/k-1}$. L'équation (III. 1.48) donne la valeur de la variable Z_k au temps t_k , obtenue par un processus de mesure. Elle suppose qu'il existe une relation linéaire entre les mesures et le vecteur de l'état. Chacun des termes W_{k-1} et V_k sont supposés être des bruits blancs de moyennes nulles et de covariances connues respectivement Q et R . La problématique peut dans ce cas être posée de la manière suivante:

1. Est t'il possible de trouver une estimation de l'état X_k qui soit une combinaison linéaire de la mesure Z_k et l'état X_{k-1} ? En plus, cette estimation doit être optimale dans le sens où la variance des carrés des erreurs doit être minimale ;
2. Est t'il possible d'obtenir, dans le même sens, une prédiction optimale de l'état et de la mesure?

Sous certaines hypothèses, la réponse à ces deux questions peut être affirmative. Ceci peut être matérialisé par l'application des équations récursives du FK, qui exige un modèle avec les composantes suivantes :

1. vecteur de variables d'état X_k ,
2. matrice de transition $\phi_{k/k-1}$ des variables d'état du temps t_{k-1} au temps t_k ,
3. l'état initial X_0 et la matrice de covariance d'erreur associée P_0 ,
4. les statistiques mutuelles et conjointes de toutes les variables aléatoires impliquées.

La modélisation de la variable d'état est donnée par l'équation d'état, équation (III. 1.47), où W_{k-1} est le vecteur erreur du système qui est supposé être une séquence bruit blanc avec une covariance connue Q_k . Les dimensions des vecteurs d'état et d'erreur du système sont tous les deux (nxI) , alors que les dimensions de la matrice de transition d'état et celle de la covariance d'erreur sont $(n \times n)$. L'équation(III. 1.48) exprime le modèle d'observation qu'on appelle également modèle de mesure. Elle permet de transformer le vecteur d'état X_k en un vecteur de mesure Z_k par le biais de la matrice de mesure H_k , une telle transformation est supposée être linéaire. Le vecteur d'erreur de mesure V_k est supposé être une séquence bruit blanc, avec une covariance connue R_k et une crosscorrélacion nulle avec la séquence W_{k-1} . Les dimensions du vecteur de mesure est celle du vecteur d'erreur de mesure sont tous les deux (mxI) , et les dimensions de H_k ainsi que la covariance de l'erreur de mesure R_k sont $(m \times n)$; m étant généralement plus petit ou égal à n . Les matrices de covariance des erreurs du système et des erreurs de mesure sont données sous forme d'espérances mathématiques, tel qu'indiqué par les équations (III. 1.49 – 51) où l'exposant T indique l'opération de transposition de matrice.

$$E [W_k W_i^T] = \begin{cases} Q_k, & i = k \\ 0, & i \neq k \end{cases} \quad (\text{III. 1.49})$$

$$E [V_k V_i^T] = \begin{cases} R_k, & i = k \\ 0, & i \neq k \end{cases} \quad (\text{III. 1.50})$$

$$E [W_k V_i^T] = 0, \forall i \text{ et } k \quad (\text{III. 1.51})$$

Dans le cas où l'on dispose de mesures jusqu'au temps t_{k-1} (temps initial), l'estimation initiale du processus en question, basée sur la connaissance du processus avant t_{k-1} pourrait être disponible et on peut la noter $\hat{X}_{k/k-1}$. L'erreur d'estimation à priori et la matrice de covariance d'erreur associée à l'estimation initiale sont données par les expressions suivantes:

$$e_{k/k-1} = X_k - \hat{X}_{k/k-1} \quad (\text{III. 1.52})$$

$$P_{k/k-1} = E[e_{k/k-1} e_{k/k-1}^T]$$

$$P_{k/k-1} = E[(X_k - \hat{X}_{k/k-1})(X_k - \hat{X}_{k/k-1})^T] \quad (\text{III. 1.53})$$

L'actualisation de l'estimation $\hat{X}_{k/k}$ au temps t_k est considérée comme une estimation à postérieure, qui incorpore la nouvelle mesure Z_k afin d'améliorer l'estimation à priori $\hat{X}_{k/k-1}$. Cette étape est accomplie en rajoutant un terme de correction, qui comprend la différence $(Z_k - H_k \hat{X}_{k/k-1})$, pondérée par un facteur K_k ; on obtient :

$$\hat{X}_{k/k} = \hat{X}_{k/k-1} + K_k(Z_k - H_k \hat{X}_{k/k-1}) \quad (\text{III. 1.54})$$

Le facteur de pondération K_k est supposé conduire à une estimation optimum $\hat{X}_{k/k}$ dans le sens des moindres carrés. Dans le but de trouver ce K_k particulier, nous allons devoir combiner ensembles, l'erreur $e_{k/k-1}$ et la matrice de covariance d'erreur $P_{k/k}$ associée à l'actualisation de l'estimation, dans les deux équations suivantes :

$$e_{k/k} = X_k - \hat{X}_{k/k} \quad (\text{III. 1.55})$$

$$P_{k/k} = E[e_{k/k} e_{k/k}^T]$$

$$P_{k/k} = E[(X_k - \hat{X}_{k/k})(X_k - \hat{X}_{k/k})^T] \quad (\text{III. 1.56})$$

En substituant l'équation (III. 1.48) dans l'équation (III. 1.54), nous obtenons l'expression suivante:

$$\hat{X}_{k/k} = \hat{X}_{k/k-1} + K_k(H_k X_k + V_k - H_k \hat{X}_{k/k-1}) \quad (\text{III. 1.57})$$

Puis, en substituant l'équation (III. 1.57) dans l'équation (III. 1.56), il est possible d'obtenir une expression générale pour la matrice de covariance d'erreur actualisée,

$$P_{k/k} = (I - K_k H_k) P_{k/k-1} (I - K_k H_k)^T + K_k R_k K_k^T \quad (\text{III. 1.58})$$

Le problème actuellement est de trouver ce vecteur particulier K_k qui minimise les termes individuels de la diagonale principale de $P_{k/k}$, puisque ces termes représentent la covariance d'erreur d'estimation des éléments du vecteur d'état qui ont été estimés. Après quelques manipulations mathématiques, nous obtenons les relations suivantes:

$$K_k = P_{k/k-1} H_k^T (H_k P_{k/k-1} H_k^T + R_k)^{-1} \quad (\text{III. 1.59})$$

Le vecteur particulier K_k , qui minimise la moyenne des carrés des erreurs d'estimation est connu par le gain du filtre de Kalman. Ce gain peut être utilisé dans le calcul de l'estimation actualisée $\hat{X}_{k/k}$ et la matrice de covariance d'erreur actualisée qui lui est associée $P_{k/k}$. Encore quelques manipulations mathématiques et nous arrivons à l'expression suivante :

$$P_{k/k} = (I - K_k H_k) P_{k/k-1} \quad (\text{III. 1.60})$$

Dans le même esprit de la procédure que nous venons de présenter, il devient possible de projeter en avant l'estimation actualisée $\hat{X}_{k+1/k}$ en utilisant la matrice de transition comme suit :

$$\hat{X}_{k+1/k} = \phi_{k+1/k} \hat{X}_{k/k} \quad (\text{III. 1.61})$$

Où la contribution du terme W_k dans l'équation (III. 1.47) n'est pas considérée, puisque sa moyenne est nulle et puisqu'il n'est pas corrélé avec les W_k précédents.

L'erreur associée est donnée par:

$$e_{k+1/k} = X_{k+1} - \hat{X}_{k+1/k} \quad (\text{III. 1.62})$$

Par substitution de l'équation (III. 1.47) et (III. 1.61) dans l'équation (III. 1.62), on obtient :

$$e_{k+1/k} = \phi_{k+1/k} e_{k/k} + W_k \quad (\text{III. 1.63})$$

Où e_k et W_k ont des cross-corrélations nulles. La matrice de covariance d'erreur est alors exprimée comme suit :

$$P_{k+1/k} = \phi_{k+1/k} P_{k/k} \phi_{k+1/k}^T + Q_k \quad (\text{III. 1.64})$$

Finalement, les formules nécessaires à la prédiction au temps t_{k+1} sont données par les équations (III. 1.54), (III. 1.59 – 61) et (III. 1.64) ", qui forment les équations récursives du filtre de Kalman multidimensionnel.

III.1.5 CRITERES D'OPTIMALISATION

Il est temps de préciser les conditions théoriques permettant une démonstration de l'expression (III. 1.59) de la correction K_k . Il faut dire qu'en général, il existe deux cas de figure, selon les hypothèses dans lesquelles on se met (**Harvey, 1989**):

- **Premier cas :**

Hypothèses :

- 1) Les bruits obéissent aux relations (III. 1.33,34,36,38) ; cependant, leur loi de probabilité n'est pas explicitée.
- 2) L'opérateur estimation est linéaire par rapport aux mesures et à l'estimation initiale [ceci implique (III. 1.45)].
- 3) Les valeurs estimées \hat{X} doivent conduire à une minimalisation de la somme de variances $(X - \hat{X})^T (X - \hat{X})$

Conséquences:

- 1) K_k est fixée par (III. 1.59).
- 2) Chaque estimation \hat{X} est égale à l'espérance mathématique de X , compte tenu des résultats de mesure obtenus :

$$\hat{X}_{k/k} = E(X_k / Z_1 Z_2 \dots Z_k) \quad (\text{III. 1.65})$$

(loi de probabilité a posteriori, c'est-à-dire compte tenu des résultats de mesure).

- 3) La minimalisation de $E(X - \hat{X})^T (X - \hat{X})$ implique celle de

$$E(X - \hat{X})^T W_k (X - \hat{X})$$

W_k étant une matrice réelle, symétrique, définie, positive arbitraire.

- **Deuxième cas**

Hypothèses :

- 1) Les lois de probabilité des vecteurs d'état $(X_0 X_1 \dots X_k)$ et des mesures $(Z_1 Z_2 \dots Z_k)$ sont gaussiennes.
- 2) La valeur estimée \hat{X}_k maximalise la probabilité a posteriori

$$P(X_k / Z_1 Z_2 \dots Z_k)$$

Conséquences :

- 1) La formule d'estimation peut être mise sous la forme linéaire (III. 1.45).
- 2) La matrice de correction à utiliser est donnée par (III. 1.59).
- 3) Chaque estimation \hat{X}_k est égale à l'espérance mathématique de X_k , compte tenu des mesures faites (loi de probabilité a posteriori) :

$$\hat{X}_{k/k} = E(X_k/Z_1Z_2\dots Z_k)$$

- 4) Il en résulte une minimalisation de la quantité

$$E(X_k - \hat{X}_{k/k})^T W_k (X_k - \hat{X}_{k/k})$$

W_k étant une matrice réelle, symétrique, définie, positive, arbitraire.

Dans les deux cas précédents, le filtre de Kalman conduit à une minimalisation de la variance $E(X - \hat{X})^T (X - \hat{X})$; et d'une manière générale, l'estimation optimale de AX (A étant une matrice carrée ou rectangulaire) est égale à $A\hat{X}$, \hat{X} étant l'estimation optimale de X .

III.1.6 STABILITE DU FILTRE VIS-A-VIS DE LA DIVERGENCE

Quand nous avons parlé de l'optimalité du filtre, nous n'avions pas fait allusion à sa stabilité vis-à-vis de la divergence. En effet, un filtre optimal n'est pas automatiquement stable et sa stabilité doit être prouvée.

On dit qu'il ya divergence du filtre, quand l'estimation qu'il nous fournit est entachée d'erreurs qui deviennent de plus en plus importantes. Le filtre devient alors instable et donc insatisfaisant (**Schlee et al., 1967**). Le critère de convergence est donné par la matrice de covariance P_k (plus précisément par sa trace) qui doit être définie positive et minimale à la fin de la dernière itération (**Grewal & Andrews, 1997**). Elle doit donc converger vers le zéro au fur et à mesure que l'on avance dans le calcul sans pour autant lui être inférieure.

Les problèmes de stabilité du filtre sont reliés à la connaissance exacte des matrices P_0, F_k, H_k, Q et R et se traduisent par le fait que l'algorithme donne l'état erroné qui peut bien fonctionner à travers quelques données, mais qui n'est plus optimal pour longtemps. A ce moment, P_k ne mesure plus exactement la variance de l'erreur de l'estimation, et on assiste à une progression de l'erreur de l'estimation du vecteur X_k dans le temps.

Ce problème peut être dû au fait que le modèle a été utilisé d'une manière erronée pouvant être engendrée par une mauvaise estimation des paramètres, comme ça peut être le résultat d'une simplification du modèle lui même qui gère la dynamique du système. Il peut être accentué, quand les termes de bruit dans l'équation d'état du système, ou celle de mesure deviennent trop petits. Dans ce cas, la covariance de l'erreur devient tellement petite que le gain devienne lui aussi petit et par conséquent, les observations qui viennent au fur et à mesure, ont un effet minime sur l'estimation.

Il ya un certain nombre de méthodes qui peuvent intervenir pour éviter l'apparition de tels problèmes. La procédure la plus utilisée est celle qui consiste à modifier les équations du filtre de sorte que les mesures récentes puissent avoir une influence plus grande sur les estimations de ce dernier et plus les mesures sont lointaines dans le temps, moins devient leur influence. Ces modifications sont apportées par des ensembles de pondération $(1, e^{-\frac{1}{\tau}}, e^{-\frac{2}{\tau}}, \dots, e^{-\frac{s}{\tau}})$ pouvant être appliqués aux observations $(Z_k, Z_{k-1}, \dots, Z_{k-s})$ où τ est la mémoire du modèle.

III.1.7 CONCLUSION

Ce chapitre nous a permis, d'exposer la méthode du filtre de Kalman linéaire en général, en accordant plus d'importance au cas discret, sachant que les processus hydrologiques que nous allons utiliser sont échantillonnés à des temps discrets.

A cet effet, le principe de la méthode ainsi que le cadre théorique (assez compliqué) qui lui est inhérent, ont été présentés d'une manière simple, afin de faciliter sa compréhension. L'accent a été mis sur l'essentiel : le modèle mathématique adopté et les formules récurrentes du filtre, en passant en revue les conditions de son application et les critères d'optimisation qui attribuent au filtre de Kalman linéaire, le mérite d'être un outil d'estimation optimale par excellence.

Le filtre de Kalman est un algorithme d'équations récurrentes optimales qui sont faciles à utiliser en calcul automatique et dont les deux équations de base regroupent toute l'information disponible. La partie déterministe de la dynamique du système est exprimée par la matrice de transition et celle de la mesure est exprimée par la matrice de mesure, tandis que la partie aléatoire est exprimée par les propriétés statistiques (valeurs moyennes, matrices de covariance) des bruits d'état et de mesures.

L'estimation fournie par le filtre de Kalman est optimale dans le sens où elle conduit à des variances d'erreur minimales ou encore, dans le cas de bruits gaussiens, à une probabilité maximale des valeurs estimées.

Enfin, l'application du filtre de Kalman à un problème particulier exige le passage par les étapes suivantes :

- 1) choix de la variable d'état du système étudié et détermination du modèle mathématique conformément aux équations (III. 1.22) et (III. 1.27).
- 2) estimation des conditions initiales.
- 3) calcul conformément aux équations récurrentes du filtre.
- 4) vérification des hypothèses d'application par l'examen des résidus et critique des résultats.

Chapitre III.2

APPLICATION DU FILTRE DE KALMAN A LA PREDICTION MULTI-SITE DES APPORTS ANNUELS

III.2.1 INTRODUCTION

Par définition, la prédiction est le comportement futur probable d'un processus. La confiance qui lui est assignée dépend de la nature du processus étudié et de la qualité de l'adéquation du modèle choisi pour mieux le représenter (**Sorenson, 1985**)

En hydrologie, ces processus sont complexes, non linéaires dépendant du temps et d'un certain nombre de paramètres incertains. Tenir compte dans un modèle mathématique, de toutes ces considérations est une tâche ardue. De ce fait, les formulations basées sur des hypothèses simplificatrices de linéarité, invariance temporelle ..etc., peuvent provoquer des différences substantielles entre les valeurs observées et les valeurs prédites (**Schlee et al., 1967**).

Ces prédictions, si elles sont valables, forment la base pour les décisions concernant la gestion et le contrôle des ressources hydriques. Clairement, la fiabilité de telles prédictions gouvernera celle de ces décisions, et si en plus cette validité est étendue au temps réel, les résultats peuvent alors être interpolés ou extrapolés à n'importe quel temps puisque la nouvelle donnée est tout-à-fait valable.

Dans le contexte d'une utilisation en temps réel (**Menahem, 1983**) il est important de distinguer entre les deux modes avec lesquels un modèle peut fonctionner :

- Mode de simulation : La sortie du modèle est basée sur son entrée. On peut éventuellement utiliser les sorties du modèle comme nouvelles entrées dans le calcul de nouvelles sorties. Cependant, il est exclu d'utiliser la sortie observée comme entrée du modèle, celle-ci peut être utilisée à des fins de calibration ou autres.
- Mode adaptatif: la sortie courante du modèle peut être basée sur les sorties observées antérieures du système. Ceci peut se faire selon deux façons :
 - 1) La sortie courante du modèle est exprimée en fonction des sorties observées antérieures en plus de l'entrée.
 - 2) La sortie courante du modèle est calculée en utilisant une forme adaptative telle que celle utilisée par le filtre de Kalman.

Dans la prédiction en temps réel, il est nécessaire de travailler dans le mode adaptif, ceci implique l'utilisation d'un modèle ayant une structure avec rétroaction, c'est à dire que la sortie $S(t)$ au temps courant t est reliée aux sorties observées précédentes :

$$S(t) = f(S_{t-1}, S_{t-2}, \dots, S_{t-n}).$$

Une autre caractéristique, souvent évoquée dans la formulation des modèles de prédiction en temps réel, est l'hypothèse de linéarité. Cependant, l'insuffisance de cette approximation qui rend linéaire la réponse très complexe d'un système hydrologique, est compensée par l'introduction d'un terme de perturbations stochastiques (résidus) qu'on additionne au modèle.

Dans ce chapitre, notre attention sera focalisée sur les modèles linéaires à temps discrets pouvant être représentés par des équations de différence plutôt que des équations différentielles. Cette restriction nous paraît assez raisonnable, sachant que les prédictions en hydrologie se font à des temps discrets.

L'objectif étant de faire des prédictions multi-sites, notre choix est orienté vers le filtre de Kalman discret. L'application est faite sur les données d'apports liquides observés à l'échelle annuelle, sachant que la procédure en gros, reste valable pour des pas de temps inférieurs et valable aussi pour la prédiction multi-site de toute autre variable hydrologique ; pourvu que le choix des variables d'état soit fait d'une manière similaire.

III.2.2 DONNEES DE LA PREDICTION MULTI-SITE

La méthode de Kalman a été appliquée aux données annuelles des apports liquides observés par l'Agence Nationale des Barrages et Transferts ainsi que l'Agence Nationale des Ressources Hydrauliques.

Ces données regroupent une dizaine de stations hydrométriques, dont cinq sont situées sur des barrages en exploitation au Nord d'Algérie (voir la section introduction générale). Les observations forment des séries chronologiques ayant une période commune de 25 ans (1968-1992).

III.2.3 HYPOTHESES

- Le système est linéaire.
- Les résidus sont des bruits blancs gaussiens.

III.2.4 FORMULATION ESPACE-ETAT DU MODELE ADOPTE

L'une des étapes les plus importantes dans l'application de la technique du Filtre de Kalman est la formulation des équations d'état et de mesure conformément à la structure d'un modèle espace état tel qu'indiqué dans le chapitre III.1.

Dans notre cas, nous allons travailler avec des séries chronologiques ce qui implique que le terme correspondant à l'entrée (F) du système, dans l'équation d'état, est nul. La variable état du système est un vecteur donnant l'apport liquide annuel observé simultanément aux dix stations hydrométriques considérées. Ainsi, pour $n=10$ les vecteurs d'état du système et de la mesure ont

pour dimensions (10x1), alors que les matrices de covariances, de transition d'état et celle de mesure sont de dimension (10x10). Par suite, l'équation d'état et celle de mesure sont respectivement formulées de la manière suivante :

- Equation d'état :

$$\begin{bmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ \cdot \\ X_{10} \end{bmatrix}_k = \begin{bmatrix} \phi_{1,1} & \phi_{1,2} & \cdot & \cdot & \cdot & \phi_{1,10} \\ \phi_{2,1} & \phi_{2,2} & \cdot & \cdot & \cdot & \phi_{2,10} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \phi_{10,1} & \phi_{2,10} & \cdot & \cdot & \cdot & \phi_{10,10} \end{bmatrix}_{k/k-1} \begin{bmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ \cdot \\ X_{10} \end{bmatrix}_{k-1} + \begin{bmatrix} W_1 \\ W_2 \\ \cdot \\ \cdot \\ \cdot \\ W_{10} \end{bmatrix}_{k-1} \quad (\text{III. 2.1})$$

Avec

– $\begin{bmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ \cdot \\ X_{10} \end{bmatrix}_k$: le vecteur état au temps k

– $\begin{bmatrix} \phi_{1,1} & \phi_{1,2} & \cdot & \cdot & \cdot & \phi_{1,10} \\ \phi_{2,1} & \phi_{2,2} & \cdot & \cdot & \cdot & \phi_{2,10} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \phi_{10,1} & \phi_{2,10} & \cdot & \cdot & \cdot & \phi_{10,10} \end{bmatrix}_{k/k-1}$: matrice de transition d'état

– $\begin{bmatrix} W_1 \\ W_2 \\ \cdot \\ \cdot \\ \cdot \\ W_{10} \end{bmatrix}_{k-1}$: vecteur des bruits du système

- Equation de mesure :

$$\begin{bmatrix} Z_1 \\ Z_2 \\ \cdot \\ \cdot \\ \cdot \\ Z_{10} \end{bmatrix}_k = \begin{bmatrix} H_{1,1} & H_{1,2} & \cdot & \cdot & \cdot & H_{1,10} \\ H_{2,1} & H_{2,2} & \cdot & \cdot & \cdot & H_{2,10} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ H_{10,1} & H_{2,10} & \cdot & \cdot & \cdot & H_{10,10} \end{bmatrix}_k \begin{bmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ \cdot \\ X_{10} \end{bmatrix}_k + \begin{bmatrix} V_1 \\ V_2 \\ \cdot \\ \cdot \\ \cdot \\ V_{10} \end{bmatrix}_k \quad (\text{III. 2.2})$$

Avec :

$$- \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ \vdots \\ Z_{10} \end{bmatrix}_k : \text{vecteur de mesure au temps } k$$

$$- \begin{bmatrix} H_{1,1} & H_{1,2} & \cdot & \cdot & \cdot & H_{1,10} \\ H_{2,1} & H_{2,2} & \cdot & \cdot & \cdot & H_{2,10} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ H_{10,1} & H_{10,2} & \cdot & \cdot & \cdot & H_{10,10} \end{bmatrix}_k : \text{matrice de mesure au temps } k$$

$$- \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ \vdots \\ X_{10} \end{bmatrix}_k : \text{vecteur d'état au temps } k$$

$$- \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ \vdots \\ V_{10} \end{bmatrix}_k : \text{vecteur des bruits de mesure au temps } k$$

III.2.5 CONDITIONS INITIALES

Une fois que la formulation du modèle est accomplie, le filtre de Kalman et l'exécution de son algorithme requiert la spécification des grandeurs suivantes :

- Vecteur d'état initial,
- Matrice de covariance d'erreur associée au vecteur d'état initial,
- Matrice de covariance des bruits du système,
- Matrice de covariance des bruits de mesures,
- Matrice de transition d'état,
- Matrice de mesure.

Ceci va être effectué selon les étapes suivantes :

III.2.5.1 Vecteur d'état initial et matrice de covariance d'erreur associée

En présence d'une information à priori, le vecteur d'état initial $\hat{\mathbf{X}}_{k/k-1}$, pour $k = 1$ est constitué par la moyenne des apports liquides annuels observés aux 10 différentes stations hydrométriques en question ($\bar{X}_1, \bar{X}_2, \dots, \bar{X}_{10}$) comme suit :

$$\hat{\mathbf{X}}_{1/0} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_{10} \end{bmatrix} \quad (\text{III. 2.3})$$

Cependant, pour spécifier la matrice de covariance d'erreur qui lui est associée $\mathbf{P}_{k/k-1}$ et qui n'est pas exactement connue, nous pouvons commencer avec des éléments ayant des valeurs assez élevées dans la diagonale principale. De cette manière, l'algorithme bénéficiera d'une certaine flexibilité afin de s'ajuster aux valeurs sensibles dans un temps relativement court.

Dans le présent travail, nous avons choisi une telle matrice (10x10) comme suit :

$$\mathbf{P}_{1/0} = \begin{bmatrix} 1000 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1000 \end{bmatrix} \quad (\text{III. 2.4})$$

Ce choix va mener à une augmentation de la matrice de covariance $\mathbf{P}_{k/k-1}$ et celle du gain du filtre \mathbf{K}_k , permettant ainsi au filtre adaptatif de pondérer plus lourdement la nouvelle information (la nouvelle mesure \mathbf{Z}_k dans l'équation 54 du chapitre III.1). Ainsi, la trace de la matrice de covariance initiale est de 10 000. Une telle valeur est supposée décroître d'une manière continue pour atteindre asymptotiquement une valeur positive et stable proche de zéro. Cette réduction est accomplie au cours de l'exécution de l'algorithme du filtre de Kalman, qui s'adapte automatiquement dès qu'il reçoit une nouvelle information objective. Dans le cas où cette valeur asymptotique de la trace de la matrice de covariance ne change pas significativement durant la procédure de calculs, cela signifie que le filtre a convergé. Nous pouvons dire que cette caractéristique attribuée à la trace de la matrice de covariance le mérite d'être une mesure fiable de la performance du filtre de Kalman.

III.2.5.2 Matrice de covariance des bruits du système et celle des bruits des mesures

De plus, du moment que l'on espère que les mesures soient moins bruitées que la dynamique du système, la matrice de covariance du bruit du système \mathbf{Q} et celle du bruit de mesure \mathbf{R} , toutes les deux (10x10) ont été choisies comme telles :

$$Q = \begin{bmatrix} 100 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 100 \end{bmatrix} \quad (\text{III. 2.5})$$

$$R = \begin{bmatrix} 50 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 50 \end{bmatrix} \quad (\text{III. 2.6})$$

III.2.5.3 Matrice de transition d'état et matrice de mesure

L'une des difficultés majeures dans l'application du filtre de Kalman est l'estimation de la matrice de transition d'état $\phi_{k/k-1}$. Il faut noter, cependant que la valeur de telle matrice n'affecte pas de manière considérable les résultats du filtre de Kalman, tel que l'ont montré les études de (**Harrison and Stevens, 1975a, b**). Par conséquent, pour l'estimation de la matrice de transition d'état, dans le présent travail on considère les inter-corrélations entre les observations des dix stations hydrométriques.

Quant à la matrice de mesure H_k , du moment que toutes les stations hydrométriques fournissent des observations, nous avons opté pour la matrice unité (10x10)

$$H_k = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix} \quad (\text{III. 2.7})$$

III.2.6 LES EQUATIONS RECURRENTES DU FILTRE DE KALMAN

Les cinq relations matricielles récursives (III. 1.54,59,60,61 et 64) indiquées au chapitre III.1 sont les équations de base dans le calcul du filtre de Kalman.

$$\begin{aligned} \hat{X}_{k/k} &= \hat{X}_{k/k-1} + K_k (Z_k - H_k \hat{X}_{k/k-1}) && \text{(mise à jour du vecteur d'état)} \\ K_k &= P_{k/k-1} H_k^T (H_k P_{k/k-1} H_k^T + R_k)^{-1} && \text{(estimation du gain)} \\ P_{k/k} &= (I - K_k H_k) P_{k/k-1} && \text{(correction de la matrice de covariance de l'erreur d'estimation)} \\ \hat{X}_{k+1/k} &= \phi_{k+1/k} \hat{X}_{k/k} && \text{(prédiction du vecteur d'état)} \\ P_{k+1/k} &= \phi_{k+1/k} P_{k/k} \phi_{k+1/k}^T + Q_k && \text{(prédiction de la covariance de l'erreur de prédiction)} \end{aligned}$$

Elles sont utilisées de la manière qui suit et qui constitue l'algorithme du filtre de Kalman.

III.2.6.1 Matrice gain du filtre

Le calcul de la matrice gain du filtre K_k est la première étape de calcul dans le présent travail. Compte tenu de toute l'information précédente et des hypothèses du filtre, la matrice gain (10x10) peut être calculée pour une prédiction à un pas en utilisant l'équation (III. 1.59), pour $k=1$ comme suit :

$$\mathbf{K}_1 = \begin{bmatrix} 1000 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1000 \end{bmatrix} \left[\begin{bmatrix} 1000 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1000 \end{bmatrix} + \begin{bmatrix} 50 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 50 \end{bmatrix} \right]^{-1} \quad (\text{III. 2.8})$$

III.2.6.2 Utilisation de la mesure pour actualiser l'estimation

C'est la seconde étape des calculs pour le même pas de temps $k=1$. L'estimation est corrigée moyennant la nouvelle mesure \mathbf{Z}_k afin d'améliorer l'estimation initiale $\hat{\mathbf{X}}_{1/0}$. Ceci est fait en faisant intervenir un terme de correction comprenant la différence $(\mathbf{Z}_k - \mathbf{H}_k \hat{\mathbf{X}}_{k/k-1})$ pondérée par le facteur \mathbf{K}_1 en concordance avec l'équation (III. 1.54) tel qu'on obtient :

$$\hat{\mathbf{X}}_{1/1} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_{10} \end{bmatrix}_{1/0} + \begin{bmatrix} k_{1,1} & k_{1,2} & \cdot & \cdot & k_{1,10} \\ k_{2,1} & k_{2,2} & \cdot & \cdot & k_{2,10} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ k_{10,1} & k_{10,2} & \cdot & \cdot & k_{10,10} \end{bmatrix}_{-1} \left(\begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_{10} \end{bmatrix}_1 - \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_{10} \end{bmatrix}_{1/0} \right) \quad (\text{III. 2.9})$$

III.2.6.3 Matrice de covariance d'erreur associée à l'estimation (actualisée) corrigée par la mesure

Par substitution de la matrice de covariance d'erreur $\mathbf{P}_{1/0}$ associée au vecteur d'état initial $\hat{\mathbf{X}}_{1/0}$, de la matrice gain \mathbf{K}_1 ainsi que la matrice de mesure \mathbf{H}_k dans l'équation (III. 1.60), nous pouvons obtenir la matrice de covariance d'erreur associée à l'estimation corrigée par la mesure, comme suit :

$$\mathbf{P}_{1/1} = \begin{bmatrix} 1000 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1000 \end{bmatrix}_{1/0} - \begin{bmatrix} k_{1,1} & k_{1,2} & \cdot & \cdot & k_{1,10} \\ k_{2,1} & k_{2,2} & \cdot & \cdot & k_{2,10} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ k_{10,1} & k_{10,2} & \cdot & \cdot & k_{10,10} \end{bmatrix}_{-1} \begin{bmatrix} 1000 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1000 \end{bmatrix}_{1/0} \quad (\text{III. 2.10})$$

III.2.6.4 Prédiction un pas en avant de l'estimation (actualisée) corrigée par la mesure

L'estimation corrigée $\hat{X}_{1/1}$ est projetée un pas en avant par l'intermédiaire de l'équation (III. 1.61), où le terme W_{k-1} est ignoré simplement parce que sa moyenne est nulle et parce qu'il est non corrélé avec les W précédents. Ceci nous conduit la prédiction un pas en avant de l'estimation corrigée $\hat{X}_{2/1}$ tel que :

$$\begin{bmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ \cdot \\ X_{10} \end{bmatrix}_{2/1} = \begin{bmatrix} \phi_{1,1} & \phi_{1,2} & \cdot & \cdot & \cdot & \phi_{1,10} \\ \phi_{2,1} & \phi_{2,2} & \cdot & \cdot & \cdot & \phi_{2,10} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \phi_{10,1} & \phi_{10,2} & \cdot & \cdot & \cdot & \phi_{10,10} \end{bmatrix}_{2/1} \begin{bmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ \cdot \\ X_{10} \end{bmatrix}_{1/1} \quad (\text{III. 2.11})$$

III.2.6.5 Matrice de covariance de l'erreur de prédiction un pas en avant

De même, en substituant la matrice de transition d'état, la matrice de covariance d'erreur associée au vecteur d'estimation corrigée $P_{1/1}$ et la matrice de covariance du bruit de système Q dans l'équation (III. 1.64) nous pouvons calculer la matrice de covariance d'erreur de prédiction un pas en avant $P_{2/1}$ associée à $\hat{X}_{2/1}$ comme suit :

$$\begin{aligned} P_{2/1} = & \begin{bmatrix} \phi_{1,1} & \phi_{1,2} & \cdot & \cdot & \phi_{1,10} \\ \phi_{2,1} & \phi_{2,2} & \cdot & \cdot & \phi_{2,10} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \phi_{10,1} & \phi_{10,2} & \cdot & \cdot & \phi_{10,10} \end{bmatrix}_{2/1} \begin{bmatrix} p_{1,1} & p_{1,2} & \cdot & \cdot & p_{1,10} \\ p_{2,1} & p_{2,2} & \cdot & \cdot & p_{2,10} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ p_{10,1} & p_{10,2} & \cdot & \cdot & p_{10,10} \end{bmatrix}_{1/1} \begin{bmatrix} \phi_{1,1} & \phi_{1,2} & \cdot & \cdot & \phi_{1,10} \\ \phi_{2,1} & \phi_{2,2} & \cdot & \cdot & \phi_{2,10} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \phi_{10,1} & \phi_{10,2} & \cdot & \cdot & \phi_{10,10} \end{bmatrix}_{2/1}^T \\ & + \begin{bmatrix} 100 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 100 \end{bmatrix} \quad (\text{III. 2.12}) \end{aligned}$$

Pour la prochaine itération dans le calcul, $\hat{X}_{2/1}$ et $P_{2/1}$ sont considérés comme conditions initiales et ainsi de suite pour le restant des itérations jusqu'à la fin des calculs.

Comme tel, avec toute l'information et les conditions précédentes, les équations multi-sites récursives du filtre de Kalman (III. 2.8) – (III. 2.12) ainsi que leurs boucles ont été calculées durant toute la période d'observation. Ce calcul a été effectué sur deux temps le premier sert à l'estimation des paramètres du modèle, le deuxième sert à la validation de ce dernier.

III.2.7 RESULTATS ET INTERPRETATIONS

La période commune d'observation entre les 10 stations hydrométriques étant de 25 ans, les observations des premières 15 années au pas de temps annuel ont été utilisées à l'estimation des paramètres du modèle. Les dix dernières années ont servi à la validation.

L'exécution consécutive des équations de Kalman sur la période d'observation représente l'application de l'approche du (FK) à la modélisation et à la prédiction des apports liquides annuels pour les stations hydrométriques mentionnées plus haut.

A la fin des calculs, il s'avère que certaines des valeurs prédites sont sous-estimées et d'autres sont surestimées. La différence entre les mesures et leurs prédictions conduit à une série de résidus (innovations) pour chaque station.

III.2.7.1 Prédictions multi-sites

Les résultats de ces prédictions sont représentés dans le Tableau III.2.1. C'est des matrices (25x10) correspondant aux prédictions durant les 25 ans (lignes) pour les 10 stations hydrométriques (colonnes). Ces résultats peuvent donc être appréciés dans les deux dimensions : temporelle si on opte pour une lecture dans le sens des lignes de chaque station, et spatiale si on opte pour une lecture en colonne de chaque année de prédiction.

- **Prédictions temporelles :** En ce qui concerne la prédiction multi-site dans la dimension temporelle, la Fig.III.2.1 en est un exemple. Elle montre, pour la station hydrométrique de AINBERDA, les apports annuels observés avec leurs prédictions durant la période (1968-1992) ainsi que l'erreur relative de la prédiction exprimée en pourcentage (%) (voir Tableau III.2.2). De la même manière, les Fig.III.2.2-10 montrent le restant des prédictions avec les erreurs relatives correspondantes en pourcentage.

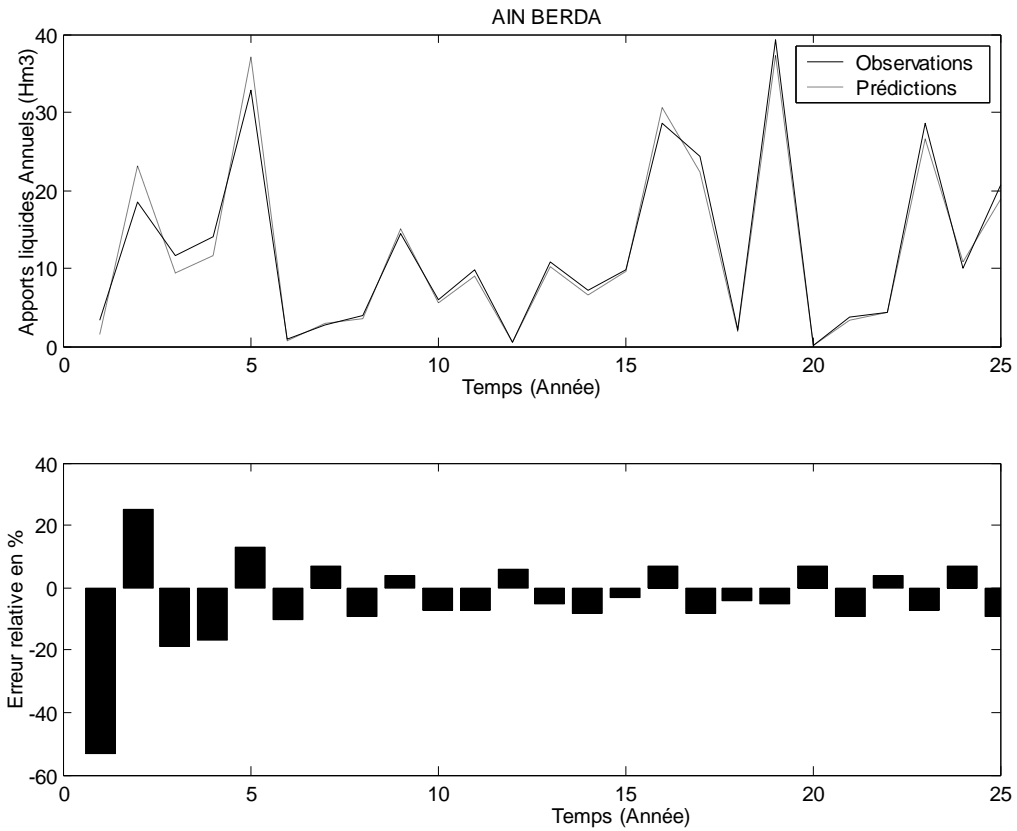


Fig. III.2.1 Prédictions obtenues à la station hydrométrique AINBERDA

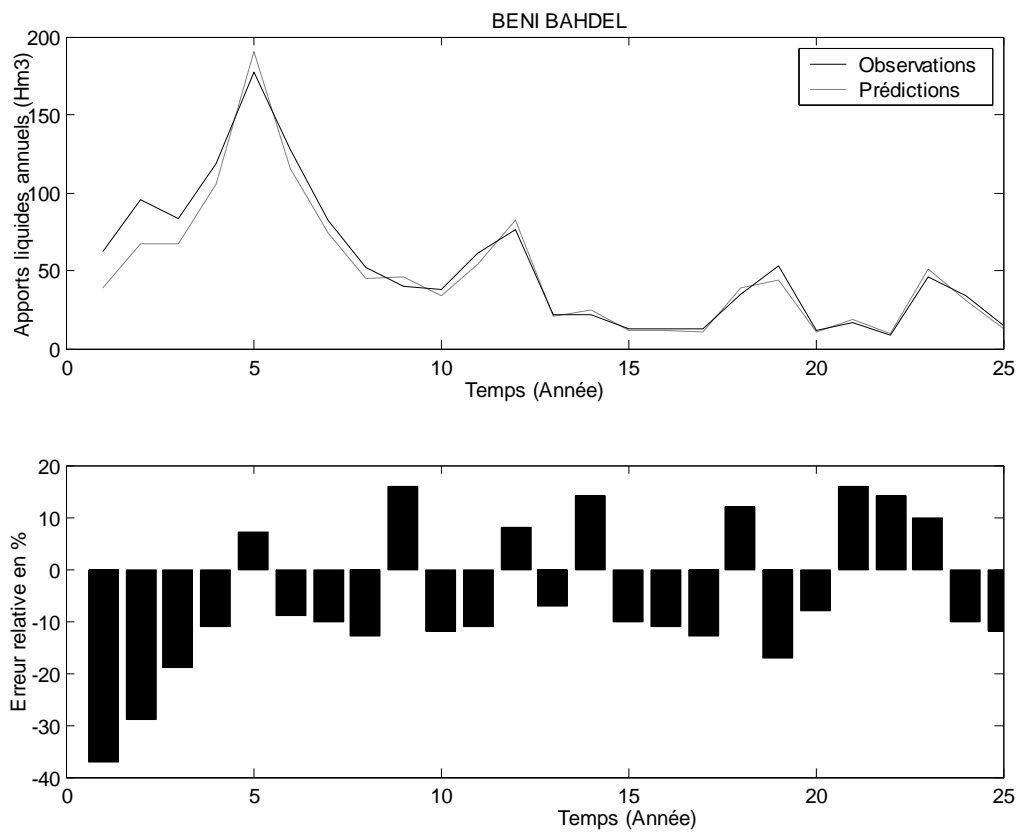


Fig. III.2.2 Prédictions obtenues à la station hydrométrique BENI BAHDEL

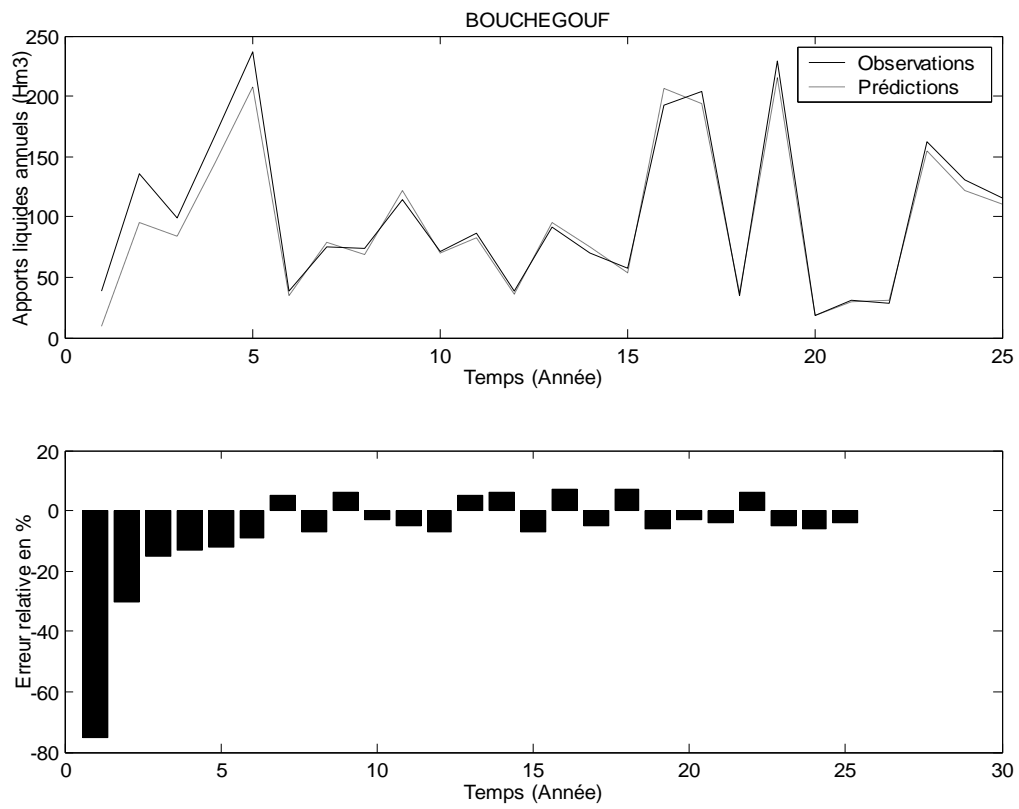


Fig. III.2.3 Prédictions obtenues à la station hydrométrique BOUCHEGOUF

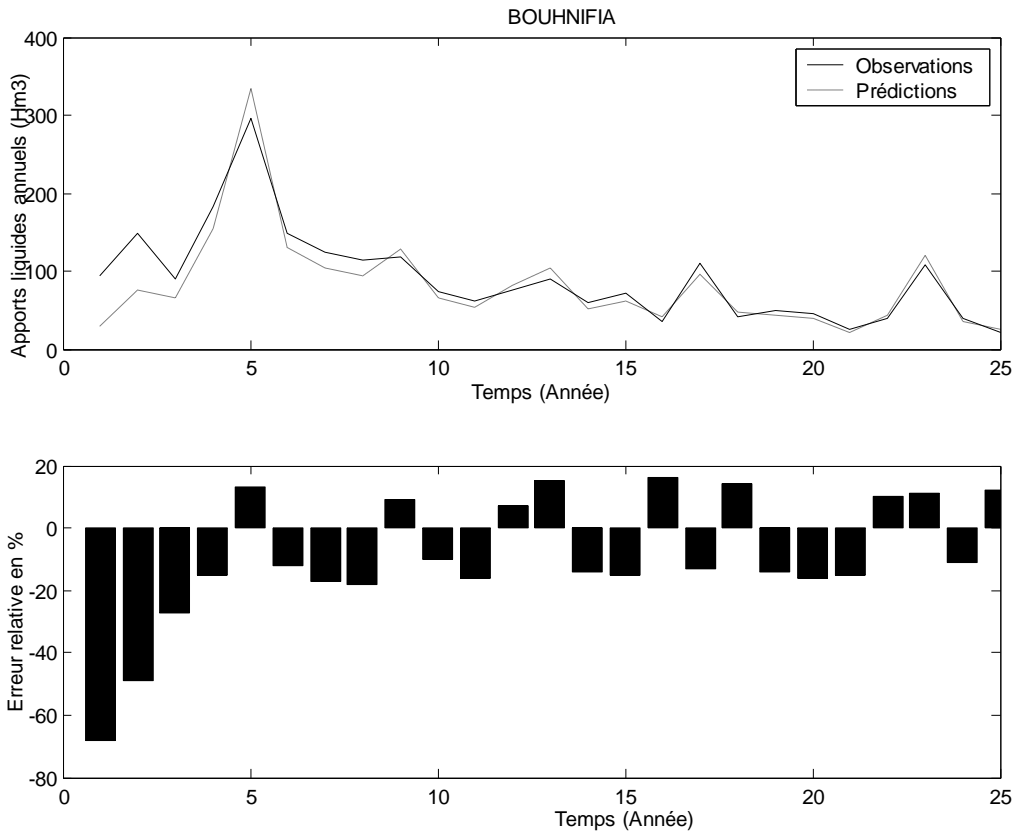


Fig. III.2.4 Prédiction obtenues à la station hydrométrique BOUHNIFIA

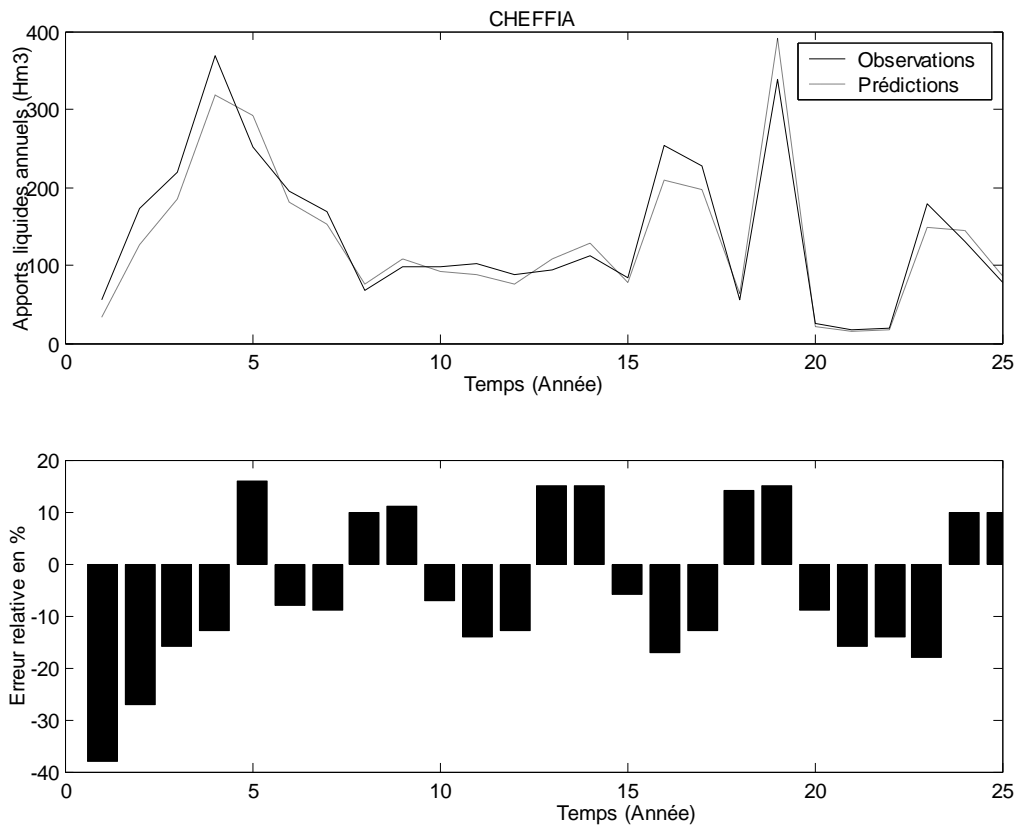


Fig. III.2.5 Prédiction obtenues à la station hydrométrique CHEFFIA

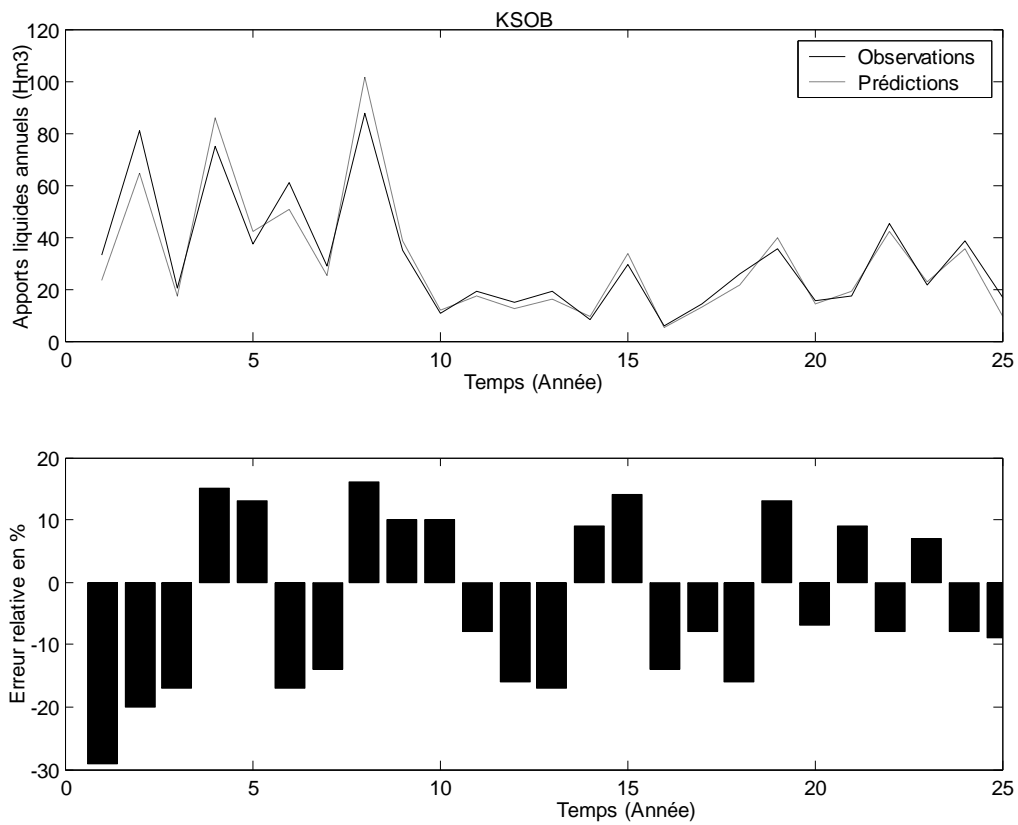


Fig. III.2.6 Prédiction obtenue à la station hydrométrique KSOB

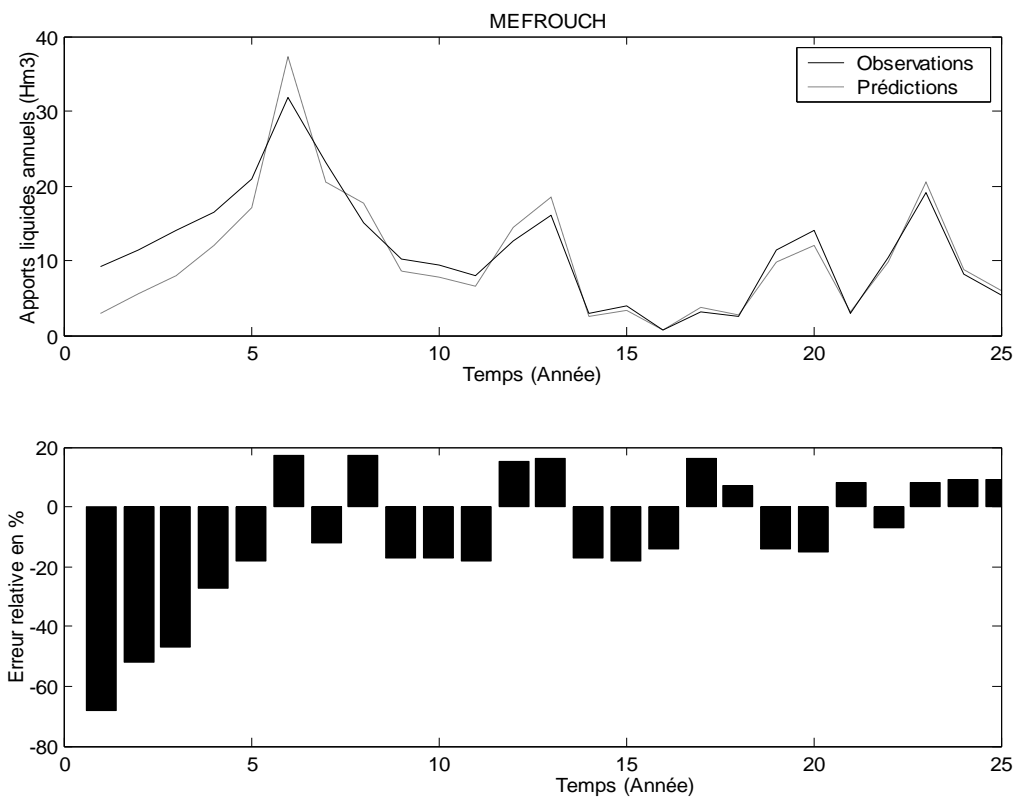


Fig. III.2.7 Prédiction obtenue à la station hydrométrique MEFROUCH

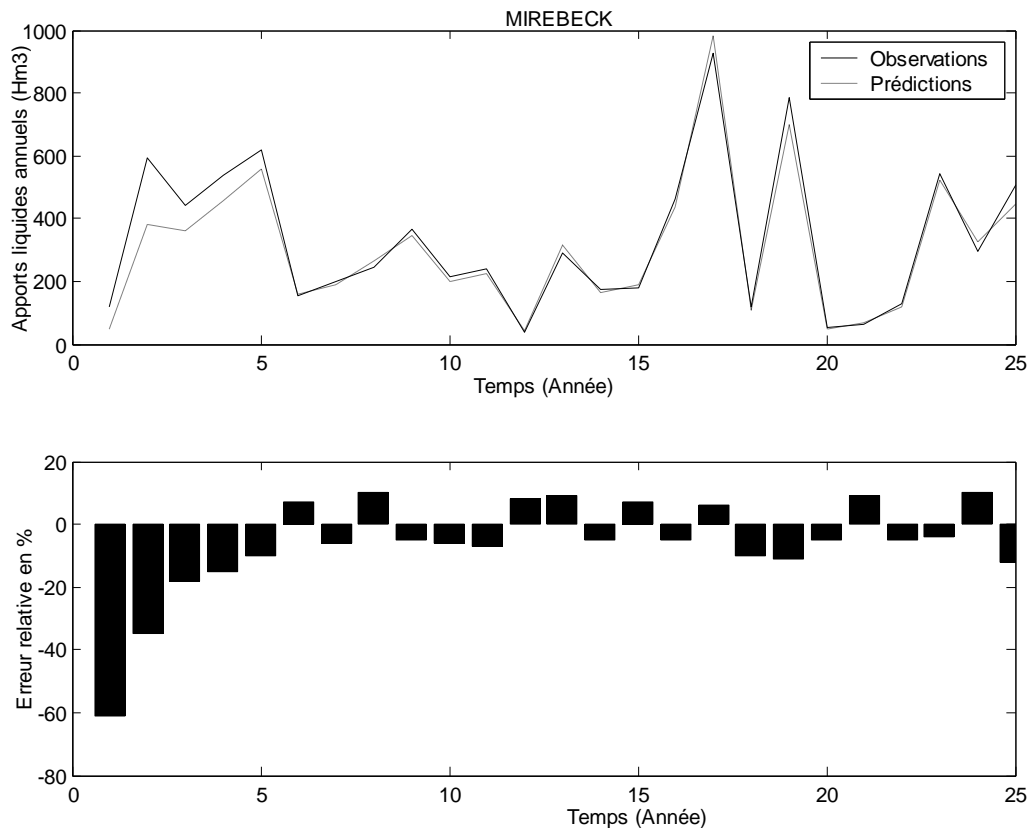


Fig. III.2.8 Prédictions obtenues à la station hydrométrique MIREBECK

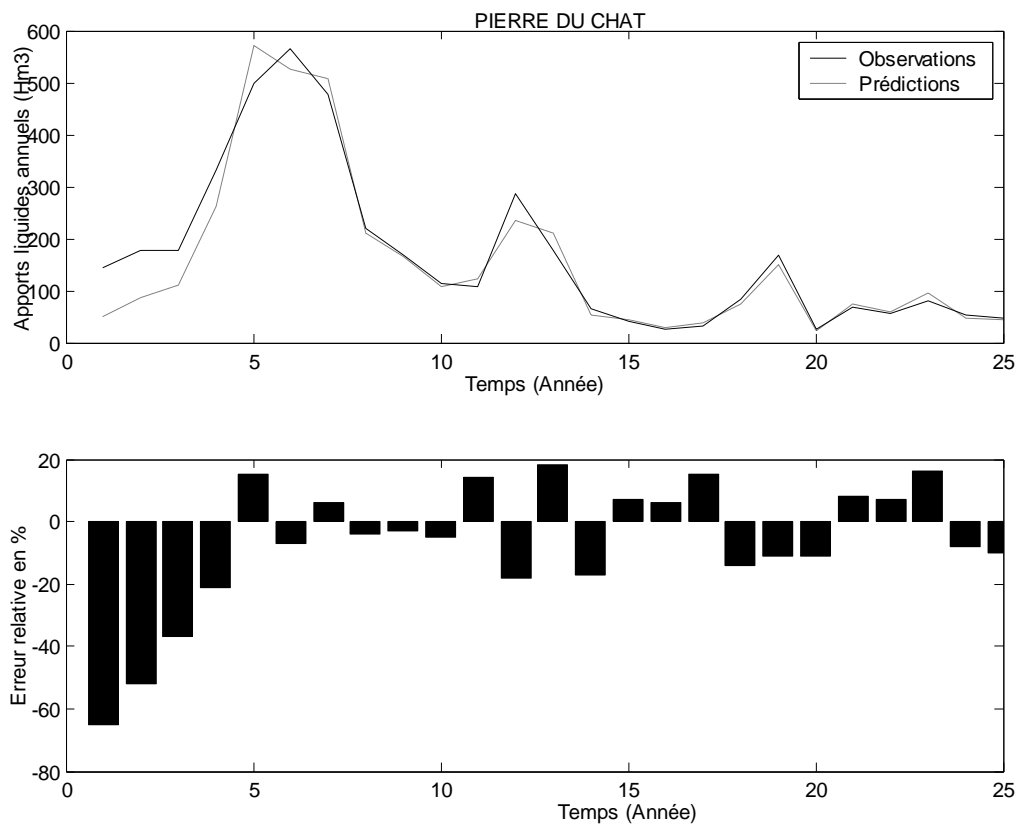


Fig. III.2.9 Prédictions obtenues à la station hydrométrique PIERRE DU CHAT

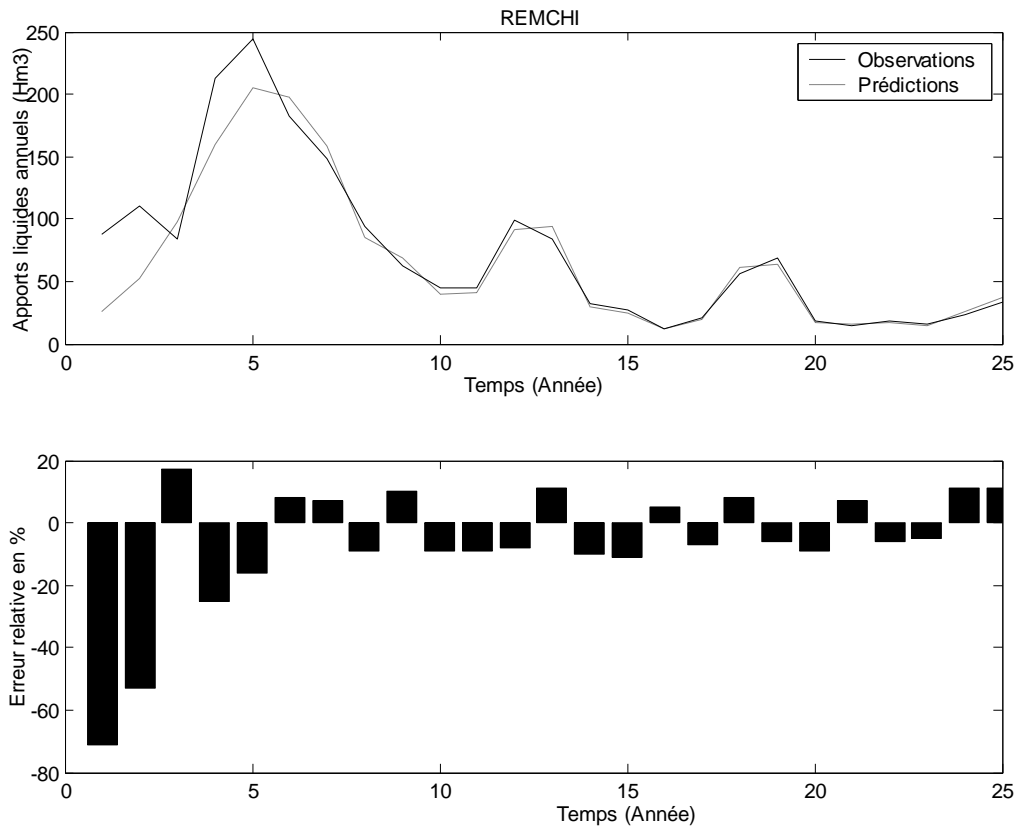


Fig. III.2.10 Prédiction obtenue à la station hydrométrique REMCHI

- **Prédictions spatiales :** Pour la prédiction multi-site dans la dimension spatiale, la Fig. III.2.11 et Fig.III.2.35 en sont également des exemples, elles montrent les prédictions obtenues simultanément dans les dix stations hydrométriques comparées aux observations enregistrées et donnent en même temps les erreurs relatives de prédiction exprimées en pourcentage. La première figure correspond au début des calculs (année 1968), nous pouvons y constater le décalage important qui existe entre les observations et leurs estimations (prédictions). Ce décalage est bien exprimé par l'erreur relative qui atteint un maximum de -75 % à la station numéro 3 (BOUCHEGOUF). La seconde figure correspond quant à elle à la fin des calculs (1992), il est évident que le décalage entre les observations et leurs prédictions est nettement plus réduit et l'erreur relative qui était maximale à BOUCHEGOUF en 1968 a atteint la valeur de -4 % en 1992.

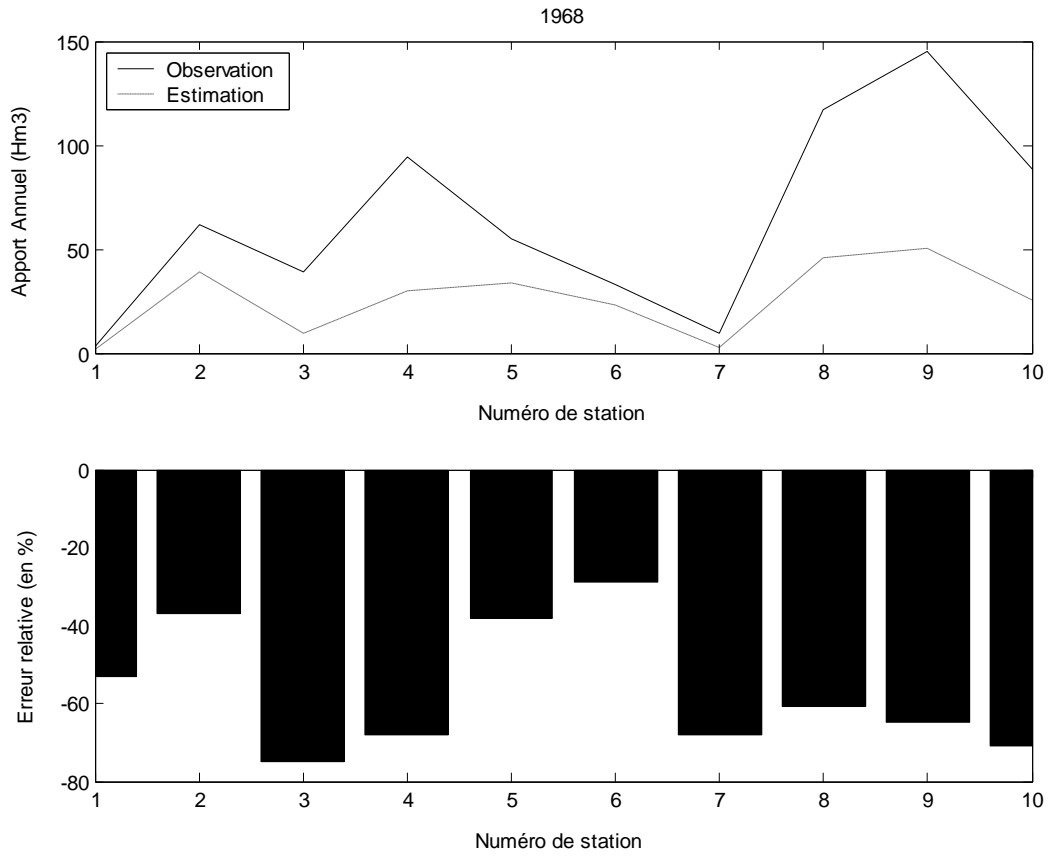


Fig.III.2.11 Prédictions des apports annuels aux dix stations hydrométriques(1968)

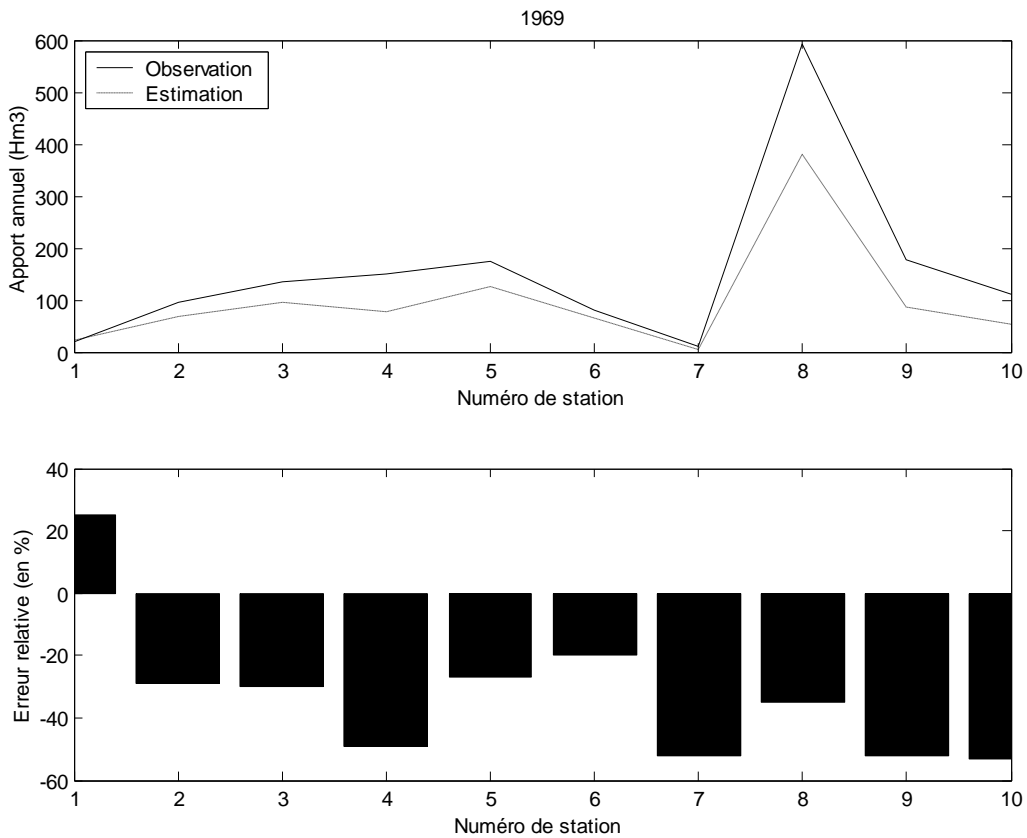


Fig.III.2.12 Prédictions des apports annuels aux dix stations hydrométriques(1969)

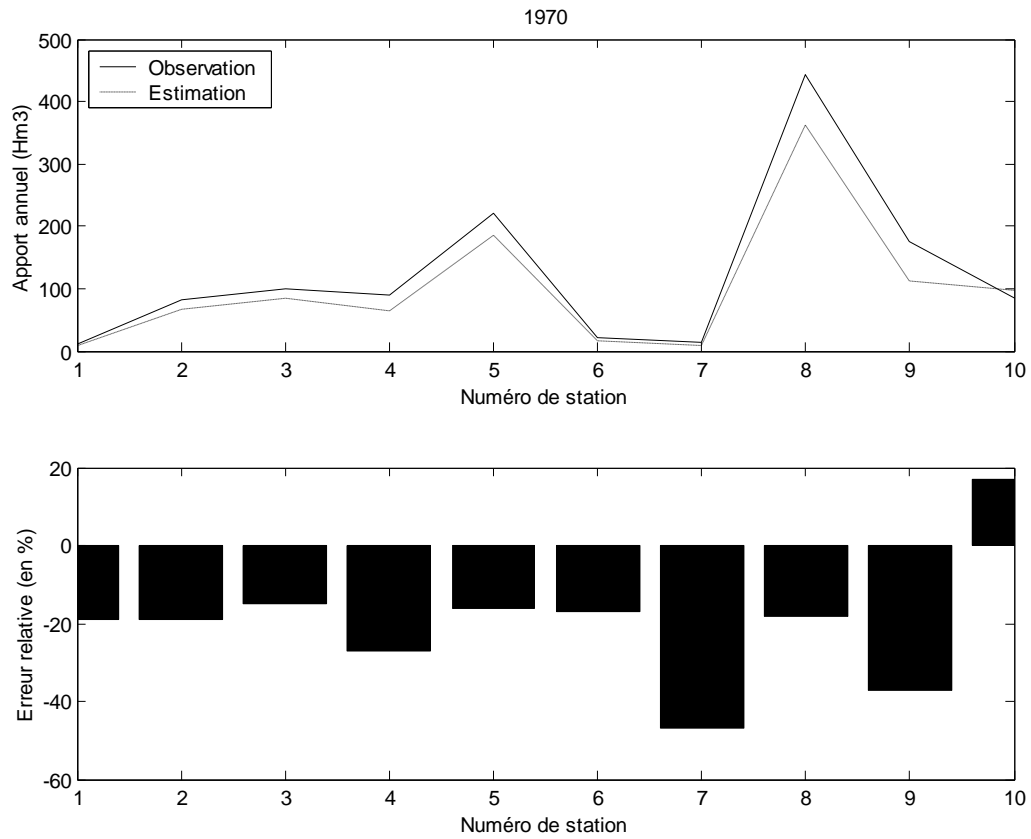


Fig.III.2.13 Prédictions des apports annuels aux dix stations hydrométriques(1970)

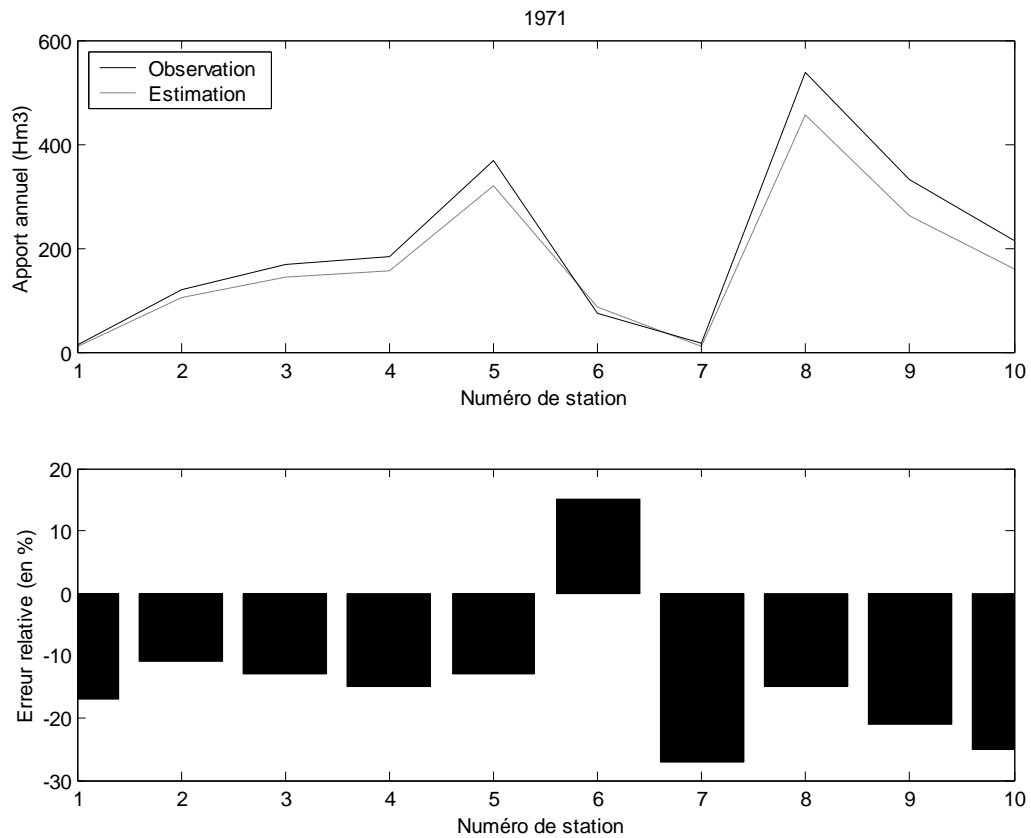


Fig.III.2.14 Prédictions des apports annuels aux dix stations hydrométriques(1971)

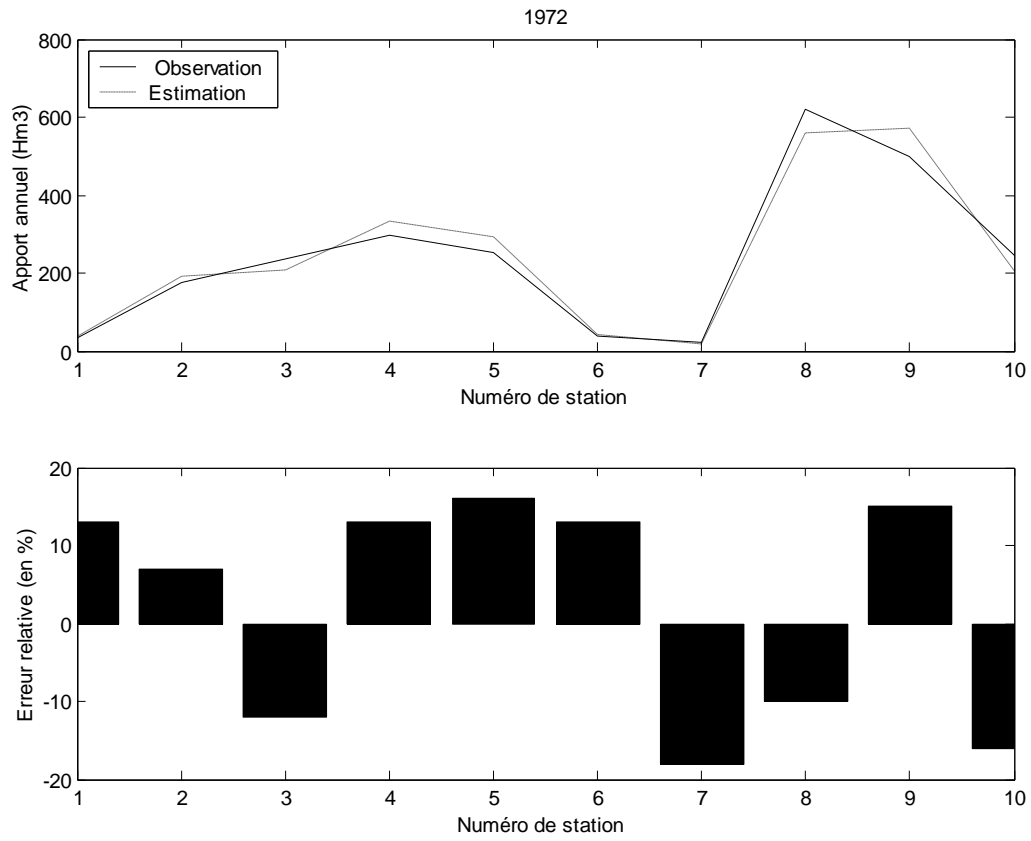


Fig.III.2.15 Prédictions des apports annuels aux dix stations hydrométriques(1972)

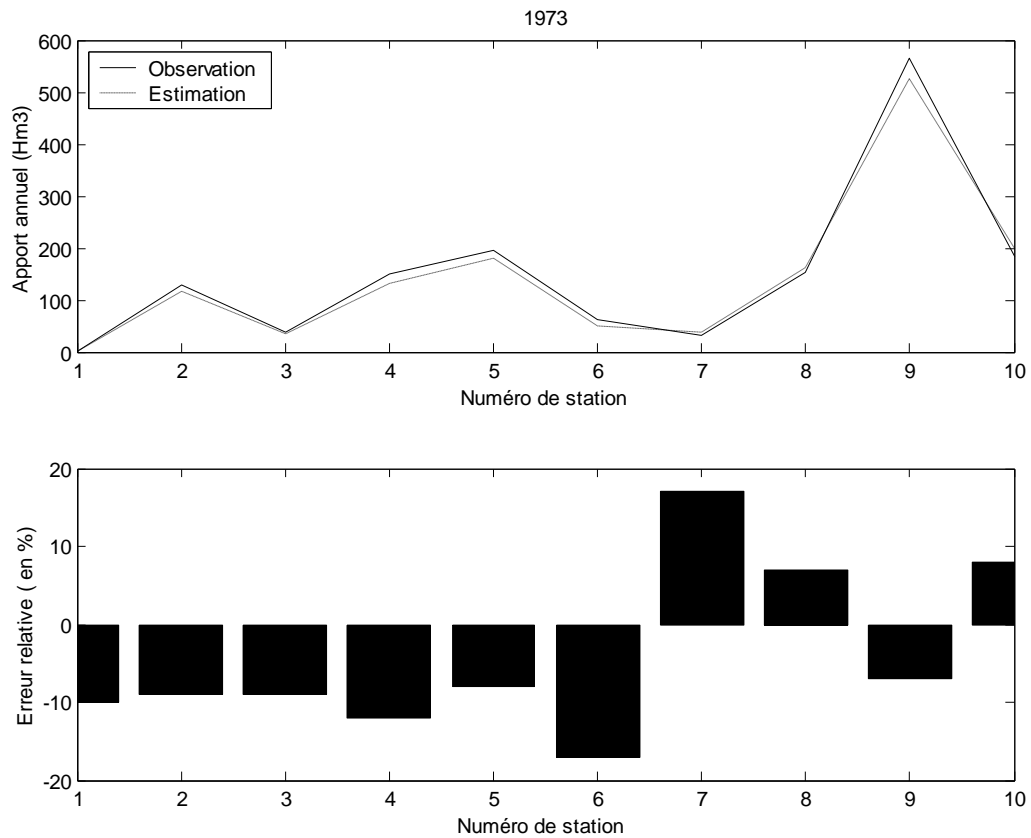


Fig.III.2.16 Prédictions des apports annuels aux dix stations hydrométriques(1973)

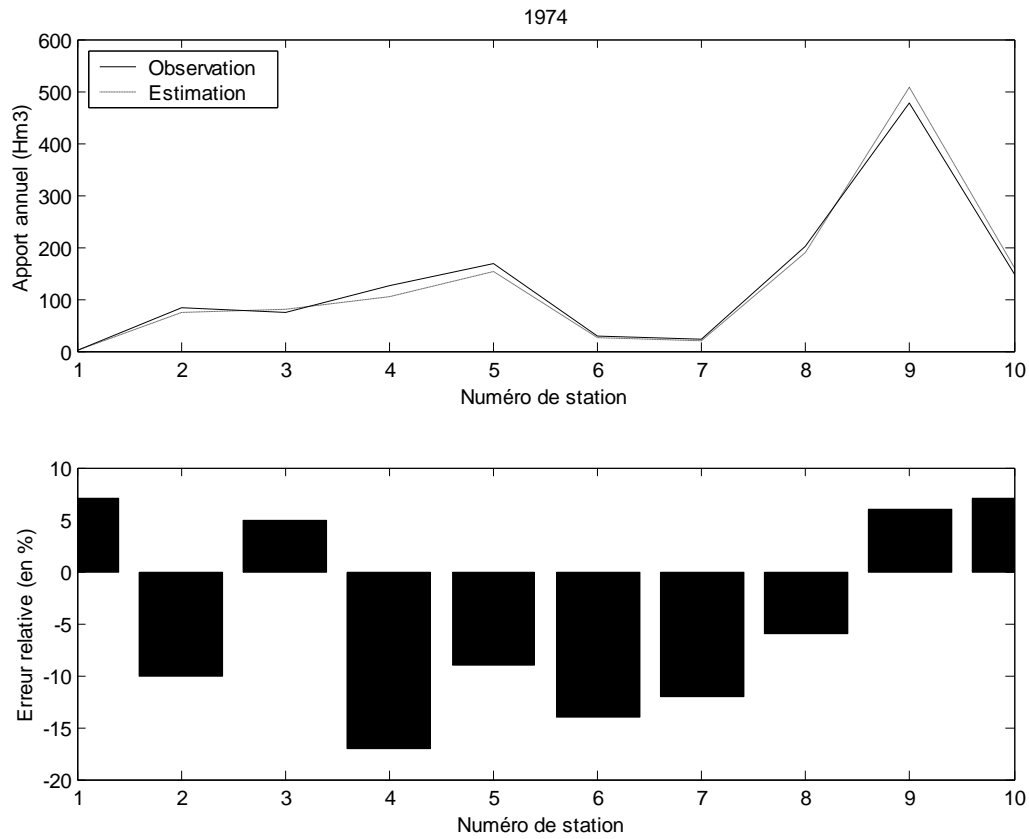


Fig.III.2.17 Prédictions des apports annuels aux dix stations hydrométriques(1974)

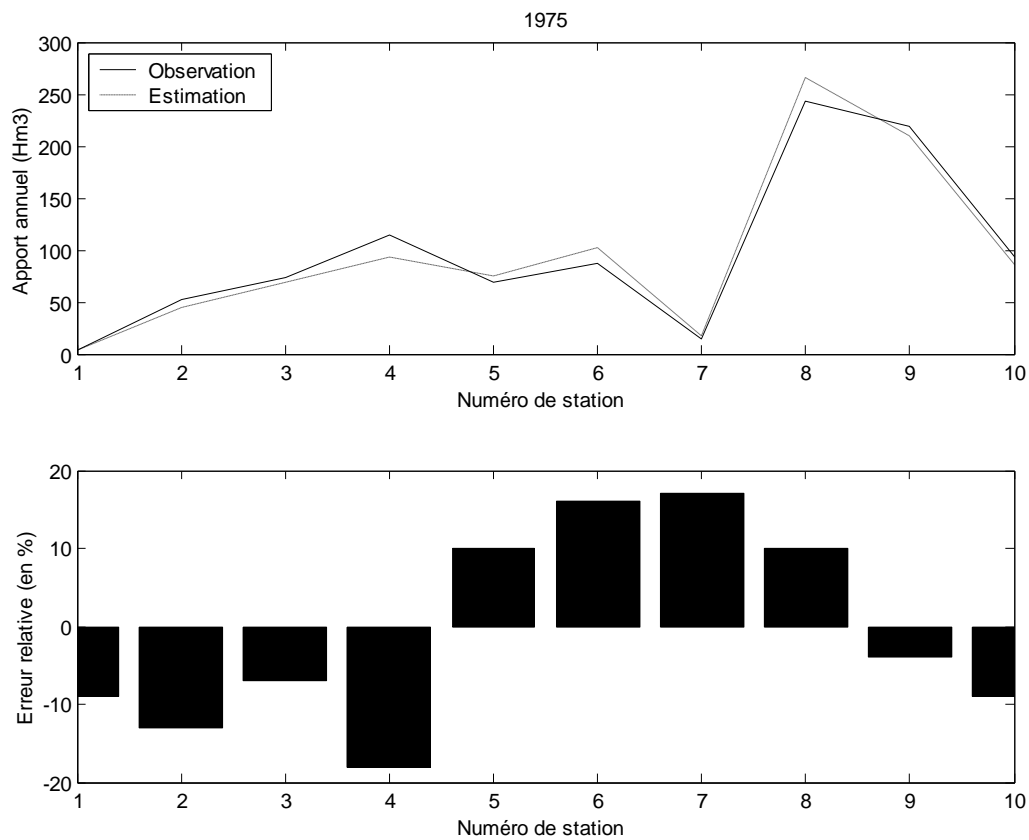


Fig.III.2.18 Prédictions des apports annuels aux dix stations hydrométriques(1975)

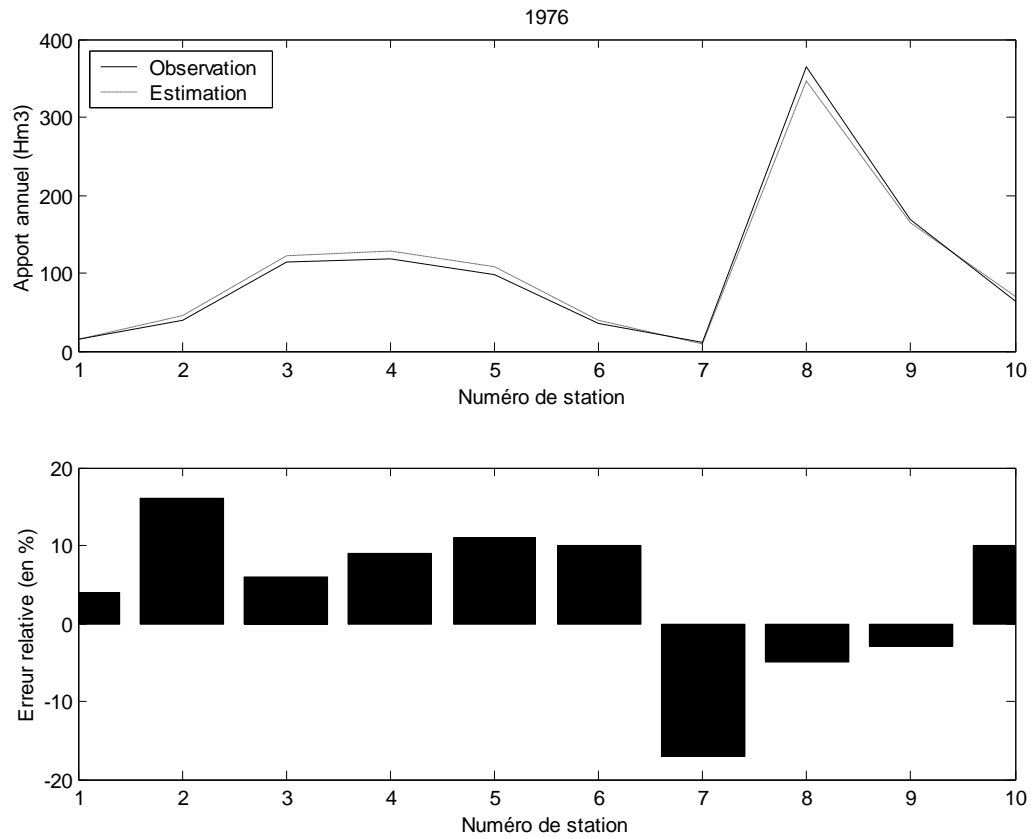


Fig.III.2.19 Prédictions des apports annuels aux dix stations hydrométriques(1976)

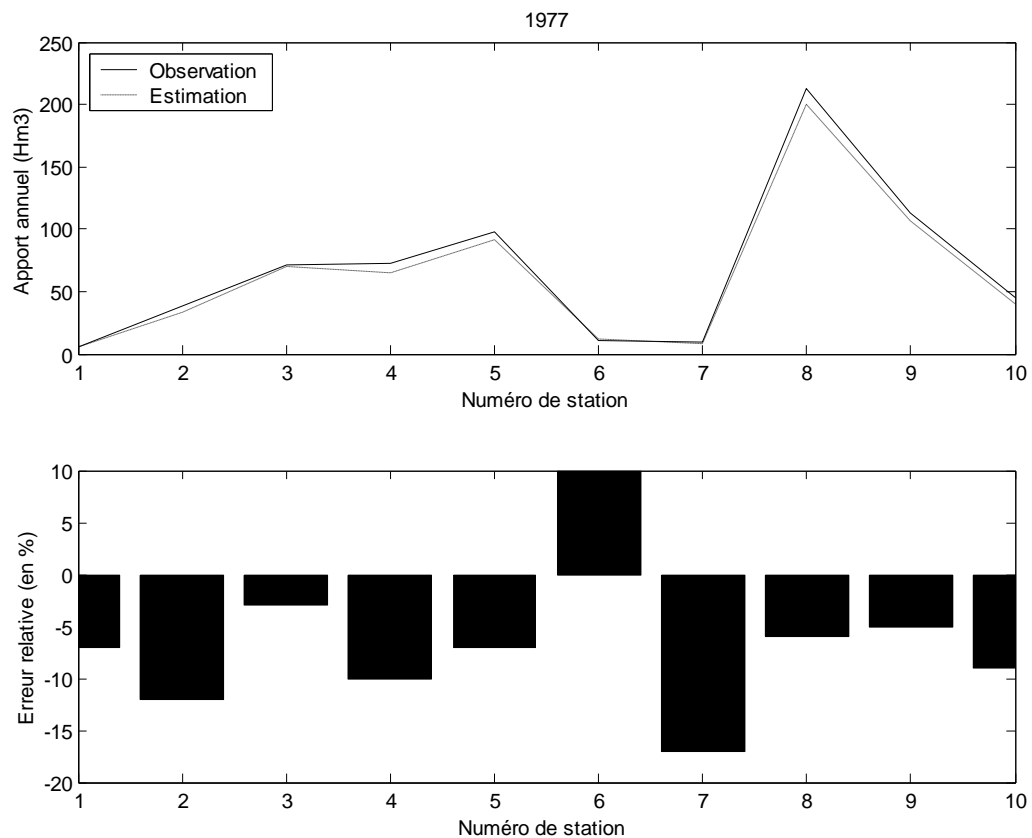


Fig.III.2.20 Prédictions des apports annuels aux dix stations hydrométriques(1977)

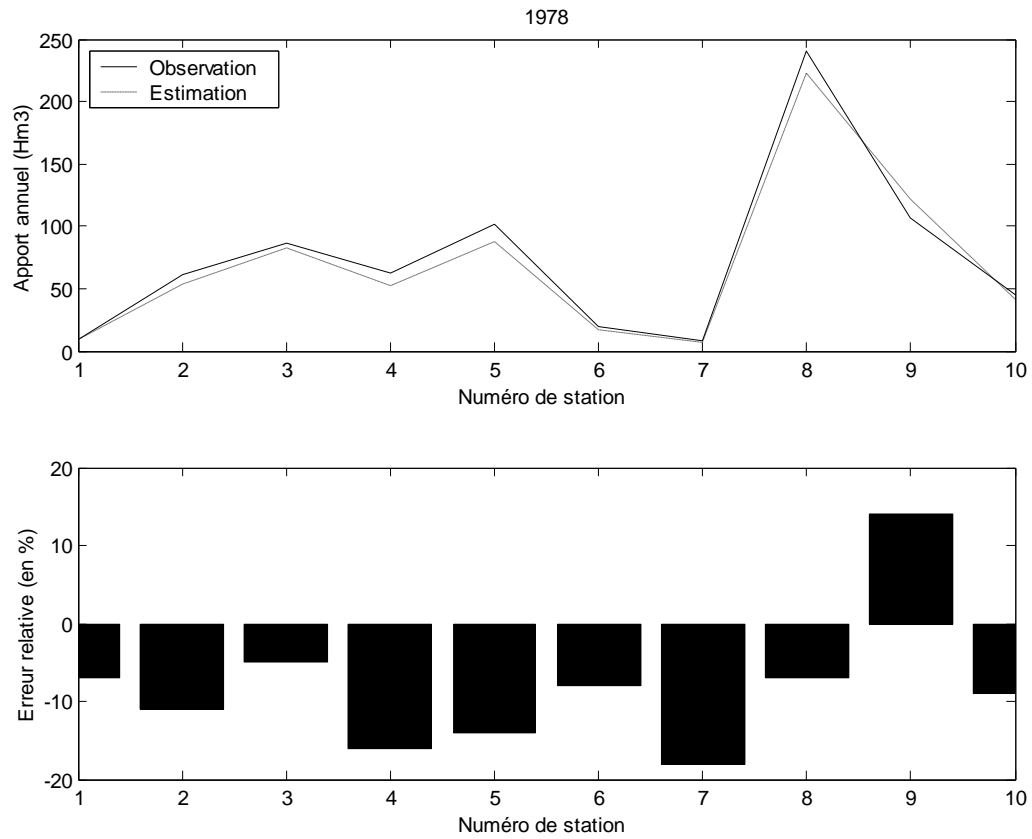


Fig.III.2.21 Prédictions des apports annuels aux dix stations hydrométriques(1978)

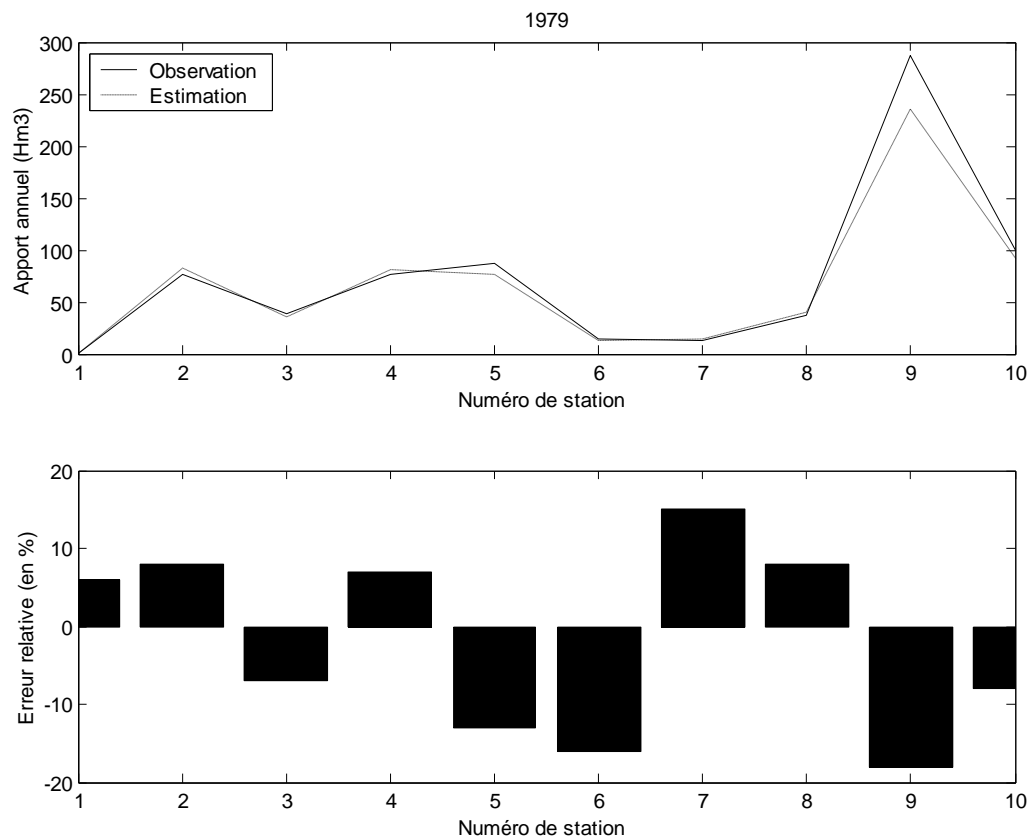


Fig.III.2.22 Prédictions des apports annuels aux dix stations hydrométriques(1979)

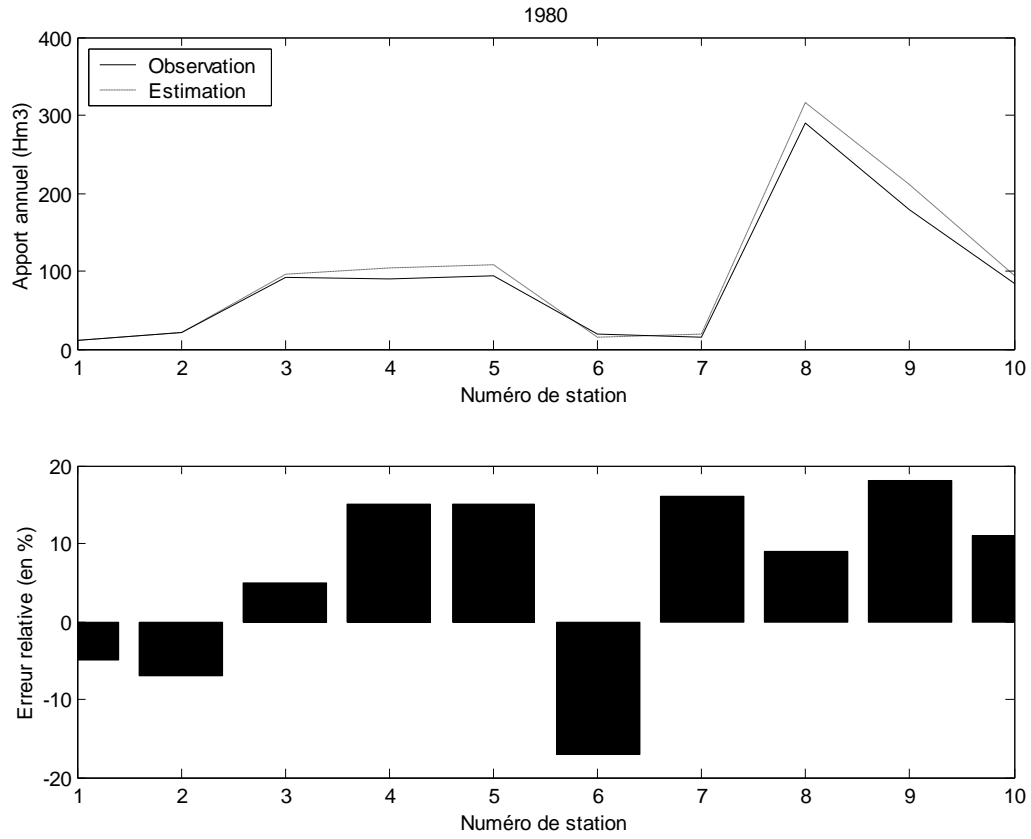


Fig.III.2.23 Prédictions des apports annuels aux dix stations hydrométriques(1980)

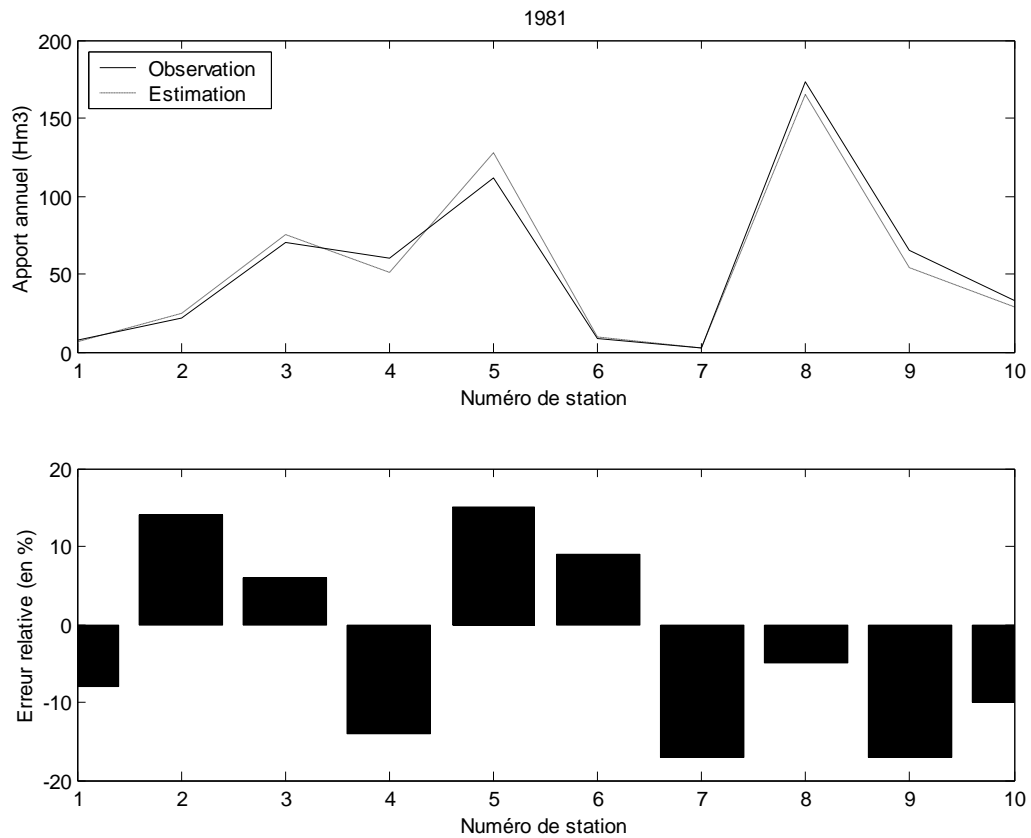


Fig.III.2.24 Prédictions des apports annuels aux dix stations hydrométriques(1981)

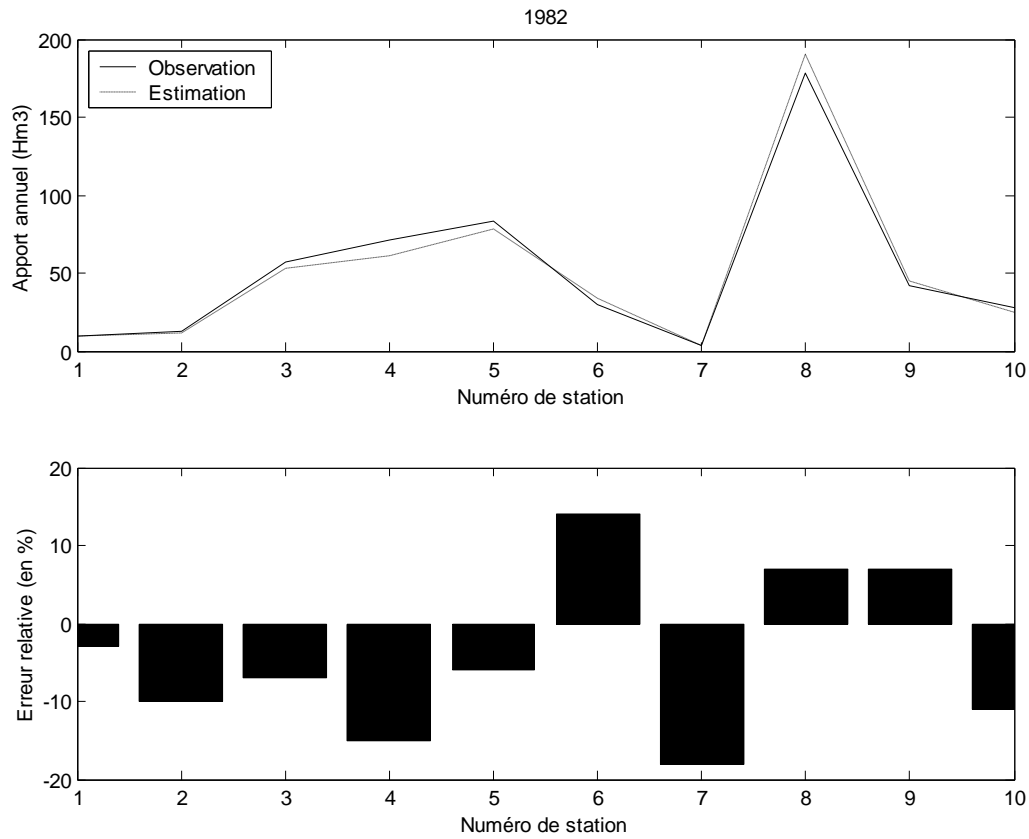


Fig.III.2.25 Prédictions des apports annuels aux dix stations hydrométriques(1982)

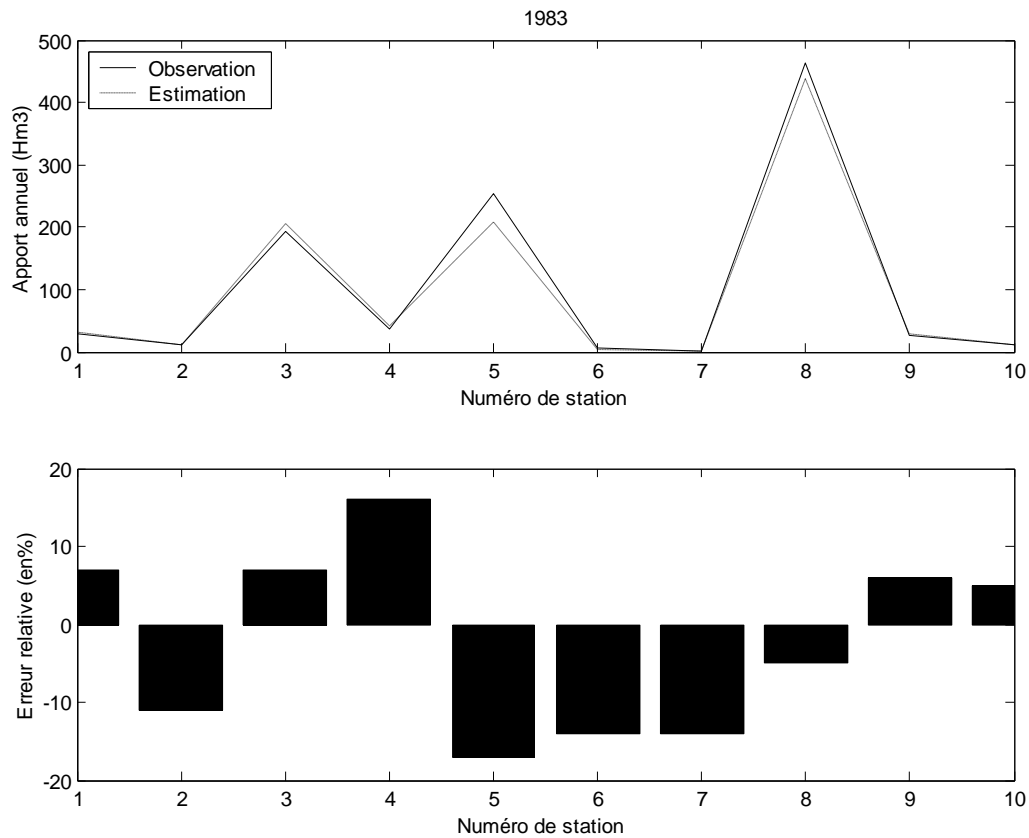


Fig.III.2.26 Prédictions des apports annuels aux dix stations hydrométriques(1983)

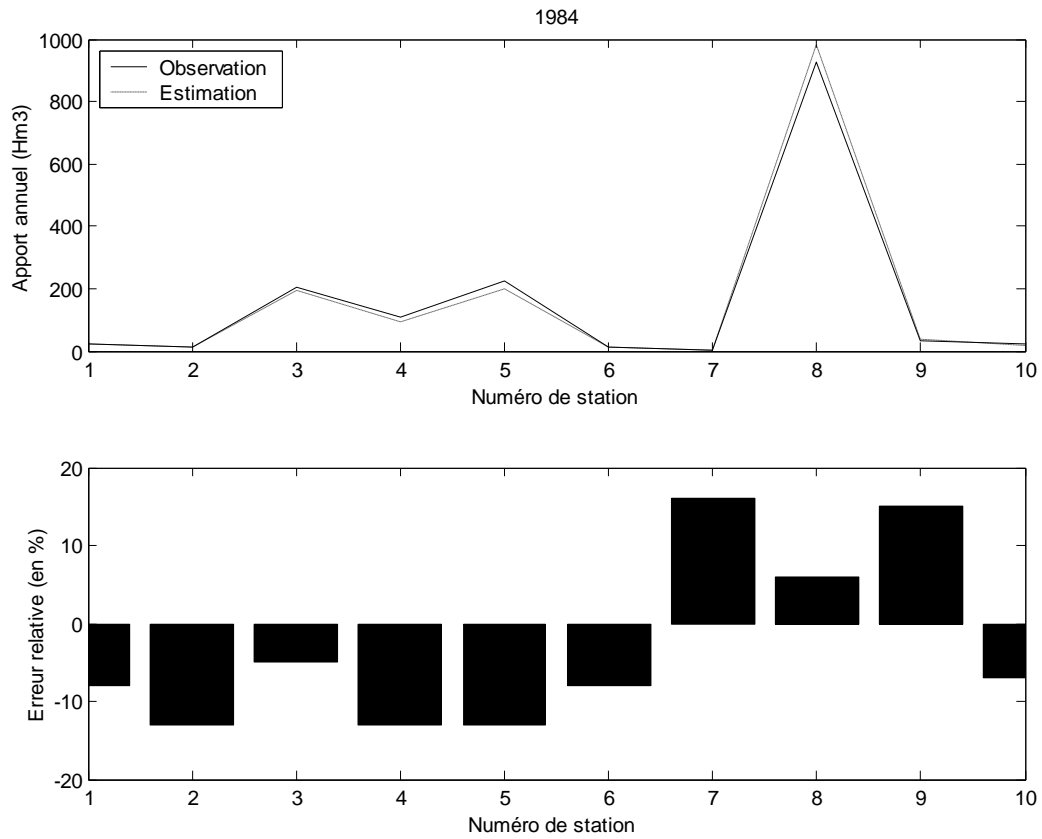


Fig.III.2.27 Prédictions des apports annuels aux dix stations hydrométriques(1984)

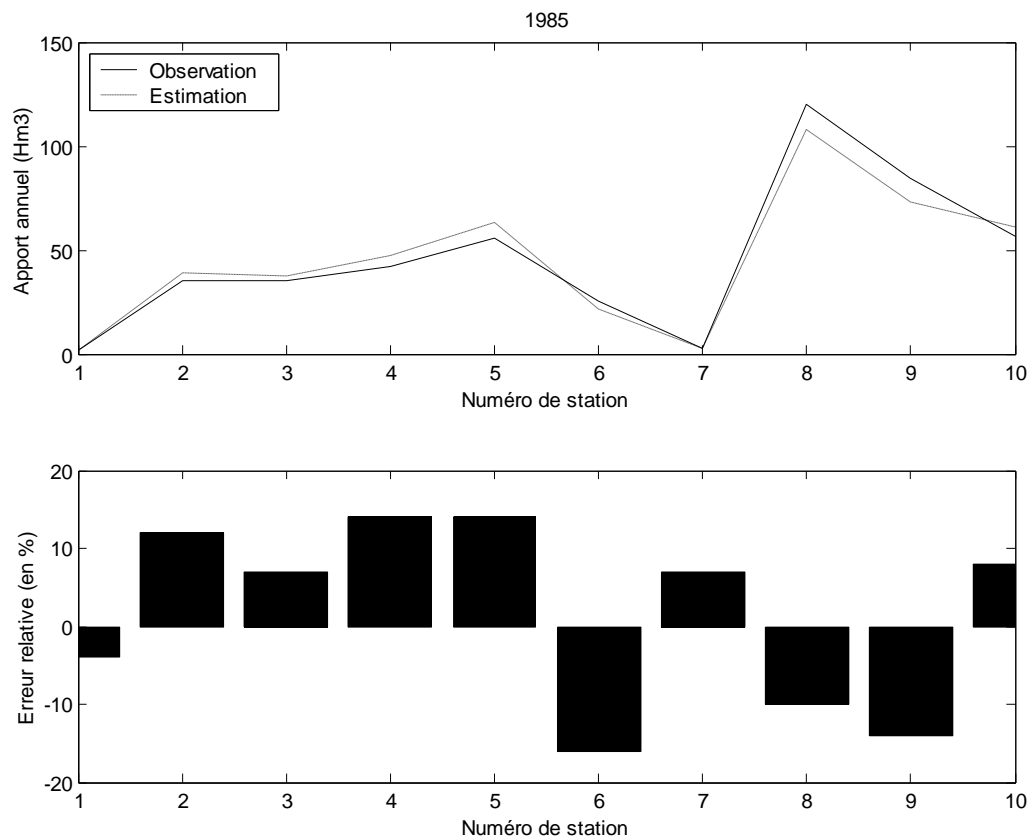


Fig.III.2.28 Prédictions des apports annuels aux dix stations hydrométriques(1985)

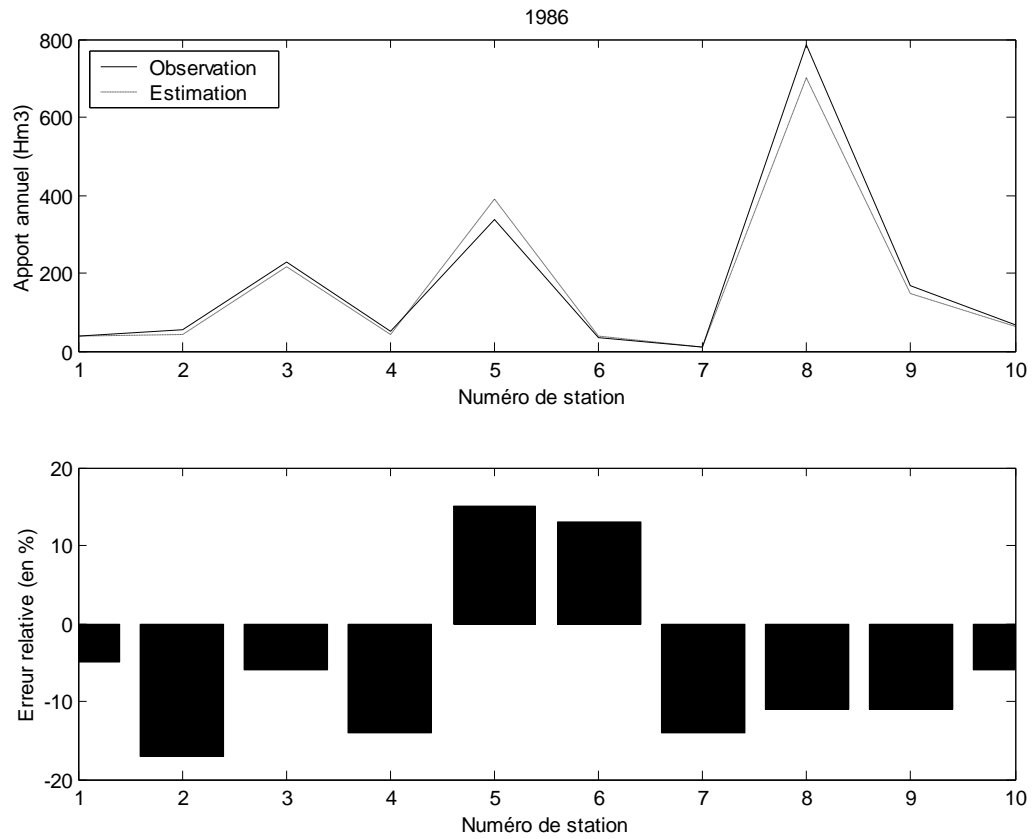


Fig.III.2.29 Prédictions des apports annuels aux dix stations hydrométriques(1986)

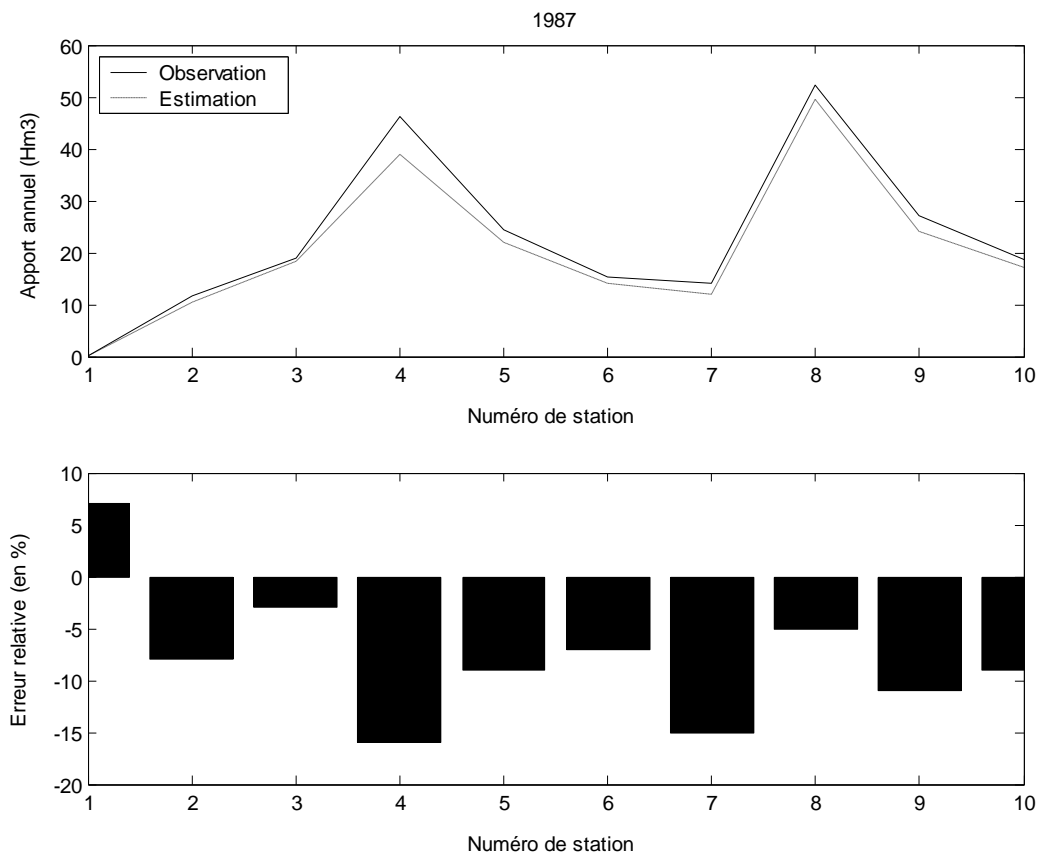


Fig.III.2.30 Prédictions des apports annuels aux dix stations hydrométriques(1987)

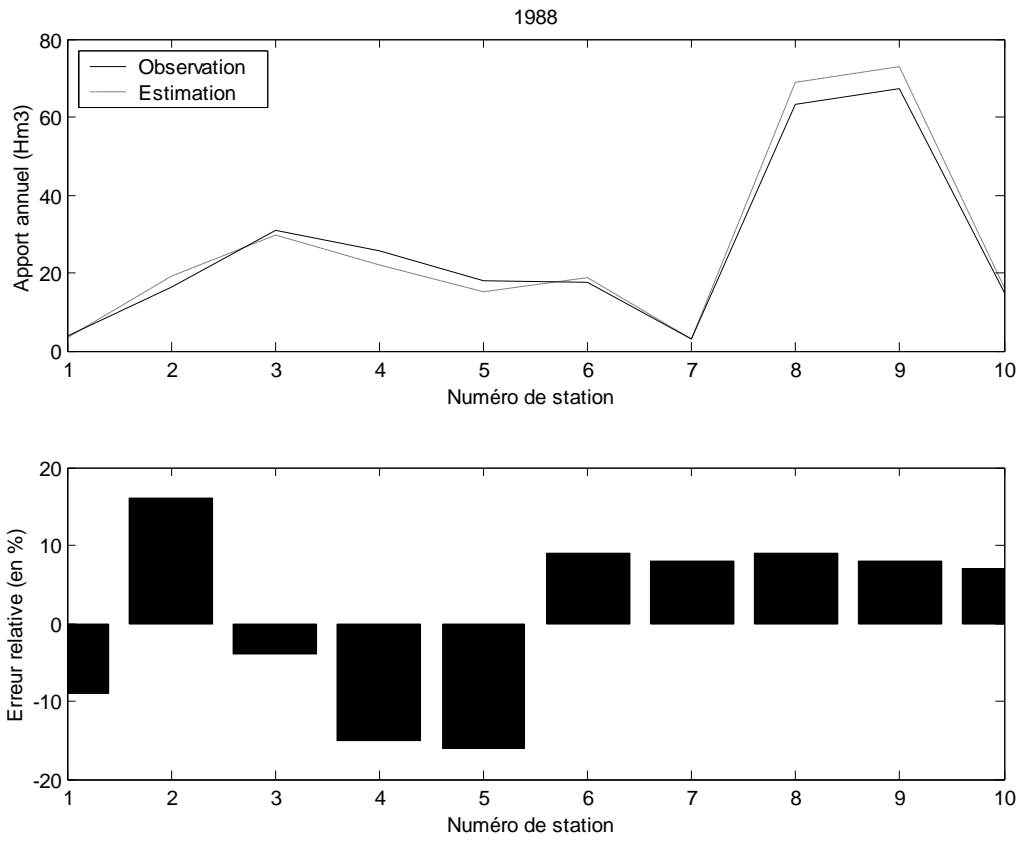


Fig.III.2.31 Prédictions des apports annuels aux dix stations hydrométriques(1988)

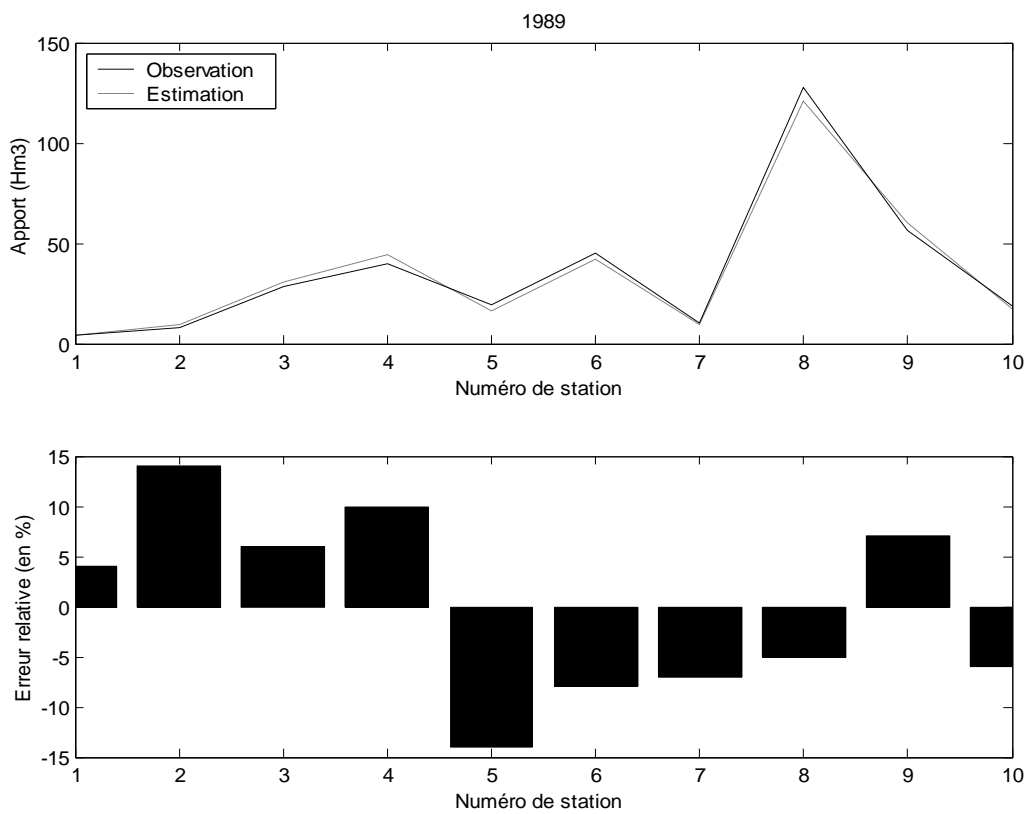


Fig.III.2.32 Prédictions des apports annuels aux dix stations hydrométriques(1989)

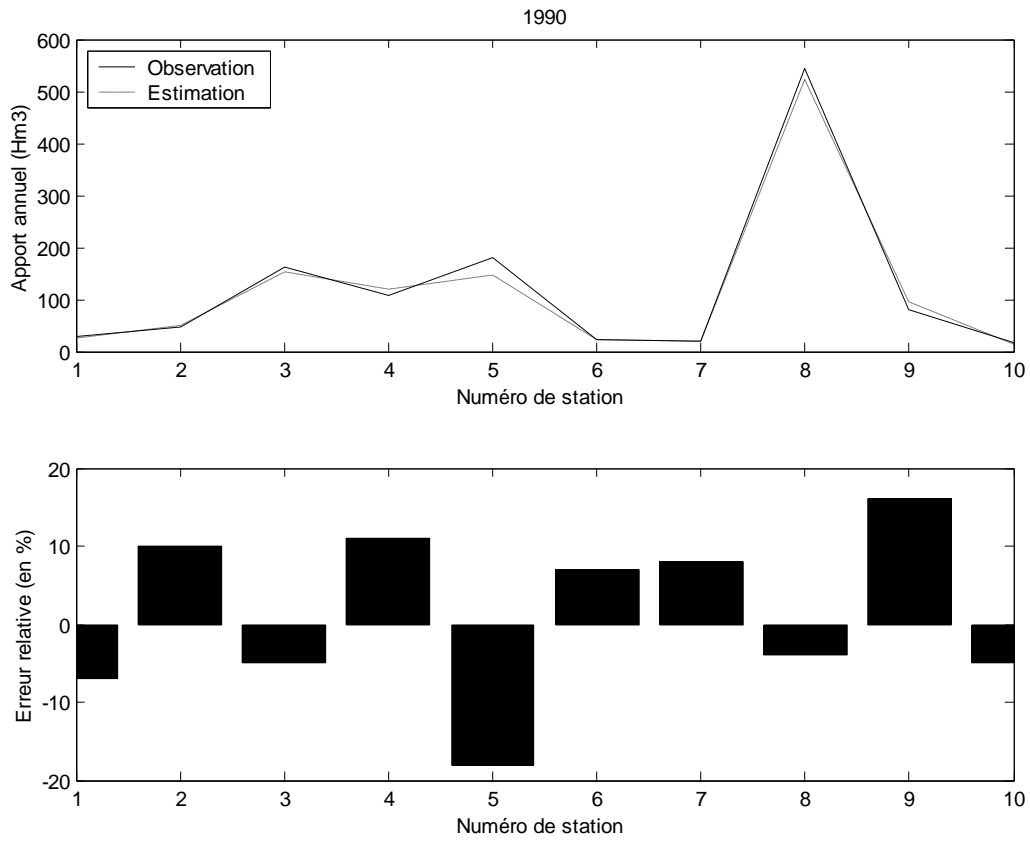


Fig.III.2.33 Prédictions des apports annuels aux dix stations hydrométriques(1990)

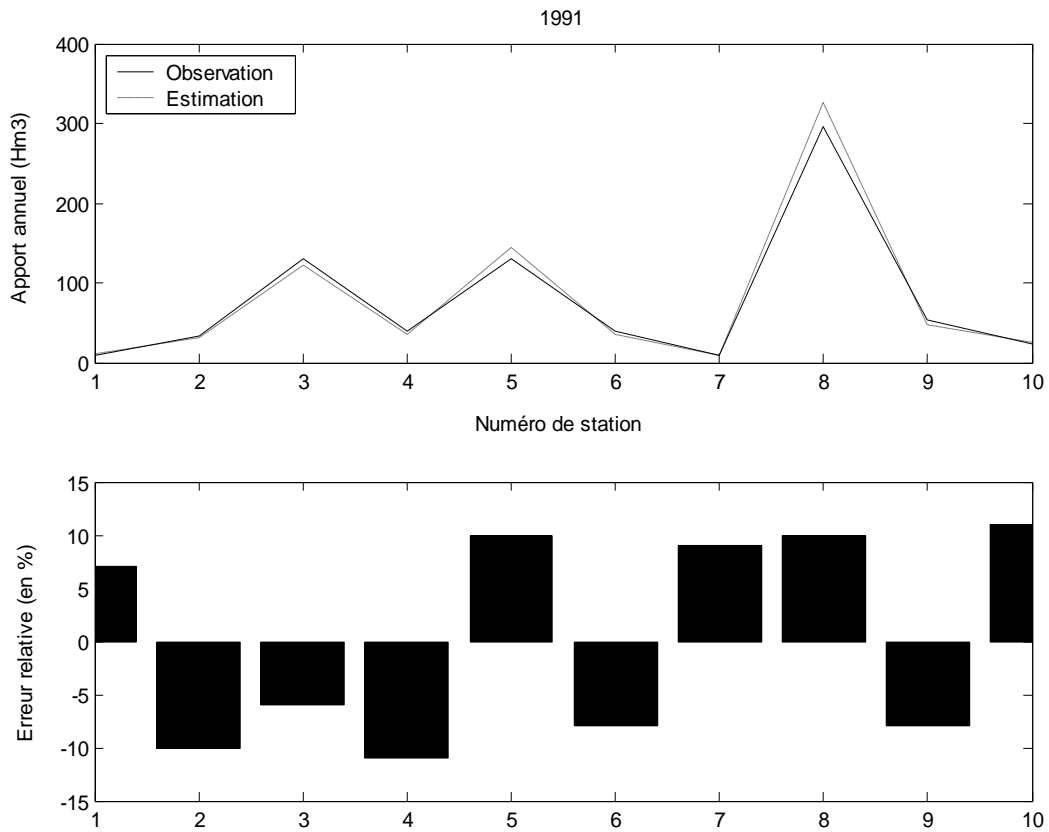


Fig.III.2.34 Prédictions des apports annuels aux dix stations hydrométriques(1991)

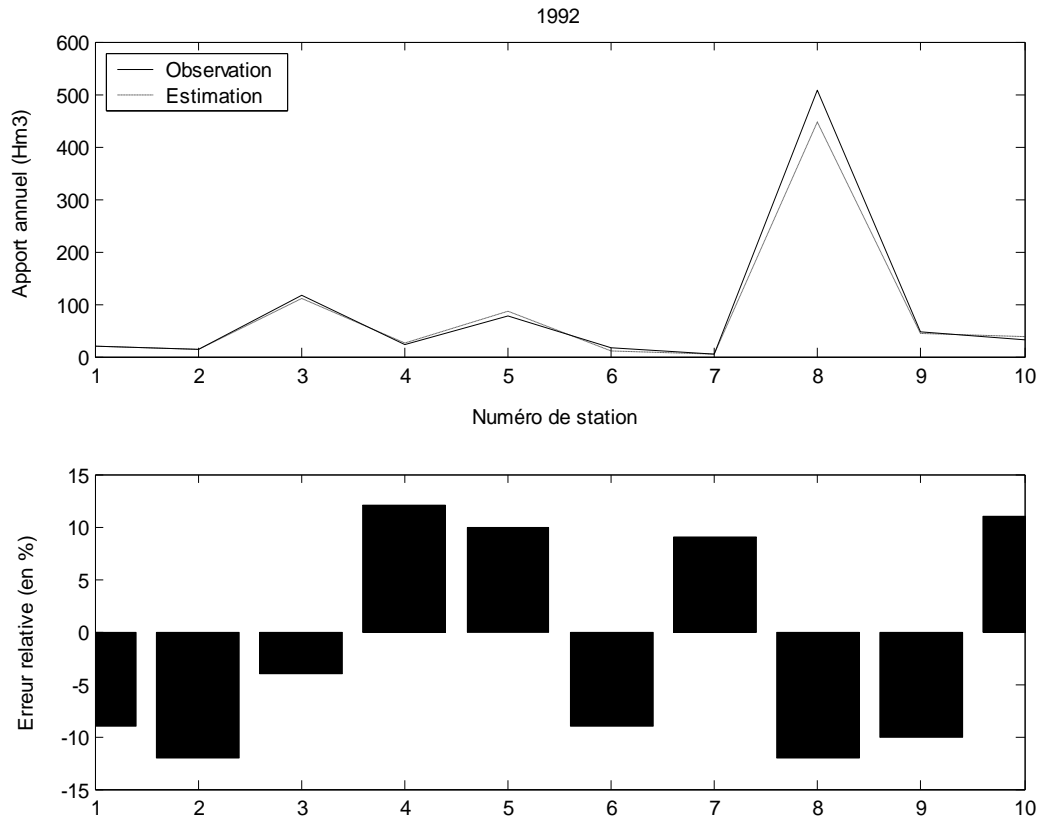


Fig.III.2.35 Prédiction des apports annuels aux dix stations hydrométriques(1992)

L'examen de toutes ces figures montre que les prédictions obtenues, aussi bien dans la dimension temporelle que spatiale, suivent de plus en plus près les valeurs observées. Cette concordance dans les variations et ce rapprochement entre les valeurs observées et leurs prédictions exprime la qualité de l'adéquation du modèle ajusté d'une part et d'autre, le pouvoir de ce dernier à prendre en charge les variations temporelles des données en s'adaptant à la nature dynamique des systèmes hydrologiques concernés. Cela montre clairement que le (FK) multi-site est vraiment efficace en matière de modélisation des apports liquides annuels.

III.2.7.2 Optimalité des résultats

Dans le Tableau III.2.2, nous présentons les différences entre les observations et les prédictions correspondantes. Les signes Positif (+) et (-) sont naturellement assignés aux différences de surestimation et sous-estimation respectivement. Dans le Tableau III.2.3 nous présentons l'erreur relative en pourcentage de ces prédictions. Il serait utile de rappeler, à ce stade, que les différences entre les observations et leurs prédictions dans les figures Fig. III.2.1-10 et Fig. III.2.11-35 sont les meilleures que l'on puisse obtenir en utilisant la théorie du filtrage linéaire et ce pour toutes les stations étudiées. Ce ci est simplement dû à l'optimalité de l'estimateur du (FK), (**Kalman, 1960**). Par ailleurs, il faut signaler que les différences relativement importantes entre les valeurs observées et leurs prédictions, durant les premières itérations, sont dues à l'adaptation du filtre. En effet

Tableau III.2.1 Prédications multi-sites des apports liquides annuels (1968-1992) aux dix stations hydrométriques.

Année	Ain berda	Béni bahdel	Boucheougouf	Bouhnifia	Cheffia	Ksob	Mefrouche	Mirebeck	Pierre du chat	Remchi
1968	1.57	38.99	9.74	30.11	33.830	23.43	2.98	45.68	50.73	25.52
1969	23.15	67.46	94.85	76.19	126.61	64.31	5.48	380.86	84.94	52.17
1970	9.47	67.15	84.53	65.55	184.76	17.04	7.89	362.62	111.28	98.17
1971	11.68	105.45	145.12	155.55	318.96	86.01	12.11	456.81	263.03	59.51
1972	37.04	190.12	208.00	334.70	291.77	41.90	17.12	558.00	571.76	204.90
1973	0.79	115.91	34.78	130.68	179.84	50.76	37.18	161.49	525.04	197.59
1974	2.87	74.09	78.85	103.50	15.34	24.85	20.41	188.20	506.72	158.16
1975	3.53	5.10	68.94	93.64	75.39	101.52	17.68	266.37	210.55	85.25
1976	14.98	46.07	121.64	127.96	108.98	38.61	8.52	346.18	164.39	69.16
1977	5.57	33.44	69.65	65.52	91.28	11.84	7.87	200.25	107.06	40.31
1978	9.08	53.99	82.33	52.58	87.93	17.33	6.61	223.24	122.02	40.74
1979	0.47	82.56	36.34	81.74	76.26	12.43	14.41	40.75	235.15	91.19
1980	10.26	20.62	95.95	103.97	107.81	15.87	18.54	316.56	210.84	93.87
1981	6.61	24.36	74.77	51.34	127.80	9.28	2.48	164.70	54.47	29.19
1982	9.57	11.70	53.28	60.77	77.86	33.78	3.24	190.38	44.57	24.46
1983	30.51	11.16	206.28	41.79	208.84	4.88	0.66	439.05	28.27	12.00
1984	22.40	10.94	193.72	95.56	196.97	13.30	3.69	983.87	37.47	19.70
1985	2.00	39.33	37.41	47.65	63.22	21.62	2.79	108.04	72.88	60.70
1986	37.32	43.62	215.63	42.88	390.29	39.76	9.76	700.37	148.94	64.16
1987	0.13	10.59	18.28	38.85	22.10	14.10	11.95	49.59	24.00	16.98
1988	3.43	19.04	29.63	21.85	15.02	18.96	3.08	69.09	72.82	15.85
1989	4.43	9.41	30.34	44.12	16.47	41.87	9.78	121.14	60.21	17.21
1990	26.49	51.09	154.14	120.67	147.62	22.99	20.45	521.80	94.66	14.50
1991	10.80	30.87	122.42	35.11	144.51	35.52	8.82	325.80	48.39	25.65
1992	18.88	13.06	110.71	24.69	86.06	9.49	5.89	446.60	42.65	36.86
1993	1.57	38.99	9.74	30.11	33.83	23.43	2.98	45.68	50.73	25.52

Tableau III.2.2 Innovations du filtre de Kalman (1968-1992).

Année	Ain berda	Béni bahdel	Boucheougouf	Bouhnifia	Cheffia	Ksob	Mefrouche	Mirebeck	Pierre du chat	Remchi
1968	-1.76	-22.90	-29.22	-63.98	-20.73	-9.75	-6.29	-71.46	-94.23	-62.48
1969	4.63	-27.55	-40.65	-73.20	-46.82	-16.38	-5.90	-212.23	-92.02	-58.83
1970	-2.22	-15.75	-14.91	-24.24	-35.19	-3.54	7.00	-79.59	-65.35	14.26
1971	-2.36	-13.03	-21.68	-27.45	-49.78	11.41	-4.38	-80.61	-69.92	-53.17
1972	4.26	12.43	-28.36	38.50	40.67	4.75	-3.70	-62.00	74.58	-39.03
1973	-0.08	-11.46	-3.44	-17.82	-15.85	-10.24	5.34	9.85	-39.51	14.63
1974	0.18	-8.23	3.75	-21.19	-1.61	-3.91	-2.78	-12.01	28.68	10.34
1975	-0.35	-6.74	-5.19	-20.55	7.04	13.85	2.54	23.77	-8.77	-8.43
1976	0.57	6.35	6.88	10.56	11.32	3.63	-1.72	-18.22	-5.08	6.28
1977	-0.41	-4.56	-2.15	-7.28	-6.76	1.09	-1.55	-12.78	-5.63	-3.98
1978	-0.68	-6.67	-4.33	-10.01	-14.07	-1.61	-1.43	-16.80	14.98	-4.03
1979	0.02	6.11	-2.73	5.34	-11.19	-2.40	1.85	2.94	-51.62	-7.92
1980	-0.54	-1.55	4.56	13.56	14.22	-3.20	2.51	26.13	32.16	9.30
1981	-0.57	2.29	4.23	-8.35	16.47	0.80	-0.50	-8.66	-11.15	-3.24
1982	-0.27	-1.30	-4.01	-10.72	-5.32	4.09	-0.71	12.12	2.90	-3.02
1983	1.99	-1.38	13.49	5.76	-44.29	-0.76	-0.10	-23.10	1.60	0.57
1984	-1.94	-1.63	-10.19	-14.28	-30.21	-1.15	0.50	55.69	4.88	-1.48
1985	-0.08	4.21	2.44	5.79	7.86	-4.02	0.18	-12.00	-11.86	4.49
1986	-1.96	-8.93	-13.76	-6.98	52.08	4.51	-1.56	-84.79	18.40	-4.09
1987	0	-0.92	-0.56	-7.40	-2.24	-1.14	-2.09	-2.61	-2.96	-1.67
1988	-0.33	2.62	-1.23	-3.85	-2.84	1.51	0.21	5.70	5.39	1.03
1989	0.17	1.15	1.71	4.01	-2.72	-3.54	-0.71	-6.37	3.93	-1.09
1990	-1.99	4.64	-8.11	11.95	-31.75	1.46	1.44	-21.17	13.05	-0.76
1991	0.70	-3.43	-7.81	-4.34	13.49	-3.04	0.71	29.61	-4.20	2.54
1992	-1.86	-1.78	-4.61	2.64	7.82	-7.04	0.48	-60.90	-4.73	3.65
1993	-1.76	-22.90	-29.22	-63.98	-20.73	-9.75	-6.29	-71.46	-94.23	-62.48

Tableau III.2.3 Erreurs relatives en % des prédictions multi-sites des apports annuels aux dix stations hydrométriques.

Année	Ain berda	Béni bahdel	Bouhegouf	Bouhnia	Cheffia	Ksob	Mefrouche	Mirebeck	Pierre du chat	Remchi
1968	-53	-37	-75	-68	-38	-29	-68	-61	-65	-71
1969	25	-29	-30	-49	-27	-20	-52	-35	-52	-53
1970	-19	-19	-15	-29	-16	-17	-47	-18	-37	17
1971	-17	-11	-13	-15	-13	15	-27	-15	-21	-25
1972	13	7	-12	13	16	13	-18	-10	15	-16
1973	-10	-9	-9	-12	-8	-17	17	7	-7	8
1974	7	-10	5	-17	-9	-14	-12	-6	6	7
1975	-9	-13	-7	-18	10	16	17	10	-4	-9
1976	4	16	6	9	11	10	-17	-5	-3	10
1977	-7	-12	-3	-10	-7	10	-17	-6	-5	-9
1978	-7	-11	-5	-16	-14	-8	-18	-7	14	-9
1979	6	8	-7	7	-13	-16	15	8	-18	-8
1980	-5	-7	5	15	15	-17	16	9	18	11
1981	-8	14	6	-14	15	9	-17	-5	-17	-10
1982	-3	-10	-7	-15	-6	14	-18	7	7	-11
1983	7	-11	7	16	-17	-14	-14	-5	6	5
1984	-8	-13	-5	-13	-13	-8	16	6	15	-7
1985	-4	12	7	14	14	-16	7	-10	-14	8
1986	-5	-17	-6	-14	15	13	-14	-11	-11	-6
1987	7	-8	-3	-16	-9	-7	-15	-5	-11	-9
1988	-9	16	-4	-15	-16	9	8	9	8	7
1989	4	14	6	10	-14	-8	-7	-5	7	-6
1990	-7	10	-5	11	-18	7	8	-4	16	-5
1991	7	-10	-6	-11	10	-8	9	10	-8	11
1992	-9	-12	-4	12	10	-9	9	-12	-10	11
1993	-53	-37	-75	-68	-38	-29	-68	-61	-65	-71

durant les premières itérations de l'algorithme de Kalman, la matrice gain K_k prend des valeurs importantes dont la conséquence est une estimation automatiquement mauvaise. Dans cette première phase, la mesure Z_k étant une information objective est plus crédible que l'estimation fournie par le modèle. En d'autres termes, on fait plus confiance à la mesure qu'au modèle. Après quelques itérations, la confiance assignée au modèle, comme mécanisme de prédiction, prend de plus en plus de l'espace, cette confiance croit et s'améliore d'une manière continue.

Dans ce cas, la valeur du gain du filtre K_k devient plus petite, par conséquent, la prédiction fournie par le modèle devient très proche de la valeur observée tel que montré par la Fig. III.2.35 pour la prédiction spatiale et par les Fig.III.2.1-10 pour la prédiction temporelle. La Fig. III.2.36 montre la variation continuellement décroissante du gain du filtre, du début à la fin des calculs.

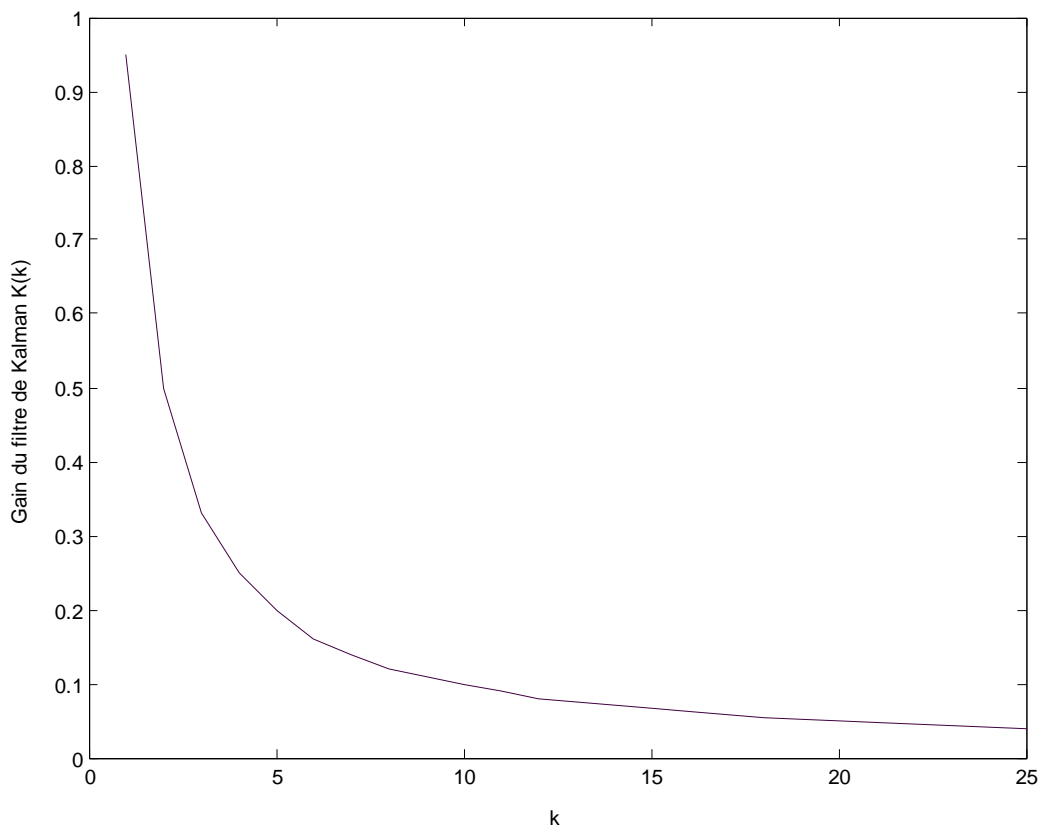


Fig. III.2.36 Gain du filtre de Kalman

L'optimalité des résultats ainsi obtenus est bien prouvée par la convergence de la matrice de covariance de l'erreur de prédiction et plus exactement par les éléments de sa diagonale majeure, (Schlee, 1967), comme c'est indiqué dans la figure Fig. III.2.37. Cette convergence vers une valeur stable devient évidente juste après les premières itérations de l'algorithme tout en restant positive. Ce résultat étant conforme à la performance espérée du (FK), vient confirmer l'adéquation du modèle ajusté au processus étudié.

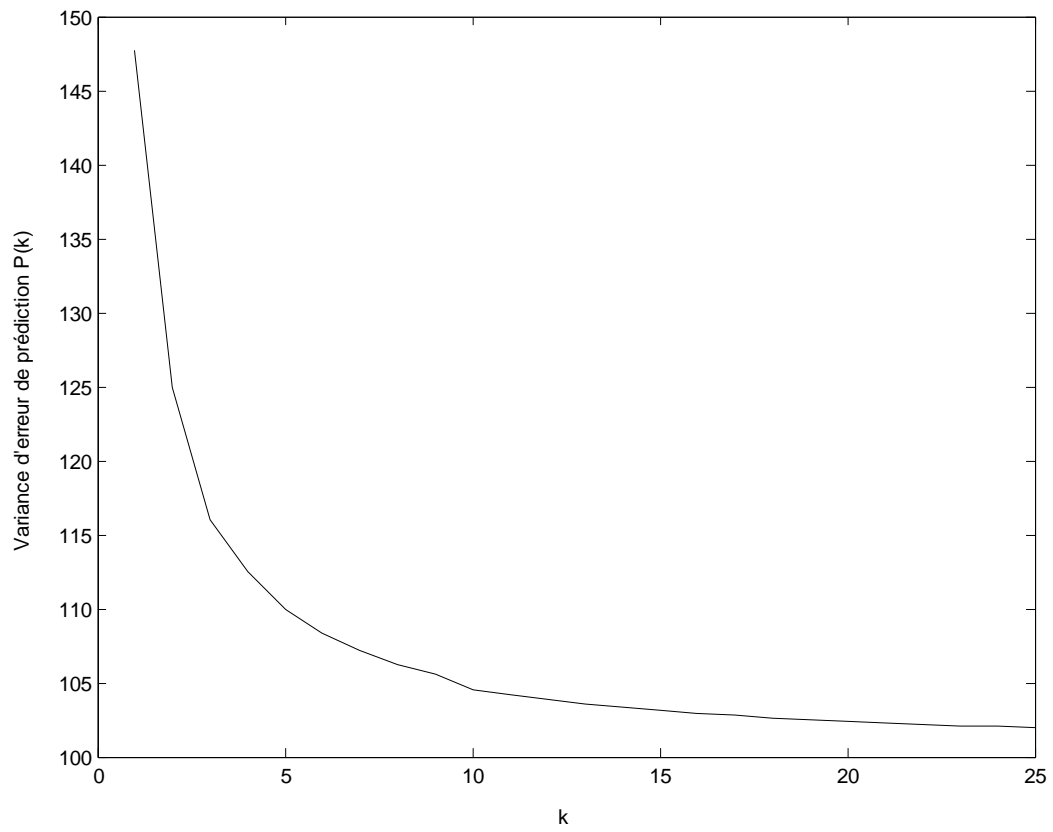


Fig. III.2.37 Covariance d'erreur de prédiction

III.2.7.3 Analyse des innovations du filtre (résidus)

Les erreurs de prédictions ou innovations du filtre pour les des dix stations étudiées, comme c'est indiqué par les Fig.III.2.38-47, présentent des valeurs trop élevées durant les cinq premiers pas par rapport au reste des valeurs. Nous rappelons que cette différence est tout à fait normale et inévitable. Elle est due principalement à l'adaptation de l'algorithme du filtre qui fait que durant les premières itérations l'on assigne plus de confiance à la mesure en tant qu'information objective qu'au modèle ajusté. Le résultat est un écart important entre l'observation et son estimation par le filtre, mais une fois que ce dernier converge tout rentre dans l'ordre : le modèle ajusté reprend beaucoup plus de confiance et les innovations qu'il produit sont par conséquent beaucoup plus petites à la fin des calculs, c'est le cas de toutes les stations étudiées comme c'est indiqué par les figures citées.

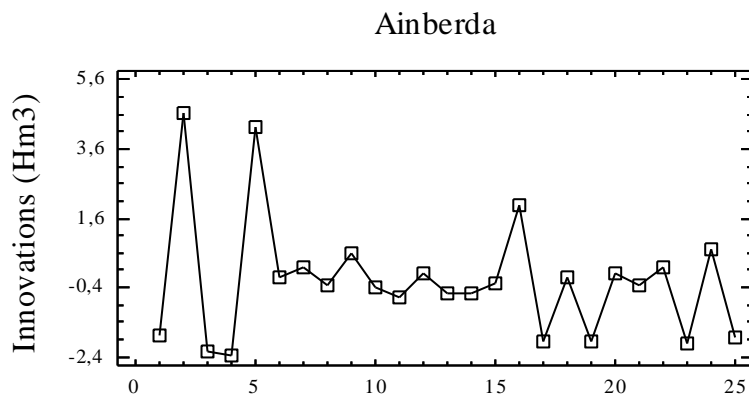


Fig.III.2.38 Innovations du FK à AINBERDA (1968-92)

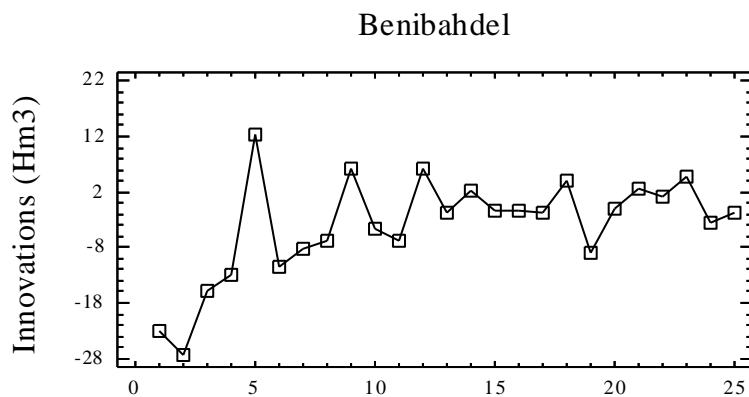


Fig.III.2.39 Innovations du FK à BENIBAHDEL (1968-92)

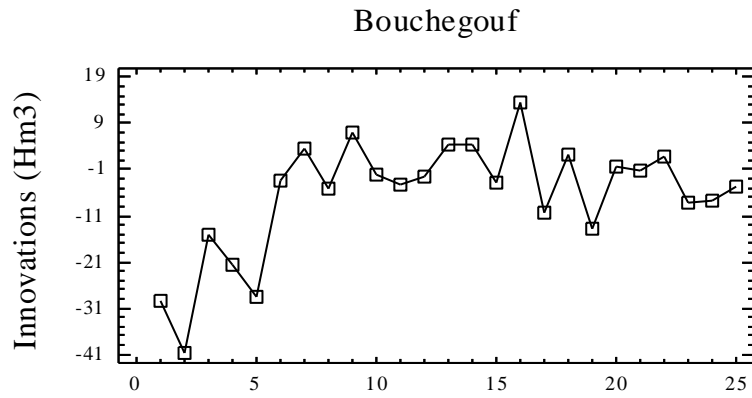


Fig.III.2.40 Innovations du FK à BOUCHEGOUF (1968-92)

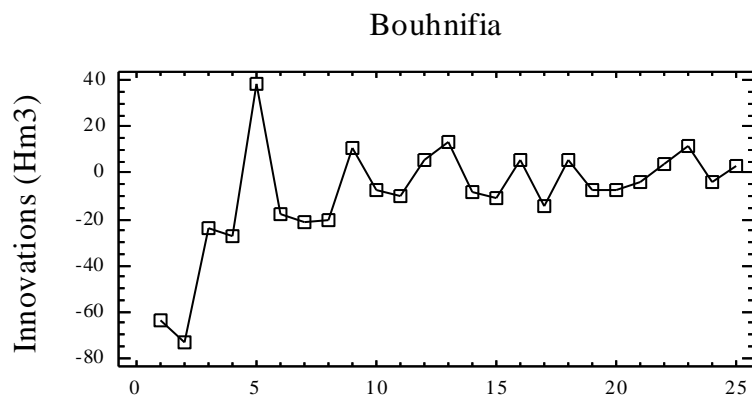


Fig.III.2.41 Innovations du FK à BOUHNIFIA (1968-92)

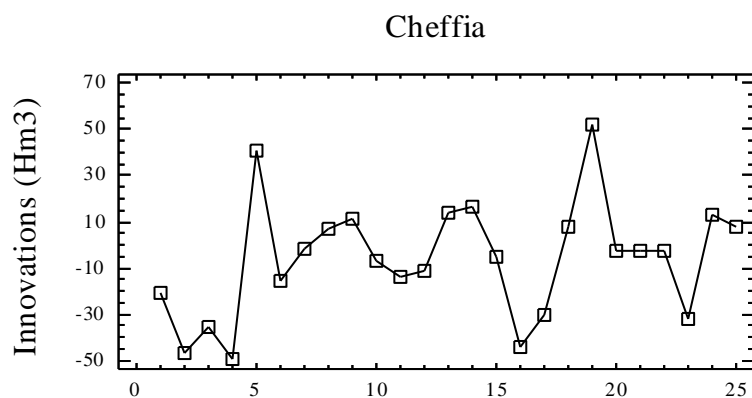


Fig.III.2.42 Innovations du FK à CHEFFIA (1968-92)

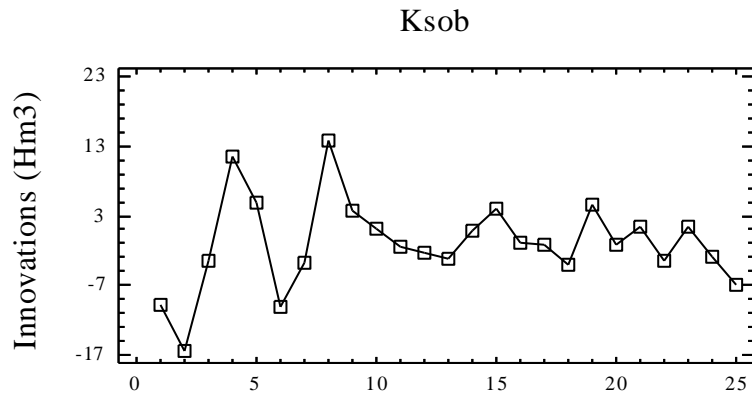


Fig.III.2.43 Innovations du FK à KSOB (1968-92)

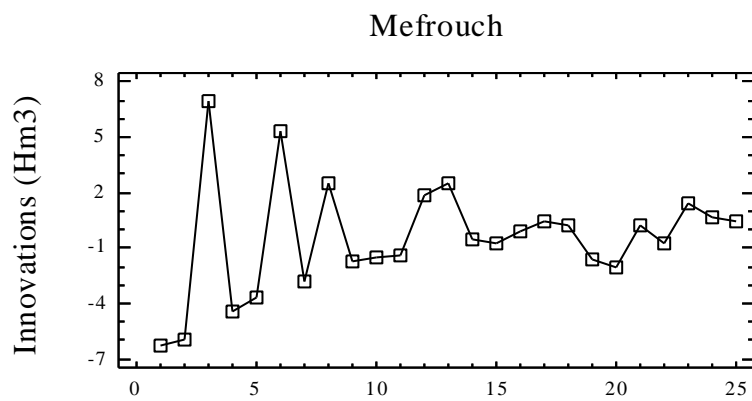


Fig.III.2.44 Innovations du FK à MEFROUCH (1968-92)

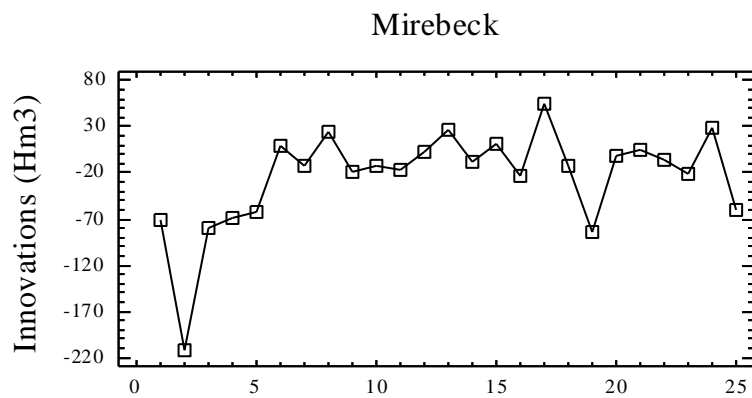


Fig.III.2.45 Innovations du FK à MIREBECK (1968-92)

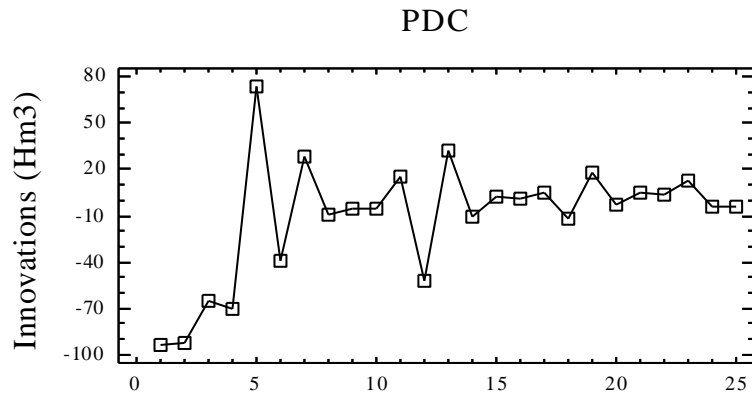


Fig.III.2.46 Innovations du FK à PIERRE DU CHAT (1968-92)

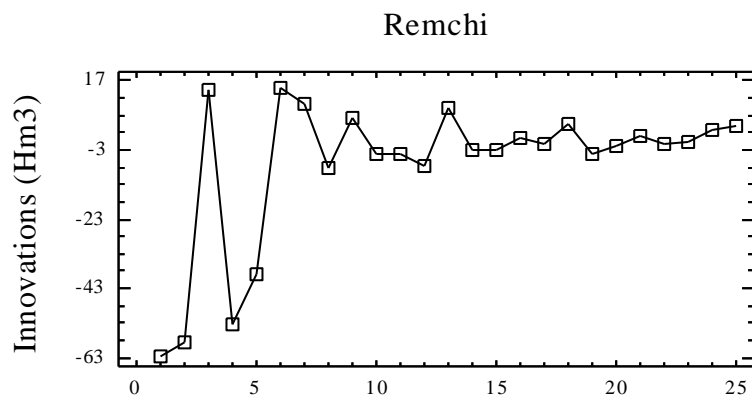


Fig.III.2.47 Innovations du FK à REMCHI (1968-92)

L'analyse de toutes ces innovations est effectuée afin de vérifier les hypothèses du filtre de Kalman concernant l'indépendance et la Normalité. Ces analyses sont effectuées uniquement sur la dimension temporelle, car d'un point de vue nombre d'observations, elle est plus intéressante.

- **L'indépendance**

Pour les prédictions temporelles, les différences entre les apports observés et leurs prédictions pour toutes les stations concernées ont été testées quant à l'indépendance. Ceci est fait moyennant l'examen visuel du tracé de la fonction d'auto-corrélation à 99% de confiance, et le test d'indépendance de Box-Pierce. Les Fig.III.2.48-57 montrent bien le tracé de cette fonction d'auto-corrélation des innovations pour les dix stations considérées.

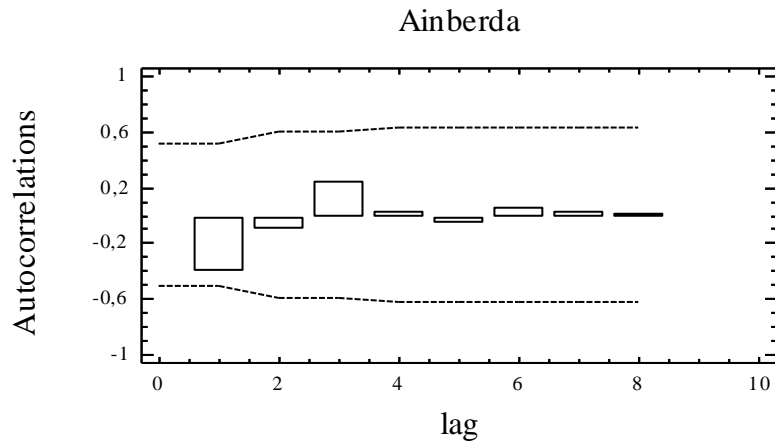


Fig.III.2.48 Fonction d'auto-corrélation des innovations à AINBERDA

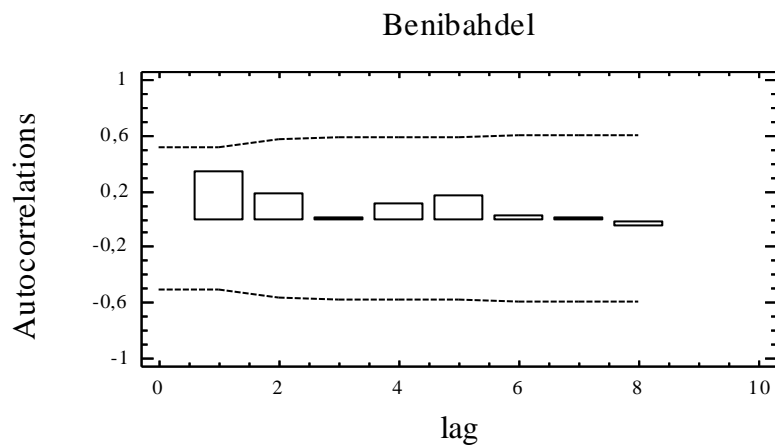


Fig.III.2.49 Fonction d'auto-corrélation des innovations à BENIBAHDEL

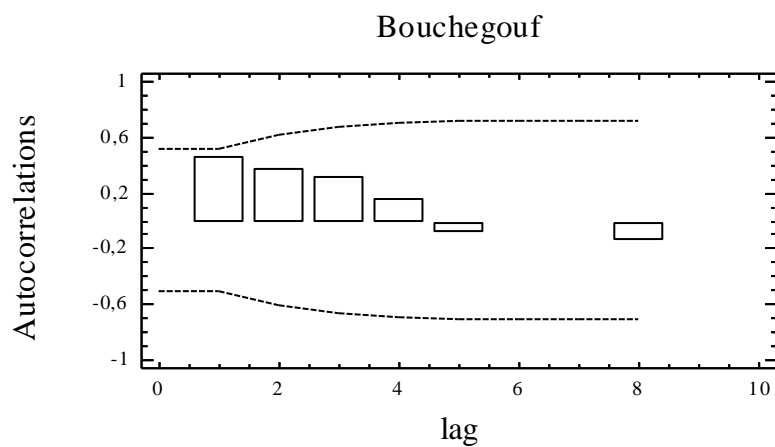


Fig.III.2.50 Fonction d'auto-corrélation des innovations à BOUCHEGOUF

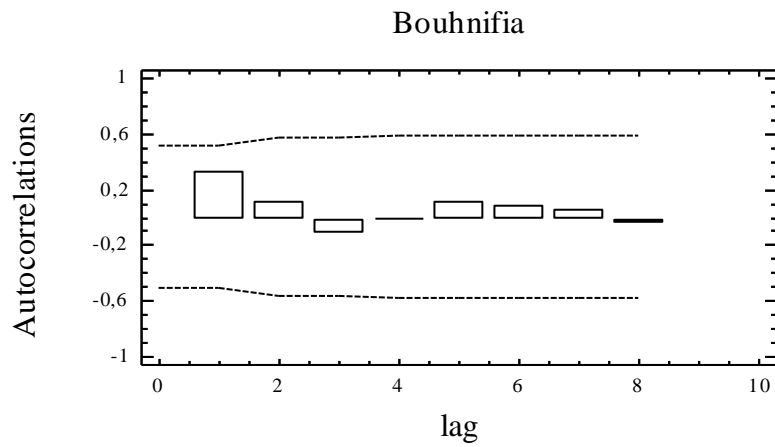


Fig.III.2.51 Fonction d'auto-corrélation des innovations à BOUHNIFIA

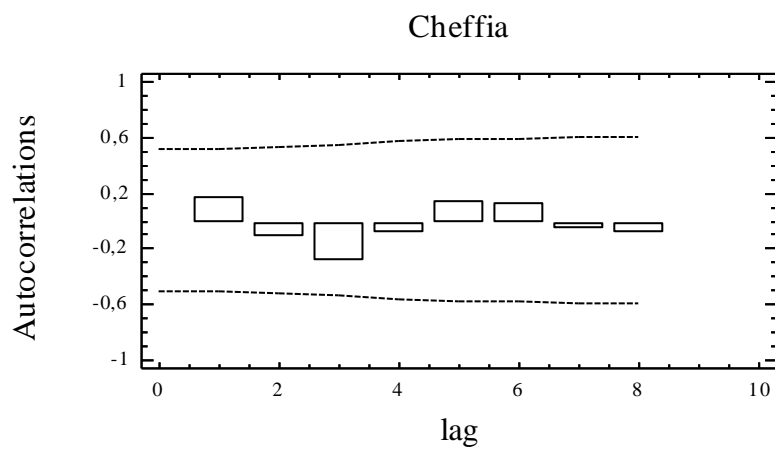


Fig.III.2.52 Fonction d'auto-corrélation des innovations à CHEFFIA

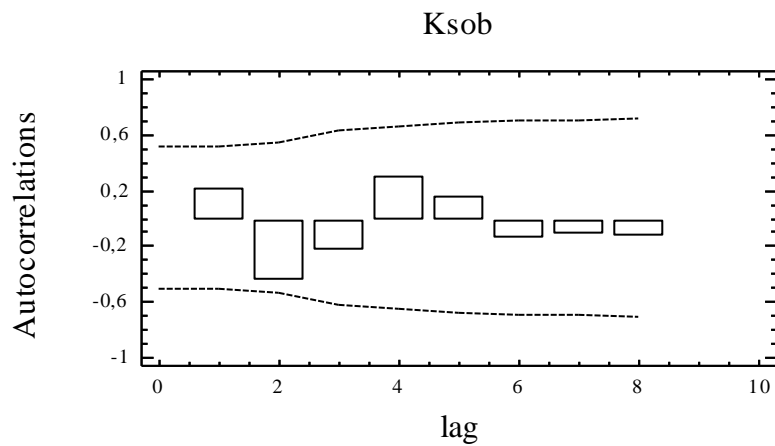


Fig.III.2.53 Fonction d'auto-corrélation des innovations à KSOB

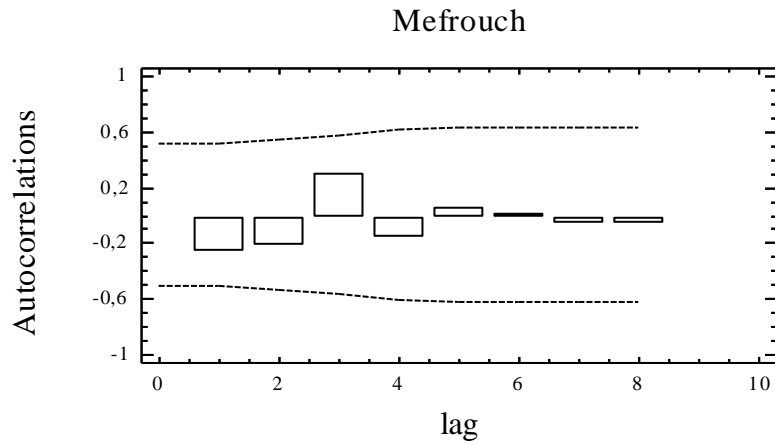


Fig.III.2.54 Fonction d'auto-corrélation des innovations à MEFROUCHE

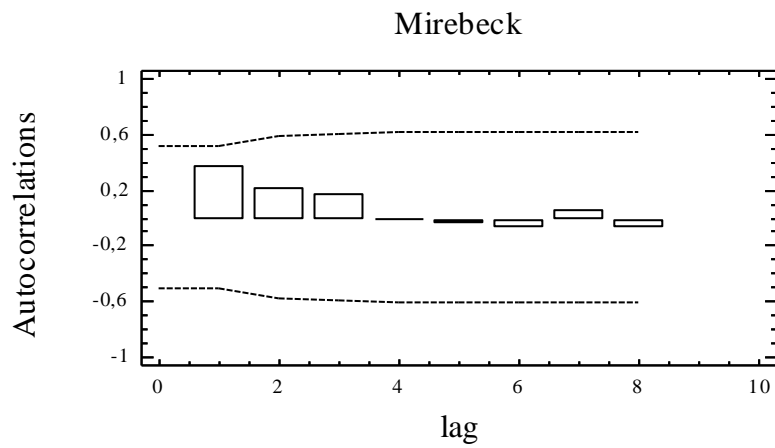


Fig.III.2.55 Fonction d'auto-corrélation des innovations à MIREBECK

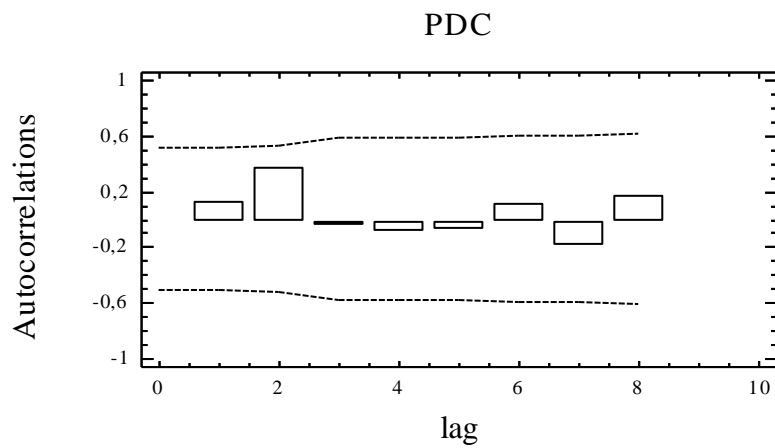


Fig.III.2.56 Fonction d'auto-corrélation des innovations à PIERRE DU CHAT

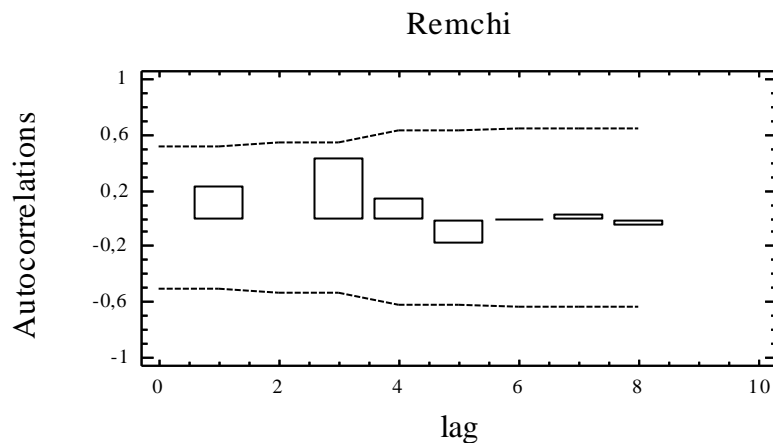


Fig.III.2.57 Fonction d'auto-corrélation des innovations à PIERRE DU CHAT

L'inspection visuelle des figures précédentes montre que ces innovations ne présentent pas d'auto-corrélations significatives, ce qui affirme l'indépendance des innovations du filtre pour la totalité des stations considérées. Les valeurs des auto-corrélations sont représentées dans les Tableaux III.2.4-13 suivants.

Tableau III.2.4 Auto-corrélations des innovations à AINBERDA

Lag	ACF	Ecart-type	-IC à 99%	+IC à 99%
1	-0,418671	0,2	-0,515167	0,515167
2	-0,113332	0,232428	-0,598696	0,598696
3	0,238853	0,234628	-0,604363	0,604363
4	0,0218999	0,244161	-0,628917	0,628917
5	-0,0642666	0,244239	-0,62912	0,62912
6	0,0439271	0,244915	-0,63086	0,63086
7	0,0285164	0,24523	-0,631671	0,631671
8	0,0105733	0,245362	-0,632012	0,632012

Tableau III.2.5 Auto-corrélations des innovations à BENIBAHDEL

Lag	ACF	Ecart-type	-IC à 99%	+IC à 99%
1	0,336535	0,2	-0,515167	0,515167
2	0,181222	0,221496	-0,570537	0,570537
3	0,00788252	0,227349	-0,585615	0,585615
4	0,108478	0,22736	-0,585643	0,585643
5	0,159242	0,229421	-0,590951	0,590951
6	0,0171814	0,233801	-0,602232	0,602232
7	0,00433155	0,233851	-0,602362	0,602362
8	-0,0632094	0,233854	-0,60237	0,60237

Tableau III.2.6 Auto-corrélations des innovations à BOUCHEGOUF

Lag	ACF	Ecart-type	-IC à 99%	+IC à 99%
1	0,451188	0,2	-0,515167	0,515167
2	0,36711	0,237246	-0,611106	0,611106
3	0,304356	0,258973	-0,667073	0,667073
4	0,145755	0,272906	-0,702961	0,702961
5	-0,0957523	0,276003	-0,710937	0,710937
6	-0,0180365	0,277328	-0,714351	0,714351
7	-0,0284582	0,277375	-0,714472	0,714472
8	-0,153512	0,277492	-0,714773	0,714773

Tableau III.2.7 Auto-corrélations des innovations à BOUHNIFIA

Lag	ACF	Ecart-type	-IC à 99%	+IC à 99%
1	0,332234	0,2	-0,515167	0,515167
2	0,10754	0,220976	-0,569197	0,569197
3	-0,116624	0,22306	-0,574564	0,574564
4	-0,00398247	0,225485	-0,580813	0,580813
5	0,104831	0,225488	-0,58082	0,58082
6	0,0820603	0,227429	-0,58582	0,58582
7	0,0495428	0,228611	-0,588863	0,588863
8	-0,0380727	0,22904	-0,589968	0,589968

Tableau III.2.8 Auto-corrélations des innovations à CHEFFIA

Lag	ACF	Ecart-type	-IC à 99%	+IC à 99%
1	0,16338	0,2	-0,515167	0,515167
2	-0,124885	0,205269	-0,528739	0,528739
3	-0,294171	0,208286	-0,536511	0,536511
4	-0,0937756	0,22429	-0,577734	0,577734
5	0,140485	0,225853	-0,58176	0,58176
6	0,1296	0,229322	-0,590695	0,590695
7	-0,0692658	0,232233	-0,598194	0,598194
8	-0,0923855	0,233058	-0,600318	0,600318

Tableau III.2.9 Auto-corrélations des innovations à KSOB

Lag	ACF	Ecart-type	-IC à 99%	+IC à 99%
1	0,212432	0,2	-0,515167	0,515167
2	-0,451763	0,208831	-0,537913	0,537913
3	-0,232681	0,244821	-0,630619	0,630619
4	0,295723	0,253513	-0,653006	0,653006
5	0,154993	0,266955	-0,687631	0,687631
6	-0,148519	0,27053	-0,696841	0,696841
7	-0,128051	0,273772	-0,705192	0,705192
8	-0,138966	0,276158	-0,711336	0,711336

Tableau III.2.10 Auto-corrélations des innovations à MEFROUCHE

Lag	ACF	Ecart-type	-IC à 99%	+IC à 99%
1	-0,260862	0,2	-0,515167	0,515167
2	-0,222417	0,213176	-0,549105	0,549105
3	0,295363	0,222264	-0,572516	0,572516
4	-0,170385	0,237446	-0,611621	0,611621
5	0,0461111	0,242287	-0,624092	0,624092
6	0,000196965	0,242638	-0,624995	0,624995
7	-0,0591262	0,242638	-0,624995	0,624995
8	-0,0587133	0,243214	-0,626478	0,626478

Tableau III.2.11 Auto-corrélations des innovations à MIREBECK

Lag	ACF	Ecart-type	-IC à 99%	+IC à 99%
1	0,364161	0,2	-0,515167	0,515167
2	0,209091	0,224965	-0,579472	0,579472
3	0,163701	0,232608	-0,59916	0,59916
4	-0,008512	0,237172	-0,610915	0,610915
5	-0,0527507	0,237184	-0,610946	0,610946
6	-0,0832641	0,237653	-0,612154	0,612154
7	0,0539376	0,238817	-0,615152	0,615152
8	-0,0746573	0,239304	-0,616406	0,616406

Tableau III.2.12 Auto-corrélations des innovations à PIERRE DU CHAT

Lag	ACF	Ecart-type	-IC à 99%	+IC à 99%
1	0,127876	0,2	-0,515167	0,515167
2	0,367217	0,203244	-0,523523	0,523523
3	-0,0360002	0,228246	-0,587923	0,587923
4	-0,0950441	0,228473	-0,588507	0,588507
5	-0,07259	0,230049	-0,592567	0,592567
6	0,102844	0,230963	-0,594922	0,594922
7	-0,19955	0,232788	-0,599622	0,599622
8	0,165526	0,239532	-0,616995	0,616995

Tableau III.2.13 Auto-corrélations des innovations à REMCHI

Lag	ACF	Ecart-type	-IC à 99%	+IC à 99%
1	0,223004	0,2	-0,515167	0,515167
2	-0,0181704	0,20971	-0,540179	0,540179
3	0,431989	0,209773	-0,540341	0,540341
4	0,134165	0,242763	-0,625318	0,625318
5	-0,197368	0,245711	-0,632911	0,632911
6	-0,0129124	0,251973	-0,64904	0,64904
7	0,0266216	0,251999	-0,649109	0,649109
8	-0,0646887	0,252112	-0,649398	0,649398

Ces résultats montrent tous des variations faibles et non significatives des auto-corrélations qui tendent vers zéro, ce qui affirme l'indépendance des innovations du filtre, pour les stations hydrométriques en question. Les résultats du test de Box-Pierce tel que c'est indiqué par le Tableau III.2.14 viennent confirmer cette indépendance. Nous rappelons que ce test est basé sur la somme des carrés des N premières auto-corrélations de la série de données et que si la probabilité P est supérieure ou égale à 0,10 l'on ne peut pas rejeter l'hypothèse que la série en question est aléatoire ou indépendante à 90% de confiance ou plus. C'est le cas de toutes les séries des innovations étudiées, d'où la confirmation de leur indépendance à 90% et plus de confiance.

Tableau III.2.14 Test de Box-Pierce pour les innovations du filtre.

Station hydrométrique	N	Statistique du test	P
Ainberda	8	6,31	0,61
Bénibahdel	8	4,68	0,79
Bouchegouf	8	12,15	0,14
Bouhnifia	8	3,92	0,86
Cheffia	8	4,68	0,79
Ksob	8	11,81	0,16
Mefrouche	8	6,07	0,63
Mirebeck	8	5,53	0,69
Pierre du chat	8	6,11	0,63
Remchi	8	7,46	0,48

Clé: N: nombre des premières auto-corrélations; P: probabilité.

• La Normalité

La Normalité des innovations (**Anderson, 1942**) a été testée uniquement dans la dimension temporelle, pour la simple raison qu'un tel ajustement n'aurait pratiquement aucun sens s'il est effectué sur une dizaine de données; ce qui est le cas de la dimension spatiale.

Le tracé de ces ajustements Normaux aux innovations des différentes stations considérées de 1968 à 1992 a été effectué à 95% de confiance et les Fig. III.2.58-67 montrent la qualité de l'adéquation de ces ajustements pour les différentes stations. Il faut noter que pour les stations AINBERDA et REMCHI l'ajustement à la loi normale ne semble pas être adéquat. L'une des raisons les plus plausibles de ce défaut de normalité pourrait être les premières valeurs qui sont dues à l'adaptation du filtre et qui sont inévitablement mauvaises au début des calculs.

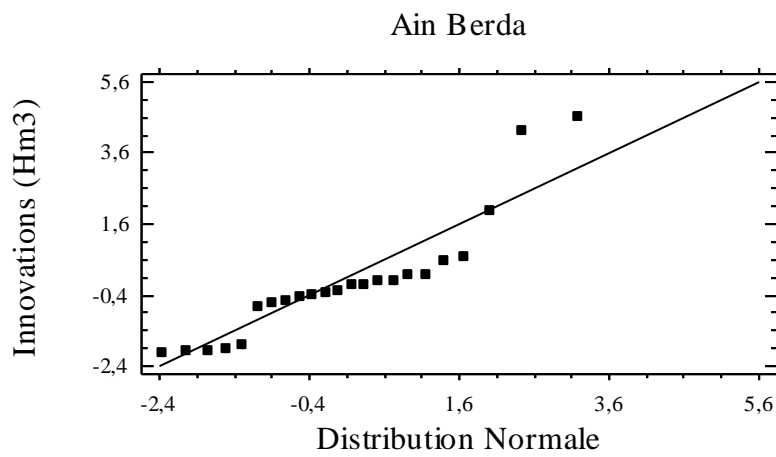


Fig.III.2.58 Ajustement Normal aux innovations de AINBERDA

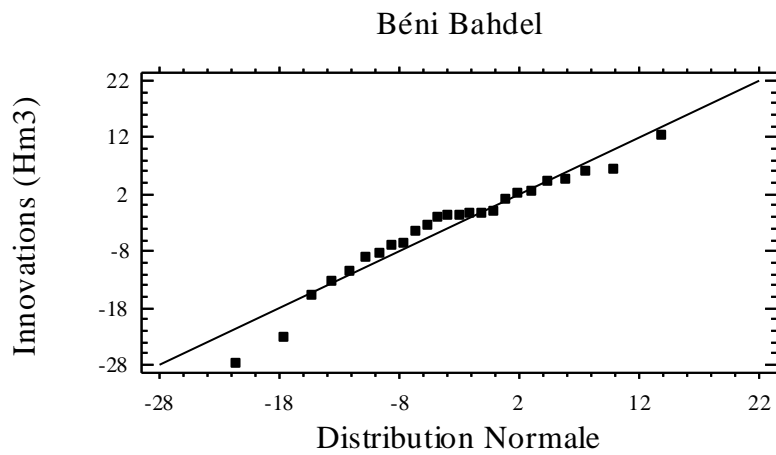


Fig.III.2.59 Ajustement Normal aux innovations de BENIBAHDEL

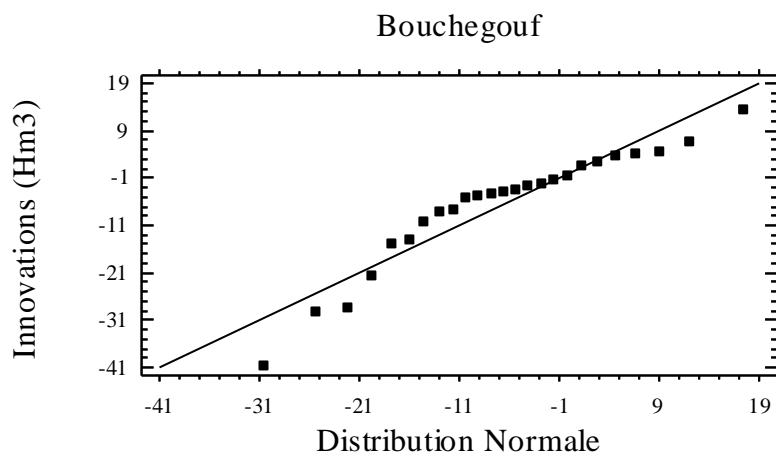


Fig.III.2.60 Ajustement Normal aux innovations de BOUCHEGOUF

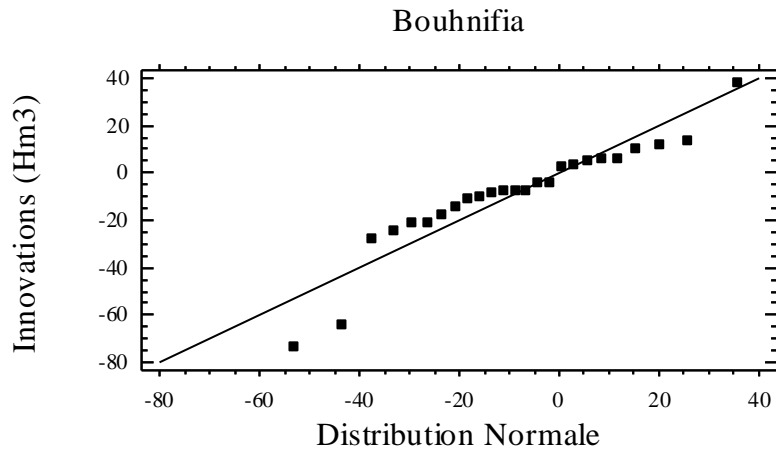


Fig.III.2.61 Ajustement Normal aux innovations de BOUHNIFIA

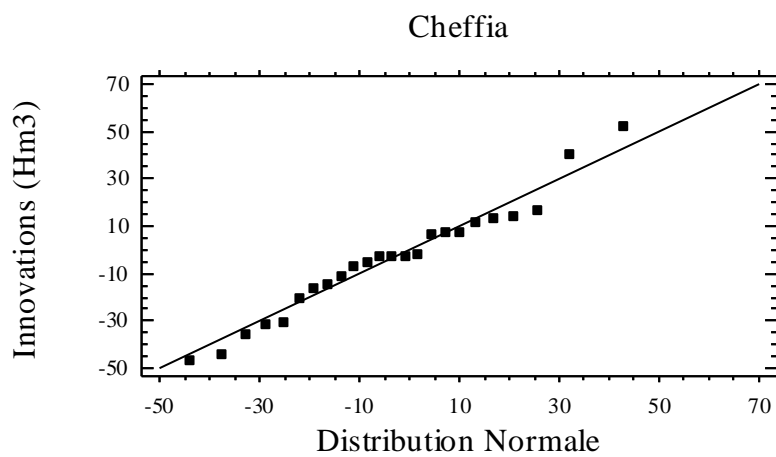


Fig.III.2.62 Ajustement Normal aux innovations de CHEFFIA

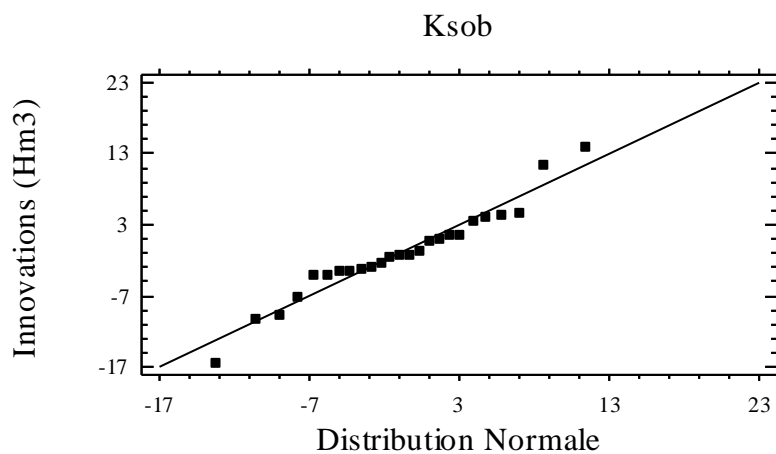


Fig.III.2.63 Ajustement Normal aux innovations de KSOB

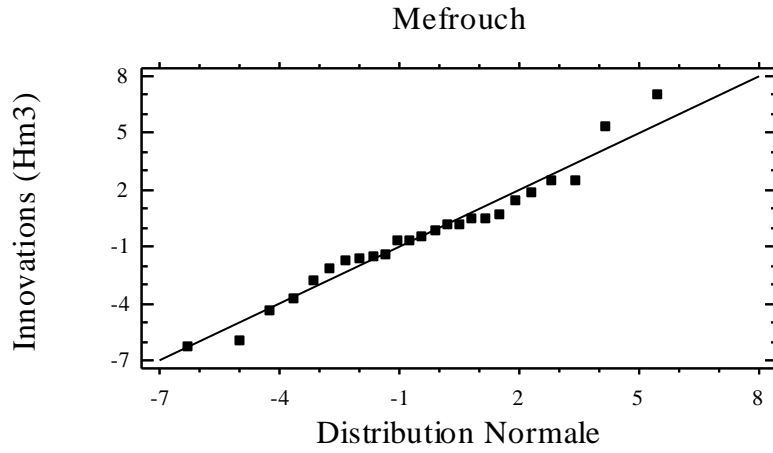


Fig.III.2.64 Ajustement Normal aux innovations de MEFROUCH

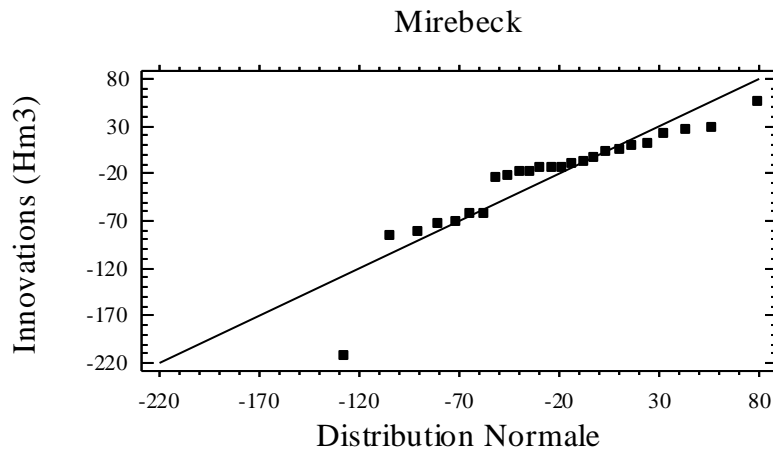


Fig.III.2.65 Ajustement Normal aux innovations de MIREBECK

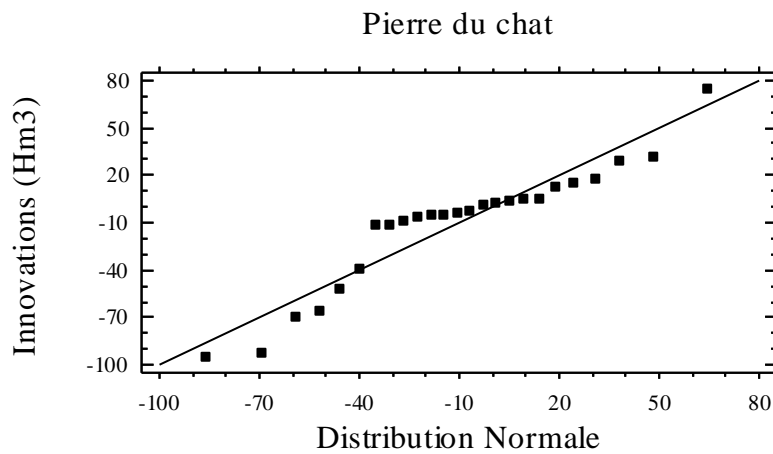


Fig.III.2.66 Ajustement Normal aux innovations de PIERRE DU CHAT

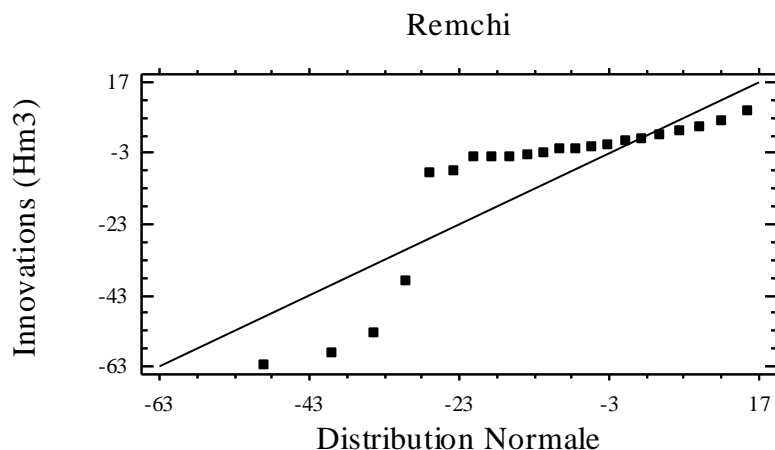


Fig.III.2.67 Ajustement Normal aux innovations de REMCHI

La qualité de l'adéquation de ces ajustements Normaux aux séries des innovations a été vérifiée par le test du Chi-deux à 95 % de confiance dont les résultats sont présentés dans le Tableau III.2.15. Ces derniers montrent effectivement qu'à l'exception des stations AINBERDA et REMCHI qui affichent une valeur de probabilité P inférieure à 0,05 toutes les autres stations présentent des innovations gaussiennes à 95 % de confiance.

Tableau III.2.15 Test d'adéquation du Chi-deux de l'ajustement normal à 95% de confiance des innovations du filtre.

Station hydrométrique	μ	σ	Statistique du test	P
Ainberda	-0,195	1,75	21,48	0,028
Ainberda(*)	-0,37	0,99	15,1	0,12
Bénibahdel	-3,92	9,16	13,64	0,25
Bouchegouf	-6,63	12,45	12,52	0,32
Bouhnifia	-8,94	23	10,28	0,50
Cheffia	-6,01	25,25	12,52	0,32
Ksob	-0,98	6,39	10,28	0,50
Mefrouche	-0,42	3,04	11,40	0,41
Mirebeck	-24,35	53,60	19,24	0,05
Pierre du chat	-10,65	39,03	14,76	0,19
Remchi	-7,44	21,63	29,32	0,002
Remchi (*)	0,65	6,01	7,3	0,69

Clée: μ : moyenne de la loi normale ajustée; σ : écart-type de la loi normale ajustée ; P: probabilité (*) : tronquée des 5 premières valeurs.

Dans le but de déterminer la cause de ce défaut de normalité afin de le corriger, nous avons procédé à l'ajustement normal de la série des innovations de AINBERDA et REMCHI en écartant les 5 premières valeurs aberrantes. Les Fig.III.2.68 et 69 confirment bien notre thèse en montrant la qualité de cet ajustement aux nouvelles séries d'innovations et le résultat du test du Chi-deux vient à son tour confirmer l'adéquation de ce dernier ce qui nous permet de confirmer la normalité des innovations pour l'ensemble des stations étudiées.

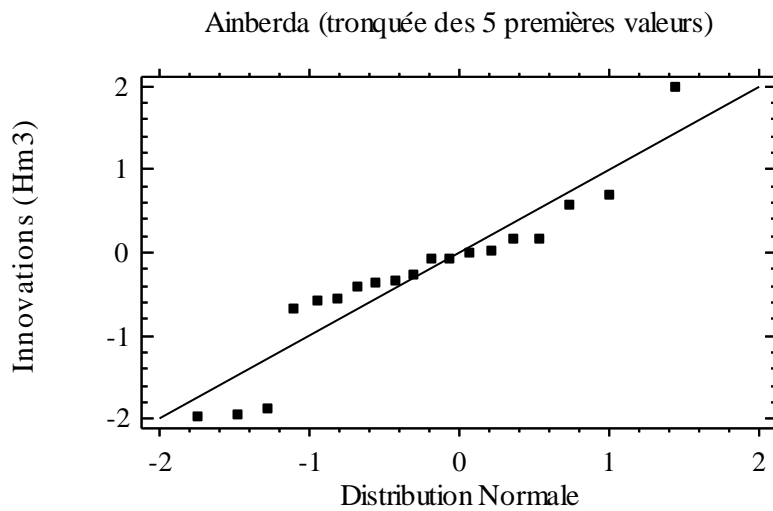


Fig. III.2.68 Ajustement Normal aux innovations de AINBERDA tronquées des 5 premières valeurs

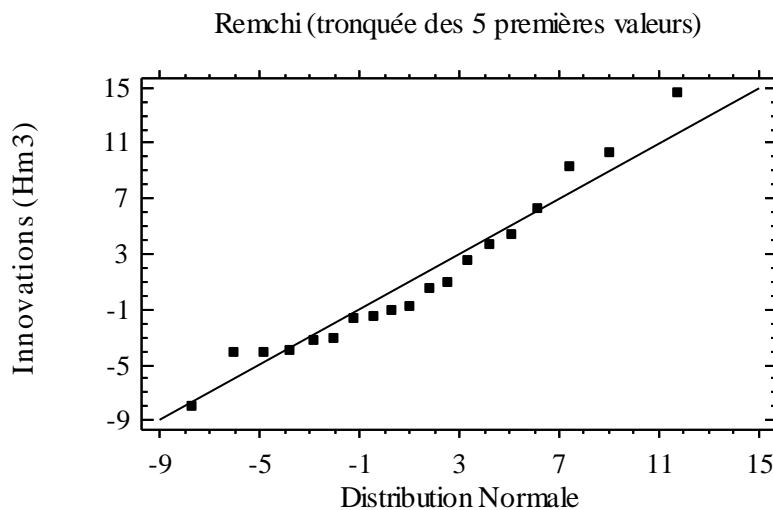


Fig.III.2.69 Ajustement Normal aux innovations de REMCHI tronquées des 5 premières valeurs

- **L'erreur relative des prédictions**

L'erreur relative des prédictions en pourcentage a également été calculée pour la totalité des stations hydrométriques et ce durant la période de 25 ans (Tableau III.2.3). A l'exception des premières années d'estimation, correspondant aux premières itérations où les valeurs relativement élevées de l'erreur relative sont dues à l'adaptation du (FK), le pourcentage d'erreur relative comme c'est montré à titre d'exemple dans la Fig. III.2.35 pour 1992, varie de -12 à la station MIREBECK et BÉNI-BAHDEL qui présente une sous-estimation à $+12$ observé à BOUHNIFIA et qui présente quant à lui une surestimation, ce qui représente en moyenne pour l'année 1992 un pourcentage de l'erreur relative de 9.8%.

De même pour la dimension temporelle, le pourcentage de l'erreur relative des prédictions à BOUCHEGOUF, tel que montré par la Fig. III.2.3, varie de - 75 en 1968 où les prédictions sont automatiquement mauvaises, à +7 en 1983 et 1985 où le (FK) avait pratiquement convergé, pour atteindre la valeur de -4 en 1992. La valeur moyenne de ce pourcentage d'erreur relative sur les 25 ans est de 10.32 alors que si l'on écarte les trois premières années où le filtre cherchait à s'adapter, nous aurons une moyenne de 5.52 %.

III.2.7.4 Moyenne et déviation standard des prédictions multi-sites

En fin, la moyenne et la déviation standard des apports liquides annuels observés et leurs prédictions, pour toutes les stations considérées ont été calculés dans les deux dimensions temporelle et spatiale, les résultats sont présentés dans les tableaux Tableau III.2.16 et Tableau III.2.17 suivants :

Tableau III.2.16 Moyenne, déviation standard et erreur relative des valeurs observés et prédites des apports liquides annuels (1968-1992).

No	Code	Station	Valeurs observées		Valeurs prédites		Erreur relative (%)	
			M (Hm3)	Dev.St (Hm3)	M (Hm3)	Dev.St (Hm3)	M	Dev.St
01	140602	Ain Berda	12.32	11.04	12.12	11.36	4.24	7.45
02	160403	Béni bahdel	52.54	42.82	44.64	41.71	2.43	7.42
03	140501	Boucheougouf	101.73	65.38	95.09	62.30	13.56	16.94
04	111504	Bouhnifia	90.82	60.01	81.87	64.14	1.43	8.40
05	31501	Cheffia	139.86	95.21	127.82	96.94	2.66	6.26
06	050902	Ksob	31.85	22.41	30.85	24	1.08	1.38
07	160701	Mefrouch	11.33	7.37	10.37	8.17	1.12	10.60
08	140601	Mirebeck	331.48	238.20	306.69	225.21	2.47	8.52
09	160801	Pierre du Chat	167.85	153.13	155.71	158.01	6.53	18.47
10	160703	Remchi	73.6	64.02	66.15	57.87	1.06	26.05
Moyenne			101.34	75.95	100.11	70.03	3.66	11.15

Tableau III.2.17 Moyenne et déviation standard des valeurs observées et prédites des apports liquides annuels aux 10 stations hydrométriques (1968-1992).

Année	Valeurs Observées		Valeurs prédites		Année	Valeurs Observées		Valeurs Prédites	
	M (Hm3)	Dev.St (Hm3)	M (Hm3)	Dev.St (Hm3)		M (Hm3)	Dev.St (Hm3)	M (Hm3)	Dev.St (Hm3)
1968	64.53	46.18	26.25	17.13	1981	55.30	54.11	54.50	54.29
1969	154.51	164.51	97.6	105.3	1982	51.59	51.78	50.96	54.60
1970	124.12	130.68	100.84	106.6	1983	102.97	153.46	98.34	143.75
1971	202.52	167.37	171.42	139.5	1984	157.75	282.92	157.76	299.47
1972	241.32	197.15	245.53	199.2	1985	45.86	35.99	45.56	32.66
1973	150.20	160.59	143.4	150.2	1986	177.67	237.73	169.27	220.08
1974	133.08	137.80	117.29	150.2	1987	22.82	15.78	20.65	13.30
1975	97.08	78.39	96.79	82.14	1988	26.06	22.41	26.87	24.55
1976	102.60	105.30	104.64	99.49	1989	35.85	36.46	35.49	35.10
1977	67.68	63.23	63.27	59.67	1990	120.57	160.02	117.44	151.98
1978	74.06	68.53	69.58	65.47	1991	76.36	88.95	78.78	98.08
1979	73.09	82.53	67.13	67.66	1992	86.12	151.87	79.48	133.48
1980	89.71	87.84	99.42	98.53	Moyenne	101.34	111.26	93.53	68.63

Ces résultats montrent encore une fois le rapprochement entre les valeurs observées et leurs prédictions, rapprochement qui est exprimé en termes de tendance centrale et de dispersion autour de cette tendance. En effet, en ce qui concerne la tendance centrale, les deux Tableaux précédents montrent les résultats suivants : pour les prédictions temporelles la moyenne (calculée sur les dix stations) des moyennes observées dans chacune de ces stations (calculée sur les 25 ans) et qui est de 101,34 est très proche de la moyenne des moyennes prédites qui a une valeur de 100,11 ; de même pour les prédictions spatiales où la valeur moyenne (calculée sur les 25 ans) des moyennes annuelles observées (calculées sur les dix stations) est de 101,34 qui est également proche de celle des valeurs prédites qui est de 93,53. En ce qui concerne la dispersion autour de cette moyenne les mêmes tableaux affichent les résultats suivants : pour les prédictions temporelles la moyenne (calculée sur les dix stations) des déviations standards observées dans chacune de ces stations (calculée sur les 25 ans) et qui est de 75,95 est proche de la moyenne des déviations standards prédites qui a une valeur de 70,03 mais elle est supérieure ; alors que pour les prédictions spatiales, la valeur moyenne (calculée sur les 25 ans) des déviations standards annuelles observées (calculées sur les dix stations) qui est de 111,26 est nettement supérieure à celle des valeurs prédites qui est de 68,63. En plus, l'erreur relative moyenne commise au sujet de la moyenne étant de 3,66 comparativement à une valeur de 11,15 au sujet de la déviation standard entre valeurs observées et prédites pour les prédictions temporelles.

Ces résultats expriment clairement que les déviations standards des valeurs observées sont plus grandes que celles des valeurs prédites, ce qui pourrait traduire une éventuelle tendance du (FK) multi-site à une sous-estimation.

III.2.8 COMPARAISON DES RESULTATS DE PREDICTION AVEC CEUX DES MODELES ARMA

Dans le but d'apprécier et de mettre en valeur la qualité des résultats de prédiction obtenus par le biais du FK, une comparaison avec ceux obtenus par les modèles ARMA s'impose. Les erreurs de prédiction constituent l'objet principal de cette comparaison. A cet effet, rappelons que pour des raisons inhérentes à la structure de dépendance interne sur laquelle repose la méthodologie ARMA, uniquement cinq parmi les dix stations étudiées ont fait l'objet d'un ajustement aux modèles ARMA. D'un autre côté, il faut également rappeler que la prédiction par le FK a été effectuée sur une période commune de 25 ans allant de 1968 à 1992 et que pour des raisons d'adaptation du filtre nous avons omis les 5 premières valeurs des erreurs de prédiction qui sont automatiquement mauvaises, ce qui nous laisse en fin de compte, 5 stations hydrométriques et 20 erreurs de prédiction (de 1973 à 1992) pour la comparaison des résultats.

Le Tableau III.2.18 résume cette comparaison en termes de caractéristiques statistiques de tendance centrale et de dispersion des erreurs de prédiction : moyenne, variance, déviation standard, minimum et maximum. Nous pouvons constater, pour chacune des caractéristiques statistiques évoquées plus haut, l'importance de l'écart existant entre les valeurs obtenues par le FK et celles obtenues par les modèles ARMA :

- **la moyenne** : les valeurs obtenues par le FK sont beaucoup plus petites que celles obtenues par les modèles ARMA et ce pour les cinq stations (Fig.III.2.70). Si l'on calcule la moyenne de cette moyenne sur les cinq stations on se retrouve avec une valeur de 1,30 pour le FK qui est 16 fois plus petite que la valeur de 21,44 pour les modèles ARMA.
- **la variance** : les valeurs obtenues par le FK sont aussi beaucoup plus petites que celles obtenues par les modèles ARMA et ce pour les cinq stations (Fig.III.2.71). La variance

moyenne qui est de 40,86 pour le FK est donc 23 fois plus petite que la valeur de 968,48 pour ARMA.

- **la déviation standard** : les valeurs obtenues par le FK sont également beaucoup plus petites que celles obtenues par les modèles ARMA et ce pour les cinq stations (Fig.III.2.72). La déviation standard moyenne est de 5,73 pour le FK donc presque 5 fois plus petite que la valeur de 27,06 pour les modèles ARMA.
- **le minimum** : les valeurs obtenues par le FK sont également beaucoup plus petites que celles obtenues par les modèles ARMA et ce pour les cinq stations (Fig.III.2.73) . Le minimum moyen est de 10,82 pour le FK donc presque 6 fois plus petit que la valeur de 64,26 pour les modèles ARMA.
- **le maximum** : les valeurs obtenues par le FK sont nettement plus petites que celles obtenues par les modèles ARMA et ce pour les cinq stations (Fig.III.2.74). Le maximum moyen est de 10,74 pour le FK donc presque 3 fois plus petit que la valeur de 31,88 pour les modèles ARMA.

Tableau III.2.18 Comparaison des erreurs de prédiction entre Filtre de Kalman (FK) et modèles ARMA

Année	Bénibahdel		Bouhniafia		Ksob		Mefrouch		Remchi	
	FK	ARMA	FK	ARMA	FK	ARMA	FK	ARMA	FK	ARMA
1973	-11.46	9,77	-17.82	-49,37	-10.24	-0,39	5.34	8,97	14.63	66,65
1974	-8.23	-1,87	-21.19	-87,81	-3.91	-7,48	-2.78	-13,08	10.34	-27,18
1975	-6.74	-26,73	-20.55	4,43	13.85	35,01	2.54	4,41	-8.43	-60,28
1976	6.35	-25,48	10.56	38,76	3.63	4,88	-1.72	-13,27	6.28	-88,28
1977	-4.56	-26,24	-7.28	-91,44	1.09	-59,27	-1.55	-5,26	-3.98	-61,56
1978	-6.67	-1,61	-10.01	-80,96	-1.61	-17,04	-1.43	-8,91	-4.03	-35,45
1979	6.11	2,62	5.34	-3,42	-2.40	-5,02	1.85	1,05	-7.92	39,17
1980	-1.55	-53,81	13.56	13,77	-3.20	-6,26	2.51	3,27	9.30	14,44
1981	2.29	-24,67	-8.35	-95,84	0.80	-14,07	-0.50	-8,96	-3.24	-35,95
1982	-1.30	-45,30	-10.72	-51,60	4.09	4,12	-0.71	-11,59	-3.02	-67,38
1983	-1.38	-33,25	5.76	-49,27	-0.76	-12,48	-0.10	-18,87	0.57	-86,11
1984	-1.63	-36,89	-14.28	31,92	-1.15	-18,14	0.50	-14,58	-1.48	-45,15
1985	4.21	-10,38	5.79	-101,08	-4.02	9,04	0.18	-9,35	4.49	1,03
1986	-8.93	-4,84	-6.98	-71,61	4.51	13,10	-1.56	4,75	-4.09	25,55
1987	-0.92	-48,20	-7.40	-46,49	-1.14	-14,03	-2.09	5,83	-1.67	-35,12
1988	2.62	-19,88	-3.85	-27,37	1.51	-18,58	0.21	-8,40	1.03	-49,18
1989	1.15	-40,21	4.01	-91,00	-3.54	22,56	-0.71	-8,69	-1.09	-74,01
1990	4.64	10,03	11.95	27,62	1.46	-2,13	1.44	-6,89	-0.76	-49,25
1991	-3.43	-26,46	-4.34	-81,72	-3.04	-4,01	0.71	-8,96	2.54	-28,81
1992	-1.78	-27,05	2.64	-84,95	-7.04	-9,94	0.48	-2,50	3.65	-8,63
Moyenne	-1,55	-21,52	-3,65	-44,87	-0,55	-5,00	0,13	-5,55	0,65	-30,27
Variance	26,08	364,09	113,50	2263,5	24,81	357,65	3,72	60,88	36,23	1761,32
Dév.Std	5,10	19,08	10,65	47,57	4,98	18,91	1,92	7,80	6,01	41,96
Minimum	-11,46	-53,81	-21,19	-101,08	-10,24	-59,27	-2,78	-18,87	-8,43	-88,28
Maximum	6,35	10,03	13,56	38,76	13,85	35,01	5,34	8,97	14,63	66,65

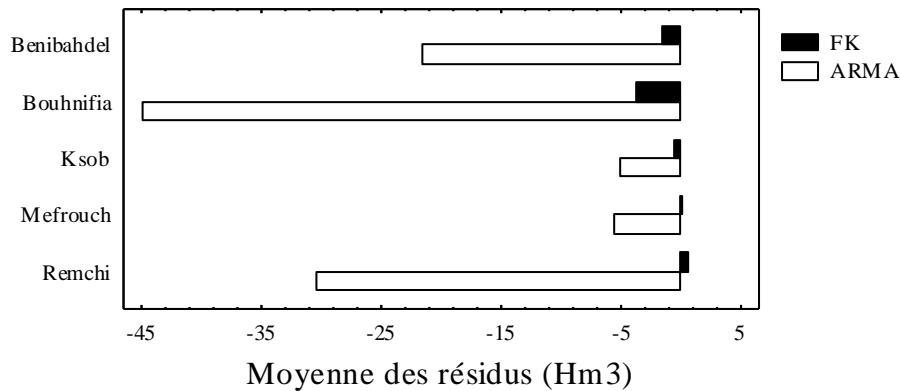


Fig.III.2.70 Comparaison de la moyenne des erreurs de prédiction entre le Filtre de Kalman (FK) et les modèles ARMA

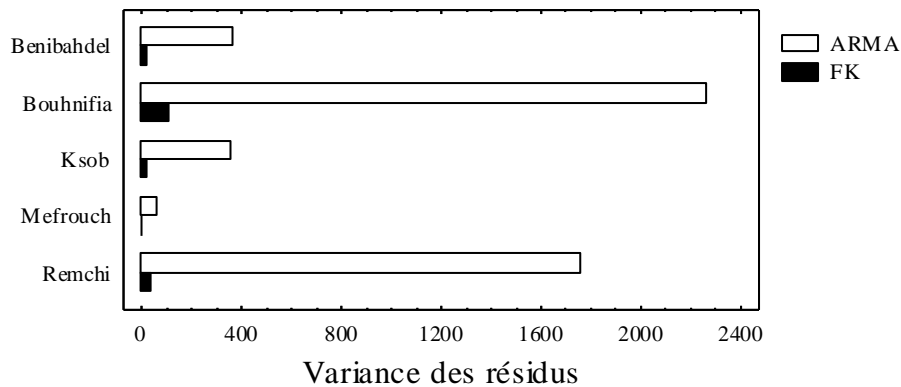


Fig.III.2.71 Comparaison de la variance des erreurs de prédiction entre le Filtre de Kalman (FK) et les modèles ARMA

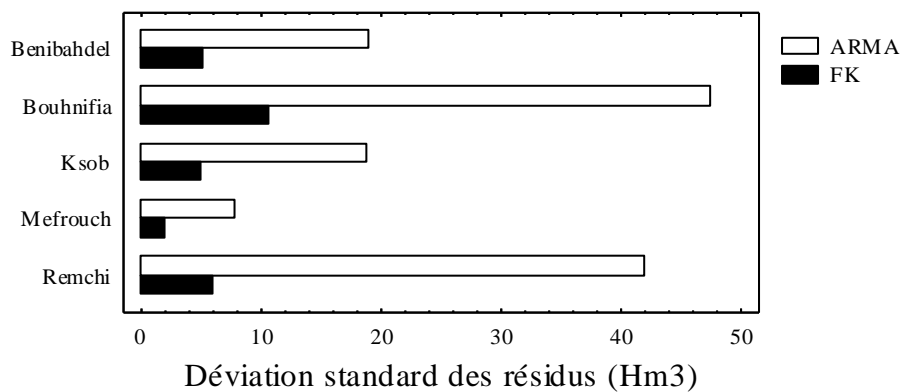


Fig.III.2.72 Comparaison de la déviation standard des erreurs de prédiction entre le Filtre de Kalman (FK) et les modèles ARMA

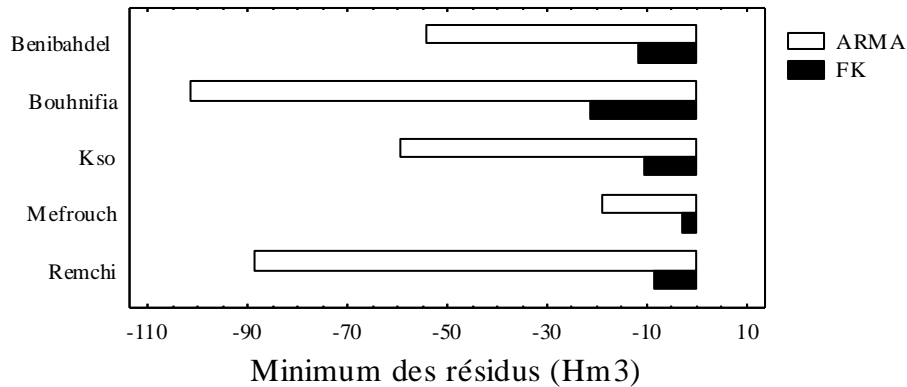


Fig.III.2.73 Comparaison du minimum des erreurs de prédiction entre le Filtre de Kalman (FK) et les modèles ARMA

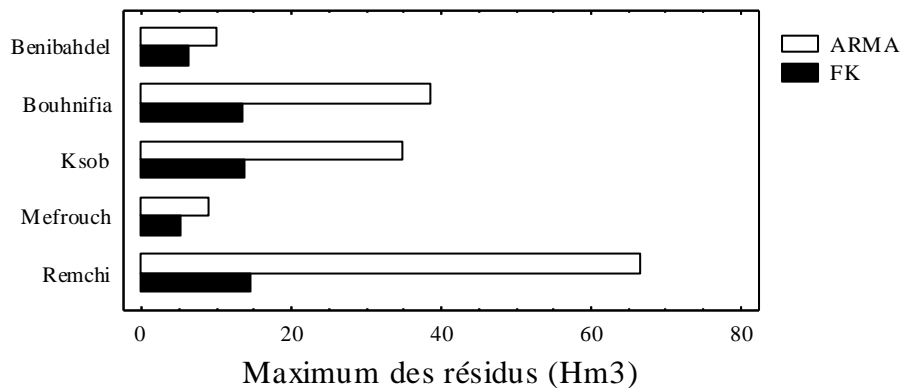


Fig.III.2.74 Comparaison du maximum des erreurs de prédiction entre le Filtre de Kalman (FK) et les modèles ARMA

Selon ces résultats, il se trouve que les erreurs de prédictions obtenues par le FK sont beaucoup plus petites en valeur absolue que celles obtenues par les modèles ARMA. Cela veut dire que l'écart entre les valeurs observées des apports annuels et leurs prédictions est beaucoup plus petit par l'usage du FK que celui des modèles ARMA. Nous pouvons en tirer que l'usage du filtre de Kalman conduit à de meilleures prédictions relativement aux modèles ARMA et que par voie de conséquences, les erreurs de prédictions obtenues dans la partie II de cette thèse ont bien été améliorées dans la partie présente par le recours au filtre de Kalman ce qui exprime la supériorité du de ce dernier quant à la prédiction des apports annuels considérés.

III.2.9 CONCLUSION

Dans la présente étude, nous avons développé un modèle stochastique linéaire multi-site, formulé en espace-état, pour satisfaire la structure d'un filtre de Kalman discret. La première équation de ce modèle exprime l'état du système hydrologique concerné, dans notre cas, c'est un vecteur donnant l'apport liquide annuel à plusieurs sites à la fois. La deuxième équation exprime la mesure de cette grandeur d'état, qui est dans ce cas un vecteur de même dimension que l'état, obtenu par observation de l'apport liquide annuel aux sites précédents.

Le développement de ce modèle a pour objectif, l'application de la technique du filtre de Kalman(FK) à la prédiction multi-site des apports liquides annuels et l'amélioration des erreurs de prédiction.

L'estimateur développé au cours de ce travail a la particularité de pouvoir s'adapter automatiquement chaque fois qu'une nouvelle information (mesure) devient disponible. Par conséquent, des prédictions optimales des apports liquides sont obtenues aussi bien dans le temps que dans l'espace.

En effet, dans le cas où l'on dispose d'une information historique au préalable, une fois que le modèle a été sélectionné, d'abord les conditions initiales concernant le vecteur d'état, les covariances des bruits du système et de la mesure doivent être déterminées, puis la matrice de transition d'état et la matrice de mesure doivent être spécifiées, pour qu'en fin les calculs puissent être amorcés.

Dans la présente étude, nous nous intéressons aux séries chronologiques des apports liquides annuels, observés à 10 stations hydrométriques au Nord d'Algérie et couvrant une période d'observation commune de 25 ans allant de 1968 à 1992.

Le travail a été fait sur deux périodes de temps. La première a été utilisée dans la procédure d'estimation des paramètres du modèle, tandis que la deuxième a servi à la validation du modèle.

L'apport liquide moyen annuel observé à chacune des 10 stations hydrométriques a été adopté comme vecteur d'état initial. Pour la matrice de covariance d'erreur initiale, nous avons opté pour de grandes valeurs dans les éléments de la diagonale principale afin de donner plus de poids à la nouvelle information (mesure) qui vient d'être disponible.

Les prédictions multi-sites des apports liquides, obtenues dans la dimension temporelle pour chacune des 10 stations hydrométriques en question, sont très proches des valeurs observées aux 10 stations sur la même période. De même pour les prédictions multi-sites dans la dimension spatiale, les valeurs observées et les prédictions correspondantes se suivent d'une manière assez proche l'une de l'autre pour toute la période d'observation, ceci indique que le filtre de Kalman multi-site est un outil efficace pour la modélisation et la prédiction des apports liquides annuels en temps et en espace.

En fin, la justesse du modèle du filtre de Kalman ou son aptitude à représenter le processus naturel des apports liquides, a été vérifiée par l'étude de l'indépendance des innovations du filtre d'une part, par le biais de leur fonction d'auto-corrélation et d'autre, par l'ajustement Normal de ces innovations.

Le pourcentage de l'erreur relative des prédictions multi-sites a également été un moyen d'apprécier cette justesse. Le pourcentage moyen de l'erreur relative de prédiction sur toute la période d'observation est inférieure à 15% ce qui est très acceptable.

En outre, le calcul des moyennes et des déviations standards pour les données observées et leurs valeurs prédites a montré que les déviations standards des valeurs observées sont plus grandes que celles des valeurs prédites, ce qui pourrait traduire une éventuelle tendance du (FK) multi-site à une sous-estimation.

La comparaison des erreurs de prédiction obtenues par les modèles ARMA dans la deuxième partie de cette thèse, avec celles obtenues par le FK a montré la supériorité du FK dans la modélisation des apports annuels. Ces erreurs de prédiction ont été améliorées d'une façon remarquable avec des rapports moyens qui sont de 1/16 pour la moyenne des erreurs de prédiction, 1/23 pour leur variance, 1/5 pour leurs déviation standard, 1/6 pour leur minimum et enfin 1/3 pour le maximum.

A la lumière de tous les résultats précédents, il est clair que la technique du FK multi-site est efficace pour la modélisation des apports liquides annuels, aussi bien dans la dimension temporelle que spatiale. Le résultat est un modèle dynamique qui présente l'avantage considérable de prendre en compte les variations dans le temps des paramètres du modèle, de la variance des bruits d'état et celle de la mesure et dont les erreurs de prédictions sont minimales.

CONCLUSION III

Dans cette troisième partie, notre attention a été focalisée sur l'étude de l'applicabilité de la technique du filtre de Kalman (FK) à la modélisation et la prédiction multi-site des apports liquides des rivières au Nord d'Algérie. L'objectif était l'obtention d'un estimateur linéaire optimal qui permet la prévision multi-site de cette variable hydrologique et l'amélioration des erreurs de prédiction.

A cet effet, nous avons choisi de travailler sur des données annuelles dont les corrélations croisées sont faibles et qui ne présentent pas d'effet saisonnier. Ce choix présente l'avantage de diminuer le degré de complexité du problème, ce qui nous a aidés à surpasser quelques unes des difficultés rencontrées lors de l'application de la méthode. Cela dit, le modèle de prédiction multi-site obtenu reste éventuellement extensible au pas de temps inférieurs, moyennant un choix judicieux de la matrice de transition, et peut être également généralisé à d'autres variables hydrologiques ce qui présentera sans doute un moyen efficace et optimal pour les gestionnaires de la ressource en eau.

Le modèle obtenu est un moyen de prédiction en ligne qui non seulement procure des prédictions multi-sites optimales mais aussi des prédictions qui tiennent en compte la nature dynamique du système hydrologique concerné : c'est un modèle dynamique qui prend en charge les variations dans le temps des paramètres du modèle, de la variance des bruits d'état et de la variance des bruits de mesure.

L'estimateur développé au cours de ce travail a la particularité de pouvoir s'adapter automatiquement, chaque fois qu'une nouvelle information (mesure) devient disponible. Cette flexibilité lui offre la possibilité d'obtenir des prédictions optimales dans le temps au-delà de l'année 1993, qui présente la fin des itérations pour le présent travail, par simple ajout de données, et dans l'espace par simple changement dans les dimensions des vecteurs et des matrices mises en œuvre par l'algorithme de calcul. Ce travail est valable pour des stations qui présentent la même période d'observation, et peut présenter la base d'un travail de cartographie.

La méthode de Kalman et les équations récurrentes qu'elle met en œuvre présentent, en plus de nombreux avantages :

- La forme récurrente des équations facilite leur exploitation par les ordinateurs ;
- La forme des équations de base du filtre se prêtent souvent à une exploitation satisfaisante des informations dont on dispose : d'une part, la dynamique du système et la nature des mesures y sont respectivement représentées par les matrices de transition d'état et celle de mesure ; d'autre part, les perturbations du système et les bruits de mesure y sont représentés par leurs propriétés statistiques (valeurs moyennes, matrices de covariance) ;
- Les optimisations de la méthode ont une signification statistique claire : variances d'erreur minimales ou encore, dans le cas de bruits gaussiens, probabilité maximale des valeurs estimées ;
- La méthode fournit non seulement une estimation des variables d'état, mais encore, par l'intermédiaire de la matrice de covariance d'erreur de prédiction $P_{k/k}$, une estimation de la confiance que l'on peut accorder à ses résultats ce qui présente un avantage majeur.

En contre partie, il faut reconnaître les difficultés inhérentes à l'application de la méthode :

- Le choix de la variable d'état et la formulation du problème selon l'équation d'état et l'équation de mesure.
- L'estimation de la matrice de transition d'état.
- L'estimation initiale des matrices ($P_{0/0}, Q_k, R_k$) est toujours entachée d'erreur pouvant perturber la performance du filtre en ce qui concerne l'estimation de la variable d'état. Il faudra essayer plusieurs valeurs et observer le comportement du filtre. Il faut également attribuer des valeurs importantes aux termes diagonaux de $P_{0/0}$ pour donner plus de poids au gain du filtre qui procèdera progressivement à leurs correction en fonction des mesures. Par ailleurs, il faut savoir qu'une estimation erronée de l'état initial ($\hat{X}_0 = E(X_0)$) est pratiquement sans importance car elle est tout de suite corrigée par le filtre en fonction de la mesure.
- Un écart pouvant exister entre le modèle mathématique utilisé et les propriétés réelles du système hydrologique concerné ; l'influence de cet écart à la longue peut provoquer la divergence du filtre, c'est-à-dire des écarts ($X_k - \hat{X}_{k/k}$) invraisemblables.

Cela dit, et en dépit de tous ces problèmes, les résultats que la méthode procure sont satisfaisants et répondent largement aux besoins et aux préoccupations pratiques envisagées, une simple comparaison de résultats le prouve. En effet, la comparaison des erreurs de prédiction du FK et des modèles ARMA a montré la supériorité du FK dans la modélisation des apports annuels : ces erreurs de prédiction ont été améliorées avec des rapports moyens qui sont de l'ordre de 1/16 pour la moyenne des erreurs de prédiction du FK par rapport aux modèles ARMA, 1/23 pour la variance, 1/5 pour la déviation standard, 1/6 pour le minimum et enfin 1/3 pour le maximum.

CONCLUSION GENERALE ET RECOMMANDATIONS FUTURES

Dans un pays semi-aride tel que l'Algérie, où la demande en eau prend de plus en plus de l'ampleur au fil des ans, les volumes d'eau véhiculés dans les cours des rivières représentent une composante très importante pour le développement social et économique. Compte tenu de ce fait, la partie Nord Algérienne a été sujette à plusieurs investigations visant l'interception et le stockage du maximum possible de quantités d'eau pluviale dans des réservoirs de stockage. La réserve hydrique dans de tels réservoirs est perçue comme un moyen potentiel et efficace pour répondre aux besoins toujours croissants en matière d'eau pour l'agriculture, l'industrie et l'utilisation domestique.

L'état de conscience au sujet de cette réalité socio-économique, a suscité la canalisation et la conjugaison de tous les efforts dans le but de répondre positivement à ces préoccupations. Le présent travail en voulant s'inscrire dans ce contexte, a apporté une contribution modeste à ces efforts. C'est une étude de modélisation qui perçoit ces apports liquides comme une variable hydrologique générée par un processus stochastique, ce qui a suscité le recours à des modèles mathématiques prenant en charge la nature stochastique de cette variable ainsi que ses changements dans le temps. Il va sans dire que de tels modèles sont indispensables à toutes les décisions importantes au sujet des ressources hydriques : ils sont utilisés en simulation pour répondre à des questions de conception et mobilisation, et aussi en prévision pour répondre à des questions de gestion. Dans le cadre de ce travail, nous nous sommes intéressés uniquement à la prévision, sachant que leur utilisation à la simulation reste toujours valable.

Dans cette perspective, et afin d'apporter une réponse aux objectifs fixés pour ce travail, nous l'avons présenté en trois parties principales. Après une introduction générale où l'on a présenté la problématique, fixé l'objet et les méthodes mises en œuvre et ciblé les objectifs, vient la première partie où l'on expose les différentes notions mathématiques et rappels théoriques indispensables à la compréhension du travail. Les deux parties suivantes ont été dédiées toutes les deux à la modélisation stochastique des apports liquides pour des fins de prévision : la deuxième utilise la méthode de Box-Jenkins qui est limitée par des contraintes de stationnarité et de normalité tandis que la troisième partie se libère de toutes les contraintes en faisant recours au filtre de Kalman discret qui dispose d'une structure rétroactive lui permettant une estimation dynamique et optimale. En fin nous terminons par une conclusion générale.

A cet effet, la première partie, avec ses trois chapitres, constitue un arrière plan théorique de la méthode de Box-Jenkins et celle du filtre de Kalman. Elle nous a permis d'introduire les notions et les concepts essentiels à la compréhension des mécanismes et des fondements mathématiques sur lesquels reposent ces méthodes. Dans le chapitre I.1, nous avons introduit les notions mathématiques et les concepts de base qui servent à la compréhension du principe de la modélisation des processus hydrologiques naturels, et présenté les instruments mathématiques qui

servent à l'analyse des processus stochastiques et leur transmission dans les systèmes dynamiques linéaires. L'accent a été mis sur l'importance de la notion de bruit blanc gaussien qui est complètement caractérisé par sa fonction d'auto-corrélation et qui présente l'une des hypothèses de base aussi bien pour la méthode de Box-Jenkins que celle du Filtre de Kalman. D'autres notions telles que les modèles Espace-état ou encore le filtrage linéaire ont également été introduites et soulignées. Dans le chapitre I.2, nous avons présenté le concept processus stochastique/série chronologique où nous avons d'abord abordé les séries chronologiques et insisté sur la particularité et l'intérêt de cet ordre temporel dans lequel elles ont été observées. Nous avons, par la même occasion, montré que cet ordre n'était pas fortuit mais qu'il était l'expression de la présence d'une relation temporelle imposée par les lois de la nature ; et que par suite de conséquences, cette relation d'antériorité attribuée aux séries chronologiques une sorte de mémoire qui fait appel à ces lois de la nature qu'on appelle « mécanismes générateurs sous-jacents » et qui sont de nature stochastique. L'établissement de ce lien nous a ensuite conduits à aborder les processus stochastiques non paramétriques et à montrer comment ces derniers peuvent être caractérisés. Cela étant fait en insistant sur la notion de stationnarité et d'ergodicité qui sont à la base de cette caractérisation stochastique. Dans le chapitre I.3, nous avons accordé un intérêt particulier à une classe de processus stochastiques linéaires : les processus de Box-Jenkins, en raison de la qualité de leur représentation des processus hydrologiques en général et des apports de rivières en particulier. Dans cette optique, nous avons d'abord introduit et explicité le lien entre série chronologiques et processus linéaires, et en faisant intervenir les différents opérateurs linéaires, nous avons montré que beaucoup de processus stochastiques avec dépendance sont obtenus par des transformations linéaires de bruits blancs. Ensuite, nous avons introduit le théorème de Wold qui est à la base de la modélisation ARMA abordée dans la deuxième partie de cette thèse. A cet effet, on définit un processus stochastique linéaire, non pas par la distribution jointe des observations ou de ses premiers moments, comme c'était le cas dans les chapitres précédents, mais à partir d'un mécanisme générateur indexé par des paramètres. Ce théorème permet la caractérisation du mécanisme générateur et stipule que tout processus stationnaire peut être représenté, de manière unique, par la somme de deux composantes indépendantes : l'une déterministe parfaitement prévisible et l'autre stochastique. Enfin, nous présentons les différents modèles de ces processus : ARMA, ARIMA et SARIMA avec leurs hypothèses et leurs caractéristiques respectives.

La deuxième partie, avec ses trois chapitres également, constitue une application pratique de la méthode de Box-Jenkins à la modélisation stochastique des apports liquides observés à 10 stations hydrométriques au Nord d'Algérie. Elle nous a permis d'obtenir des modèles de prédiction au pas de temps annuel et mensuel pour chacune des 10 stations considérées dont les résultats étaient satisfaisants. Dans le chapitre II.1, nous avons exposé la méthodologie de Box-Jenkins pour la modélisation stochastique des séries chronologiques avec les étapes principales correspondantes, et les outils mathématiques indispensables à l'accomplissement de chacune de ces dernières. Un accent particulier a été mis sur les hypothèses de normalité et de stationnarité sur lesquelles reposent les fondements théoriques de la méthode. La structure de dépendance interne des données sur laquelle se base cette méthode a été mise en valeur ainsi que le caractère itératif de la procédure. Enfin, nous avons exposé comment le modèle ARMA résultant peut être mis en œuvre pour faire de la prédiction tout en indiquant les intervalles de confiance de ces dernières. Dans le chapitre II.2 et II.3, nous avons exposé en détail et pour chacune des 10 stations, comment la méthode de Box-Jenkins a été mise en œuvre dans le but de trouver le modèle le plus adéquat aux données observées. Le chapitre II.2 a été consacré à l'application de la dite méthode aux données annuelles qui sont généralement stationnaires, normalement distribuées avec une faible auto-corrélation dans leur structure de dépendance interne, tandis que le chapitre II.3 concernait les données mensuelles qui sont, au contraire caractérisées par une instabilité saisonnière au niveau de la variance, par une distribution non normale et par une forte auto-corrélation régulière et saisonnière dans la structure

de dépendance interne. Les résultats obtenus dans ces deux chapitres ont conduit à la détermination de modèles pratiques, simples et parcimonieux qui représentent le plus possible le processus générateur de la variable hydrologique en question et reflètent l'ampleur de la variabilité aléatoire de celle-ci. Cependant, il a fallu reconnaître que malgré les résultats satisfaisants qu'elle peut produire, l'usage de cette classe de modèles reste limité par certaines contraintes qui font de ces derniers loin d'être l'instrument mathématique idéal pour les processus hydrologiques et ce pour plusieurs raisons : (1) l'une de ces raisons principales est que ces modèles reposent largement sur l'hypothèse de Normalité alors qu'il est connu que les processus hydrologiques, particulièrement à un pas inférieur à l'annuel, s'éloignent de la distribution Normale et tendent vers des distributions de type puissance. L'autre hypothèse d'appui de ces modèles est la stationnarité alors que le comportement saisonnier exhibé par les processus à une échelle inférieure à l'année est complexe (type de distribution et structure de dépendance qui changent au sein de l'année) et ne peut pas être pris en charge par les techniques de dessaisonnalisation de Box-Jenkins (**Koutsoyannis 2006**). (2) Les modèles Box-Jenkins qui se basent essentiellement sur la structure de dépendance des séries chronologiques ont tous une structure de dépendance de court terme, c'est-à-dire une fonction d'auto-corrélation qui diminue exponentiellement avec le pas de temps, alors qu'à l'opposé les processus hydrologiques montrent une dépendance de long terme, c'est-à-dire une décroissance en puissance connue par le phénomène de Hurst (**Koutsoyannis 2005b**). (3) A l'exception des modèles simplifiés AR(1) ou ARMA(1,1) les modèles ne sont pas parcimonieux car ils font intervenir plusieurs paramètres estimés à partir des données. En plus, les échantillons statistiques tels que les séries chronologiques ne permettent pas l'estimation fiable de plusieurs paramètres. Ceci est particulièrement le cas des processus hydrologiques avec une dépendance de long terme puisque ce comportement induit des incertitudes importantes par rapport aux processus avec une dépendance de court terme. (4) De nouveau, à l'exception des modèles AR(1) et ARMA(1,1), d'autres processus de la même famille ne possèdent pas de sens physique et sont en plus utilisés dans un concept de boîte noire.

Ces limites et contraintes de la méthode de Box-Jenkins, malgré les résultats satisfaisants qu'elle procure, ont suscité l'appel à d'autres outils mathématiques qui répondent mieux à la nature dynamique et au comportement saisonnier des processus hydrologiques, c'est l'une des raisons qui ont justifié le recours au filtre de Kalman par la suite.

Au même titre que la partie précédente, la troisième partie avec ces deux chapitres, constitue une étude de modélisation stochastique des apports liquides pour des fins de prédiction à la différence que le modèle stochastique résultant possède les propriétés suivantes :

- 1) Il permet d'obtenir des prédictions multi-sites ;
- 2) Il est dynamique avec une prise en charge de la nature variable dans le temps du processus étudié.

La méthode préconisée dans cette partie est celle du filtre de Kalman linéaire discret dont le résultat est l'obtention d'un estimateur optimal en ligne qui permet la prévision de cette variable hydrologique à plusieurs sites en même temps. Pour cela, après une introduction retraçant l'état de l'art de la méthode et de ses applications aux domaines différents des sciences de l'ingénieur, le chapitre III.1 nous a permis d'exposer la méthode du filtre de Kalman discret. Ainsi, le principe de la méthode avec son cadre théorique y sont présentés. L'accent a été mis sur l'essentiel : le modèle mathématique adopté et les formules récurrentes du filtre, en passant en revue les conditions d'optimisation qui font le mérite du filtre de Kalman. Le chapitre III.2 nous a permis d'étudier l'applicabilité de la technique du filtre de Kalman à la modélisation et la prédiction multi-site des apports liquides annuels observés à 10 stations hydrométriques au Nord d'Algérie; la période d'observation commune étant 25 ans (1968 à 1992). A cet effet, nous avons expliqué pourquoi nous avons choisi de travailler sur des données annuelles et comment ce choix nous a aidés à surpasser quelques unes des difficultés rencontrées lors de l'application de la méthode. Nous avons

montré comment le modèle mathématique a été formulé et comment les conditions initiales concernant le vecteur d'état, la covariance des bruits du système et la covariance des bruits de mesure ont été déterminées, puis comment la matrice de transition d'état et la matrice de mesure ont été spécifiées, et en fin comment les calculs ont été effectués. Les résultats obtenus ont été discutés en détail, par la suite et leur qualité était bien mise en exergue par l'étude des erreurs de prédiction qui affichait un pourcentage moyen de l'erreur relative très acceptable (15%). A la suite de cela, nous y avons indiqué comment le modèle de prédiction multi-site obtenu pourrait éventuellement être extensible au pas de temps inférieurs, et comment il pourrait être généralisé à d'autres variables hydrologiques ce qui présenterait sans doute un moyen efficace et optimal pour les gestionnaires de la ressource en eau. Aussi, nous avons montré comment le modèle obtenu était un moyen de prédiction en ligne qui non seulement procure des prédictions multi-sites optimales mais aussi des prédictions tenant en compte la nature dynamique du système hydrologique concerné (variations temporelles des paramètres du modèle, des variances des bruits d'état et de mesure, dues aux changements dans les infiltrations, l'humidité du sol et toute autre variable pouvant affecter le cours des apports liquides). Par la même occasion, nous avons indiqué comment l'estimateur développé au cours de ce travail pourrait être à la base d'une cartographie en montrant la particularité de ce dernier de pouvoir s'adapter automatiquement, chaque fois qu'une nouvelle information (mesure) devient disponible et en insistant sur sa flexibilité qui lui permet une extension dans les deux dimensions : le temps et l'espace.

A la fin de cette partie, il a fallu reconnaître certaines difficultés rencontrées dans l'application pratique de cette méthode. La première relevait de la disponibilité d'une période d'observation commune entre les stations hydrométriques, ce qui a limité le nombre de stations étudiées à 10 et conduit au renoncement à la cartographie prévue au début de ce travail. La deuxième difficulté était inhérente à l'application propre de la méthode de Kalman concernant essentiellement l'estimation des conditions initiales des matrices ($P_{0/0}$, Q_k , R_k) et l'estimation de la matrice de transition d'état, sachant que d'autres problèmes pouvaient surgir tel que la divergence du filtre par exemple ou encore plus des problèmes dans la formulation espace-état du problème. En contre partie, quelques uns des avantages de la méthode ont été mis en valeur, particulièrement l'exploitation de toutes les informations disponibles sur le système et ses mesures : d'un côté la dynamique du système et la nature des mesures sont respectivement représentées par la matrice de transition d'état et celle de la mesure, d'un autre côté les perturbations du système et de la mesure sont représentées par leurs propriétés statistiques (valeurs moyennes et matrices de covariance). Un autre avantage de la méthode a été souligné à savoir la procuration d'une estimation optimale en plus de la covariance d'erreur de cette dernière qui présente la confiance qui pourrait lui être accordée. La comparaison entre le filtre de Kalman et la méthode de Box-Jenkins concernant les erreurs de prédiction a confirmé un des aspects de cette optimalité en affichant des caractéristiques statistiques moyennes beaucoup plus réduites pour le filtre de Kalman. Ces résultats montrent la robustesse et soulignent la supériorité du filtre de Kalman dans la modélisation et la prédiction des apports considérés.

Pour les futures applications du FK, les recommandations suivantes devraient être considérées :

- (1) En vertu de la nature dynamique du modèle développé au cours de cette thèse, l'introduction de paramètres additionnels tels que les précipitations par exemple, dans l'équation d'état est fortement recommandée pour l'amélioration de la performance du modèle.
- (2) L'incorporation de plus de stations hydrométriques est aussi fortement recommandée, ceci requière la disponibilité d'autres stations dans la même région ayant la même période d'observation, ou une période d'observation commune avec les stations déjà considérées.

Quand le nombre de ces stations sera assez important les résultats obtenus peuvent être cartographiés et les prédictions spatiales peuvent être interpolées/extrapolées avec les erreurs de prédiction correspondantes. Ceci sera d'une grande importance pour les plans de gestion de la ressource en eau au pas annuel.

- (3) Les prédictions mensuelles, hebdomadaires, ou encore journalières sont aussi fortement recommandées par couplage de modèles de désagrégation au FK. Dans ce cas, il faut tenir compte de l'effet saisonnier caractérisant de telles séries de données, chose qui n'a pas été considérée ici simplement parce que les données annuelles ne sont généralement pas sujettes à un tel effet.

Enfin, il faut mentionner que l'approche suivie au cours de ce travail de recherche que se soit pour la méthode de Box et Jenkins ou encore celle du filtre de Kalman reste valable est applicable pour n'importe qu'elle autre variable hydrologique.

REFERENCES BIBLIOGRAPHIQUES

- [1] Adenstedt, R. K. (1974) On large sample estimation for the mean of a stationary random sequence. *Ann. Statist.***2**, 1095–1107.
- [2] Ahsan, M. & Connor, K. (1994) A reappraisal of the Kalman filtering technique, as applied in river flow forecasting. *J. Hydrol.*, **161**, 197-226.
- [3] Akaike, H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control*. **19**(6), 716–723.
- [4] Alpuim, T. & Barbosa, S. (1999) The Kalman filter in the estimation of area precipitation. *Environmetrics*. **10**(4), 377 – 394.
- [5] Amorocho, J. & Orlob, G.T. (1961) *Nonlinear analysis of hydrologic systems*. University of California, Water Ressources Center, contribution N°40, 147p.
- [6] Anderson, O.D. (1976) *Time Series Analysis and Forecasting : The BOX JENKINS Approach*. Butterworth, London.
- [7] Anderson, R. L. (1942) Distribution of the serial correlation coefficient. *Ann. Math.Stat.* **13**, 96-101.
- [8] Anderson, T. W. (1971) *The Statistical Analysis of Time Series*. John Wiley & Sons, NewYork.
- [9] Awwad, H. M. & Valdés, J. B. (1992) Adaptive parameter estimation for multisite hydrologic forecasting. *J. Hydr. Engrg., ASCE*, **118**(9), 1201-1221.
- [10] Awwad, H. M., Valdés, J. B. & Restrepo, P. J. (1994) Streamflow Forecasting for Han River Basin, Korea. *J. Water Resour. Plng. and Mgmt.*, **120**(5), 651-673.
- [11] Azencott, Dachuna-Castelle (1989) *Séries d'observations irrégulières*. Masson, Paris.
- [12] Beran, J. (1994) Statistics for Long-Memory Processes, vol. 61 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, New York, USA.
- [13] Beran, J., (1994a) *Statistics for Long Memory Processes*. Chapman and Hall, New York, USA.
- [14] Berberian, S.K. (1961) *Introduction to Hilbert Space*. Oxford University Press, New York, USA.
- [15] Bergman, M. J. & Delleur, J. W. (1985) Kalman Filter Estimation and Prediction of daily Stream Flows: I. Review, Algorithm, and Simulation Experiments, *Water Resources Bulletin*. **21**(5), 815-826.
- [16] Bertino, L., Evensen, G. & Wackernagel, H. (2002) Combining geostatistics and Kalman filtering for data assimilation in an estuarine system. *Inverse Problems*, **18**,1–23.
- [17] Bialkovski, W.L. (1983) Application of Kalman Filters to the Regulation of Dead Time Processes. *IEEE*, Vol.AC-**28**, NO 3.
- [18] Bloomfield, P. (1976) *Fourier Analysis of Time Series : An Introduction*. John Wiley & Sons, New York, USA.
- [19] Bloomfield, P. (1992) Trends in global temperature. *Clim. Change* **21**, 1–16.
- [20] Boone, L. (2000) *Comparing Semi-structural Methods to Estimate Unobserved Variables: The HPMV and Kalman Filters Approaches*. Economics Department. Working Paper, OECD, n° 240.

- [21] **Boukharouba, K.** & Kettab A. (2009) Kalman filter technique for mutli-site modelling and streamflow prediction in Algeria. *Journal of Food, Agriculture & Environment*, Vol.7(2), 671-677.
- [22] **Boukharouba, K.** & Kettab, A. (2006) ARMA Modelling for Identification and Simulation of Liquid and Solid Discharges (Application to Foum-El-Kerza dam in Algeria). Watmed 3, Tripoli, Lebanon, 1-3 November 2006.
- [23] **Boukharouba, K.** & Kettab A., (2006) Application of Kalman Filter to multisite annual discharge prediction. 22nd International Congress On Large Dams, Barcelona 18 - 23 June 2006, Spain. Dans: *Dams and Reservoirs, Societies and Environment in the 21st Century* - Berga *et al.* (eds)©2006, Taylor & Francis Group, London, ISBN 0 415 40423.
- [24] **Boukharouba, K.** & Kettab, A. (2005a) Stochastic modelling formulation of the waste problem in Foum-el-kherza dam in Algeria. International conference on Waste recycling IX, 17-19 November 2005, Exhibition and Conference Center, Wawel Hill 9 , Cracow, Poland.
- [25] **Boukharouba, K.** & Kettab, A. (2005b) Les processus de type ARMA pour la simulation stochastique des débits liquides de l'Oued Labiod en Algérie. 21e congrès régional de l'est du Canada de l'acqe (Association canadienne pour la qualité de l'eau), 4 novembre 2005, Institut national de la recherche scientifique, Centre Eau, Terre et Environnement (INRS-ETE), Québec, QC, CANADA.
- [26] **Boukharouba, K.** & Kettab, A. (2003a) Contribution à la gestion des cours d'eau par Simulation Stochastique (Etude de cas). 7^{ième} Conférence Internationale des Limnologues d'Expression Française, CILEF-2003, 27 juillet-1^{er} août, 2003, Montréal, Québec, Canada.
- [27] **Boukharouba, K.** & Kettab, A. (2003b) The sequential simulation for model parameter estimation and short term prediction (case study). The 2nd International Conference on Wadi Hydrology, 1/4 Juil, Amman, Jordanie.
- [28] **Boukharouba, K.** & Kettab, A. (2003c) Box-Jenkins Modeling of annual River Flows (case study). The 2nd International Conference on Wadi Hydrology, 1/4 Juil, Amman, Jordanie.
- [29] **Boukharouba, K.**, Chebbah, M., & Kettab, A. (2002) Etude statistique et analyse de la série chronologique des débits liquides de l'Oued Labiod-Biskra. 4^e Symposium International sur l'eau, Cannes 3/7 juin 2002, Palais des Festivals, France.
- [30] **Boukharouba, K.** & Kettab, A. (2000) Adaptation and application of the unbiased recursive prediction algorithm of Kalman Filter to the prediction of annual Flows. ACE, International Conference, 1-4 novembre 2000, western university, chypre, TRNC.
- [31] **Boukharouba, K.** & Mokrane, B. (1999) Modélisation stochastique des apports annuels du barrage de béni bahdel en vue de la simulation. Séminaire national sur l'eau : enjeux et stratégies, sidi bel abbes, Algérie.
- [32] **Boukharouba, K.** & Mokrane, B. (1997) prévision à court terme des débits liquides annuels du barrage de béni bahdel. 2^e séminaire national sur l'eau, Université de Biskra.

- [33] **Boukharouba, K.** (1997) Modélisation stochastique des apports liquides annuels : simulation par les modèles ARMA et prédiction par le filtre de Kalman. Thèse de Magister, département de l'hydraulique, Ecole Nationale Polytechnique d'Alger.
- [34] Box, P. G. & Jenkins, P. M. (1970) *Time Series Analysis*. Holden Day, San Francisco, USA.
- [35] Brammer, R.F. *et al.* (1983) Bathymetric and Oceanographic Applications of Kalman Filtering Techniques. *IEEE*, Vol. AC-28, NO.3.
- [36] Bras, R. L. & Rodriguez-Iturbe, I. (1985) *Random Functions in Hydrology*. Addison-Wesley, New York, USA.
- [37] Bras, R. L. & Rodriguez-Iturbe, I. (1993) *Random Functions and Hydrology*. Chapters 8 and 9. Addison-Wesley, Reading, Massachusetts, USA.
- [38] Brillinger, D. R. & Krishnaiah, P. R., eds (1983), *Handbook of Statistics 3: Time Series in the Frequency Domain*. North-Holland, Amsterdam.
- [39] Brillinger, D. R. (1975) *Time Series: Data Analysis and Theory*. Holt, Rinehart & Winston, New York, USA.
- [40] Brillinger, D. R. (1976) Some history of the data analysis of time series in the United States. In (*Owen 1976*).
- [41] Brockwell, P.J. & Davis, R.A. (1996) *Introduction to time series and forecasting*. Springer texts in statistics, Springer, New York, USA.
- [42] Brockwell, P.J., Davis, R.A. (1990) *Time Series : Theory and Methods*. (2nd ed.), Springer, New York, USA.
- [43] Brown, R.G. & Hwang, P.Y.C. (1997) *Introduction to random signals and applied Kalman filtering: with MATLAB exercises and solutions* (3rd Ed.) John Wiley and sons Inc., New York, USA.
- [44] Chatefield, C. (2001) *Time series forecasting*. Chapman and Hall, New York, USA.
- [45] Chiu, C.L. (1978) *Application of Kalman filtering of American Geophysical Union*. Chapman Conference on Applications of Kalman filtering theory and technique to hydrology, hydraulics and water resources. University of Pittsburgh, Pittsburgh, Pennsylvania, USA, 738p.
- [46] Chow, V. T. (1964) Statistical and probabilistic analysis of hydrologic data. In: *Handbook of Hydrology* (ed. by V. T. Chow), McGraw-Hill, New York, USA.
- [47] Chui, C.K., & Chen, G. (1987) *Kalman filtering with real-time applications*. 191 pp.
- [48] Clement, M. P. & Hendry, D. F. (2000) *Forecasting Economic Time Series*. Cambridge University continentale Vol. 1, 57-70.
- [49] Cohn, T. A. & Lins, H. F. (2005) Nature's style: naturally trendy. *Geophys. Res. Lett.* **32**(23) L23402, doi:10.1029/2005GL024476.
- [50] Cox, D.R. (1981) Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics*, **8**, 93–115.
- [51] Davidson, J. (1994) *Stochastic Limit Theory*. Oxford University Press, New York, USA.
- [52] Doornik, J.A. & Ooms, M. (2003) Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models. In: *Computational Statistics and Data Analysis*, **41**, 333–348.
- [53] Droesbeke, J.-J. & Tassi, P. (1990) *Histoire de la statistique*. Collection Que sais-je ?, no2527, Presses Universitaires de France, Paris.
- [54] Droesbeke, J.-J., Fichet, B. & Tassi, P., eds (1989) *Séries chronologiques : théorie et pratique des modèles ARIMA*. Economica, Paris.
- [55] Fortin, V., Ouarda, T. Rasmussen & Bobée, B. (1997) Revue bibliographique des méthodes de prévision des débits. *Revue des Sciences de l'Eau*, **4**: 461-487.
- [56] Fuller, W. A. (1976) *Introduction to Time Series*. John Wiley, New York.
- [57] Gelb, A. (1974) *Applied Optimal Estimation*. Cambridge, MA: MIT Press.
- [58] Georgakakos, K. P. & Smith, G. F. (1990) On improved operational hydrologic forecasting, Results from a WMO realtime forecasting experiment. *J. Hydrol.* **114**, 17–45.

- [59] Geweke, J. & Porter-Hudak, S. (1983) The estimation and application of long-memory time series models. *Journal of Time Series Analysis*, **4**, 221-238.
- [60] Goodwin & Sign (1985) *Predictive Control*. Englewood Cliffs, Prentice-Hall, 540p.
- [61] Gourieroux, C. & Monfort, A. (1990) *Séries temporelles et modèles dynamiques*. Economica, Paris.
- [62] Granger, C.W.J. (1986) *Forecasting Economic Time Series*. Orlands, Academic press, 338p.
- [63] Granger, N. (1977) *Forecasting economic time series*. Academic Press, USA.
- [64] Grewal, M. S. & Andrews, A.P. (1997) *Kalman filtering theory and practice*. (4th edn.) Prentice Hall Information and System Sciences Series, Englewood Cliffs, N.J, USA.
- [65] Grewal, M. S. & Andrews, A.P. (2001) *Kalman Filtering Theory and Practice Using MATLAB* (2nd edn.). John Wiley & Sons, Inc. New York, NY USA.
- [66] Guy, R. (1983) A Kalman Filtering Approach to Natural Gamma Ray Spectroscopy in Well Logging. *IEEE*, Vol.AC-28, NO.3, 372-380.
- [67] Hamilton, J. .D. (1994) *Time Series Analysis*. Princeton University Press, Princeton.
- [68] Hannan, E. (1970) *Multiple Time Series*. John Wiley, New York.
- [69] Harrison, P.J., & Stevens, C.F. (1975b) Bayesian forecasting in action: Case studies. University of Warwick, working paper No. 14.
- [70] Harrison, P.J., and Stevens, C.F. (1975a) *Bayesian forecasting*. University of Warwick, working paper No.13.
- [71] Hartnack, J. & Madsen, H. (2001) Data assimilation in river flow modelling. Dans: *Proc. Fourth DHI Software Conf.* 6–8 June 2001, Scanticon Conf. Centre, Helsingør, Denmark.
- [72] Harvey, A. C. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.
- [73] Harvey, A.C. (1981) *The Econometric Analysis of Time Series*. Philip Allan, Oxford.
- [74] Haslett, J. & Raftery, A. E. (1989) Space–time modelling with long-memory dependence: assessing Ireland’s wind power resource. *Appl. Statist.* **38**(1), 1–50.
- [75] Hendry, D. F. & Clements M. P. (1998) *Forecasting Economic Time Series*. Cambridge.
- [76] Hendry, D. F. (1993) *Dynamic Econometrics*. Oxford University Press, New York, USA.
- [77] Hino, M. (1973) *On line prediction of hydrologic systems*. Proc. 15th Conf. Int. Assoc. Hydrol. Res. Istanbul, 121-129.
- [78] Hipel, K.W. & McLeod, A.I. (1994) *Time series modelling of water resources and environmental systems*. Elsevier, Hollande.
- [79] Hirsch, R. M., Helsel, D. R., Cohn, T. A. & Gilroy, E. J. (1993) Statistical analysis of hydrologic data. In: *Handbook of Hydrology* (ed. by Maidment, R. D.), Chapter 17, 17.1–17.55. McGraw-Hill, New York, USA.
- [80] Hosking, J. R. M. (1996) Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series. *J. Econometrics* **73**, 261–284.
- [81] Husain, T. (1985) Kalman filter estimation model in flood forecasting. *Adv. Water Resour.* **7**(2), 15–21.
- [82] Jacobs, O. L. R. (1993) *Introduction to Control Theory* (Second edn.). Oxford University Press.
- [83] Jazwinski, A. H. (1969) Adaptive Filtering. *Automatica*. Vol.5, 475-485.
- [84] Jazwinski, A. H. (1970) *Stochastic processes and filtering theory*. Academic Press, New York, 376 p.
- [85] Jimenez, C., McLeod, A. I. & Hipel, K. W. (1989) Kalman filter estimation for periodic autoregressive-moving average models, *Stochastic Hydrol. Hydraul.* **3**: 227-240.
- [86] Johnston, J. (1984) *Econometric Methods*. McGraw Hill, New York, USA.
- [87] Kailath, T. (1977) A view of three decades of linear filtering theory. Dans: Linear Least-Squares Estimation. Number 17 (ed. Kailath, T.) dans : *Benchmark Papers in Electrical Engineering and Computer Science*. Dowden, Hutchison & Ross, Stroudsburg, PA, 10–45.

- [88] Kalman, R. E. & Bucy, R. S. (1961) New Results in Linear Filtering and Prediction Theory. Trans. ASME. Ser. D. *J. Basic Eng.* Vol. **83**, 95-108.
- [89] Kalman, R. E. (1960) A new approach to linear filtering and prediction problems. Trans. ASME. Ser. D. *J. Basic Eng.* Vol. **82**, 35-45.
- [90] Kendall, M. & Ord, J.K. (1990) *Time series*. J. Wiley and sons, New-York.
- [91] Kettab, A., Mitiche, R. & Bennaçar, N. (2008) De l'eau pour un développement durable : enjeux et stratégies. *Revue des Sciences de l'Eau*, N° spécial 20eme anniversaire, 137-146.
- [92] King, M. L. (1987) Testing for autocorrelation in linear regression models : A survey. In (King & Giles 1987), 19–73.
- [93] Kottegoda, N. T. (1980) *Stochastic Water Resources Technology*. Macmillan Press, London, UK.
- [94] Koutsoyiannis, D. & Georgakakos, A. (2006) Lessons from the long flow records of the Nile: determinism vs indeterminism and maximum entropy. In: *20 Years of Nonlinear Dynamics in Geosciences* (ed. by Tsonis, A. A. & Elsner, J. B.) (Rhodes, Greece, 11–16 June 2006; Aegean Conferences Series – Vol. 24)
- [95] Koutsoyiannis, D. & Montanari, A. (2007) Statistical analysis of hydroclimatic time series: Uncertainty and insights. *Water. Resour. Res.* **43**(5), W05429,1–9.
- [96] Koutsoyiannis, D. (2000) A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series. *Water. Resour. Res.* **36**(6), 1519–1533.
- [97] Koutsoyiannis, D. (2001) Coupling stochastic models of different time scales. *Water Resour. Res.* **37**(2), 379–392.
- [98] Koutsoyiannis, D. (2002) The Hurst phenomenon and fractional Gaussian noise made easy. *Hydrol. Sci. J.* **47**(4), 573–596.
- [99] Koutsoyiannis, D. (2005a) Uncertainty, entropy, scaling and hydrological stochastics, 1. Marginal distributional properties of hydrological processes and state scaling. *Hydrol. Sci. J.* **50**(3), 381–404.
- [100] Koutsoyiannis, D. (2005b) Uncertainty, entropy, scaling and hydrological stochastics, 2. Time dependence of hydrological processes and time scaling. *Hydrol. Sci. J.* **50**(3), 405–426.
- [101] Koutsoyiannis, D. (2006) On the quest for chaotic attractors in hydrological processes. *Hydrol. Sci. J.* **51**(6), 1065–1091.
- [102] Koutsoyiannis, D., Huaming, Y. & Georgakakos, A. (2008) Medium-range flow prediction for the Nile: a comparison of stochastic and deterministic methods. *Hydrological sciences–Journal–des Sciences Hydrologiques* **53**(1).
- [103] Lachtermacher, G. & Fuller, J.D. (1994) Back propagation in hydrological time series forecasting. In: *Stochastic and statistical methods in hydrology and environmental engineering*. Time series analysis in hydrology and environmental engineering. (Éditeurs: Hipel, K.W. , McLeod, A.I., Panu, U.S. & Singh, V.P.) Kluwer Academic Publishers, Dordrecht, Hollande, 229–242.
- [104] Langousis, A. & Koutsoyiannis, D. (2006) A stochastic methodology for generation of seasonal time series reproducing overyear scaling. *J. Hydrol.* **322**, 138–154.
- [105] Latif, A.M. (1999) A Kalman filter approach to multisite precipitation modeling in meteorology. Thèse de Ph. D., Istanbul Technical University, Institute of Science and Technology, 125 pp.
- [106] Lettenmaier, D.P., & Wood, E.F. (1993) Hydrologic forecasting. Dans: *Handbook of hydrology*. (ed. Par Maidment, D.R. McGraw-Hill), New York, USA.
- [107] Lewis, F. L. (1986) *Optimal Estimation with an Introductory to Stochastic Control Theory*. John Wiley & Sons, Inc.
- [108] Lutkepohl, H. (1991) *Introduction to Multiple Times Series*. Springer, Berlin.
- [109] Maddala, G.S. (1977) *Econometrics*. McGraw Hill, New York, USA.

- [110] Makridakis, S., Wheelwright, S.C., & Hyndman, R.J.(1998) *Forecasting : methods and application*. Wiley, J. and sons, New-York, USA.
- [111] Mandelbrot, B.B., & Wallis, J. (1968) N. Joseph and operational hydrology, *Water Resources Research*, **4**, 909–918.
- [112] Maybeck, P. S. (1979) *Stochastic models, estimation, and control* (Vol. 1). Academic Press, Inc., New York, USA.
- [113] McLaughlin, D. (2002) An integrated approach to hydrologic data assimilation: interpolation, smoothing, and filtering. *Adv. Water Resour.* **25**, 1275–1286.
- [114] Menahem, M. S. (1983) On the Feasibility of Real-Time Prediction of Aircraft Carrier Motion at Sea. *IEEE* .Vol. AC-**28**, 350-355.
- [115] Montanari, A., Rosso, R. & Taqqu, M. S. (1997) Fractionally differenced ARIMA models applied to hydrologic time series. *Water Resour. Res.* **33**(5), 1035–1044.
- [116] Moradkhani, H., Sorooshian, S., Gupta, H. V. & Houser, P. R. (2005) Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Adv. Water Resour.* **28**, 135–147.
- [117] Muluye, G. & Coulibaly, P. (2007) Seasonal reservoir inflow forecasting with low-frequency climatic indices: a comparison of data-driven methods. *Hydrol. Sci. J.* **52**(3), 508–522.
- [118] Nerlove, M., Grether, D. & Carvalho, J. L. (1979) *Analysis of Economic Time Series : A Synthesis*. Academic Press, New York, USA.
- [119] Neveu, J. (1964) *Bases mathématiques du calcul des probabilités*. Masson, Paris.
- [120] Ngan, P. & Russel, S. O. (1986) Example of Flow Forecasting with Kalman Filter. *Journal of Hydraulic Engineering*, **112**(9), 818-832.
- [121] Ouachani, R., Bargaoui, Z. & Taha, O. (2007) Intégration d'un filtre de Kalman dans le modèle hydrologique HBV pour la prévision des débits. *Hydrol. Sci. J.* **52**(2), 318-337.
- [122] Owen, D. B., ed. (1976) *On the History of Statistics and Probability*. Marcel Dekker, New York, USA.
- [123] Pankratz , A. (1983) *forecasting with univariate Box-Jenkins models*. Wiley, NewYork.
- [124] Pankratz , A. (1991) *forecasting with dynamic regression models*. Wiley, NewYork.
- [125] Papoulis, A. (1991) *Probability, Random Variables, and Stochastic Processes* (third edn). McGraw-Hill, New York, USA.
- [126] Parzen, E. (1974) Some recent advances in time series modelling. *IEEE Transactions on Automatic Control*, AC-**19** , 723–729.
- [127] Pena D., Tiao G. C. & Tsay R. S. (2001) *A Course in Time Series Analysis*. John Wiley, NewYork.
- [128] Pindyck, R.S. & Rubinfeld, D.L. (1981) *Econometric Models and Economic Forecasts*. McGrawHill, New York, USA.
- [129] Pourahmadi, M. (2001) *Foundations of Time Series Analysis and Prediction Theory*. John Wiley, New York, USA.
- [130] Priestley, R. (1981) *Spectral Analysis and Time Series*. Academic Press, London.
- [131] Robinson, P.M. (1994a) Time series with strong dependence. In: Sims, C.A. (Ed.), *Advances in Econometrics*, Sixth World Congress, vol. 1. Cambridge University Press, Cambridge.
- [132] Salas, J. D. (1993) Analysis and modeling of hydrologic time series. In: *Handbook of Hydrology* (ed. by D. Maidment), Chapter 19, 1–19. McGraw-Hill, New York, USA.
- [133] Sargent, T.J. (1987) *Macroeconomic Theory*. Academic Press, Orlando.
- [134] Schlee, F. H *et al.* (1967) Divergence in the Kalman Filter. *AIAAJ*. Vol.**5**, 1114 –1120.
- [135] Schreider, S. Y., Young, P. C. & Jakeman, A. J. (2001) An application of the Kalman filtering technique for streamflow forecasting in the upper Murray Basin. *Math. Comput. Modell.* **33**, 733–743.

- [136] See, L., Solomatine, D., Abrahart, R. & Toth, E. (2007) Hydroinformatics: computational intelligence and technological developments in water science applications – Editorial. *Hydrol. Sci. J.* **52**(3), 391–396.
- [137] Sen, Z. & Latif, A.M. (2002) Multisite Kalman Filtering Application to Turkish Precipitation Records. Proc. International Conference on Environmental problems of the mediterranean region (EPMR), Near-East University, Lefkosa, Turkish Republic of Northern Cyprus.
- [138] Sen, Z. (1980) Adaptive Fourier analysis of periodic stochastic hydrological sequences. *J. Hydrol.*, **46**, 239–249.
- [139] Şen, Z. (1991) Adaptive prediction of hydrologic series by Walsh-Kalman model, *J. Hydrol.* **122**, 221-234.
- [140] Şen, Z., Altunkaynak, A., ASCE, M. & Özger, M. (2004) Sediment concentration and its prediction by perceptron Kalman filtering procedure. *J. Hydraul. Eng.* **130** (8), 816-826.
- [141] Sorenson, H. W. (1970) Least-Squares estimation: from Gauss to Kalman. *IEEE Spectrum*, **7**, 63-68.
- [142] Sorenson, H.W. (1985) Kalman Filtering Theory and Applications. *IEEE*, 90-126.
- [143] Stedinger, J. R., Vogel, R. M. & Foufoula G. (1993) Frequency analysis of extreme events. In: *Handbook of Hydrology* (ed. by Maidment, D. R.), Chapter 18. McGraw-Hill, New York, USA.
- [144] Stephenson, D. B., Pavan, V. & Bojariu, R. (2000) Is the North Atlantic Oscillation a random walk? *Int. J. Climatol.* **20**, 1–18.
- [145] Stone, C.J. (1982) Local asymptotic admissibility of a generalization of Akaike's model selection rule. *Journal of the Royal Statistical Society* (London), A-**34** : 123–234.
- [146] Tang, Z., Almeida, C., & Fishwick, P.A. (1991) Time series forecasting using neural networks vs. Box–Jenkins methodology. *SIMULATIONS*, **57**(5), 303–310.
- [147] Thomas, H. A. & Fiering, M. B. (1962) Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. In: *Design of Water Resource Systems* (ed. by A. Maass *et al.*), 459–493. Harvard University Press, Cambridge, USA.
- [148] Todini, E. & Boillot, D. (1975) A rainfall- runoff kalman filter model. Dans: *System Simulation in Water Resources* (ed. par Vansteenkiste, G.C.), Amsterdam , North-Holland.
- [149] Tong, H. (1983) *Threshold models in nonlinear time series analysis*. Springer-Verlag, Berlin, Germany.
- [150] Tong, H., & Lim, K.S. (1980) Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society*, B-**42** , 245–292.
- [151] Troch, A., Paniconi, C. & McLaughlin, D. B. (2003) Catchment-scale hydrological modeling and data assimilation. *Adv. Water Resour.* **26**, 131–135.
- [152] Vanmarcke, E. (1983) *Random Fields*. The MIT Press, Cambridge, Massachusetts, USA.
- [153] Vaté, M. (1993) *Statistique chronologique et prévision*. Economica, Paris.
- [154] Von Storch, H., von Storch, J-S. & Müller, P. (2001) Noise in the climate system—ubiquitous, constitutive and concealing. In: *Mathematics Unlimited—2001 and Beyond* (ed. by Engquist, B. & Schmid, W.). Springer, Berlin, Germany.
- [155] Wallace, J.N. & Clarke, R. (1983) The Application of Kalman Filtering Estimation Techniques in Power Station Control Systems. *IEEE*, Vol.AC-**28**, NO.3, 416-426.
- [156] Weerts, A. & El Serafy, G. (2006) Particle filtering and ensemble Kalman filtering for State updating with hydrological conceptual Rainfall Runoff Models. *Water Resour. Res.* **42**, W09403, doi:10.1029/2005WR004093.
- [157] Welch, G. & Bishop, G. (2003) An introduction to the Kalman filter. TR 95-041 Department of Computer Science, University of North Carolina, USA.
- [158] Welch, G., Bishop, G., Vicci, L., Brumback, S., Keller, K., & Colucci, D. (2001) High-Performance Wide-Area Optical Tracking -The HiBall Tracking System. *Presence:Teleoperators and Virtual Environments*, **10**(1).

- [159] Winters, P.R. (1960) Forecasting sales by exponentially weighted moving averages. *Management Sciences*, **6**, 324–342.
- [160] WMO (1994) Guide to hydrological practices. WMO-No. 168, Genève, Suisse.
- [161] Wood, E. & O'Connell, P. E. (1985) Real-time forecasting. Dans: *Hydrological Forecasting* (ed. par Anderson, M. G. & Burt, T. P.), 505–558. Wiley & Sons, Inc., New York, USA.
- [162] Yevjevich, V. (1971) *Stochastic Processes in Hydrology*. Water Resources Pub., Fort Collins, Colorado, USA.
- [163] Yevjevich, V. (1972) *Probability and Statistics in Hydrology*. Water Resources Publications, Fort Collins, Colorado, USA.