

**REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET  
POPULAIRE**

MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE  
SCIENTIFIQUE

*Ecole Nationale Polytechnique*

Département Hydraulique



**Mémoire de Master**

Thème :

Hybridation Analyse en Composantes Principales Neuro-flou  
pour l'optimisation de l'énergie électrique de la STEP de

Boumerdes

Soraya ABDELAZIZ

Sous la direction de Mr. A.LEFKIR

Docteur ENSTP

**ENP 2016**

E.N.P. 10, Avenue Hassen Badi, BP 182 El-Harrach, Alger 16200

*Je dédie ce travail à ma petite sœur Sonia en lui souhaitant la guérison*

## **REMERCIEMENTS :**

Je tiens tout d'abord à exprimer mes remerciements les plus chaleureux à Monsieur A.LEFKIR qui m'a encadrée dans la réalisation de ce mémoire et m'a fait bénéficier de ses conseils éclairés qui m'ont grandement facilité ma tâche.

De même, je remercie l'ensemble du jury qui a bien voulu étudier mon mémoire et l'évaluer avec toute la compétence qui est la sienne.

A cette occasion, je mentionne aussi avec une particulière affection mon père, monsieur Abdelaziz Antar, ma mère, Abdelaziz Fella, mon frère, Fouzi et ma sœur Radhia et son mari Hamza qui, indirectement mais efficacement m'ont apporté leur soutien moral et ont contribué à me soutenir dès l'entame de mon mémoire.

Je remercie également le reste de ma famille avec à leur tête mon oncle Aiouache Salah dont l'aide généreuse et le soutien ne se sont jamais démentis de même que sa femme Latifa, mon autre oncle Abdelaziz Mohamed et son épouse, avec une mention particulière à mon adorable grand-mère Habiba.

Soraya

## ملخص

إن تنقية المياه المستعملة تعتبر وسيلة ضرورية للحد من ارتفاع تركيز المواد الملوثة. المشاكل البيئية التي يسببها صرف المياه من محطات التنقية وكذا ارادة التحسين والحفاظ على جودة الوسط المستقبل تستلزم رفع أداء المحطات من ناحيتي الجودة والفعالية

الهدف من هذه الدراسة هو مساعدة مسيري المحطة على إتخاذ القرار الأمثل من حيث إستهلاك الطاقة بواسطة نموذج التقارن للتحليل المكون الرئيسي و الذكاء الإصطناعي

**الكلمات المفتاحية :** معالجة مياه الصرف الصحي، تحليل المكون الأساس ، التقارن للتحليل المكون الأساسي و الذكاء الإصطناعي، تصميم الطاقة

## Abstract

The wastewater treatment is essential today to cope with the increasing concentrations of pollutants. The awareness of the environmental problems caused by discharges from wastewater treatment plants and the will to improve and preserve the quality of receiving waters has led the obligation to increase the performance of treatment plants , both in terms of quality and reliability . The purpose of our work is to establish a Neuro fuzzy CPA coupling model based on the best experiences to optimize electrical energy and help managers to reduce operating costs.

**keyword :** Wastewater treatment , principal component analysis (PCA) , coupling PCA Neuro-Fuzzy , energy optimization.

## Résumé

L'épuration des eaux usées est aujourd'hui indispensable pour faire face à l'augmentation des concentrations substances polluantes. La prise de conscience des problèmes environnementaux causés par les rejets des stations d'épuration et la volonté d'améliorer et de préserver la qualité des milieux récepteurs a entraîné l'obligation d'augmenter les performances des stations d'épuration, tant en terme de qualité de traitement qu'en terme de fiabilité. L'objet de notre travail, est d'établir un modèle couplage ACP Neuro flou basé sur les meilleures expériences afin d'optimiser l'énergie électrique et d'aider les gérants à réduire les coûts d'exploitation.

**Mot clé :** Epuration des eaux usées, analyse en composantes principales (ACP), couplage ACP Neuro-floue, optimisation de l'énergie.

## Sommaire

Liste des figures

Liste des tableaux

Introduction générale.....	9
1. Théorie de l'Analyse en composantes principales et Neuro-flou.....	11
1.1. Analyse en Composantes Principales.....	11
1.2. Objectifs.....	14
1.3. Procédé d'application de l'ACP.....	14
1.4. Théorie de la Neuro-flou.....	15
1.5. Introduction.....	15
1.6. Identification floue.....	15
1.7. Définitions de base.....	16
1.7.1. Les réseaux de neurones.....	16
1.7.2. La logique floue.....	16
1.7.3. Les sous-ensembles flous.....	16
1.7.4. La fonction d'appartenance.....	17
1.7.5. Les variables linguistiques.....	17
La fuzzification.....	18
La défuzzification.....	18
1.7.6. Synoptique d'un système flou.....	18
Le critère de Nash Sutcliffe (Nash et Sutcliffe, 1970).....	19
L'erreur quadratique moyenne (RMSE: Root Mean Square Error).....	19
Percent bias PBIAS.....	20
Le coefficient de détermination.....	20
2. Couplage ACP Neuro-flou pour l'optimisation de l'énergie de la STEP de Boumerdes....	22
2.1. Introduction.....	22
2.2. La modélisation de l'énergie électrique en utilisant l'Analyse en composantes principales.....	22
2.2.1. Elaboration du modèle de l'Analyse en composantes principales.....	23
2.2.2. Interprétation de la représentation graphique et des critères de validation.....	25
2.3. La modélisation en utilisant la Neuro-flou.....	26
2.3.1. Elaboration du modèle Neuro-flou.....	26
2.3.2. Interprétation de la représentation graphique et des critères de validation.....	27

2.4. La modélisation en utilisant le couplage analyse en composantes principale Neuro-flou.	28
2.4.1. Elaboration du modèle couplage ACP Neuro-flou.	28
2.4.2. Interprétation de la représentation graphique et des critères de validation.	30
2.5. Conclusion.	31
3. Optimisation de l'énergie électrique par Couplage ACP Neuro-flou.	32
3.1. Introduction.	32
3.2. La sélection des données sans prendre en compte les « WQI ».	32
3.2.1. 1ere sélection.	32
3.2.2. 2eme sélection.	33
3.3. La sélection des données en prenant en compte les « WQI ».	34
3.3.1. 1ere sélection (Avec NH4).	34
Elaboration du modèle d'optimisation.	35
3.3.2. La deuxième sélection (Sans NH4).	38
Elaboration du modèle d'optimisation.	39
3.4. Le gain en énergie électrique à partie du modèle couplage ACP Neuro flou de la deuxième sélection.	42
Conclusion générale	45
Bibliographie.	46

## Liste des figures

Figure 1 : résultats de la régression linéaire du modèle ACP.....	25
Figure 2 : résultats de la régression linéaire du modèle NF.....	27
Figure 3 : Résultats de la régression linéaire du modèle couplage ACP-NF.....	30
Figure 4 : Représentation de l'intervalle de la première sélection.....	32
Figure 5 : Régression linéaire (Energie observée brute / Energie simulée brute) calage.....	37
Figure 6 : Régression linéaire (Energie observée brute / Energie simulée brute) validation...	38
Figure 7 : Régression linéaire (Energie observée brute / Energie simulée brute) calage.....	41
Figure 8 : Régression linéaire (Energie observée brute / Energie simulée brute) validation...	42
Figure 9 : Comparaison entre l'énergie simulée et l'énergie observée.....	43
Figure 10 : Gain moyen annuel journalier en matière d'énergie.....	43

## **Liste des tableaux**

Tableau 1 : Critères de validation du modèle ACP.....	25
Tableau 2 : Critères de validation du modèle NF.....	27
Tableau 3 : Critères de validation du modèle couplage ACP-NF.....	30
Tableau 4 : Critère de validation de la première sélection.....	37
Tableau 5 : Critères de validation de la deuxième sélection.....	41



## Introduction générale

Dans la première partie de notre travail, nous avons fait une modélisation de l'énergie électrique de la STEP de Boumerdes par trois différents modèles : Analyse en Composantes Principales, Neuro-flou et Couplage ACP Neuro-flou.

Pour cela nous avons commencé par faire un choix de variables en nous basant sur l'Analyse en Composantes Principales (ACP). Les résultats obtenus nous ont permis de faire notre choix de variables sur lesquelles se basera la suite de notre étude.

Afin d'obtenir le meilleur modèle qui représentera notre phénomène de traitement d'eaux usées, nous avons divisé notre matrice de données en deux parties : une partie sera réservée à l'apprentissage de notre modèle et contiendra (318) observations afin de permettre au modèle de pallier tous les cas possibles rencontrés lors de la validation, et une partie réservée à la validation contenant (51) observations qui sont celles de l'année 2013 qui est l'année de transition entre le travail dans les normes et le travail en surcharge de la STEP.

En premier lieu, nous avons établi un modèle linéaire par Analyse en Composantes Principales (ACP). En le validant nous n'avons pas obtenu des résultats satisfaisants et cela par les deux méthodes : Graphique et numérique. Afin d'essayer d'améliorer les résultats de la modélisation de ce phénomène épuratoire très complexe, nous avons utilisé le modèle ANFIS intégré dans MATLAB, ce qui a permis d'améliorer les résultats. La vérification graphique ainsi que la vérification numérique ont montré que le modèle ANFIS représente mieux notre phénomène.

Dans le but de chercher à minimiser les coûts d'épuration des eaux, tout en maintenant un niveau de performance adéquat, nous avons pensé à combiner entre les avantages de l'Analyse en Composantes Principales (ACP) et le modèle ANFIS afin d'avoir une meilleure modélisation de notre phénomène. Les résultats obtenus ont permis de montrer que notre dernier modèle est le meilleur modèle permettant de modéliser l'énergie électrique de la STEP de Boumerdes.

Dans cette deuxième partie du travail, nous avons commencé par tester nos trois modèles sur d'autres données de calage et de validation, en changeant à chaque fois l'année cette dernière, afin d'essayer d'améliorer d'avantage nos résultats.

Ensuite en utilisant les résultats de l'optimisation de l'énergie électrique de la STEP de Boumerdes, nous avons essayé d'établir un modèle de couplage ACP Neuro-flou pour chacune de nos sélections de meilleures expériences.

Les variables explicatives présent en considération dans cette deuxième partie de validation sont : Indice de pollution (IP), Débit recirculé (DR), Température (T), Débit (D), Abatement de la pollution (AB).

# 1. Théorie de l'Analyse en composantes principales et Neuro-flou

## 1.1. Analyse en Composantes Principales

L'Analyse en Composante Principale (ACP) est une méthode d'analyse des données dite méthode factorielle. Cette méthode permet la présentation géométrique d'un tableau rectangulaire composé de  $p$  variables et de  $n$  individus.

Les données sur lesquelles on applique une ACP sont obtenues soit d'une observation d'un échantillon ou de toute une population.

A partir de la représentation graphique, on peut définir une liaison linéaire entre certaines variables formant un groupe entre elles et donc définir les variables qui sont fortement corrélées ou au contraire, celles aussi qui ne sont pas corrélées entre elles.

La projection des individus sur des axes factoriels permet de voir les individus qui se ressemblent et ceux qui se distinguent.

Enfin, comme pour toute méthode descriptive, réaliser une ACP n'est pas une fin en soi. L'ACP servira à mieux connaître les données sur lesquelles on travaille, à détecter éventuellement des valeurs suspectes, et aidera à formuler des hypothèses qu'il faudra étudier à l'aide de modèles et d'études statistiques.[1]

- Notion d'individu et de caractère

### **Individu**

L'individu est défini comme étant un élément de la population étudié. Il peut être une année d'observation ou une autre unité de temps.

Un individu est défini dans l'espace  $R^P$ , par le vecteur :

$$X_i = \{X_{ij} / j = 1 \text{ à } P\} \quad (I.1)$$

Le terme  $X_{ij}$  est un nombre réel qui représente la mesure de la variable  $X_j$  relevée sur l'individu  $i$ . [17]

### **Caractère**

Sur un individu, on relève un certain nombre de caractères, dits aussi variables, désignant en général un paramètre intervenant dans un phénomène complexe à étudier. [2]

Le caractère (ou variable) " $j$ " est décrit par le vecteur appartenant à  $R^N$  :

$$X_j = \{X_{ij} / i = 1 \text{ à } N\} \quad (I.2)$$

- **Définition algébrique**

L'analyse en composantes principales se fait sur une matrice de P variables et N individus. Cette dernière est dite matrice de données, qui résulte du croisement « NxP ».

La matrice de données est notée par [X] et elle est décrite comme suit :

$$[X] = \begin{pmatrix} X_{11} & X_{12} & \dots & \dots & X_{1j} & \dots & \dots & X_{1P} \\ X_{21} & X_{22} & \dots & \dots & X_{2j} & \dots & \dots & X_{2P} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ X_{i1} & X_{i2} & \dots & \dots & X_{ij} & \dots & \dots & X_{iP} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ X_{N1} & X_{N2} & \dots & \dots & X_{Nj} & \dots & \dots & X_{NP} \end{pmatrix} \quad (I.3)$$

Une fois qu'on a la matrice des données, on pourrait définir les paramètres statistiques pour chaque variable :

**La moyenne**

$$\overline{X}_j = \frac{1}{N} \sum_{i=1}^N X_{ij} \quad (I.4)$$

$\overline{X}_j$  : Moyenne de la j<sup>ième</sup> variable.

**Ecart type**

$$S_j = \left[ \frac{1}{N} \sum_{i=1}^N (X_{ij} - \overline{X}_j)^2 \right]^{1/2} \quad (I.5)$$

S<sub>j</sub> = Ecart type de la j<sup>ième</sup> variable.

N : Nombre d'individus.

**Le coefficient de covariance**

$$Cov(X_j, X_k) = \frac{1}{N} \sum_{i=1}^N (X_{ij} - \overline{X}_j) * (X_{ik} - \overline{X}_{ik}) \quad (I.6)$$

**Le coefficient de corrélation**



Un second axe est obtenu après projection des N points sur un hyperplan orthogonal au premier axe, tel que la dispersion des projections des N points sur celui-ci soit toujours maximale, et le processus se réitère P fois. [2]

## 1.2.Objectifs

L'Analyse en Composantes Principales a pour objectifs :

- ✓ Avoir une corrélation entre les variables grâce à une représentation des données brutes et fournir des outils simples et lisibles.
- ✓ Essayer d'examiner les liens existants entre les variables afin de faciliter leur interprétation.
- ✓ Essayer de réduire au maximum le nombre de dimensions étudiées tout en gardant l'ensemble original et en utilisant les relations détectées entre les variables.

## 1.3.Procédé d'application de l'ACP

Afin d'appliquer une Analyse en Composantes Principales, il faut procéder comme suit :

1. Formation de la matrice des données brutes [X] (composée par le croisement de N individus x P variables).
2. Calcul des paramètres statistiques :(moyenne, écart type).
3. Calcul des variables centrées réduites

$$Y_j = \frac{X_j - \overline{X_j}}{\sigma_{X_j}}$$

4. Calcul des coefficients de corrélation :

$$Cor(X_j, X_k) = \frac{Cov(X_j, X_k)}{S_j * S_k} = \frac{\sum_{i=1}^N (X_{ij} - \overline{X_j}) * (X_{ik} - \overline{X_k})}{\left[ \sum_{i=1}^N (X_{ij} - \overline{X_j})^2 * \sum_{i=1}^N (X_{ik} - \overline{X_k})^2 \right]^{1/2}}$$

5. Calcul des valeurs propres et des vecteurs propres de la matrice de corrélation obtenue.
6. Calcul des composantes principales, avec :  
CP = matrice des variables centrées réduites (NxP) x matrice des vecteurs propres (NxN).

7. Calcul des composantes principales centrées réduites

$$CP'(i, j) = \frac{CP(i, j)}{\sqrt{Vp(j)}}$$

Avec  $V_P(j)$  : la valeur propre correspondant à l'individu  $j$ .

8. Calcul de la matrice de corrélation entre CP'et XCR ( $\text{cor}(\text{CP}', \text{XCR})$ ).

9. Tracé des cercles de corrélation.

## **1.4. Théorie de la Neuro-flou**

### **1.5. Introduction**

Ces dernières années, le nombre et la variété des applications de la logique floue ont augmenté de façon significative, elle a connu un succès remarquable dans la description des systèmes complexes. Ces applications vont des produits de consommation à la reproduction des phénomènes naturels ...

Pour les phénomènes naturels la problématique de prédiction peut être formalisée de la manière suivante : étant donnée une situation caractérisée à l'instant « t » (entrée), l'outil de prédiction choisi doit prédire l'évolution de cette situation à l'instant « t+prévision » (sortie). Des développements actuels sur ce type de problématique montrent que les performances des neuro-flous dépassent celles d'autres méthodes aussi bien en terme de précision des résultats que d'efficacité d'apprentissage (Wang W.Q.et al.,2004) surtout pour les systèmes complexes à comportement non stationnaire et non linéaire.

Dans cet ensemble, notre travail traite plus spécifiquement du système ANFIS (Adaptative Neuro-Fuzzy Inference System) proposé par (Jang, 1993).

### **1.6. Identification floue**

L'identification floue est un domaine extrêmement vaste, où chaque jour ou presque, une nouvelle approche est proposée. Certaines approches confinent à la théorie des réseaux de neurones.

L'idée générale d'un modèle neuro-flou est d'intégrer dans un seul concept les avantages des deux domaines : Logique floue et Réseaux de neurones.

- De la logique floue : Introduction de connaissances à priori afin de réduire l'espace des paramètres à optimiser.
- Des réseaux de neurones : utilisation de la capacité d'apprentissage afin d'améliorer les performances.

Ainsi, nous sommes dans l'obligation de présenter dans ce qui suit plusieurs définitions qui nous aiderons à la compréhension du système ANFIS.

## 1.7. Définitions de base

### 1.7.1. Les réseaux de neurones

Les réseaux de neurones artificiels sont des réseaux fortement connectés de processeurs élémentaires fonctionnant en parallèle. Chaque processeur élémentaire calcule une sortie unique sur la base des informations qu'il reçoit. Toute structure hiérarchique du réseau est évidemment un réseau, elle est dotée d'une fonction particulière.

Chaque neurone des différents réseaux reçoit un nombre variable d'entrées en provenance de neurones amont. A chacune de ces entrées est associé un poids  $w$  représentatif de la force de connexion. Chaque processeur élémentaire est doté d'une sortie unique, qui se ramifie ensuite pour alimenter un nombre variable de neurones aval.

### 1.7.2. La logique floue

La logique floue est une extension de la logique booléenne par Lofti Zadeh en 1965, donc elle représente une généralisation des ensembles classiques. En introduisant la notion de degré dans la vérification d'une condition, permettant ainsi d'avoir un état que vrai ou faux. La logique floue confère une flexibilité appréciable aux raisonnements de l'utilisateur, ce qui rend possible la prise en compte des imprécisions et des incertitudes.

- **Imprécisions** désigne les connaissances qui ne sont pas perçues ou définies nettement.
- **Incertitudes** désigne les connaissances dont la validité est sujette à question.

Un des intérêts de la logique floue pour formaliser le raisonnement humain est que les règles sont énoncées en langage naturel.

### 1.7.3. Les sous-ensembles flous

La logique floue repose sur la théorie des ensembles flous, qui est une généralisation de la théorie des ensembles classiques [Zadeh, 1965]. Par abus de langage, nous utiliserons indifféremment les termes sous-ensembles flous et ensembles flous.

*Soit  $x$  un ensemble. Un sous-ensemble flou  $A$  de  $x$  est caractérisé par une fonction d'appartenance.  $F^A = x \rightarrow [0, 1]$ . La fonction d'appartenance est notée  $\mu_A$ .*

*Cette fonction d'appartenance est l'équivalent de la fonction caractéristique d'un ensemble classique.*



*Le sous-ensemble flou est défini sur un univers de discours X. L'univers de discours est l'ensemble des valeurs que peut prendre la variable floue x est  $\mu_A(x)$  est le degré d'appartenance de l'élément x à l'ensemble flou A.[3]*

#### **1.7.4. La fonction d'appartenance**

La fonction d'appartenance représente le degré d'appartenance d'une valeur à un état donné, elle peut avoir différentes formes (en triangle, en trapèze...).

#### **1.7.5. Les variables linguistiques**

Soit x une variable (débit solide), X la plage de valeur de la variable et  $A_x$  un ensemble fini ou infini de sous-ensembles flous. Une variable linguistique correspond au triplet (x, X,  $A_x$ ).

- **Le raisonnement en logique floue**

En logique classique, les raisonnements sont de la forme :

$$\begin{cases} \text{Si } p \text{ alors } q \\ p \text{ vrai alors } q \text{ vrai} \end{cases}$$

En logique floue, le raisonnement flou, également appelé raisonnement approximatif, se base sur des règles floues qui exprimées en langage naturel en utilisant les variables linguistiques.

Une règle floue aura cette forme :

*Si  $x \in A$  et  $y \in B$  alors  $z \in C$ , avec A, B et C des ensembles flous.*

Par exemple :

Si (le débit est élevé), alors (la concentration des MES est élevée).

La variable « débit » appartient à l'ensemble flou « élevé » à un degré qui dépend du degré de validité de la prémisse, autrement dit du degré d'appartenance de la variable « débit » à l'ensemble flou « élevé ». L'idée sous-jacente est que plus les propositions en prémisse sont vérifiées, plus l'action préconisée pour les sorties doit être respectée.

- **Système d'inférence flou (SIF)**

Une simple règle floue ne suffit pas à elle seule de représenter un problème complexe. Un ensemble de règles doivent être établies pour bien décrire notre système et simplifier son utilisation. L'association de ces ensembles de règles avec des mécanismes de décision permettent de déduire de nouvelles connaissances, elles constituent des systèmes d'inférences floues.

Un système d'inférence flou est essentiellement composé de deux blocs fonctionnels différents :

- Une base de connaissance constituée d'une base de règles floues et d'une base de données définissant les fonctions d'appartenance des ensembles flous.
- Un mécanisme d'inférence flou qui détermine la sortie du système sous la forme d'un ensemble flou.

On y ajoute deux blocs :

- Une interface de fuzzification qui transforme des entrées ponctuelles en ensembles flous
- Une interface de défuzzification qui transforme la sortie floue en sortie ponctuelle.

#### ➤ **La fuzzification**

C'est la première étape du traitement flou, les variables d'entrées se trouvent caractérisées à l'aide des sous-ensembles flous. La fuzzification est l'étape où les variables d'entrées « physiques » subissent une conversion artificielle et se transforment en des variables « linguistique ou floues ».

#### ➤ **La défuzzification**

La dernière étape est celle de la défuzzification, elle transforme la partie floue résultant de l'inférence en une valeur numérique. Le but de cette étape est de transformer l'ensemble flou en valeurs non floues et de trouver une sortie précise correspondant le mieux à la caractérisation floue.

La technique la plus utilisée est celle du centre de gravité. Dans notre cas la valeur simulée (concentration en MES) correspond à la projection du centre de gravité (défuzzification centroïde) de la surface d'inférence sur l'abscisse de la série des concentrations observées (Lefkir et al, 2006).

### **1.7.6. Synoptique d'un système flou**

- Les inputs ;

- La fuzzification elle correspond aux deux variables linguistiques ;
- Le moteur d'inférence est constitué du choix des opérateurs flous;
- La base de connaissances floues est l'ensemble des règles floues;
- La défuzzification;
- Les outputs correspondent à la décision finale ;

Ainsi, la logique floue permet de mettre en place des systèmes d'inférence dont les décisions sont sans discontinuités, flexibles et non linéaires, plus proche du comportement humain que ne l'est la logique classique.

### ➤ **Le critère de Nash Sutcliffe (Nash et Sutcliffe, 1970)**

Il donne une indication sur la robustesse (pouvoir prédictif) des modèles utilisés. Ainsi, il est utilisé comme un outil d'aide à la décision.

Le critère est le suivant :

$$NASH = 1 - \frac{\sum_{t=1}^T (Q_0^t - Q_m^t)^2}{\sum_{t=1}^T (Q_0^t - \overline{Q_0})^2}$$

Avec :  $Q_0$  représente les valeurs observées ;

$Q_m$  sont les valeurs simulées est  $Q_0^t$  est l'observé à l'instant  $t$  ;

$\overline{Q_0}$  est la moyenne des valeurs observées.

La valeur du NASH varie entre « -∞ » à « 1 » et est généralement présenté en pourcentage.

- Une efficacité de NASH=1 correspond à une modélisation parfaite où les simulées = les observées.
- Une efficacité de NASH=0 indique que les prédictions du modèle sont aussi précises que la moyenne des données observées.
- Une efficacité de NASH<0 indique que les valeurs simulées conduisent à de moins bons résultats et que la moyenne observée est meilleur prédicateur que le modèle.

### ➤ **L'erreur quadratique moyenne (RMSE: Root Mean Square Error)**

La mesure de l'erreur quadratique moyenne est aussi un critère souvent utilisé pour mesurer l'écart entre la simulation et l'observation, en donnant notamment une idée de la

dispersion entre les deux et cela dans l'unité des variables étudiées. Plus le RMSE sera faible et plus l'écart entre la simulation et les observations sera faible.

Le critère est le suivant :

$$RMSE = \sqrt{\frac{1}{n_0} \sum_{t=1}^T (Q_m^t - Q_0^t)^2}$$

Le RMSE est couramment utilisé pour d'écrire l'erreur statistique. La performance du modèle est retenue selon une valeur minimale du RMSE.

### ➤ **Percent bias PBIAS**

Le pourcentage du bias mesure la tendance moyenne des données simulées à être plus grandes ou plus petites que leurs homologues observées. Le PBIAS prend sa valeur optimale quand il est égal à zéro.

Si le PBIAS <0 il indique que les valeurs simulées sont plus élevés que ceux observés, ainsi on peut dire qu'il y'a une surestimation des valeurs.

Si le PBIAS >0 il indique que les valeurs simulées sont inférieurs à ceux observés, ainsi on peut dire qu'il y'a une sous-estimation des valeurs par le modèle.

$$PBIAS = \frac{\sum_{t=1}^T (Q_0^t - Q_m^t) * 100}{\sum_{t=1}^T Q_0^t}$$

### ➤ **Le coefficient de détermination**

Il est défini comme étant la valeur élevée au carré du coefficient de corrélation de Pearson. Donc il estime la dispersion entre la série des observées et des simulées, plus il sera important plus l'accord entre les deux séries de valeurs le sera aussi.

Il nous informe également sur le degré de corrélation entre deux variables à étudier, si sa valeur est égale à « 0 » cela signifie qu'il n'y a aucune corrélation entre les deux variables.

La valeur maximale est de « 1 » et elle représente la valeur optimale.

Le R<sup>2</sup> est donné par la relation :

$$R^2 = \left[ \frac{\sum_{t=1}^T (Q_m^t - \bar{Q}_0) * (Q_m^t - \bar{Q}_m)}{\sqrt{\sum_{t=1}^T (Q_0^t - \bar{Q}_0)^2} * \sqrt{\sum_{t=1}^T (Q_m^t - \bar{Q}_m)^2}} \right]$$

L'inconvénient du  $R^2$  est qu'il quantifie la dispersion des points "observés-simulés" et ne donne aucune information sur la qualité du modèle, donc même si les valeurs simulées sont sous-estimées ou surestimées mais que leur dispersion est identique à celle des observées, on aura un  $R^2$  proche de « 1 ». Comme solution à ce problème on définit ci-dessous le coefficient de détermination pondéré.

## **2.Couplage ACP Neuro-flou pour l'optimisation de l'énergie de la STEP de Boumerdes**

### **2.1.Introduction**

La modélisation par régression est une technique statistique permettant de modéliser la relation qui existe entre des variables explicatives (notées  $X_i$ ) et une variable à expliquer (notée  $Y$ ).

Dans le projet de fin d'études intitulé « Couplage Analyse en composantes principales Neuro-flou pour la modélisation de l'énergie électrique de la STEP de Boumerdes », nous avons établi trois différents modèles soient :

- Régression par ACP
- Régression par Neuro-Floue
- Régression par couplage ACP Neuro-floue

Les résultats obtenus par ces trois différents modèles nous ont permis de conclure sur le meilleur modèle à utiliser pour la modélisation de l'énergie électrique et cela en se basant sur les différents critères de validation ainsi que la représentation graphique.

### **2.2.La modélisation de l'énergie électrique en utilisant l'Analyse en composantes principales**

L'Analyse en Composantes Principales (ACP) est une méthode statistique qui permet de réduire la dimension d'une matrice de données. En effet, elle transforme un premier jeu de données en un second jeu de plus petite dimension composé de nouvelles variables qui sont des combinaisons linéaires des variables originelles.[1]

Dans un premier lieu, nous avons centré et réduit notre matrice de données (calage et validation) pour ce faire, on a pris la moyenne et l'écart type de la partie calage uniquement car on suppose qu'on ne possède pas les données de la partie validation.

Une analyse en composante principale (ACP) a été effectuée seulement sur les données de la partie calage de la STEP. A partir de la matrice des cosinus carrés des variables on a pu déduire que :

- L'axe F1 explique respectivement le rabattement des MES, DBO5 et DCO.
- L'axe F2 explique l'énergie.
- L'axe F3 explique le rabattement de NH4.

### 2.2.1. Elaboration du modèle de l'Analyse en composantes principales

Dans le but d'obtenir la matrice de passage, on a procédé comme suit :

- ✓ A partir des résultats de l'ACP effectuée sur les cinq paramètres (MES, DBO5, DCO, NH4 et Energie) et depuis le tableau des coordonnées des observations, on a effectué à l'aide de XLSTAT une régression multiple entre chaque axe (F1, F2, F3, F4 et F5) et les données centrées réduites (MES, DBO5, DCO, NH4 et Energie) et cela axe par axe.
- ✓ Une fois qu'on obtiendra l'équation de chaque axe en fonction des paramètres centrés réduits, on aura une matrice dite A.
- ✓ A l'aide du logiciel Matlab, on calculera l'inverse de A et qui doit être égale à sa transposée.
- ✓ L'inverse de la matrice A est notre matrice de passage.

Les résultats obtenus de cette démarche sont les suivants :

- ✓ Les équations obtenues à partir de la régression multiple de chaque axe sont les suivantes :

$$F1 = 0,453 * RMES \text{ CR} + 0,436 * RDBO5 \text{ CR} + 0,564 * RDCO \text{ CR} + 0,415 * RNH4 \text{ CR} + 0,339 * \text{Energie CR}$$

$$F2 = 0,107 * RMES \text{ CR} - 0,595 * RDBO5 \text{ CR} - 0,14 * RDCO \text{ CR} + 0,056 * RNH4 \text{ CR} + 0,784 * \text{Energie CR}$$

$$F3 = -0,632 * RMES \text{ CR} - 0,041 * RDBO5 \text{ CR} - 0,026 * RDCO \text{ CR} + 0,774 * RNH4 \text{ CR} - 0,005 * \text{Energie CR}$$

$$F4 = -0,522 * RMES \text{ CR} + 0,554 * RDBO5 \text{ CR} - 0,028 * RDCO \text{ CR} - 0,394 * RNH4 \text{ CR} + 0,515 * \text{Energie CR}$$

$$F5 = -0,335 * RMES \text{ CR} - 0,386 * RDBO5 \text{ CR} + 0,814 * RDCO \text{ CR} - 0,266 * RNH4 \text{ CR} - 0,082 * \text{Energie CR}$$

- ✓ La matrice A est donc la suivante :

	F1	F2	F3	F4	F5
RMES	0,453	0,436	0,564	0,415	0,339
RDBO5	0,107	-0,595	-0,14	0,056	0,784
RDCO	-0,632	-0,041	-0,026	0,774	-0,005
RNH4	-0,522	0,554	-0,028	-0,394	0,515
Energie	-0,335	-0,386	0,814	-0,266	-0,082

✓ L'inverse de la matrice A : (Matrice de passage)

Les résultats obtenus par MATLAB sont les suivants

	F1	F2	F3	F4	F5
RMES	0,454	0,107	-0,631	-0,521	-0,334
RDBO5	0,435	-0,593	-0,041	0,554	-0,386
RDCO	0,563	-0,140	-0,027	-0,029	0,813
RNH4	0,415	0,057	0,773	-0,394	-0,266
Energie	0,339	0,782	-0,005	0,515	-0,083

Depuis la matrice de passage on a écrit les  $F_i$  (F1, F2 et F3) en fonction des  $X_i$  (RMES, RDBO5, RDCO, RNH4, Energie).

$$F1 = 0,453*RMES + 0,435*RDBO5 + 0,564*RDCO + 0,415*RNH4 + 0,339*E$$

$$F2 = 0,108*RMES - 0,594*RDBO5 - 0,14*RDCO + 0,056*RNH4 + 0,783*E$$

$$F3 = -0,632*RMSE - 0,042*RDB5 - 0,027*RDCO + 0,774*RNH4 - 0,005*E$$

On a pris seulement les trois premières composantes car avec ces composantes on a une explication de l'information de 79%

Ensuite on a écrit les  $X_i$  en fonction des  $F_i$  (F1, F2 et F3). Entre autre la variable à expliquer l'énergie.

$$RMES = 0,453*F1 + 0,108*F2 - 0,632*F3$$

$$RDBO5 = 0,435*F1 - 0,594*F2 - 0,042*F3$$

$$RDCO = 0,564*F1 - 0,14*F2 - 0,027*F3$$

$$RNH4 = 0,415*F1 + 0,056*F2 + 0,774*F3$$

$$E = 0,339*F1 + 0,783*F2 - 0,005*F3$$

En remplaçant F1, F2 et F3 avec leurs équations on obtiendra l'équation de l'énergie en fonction des (5) variables.

$$E_{SIM} = 0,24*RMES_{CR} - 0,32*RDBO5_{CR} + 0,082*RDCO_{CR} + 0,18*RNH4_{CR} + 0,73*Energie_{CR}$$

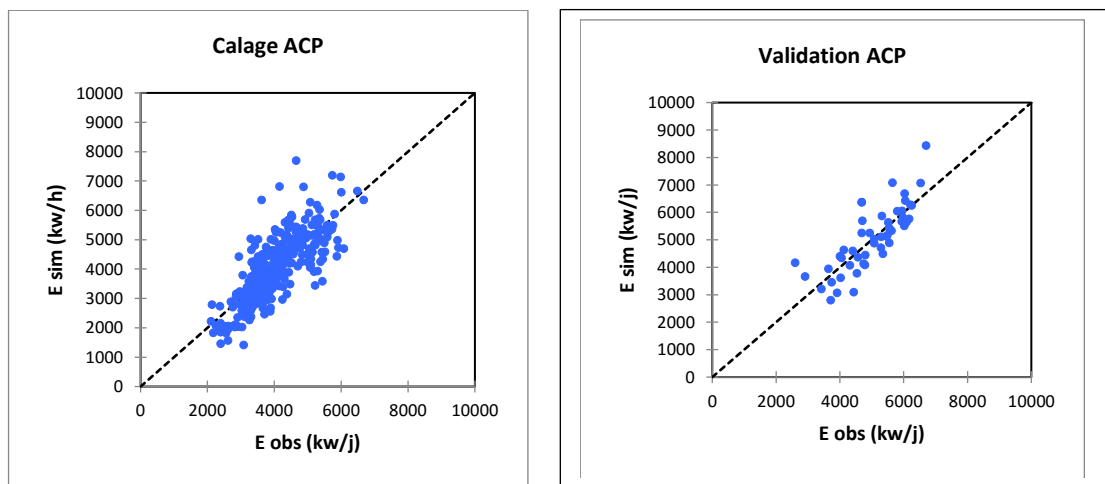
➤ Modélisation de l'énergie

Une fois qu'on a obtenu le modèle de l'ACP, on remplacera les variables (RMES, RDBO5, RDCO, RNH4 et Energie) par leurs valeurs dans la partie calage et validation. On obtiendra par la suite l'énergie simulée par ACP centrée réduite.



Afin de faire une régression linéaire entre l'énergie brute observée et l'énergie brute simulée des deux parties calage et validation, on doit faire le passage de l'énergie simulée centrée réduite à l'énergie simulée brute. Pour ce faire on multiplie l'énergie simulée centrée réduite par l'écart type de la partie calage et on ajoute la moyenne de la partie calage (moy= 0,37 et Ect= 0,17).

### 2.2.2. Interprétation de la représentation graphique et des critères de validation



**Figure 1 : résultats de la régression linéaire du modèle ACP**

A partir de la représentation graphique des résultats obtenus dans la partie calage, On a remarqué qu'une partie des points est proche la première bissectrice, par contre, pas mal de points valeurs sont visiblement très loin de cette dernière.

Pour la partie validation, on a pu voir que l'échantillon est bien étendu. Une grande partie des points sont proches de la première bissectrice mais certains points sont loin de cette dernière, donc on a pu juger donc que le modèle obtenu par ACP est de qualité moyenne.

Dans le but de confirmer ces résultats, on a passé au calcul des critères de validations.

	Calage	Validation
R <sup>2</sup>	60,6%	63,86%
PBIAS	-1,09	-6,4

**Tableau 1 : Critères de validation du modèle ACP**

Pour la partie calage, une fois qu'on a calculé le coefficient de corrélation (R<sup>2</sup>) on a obtenu une valeur de (60,6%) et donc la corrélation entre l'énergie observée et l'énergie simulée par ACP est moyenne.

Le  $PBIAS < 0$  ceci montre que certaines valeurs simulées sont supérieures aux valeurs observées.

On peut conclure que la qualité du modèle issu de l'ACP pour la partie calage est moyenne et ceci confirme l'interprétation de la distribution du nuage de points autour de la courbe de tendance.

Pour la partie validation, le coefficient de corrélation est de (63,86%) pour la partie validation, on remarque une légère augmentation par rapport au  $R^2$  obtenu dans la partie calage et donc la corrélation entre l'énergie observée brute et l'énergie simulée brute est moyenne.

Le  $PBIAS$  est toujours inférieur à zéro et ceci montre que dans les valeurs de la partie validation, certaines sont surestimées.

On peut conclure que la modélisation par ACP nous donne une simulation de l'énergie de qualité moyenne.

### **2.3.La modélisation en utilisant la Neuro-flou**

Le modèle ANFIS (Système d'inférence flou à base de réseaux de neurones adaptifs) est un modèle qui combine les techniques floues et neuronales en formant un réseau à apprentissage supervisé, et donc il est appliqué durant la phase calage et testé en période de validation.

#### **2.3.1.Elaboration du modèle Neuro-flou**

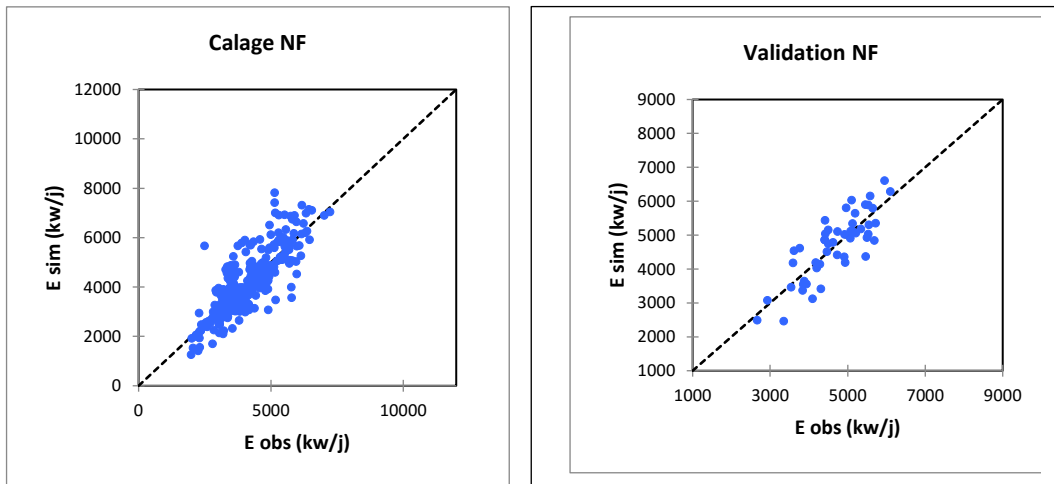
Pour ce faire, on a appliqué le modèle sur des données centrées réduites ensuite on a fait le chemin inverse pour obtenir l'énergie simulée brute.

Pour générer un système d'inférence flou, c'est-à-dire notre modèle, nous avons fait usage de l'outil de génération de systèmes neuro-flou dans Matlab « The Neuro-Fuzzy Design Toolbox ».

Après avoir complété l'apprentissage du modèle, on a procédé à son utilisation dans un programme qui a comme sorties les deux colonnes de l'énergie simulée par Neuro-floue centrée réduite de la partie calage ainsi que celle de la partie validation.

Afin de revenir à des données de l'énergie simulée brute on a multiplié les résultats obtenus dans la partie du calage ainsi que la validation par l'écart type de la partie calage et on ajoute la moyenne de la partie calage.

### 2.3.2. Interprétation de la représentation graphique et des critères de validation



**Figure 2 : résultats de la régression linéaire du modèle NF**

A partir de la représentation graphique des données du calage, on a remarqué que certains points sont très loin de la première bissectrice, cela nous mène à dire que ces points n'ont pas été bien simulés et que l'écart entre l'énergie observé et l'énergie simulée est assez important.

Pour les données de la partie validation, les points sont proches de la première bissectrice. Le modèle Neuro-floue a bien simulé l'énergie.

Dans le but de confirmer ces résultats, on a passé au calcul des critères de validations.

	Calage	Validation
$R^2$	66%	68%
PBIAS	-5,02	1,1

**Tableau 2 : Critères de validation du modèle NF**

Une fois qu'on a calculé le coefficient de corrélation ( $R^2$ ) on a obtenu une valeur de (66%) et donc la corrélation entre l'énergie observée et l'énergie simulée par Neuro-floue est moyenne avec une légère amélioration par rapport à celle obtenue par ACP.

On remarque que la valeur du  $PBIAS < 0$  ceci montre que le modèle surestime l'énergie.

On a pu conclure que la qualité du modèle Neuro-floue pour la partie calage est moyenne mais il s'est amélioré par rapport au modèle issu de l'ACP et ceci confirme l'interprétation de la distribution du nuage de points autour de la courbe de tendance.

Le coefficient de corrélation est de (68%) pour la partie validation, on remarque une augmentation par rapport au R<sup>2</sup> obtenu dans la partie calage et donc la corrélation entre l'énergie observée et l'énergie simulée est meilleure. Par contre le PBIAS>0 donc le modèle sous-estime l'énergie.

On a pu conclure que la modélisation par Neuro-floue nous donne une bonne simulation de l'énergie si on ne prend pas en considération la surestimation de certaines valeurs.

## **2.4.La modélisation en utilisant le couplage analyse en composantes principale Neuro-flou**

On a vu précédemment que le modèle Neuro-floue a donné de meilleurs résultats par rapport au modèle issu de l'ACP et afin d'essayer d'améliorer davantage les résultats on a pensé au couplage ACP Neuro-floue.

### **2.4.1.Elaboration du modèle couplage ACP Neuro-flou**

D'après les résultats de l'application de l'ACP, on remarque que l'énergie est bien représentée sur l'axe F2. Donc on prendra l'axe F2 comme variable à expliquer et les autres axes (F1, F3, F4 et F5) comme variables explicatives.

Afin d'obtenir notre matrice de données (calage et validations) on a procédé comme suit :

- ✓ Les données de la parties calage sont ceux obtenues par ACP (matrice des coordonnées des observations)
- ✓ Les données de la partie validation ont été calculées à partir de la matrice de passage :

$$F1= 0,453*RMSE + 0,435*RDBO5 + 0,564*RDCO + 0,415*RNH4$$

$$F3= -0,632*RMES - 0,042*RDBO5 - 0,027*RDCO + 0,774*RNH4$$

$$F4= -0,522*RMES + 0,554*RDBO5 - 0,028*RDCO - 0,394*RNH4$$

$$F5= -0,334*RMES - 0,386*RDBO5 + 0,813*RDCO - 0,266*RNH4$$

Pour les (4) axes on a pris E=0 et qui est égale à l'énergie moyenne car on travaille avec des données centrées réduites.

$$F2= 0,108*RMES - 0,594*RDBO5 - 0,140*RDCO + 0,056*RNH4 + 0,783*E$$

Comme l'énergie est représentée par l'axe (2) donc on l'a pris en considération dans l'équation du F2.

En remplaçant RMES, RDBO5, RDCO, RNH4 et l'énergie par leurs valeurs on obtiendra la matrice des données de la partie validation.

Ensuite on passe à la simulation de l'axe F2 par le modèle Neuro-floue.

On génère un système d'inférence flou à l'aide de MATLAB, c'est-à-dire notre modèle, On fait usage de l'outil de génération de systèmes neuro-flou « The Neuro-Fuzzy Design Toolbox ».

On fait les jeux de données de calage et de validation, et on choisit une fonction d'appartenance et le nombre de règles floues.

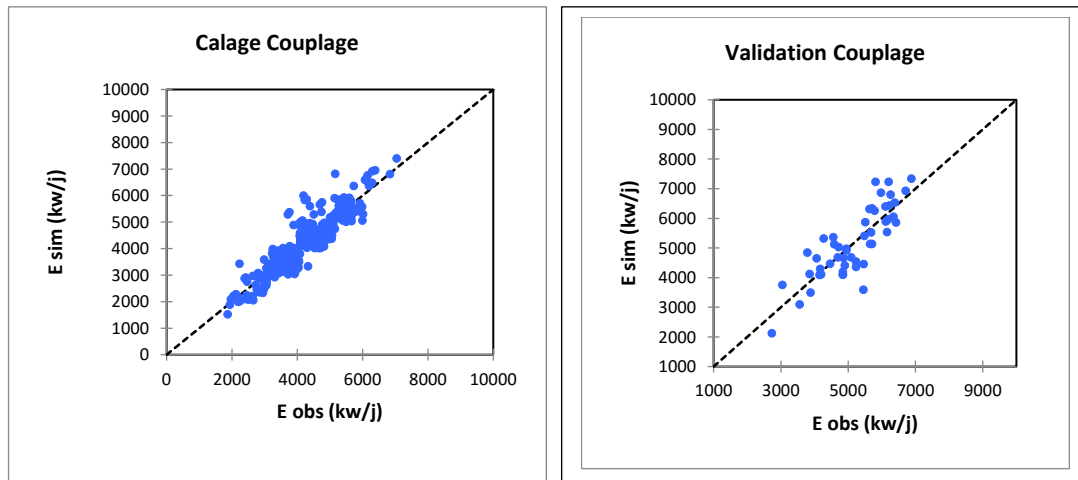
Après avoir complété l'apprentissage du modèle, on procède à son utilisation dans un programme qui a comme sorties les deux colonnes de F2 simulée par Neuro-floue de la partie calage ainsi que celle de la partie validation.

A partir de la matrice de passage, on obtient le modèle suivant qui nous permettra de simuler l'énergie à partir d'un couplage ACP Neuro-floue pour la partie calage ainsi que pour la partie validation :

$$E_{\text{sim}} = 0,339 * F1 + 0,783 * F2_{\text{sim}} - 0,005 * F3 + 0,515 * F4 - 0,083 * F5$$

Afin de faire une régression linéaire entre l'énergie brute observée et l'énergie brute simulée des deux parties calage et validation, on doit faire le passage de l'énergie simulée centrée réduite à l'énergie simulée brute. Pour ce faire on multiplie l'énergie simulée centrée réduite par l'écart type de la partie calage et on ajoute la moyenne de la partie calage (moy= 0,37 et Ect= 0,17).

## 2.4.2. Interprétation de la représentation graphique et des critères de validation



**Figure 3 : Résultats de la régression linéaire du modèle couplage ACP-NF**

D'après la représentation graphique des résultats de la partie calage on a vu que les valeurs simulées sont bien regroupées autour de la première bissectrice. Ceci montre que le modèle couplage ACP Neuro-floue a permis de bien simuler l'énergie et que l'écart entre l'énergie simulée et l'énergie observée est très petit pour la majorité des points de l'échantillon.

D'après la représentation graphique des résultats de la modélisation par couplage ACP Neuro-floue de la partie validation, on a vu que les points sont de plus en plus proches de la première bissectrice par rapport aux deux derniers modèles ( ACP et Neuro-floue) et donc on peut dire que le modèle couplage ACP Neuro-floue nous a permis d'améliorer d'avantage les résultats.

Dans le but de confirmer ces résultats, on a passé au calcul des critères de validations.

	Calage	Validation
R <sup>2</sup>	82,43%	70%
PBIAS	-1,95	-26,3

**Tableau 3 : Critères de validation du modèle couplage ACP-NF**

Une fois qu'on a calculé le coefficient de corrélation (R<sup>2</sup>) on a obtenu une valeur de (82,43%) qui est une très bonne valeur pour le coefficient de corrélation et qui indique que les valeurs simulées par couplage ACP Neuro-floue sont très proches des valeurs observées de l'énergie.

Le PBIAS est toujours négatif et ceci montre que les certaines valeurs sont surestimées par le modèle couplage ACP Neuro-flou.

Le  $R^2$  de la validation a diminué par rapport au calage (70%) mais il est quand même meilleur par rapport aux deux autres modèles (ACP et Neuro-floue). Cette diminution peut être expliquée par le fait que le vecteur F2 a été calculé à partir de la matrice de passage et donc il est moins précis que celui de la partie calage qui a été obtenu directement de l'application de l'ACP.

Le PBIAS est négatif et donc ce modèle surestime certaines valeurs de la partie validation.

On peut conclure que le modèle couplage ACP Neuro-floue a donné les meilleurs résultats de la simulation de l'énergie.

## **2.5.Conclusion**

Après analyse des résultats obtenus par les trois modèles soit : ACP, Neuro-floue ou couplage ACP Neuro-floue, on peut se prononcer sur le modèle à adopté pour la modélisation de l'énergie. Le modèle couplage ACP Neuro-floue est celui à retenir car il permet de combiner entre l'avantage de l'ACP qui est la réduction de la dimension d'une matrice de données et l'avantage de la Neuro-floue qui est la grande robustesse dans la prédiction des phénomènes complexes (traitement biologique).

De manière générale le modèle couplage ACP Neuro-floue réussit à améliorer les résultats des deux modèles ACP et Neuro-floue et présente des résultats très satisfaisants.

### 3.Optimisation de l'énergie électrique par Couplage ACP Neuro-flou

#### 3.1.Introduction

Dans le projet de fin d'étude de Mr.DJAOUI Toufik intitulé « Optimisation de la consommation de l'énergie électrique de la STEP de Boumerdes », et plus précisément dans le chapitre « élaboration du modèle d'optimisation » un modèle d'optimisation de la consommation énergétique basé sur l'historique de gestion de la STEP de Boumerdes durant les années « 2008-2016 ».Ce modèle a été créé grâce à une application codée sous « Matlab ».

#### 3.2.La sélection des données sans prendre en compte les « WQI »

Avant de faire un choix sur les expériences jugées bonnes, un classement de toutes les expériences de la station de Boumerdes a été fait et une matrice de 237 lignes a été obtenue.

Deux sélections ont été choisies, une basée sur le rendement épuratoire de la station par rapport au rendement optimal et l'autre sur la consommation énergétique des aérateurs.

##### 3.2.1. 1ere sélection

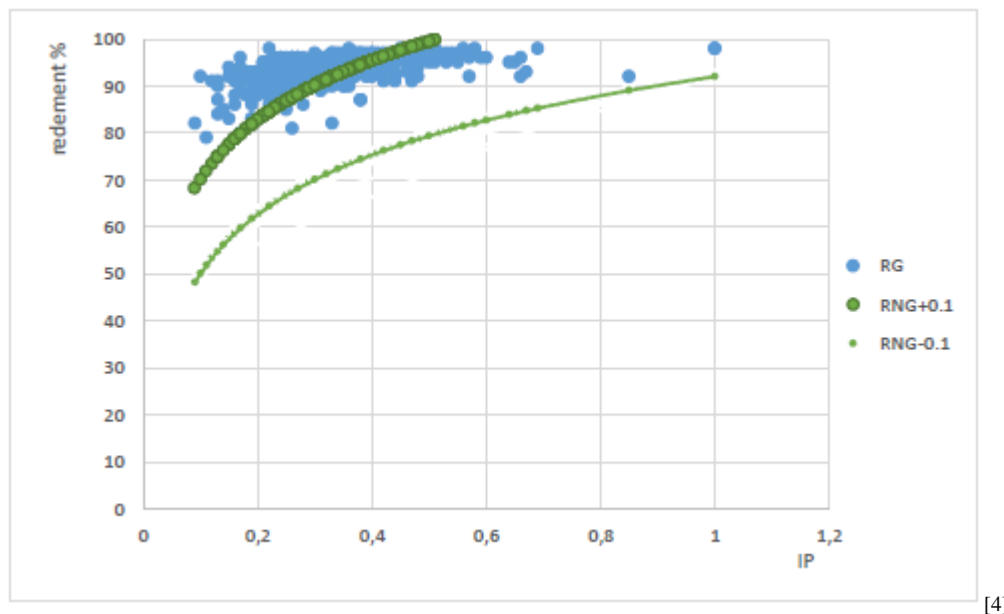


Figure 4 : Représentation de l'intervalle de la première sélection

La figure ci-dessus représente le rendement global (bleu) de 2008 à 2016 et un intervalle de 0,1 autour du rendement global lié à la norme (vert) en fonction de l'indice de pollution.

Les points bleus se trouvant dans l'intervalle vert jugés pertinents seront sélectionnés pour établir notre modèle.



### 3.2.2. 2eme sélection

Pour cette deuxième sélection on a choisi les expériences ayant le même indice de pollution comme meilleures expériences.

La sélection des meilleures expériences a été faite grâce à un programme codé sous « Matlab ».

Une fois ces deux étapes de sélections terminées, on aboutit aux résultats suivants :

- **1<sup>er</sup> Cas avec un intervalle de sélection :  $\pm 0,1$  du rendement optimal**

Le nombre de sélections jugées bonnes par rapport au rendement proche du rendement optimal et avec une consommation énergétique minimale est de 36 journées.

En utilisant le résultat obtenu précédemment, un modèle de régression multiple en utilisant « XLSTAT » a été établi.

<i>les modèles de la régression multiple</i>		
<u><i>modèle A1</i></u>		
$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_pX_p + \varepsilon$		
$E = 0,116 + 0,022 * DR + 0,01 * T - 0,002 * D - 0,014 * RB + 0,033 * IP$		
$R^2$	$\overline{R}^2$	$F$
0.382	0.31	12.65
<u><i>modèle A2</i></u>		
$Y = b_0 + b_1X_1 + b_1X_1^2 + \dots + b_pX_p + b_pX_p^2 + \varepsilon$		
$E = 0,004 + 0,03 * DR + 0,18 * T - 0,12 * D - 0,01 * RB - 0,16 * IP - 0,002 * DR^2 - 0,01 * T^2 + 0,01 * D^2 - 0,001 * RB^2 + 0,15 * IP^2$		
$R^2$	$\overline{R}^2$	$F$
0.67	0.55	6.57

[4]

Les résultats obtenus ont été jugés mauvais, donc un choix de réduction du seuil de la 1ere sélection a été pris.

- **2<sup>ème</sup> cas avec un intervalle de sélection :  $\pm 0,05$  du rendement optimal**

Le nombre d'expériences jugées bonne a été réduit à 14 journées.

En utilisant le résultat obtenu précédemment, un modèle de régression multiple en utilisant « XLSTAT » a été établi.

<i>les modèles de la régression multiple</i>		
<i>modèle B1</i>		
$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_pX_p + \varepsilon$		
$E = 0,073 + 0,044 * DR - 0,017 * T + 0,007 * D + 0,010 * AB + 0,032 * IP$		
$R^2$	$\overline{R}^2$	$F$
0.49	0.21	1.67
<i>modèle B2</i>		
$Y = b_0 + b_1X_1 + b_1X_1^2 + \dots + b_pX_p + b_pX_p^2 + \varepsilon$		
$E = -0,29 + 0,24 * DR + 0,06 * T - 0,22 * D + 0,03 * RB + 0,87 * Ip - 0,03 * DR^2 - 0,004 * T^2 + 0,02 * D^2 - 0,003 * RB^2 - 0,82 * Ip^2$		
$R^2$	$\overline{R}^2$	$F$
0.78	0.44	8.10

[4]

On remarque une certaine amélioration dans les résultats des critères de validations mais toujours faibles.

Le raisonnement par indice de pollution a donné une formule qui ne prend pas considération le poids de la contribution de chaque paramètre dans la caractérisation de l'effluent et donc il ne permet pas de trouver un modèle de régression adapté aux problématiques étudiées.

Une fois que les sous indices de pollutions ont été déterminé soient :  $I_{DBO5}$ ,  $I_{DCO}$ ,  $I_{MES}$ ,  $I_{NH4}$ . On passe à la détermination des WQI (Water Quality Indice).

Le calcul de ce dernier a permis de faire un choix sur les expériences jugées bonnes on a abouti à deux sélections : une sélection contenant (29) bonnes expériences en prenant en compte le  $NH_4$  et une autre sélection contenant (35) bonne expériences sans prendre en considération le  $NH_4$ .

### 3.3.La sélection des données en prenant en compte les « WQI »

#### 3.3.1.1ere sélection (Avec $NH_4$ )

Dans le cas d'un intervalle de sélection de  $\pm 0,1$  du rendement optimal.

	DR	T	D	AB	IP	E
13/11/2011	4,42	5,29	4,88	0,55	0,12	0,3
16/08/2009	3,68	6,94	5,41	1,27	0,15	0,18
13/08/2012	4,6	6,69	4,63	1,94	0,17	0,23

01/02/2012	4,74	3,8	4,29	1,45	0,18	0,27
10/09/2012	4,38	6,31	5,96	1,53	0,19	0,2
01/08/2012	4,19	6,71	4,85	1,81	0,2	0,18
12/11/2012	4,08	5,08	5,55	1,42	0,21	0,18
26/09/2012	4,71	6,22	5,65	1,22	0,22	0,2
01/07/2012	4,53	6,58	4,83	1,48	0,23	0,19
08/07/2012	4,66	6,64	5,53	0,98	0,24	0,2
12/07/2009	3,7	6,63	4,7	2,33	0,25	0,19
27/03/2011	2,76	4,68	5,09	0,94	0,26	0,15
01/02/2009	3,49	3,94	5,29	5,28	0,27	0,16
24/06/2012	5,44	6,25	5,39	2,26	0,28	0,19
10/04/2011	2,91	4,96	4,78	1,1	0,29	0,17
05/04/2009	1,69	4,6	4,7	5,44	0,3	0,19
12/04/2009	1,98	3,95	4,85	2,05	0,31	0,16
12/05/2014	5,22	5,62	5,88	2,56	0,32	0,25
19/04/2009	3,5	4,32	5,06	2,92	0,33	0,2
20/03/2011	3,08	4,42	5,22	2,81	0,34	0,16
10/05/2009	3,66	5,29	5,34	3,98	0,35	0,19
24/05/2009	2,63	5,65	4,79	1,19	0,36	0,19
09/08/2011	5,55	6,68	5,08	6,53	0,37	0,23
19/05/2014	5,35	5,51	5,76	7,15	0,38	0,26
16/02/2009	3,17	4	4,24	2,11	0,39	0,16
01/03/2009	3,46	4,06	4,67	3,04	0,41	0,19
10/03/2009	3,51	4,25	4,1	5,94	0,44	0,22
17/05/2011	5,11	5,49	2,97	4,04	0,46	0,33
25/01/2009	3,37	4,04	4,26	6,32	0,53	0,22

Au total nous avons obtenus 29 journées sélectionnées jugées des bonnes expériences où on a un rendement proches du rendement optimal et une consommation en énergie électrique minimale.

### • **Elaboration du modèle d'optimisation**

Dans cette partie nous allons exploiter les matrices résultantes de la sélection des données afin d'établir notre modèle de couplage ACP Neuro-flou le moins contraignant permettant de simuler la consommation en énergie électrique des aérateurs avec la plus grande facilité.

En appliquant une ACP sur les données centrés réduits de la partie calage (les expériences jugées bonnes), on a pu voir que l'énergie est bien représentée sur l'axe F2. Donc on prendra l'axe F2 comme variable à expliquer et les autres axes (F1, F3, F4, F5 et F6) comme variables explicatives.

Afin d'obtenir notre matrice de données (calage et validations) on a procédé comme suit :

- ✓ Les données de la parties calage sont ceux obtenues par ACP (matrice des coordonnées des observations)
- ✓ Les données de la partie validation ont été calculées à partir de la matrice de passage :

$$F1 = -0,287*DR - 0,492*T - 0,414*D + 0,432*AB + 0,562*IP$$

$$F3 = 0,151*DR + 0,134*T + 0,67*D + 0,577*AB + 0,27*IP$$

$$F4 = -0,138*DR + 0,763*T - 0,435*D + 0,031*AB + 0,281*IP$$

$$F5 = 0,448*DR - 0,167*T + 0,09*D - 0,585*AB + 0,615*IP$$

$$F6 = -0,549*DR + 0,253*T + 0,39*D - 0,236*AB + 0,367*IP$$

Pour les (5) axes on a pris  $E=0$  et qui est égale à l'énergie moyenne car on travaille avec des données centrées réduites.

$$F2 = 0,612*DR + 0,257*T - 0,174*D + 0,285*AB + 0,139*IP + 0,655*E$$

Comme l'énergie est représentée par l'axe (2) donc on l'a pris en considération dans l'équation du F2.

En remplaçant DR, T, D, AB, IP et l'énergie par leurs valeurs on obtiendra la matrice des données de la partie validation.

Ensuite on passe à la simulation de l'axe F2 par le modèle Neuro-floue.

On génère un système d'inférence flou à l'aide de MATLAB, c'est-à-dire notre modèle, On fait usage de l'outil de génération de systèmes neuro-flou « The Neuro-Fuzzy Design Toolbox ».

On fait les jeux de données de calage et de validation, et on choisit une fonction d'appartenance et le nombre de règles floues.

Après avoir complété l'apprentissage du modèle, on procède à son utilisation dans un programme qui a comme sorties les deux colonnes de F2 simulée par Neuro-floue de la partie calage ainsi que celle de la partie validation.

A partir de la matrice de passage, on obtient le modèle suivant qui nous permettra de simuler l'énergie à partir d'un couplage ACP Neuro-floue pour la partie calage ainsi que pour la partie validation :

$$E_{\text{sim}} = 0,044 * F1 + 0,655 * F2_{\text{sim}} - 0,324 * F3 - 0,36 * F4 - 0,205 * F5 + 0,541 * F6$$

Le calcul des critères de validation pour la partie calage ainsi que pour la validation sont données par le tableau suivant :

	Calage	Validation
$R^2$	47,2%	11,6%
$\overline{R^2}$	45,3%	11,2%

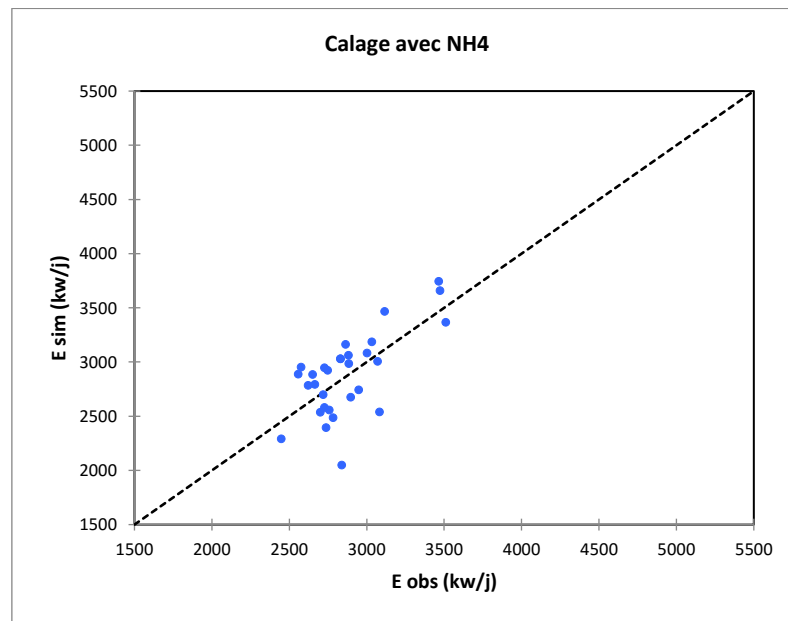
**Tableau 4 : Critère de validation de la première sélection**

La valeur du  $R^2$  et du  $\overline{R^2}$  ( $R^2$  ajusté) sont insatisfaisants et donc on conclure que le modèle couplage ACP Neuro-flou établi sur cette sélection n'a pas donné les meilleurs résultats.

Une fois qu'on a calculé l'énergie simulée par couplage ACP Neuro-flou des expériences jugées bonnes ainsi que l'énergie simulée des expériences non sélectionnées, on a fait une régression linéaire entre les énergies observées brutes et les énergies simulées brutes.

Les résultats de la régression sont représentés dans les figures suivantes :

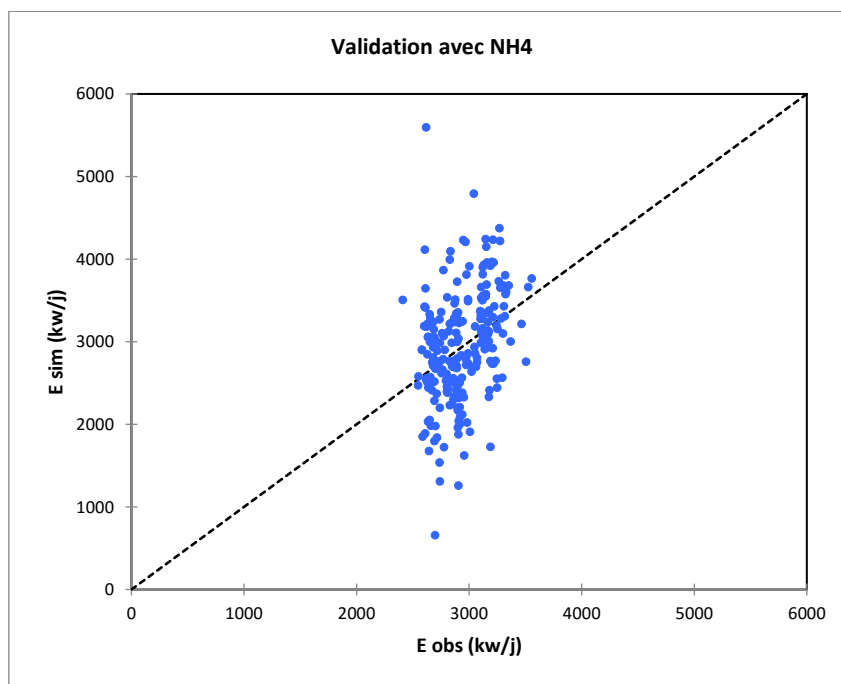
- **Calage (expériences jugées bonnes)**



**Figure 5 : Régression linéaire (Energie observée brute / Energie simulée brute) calage**

La représentation graphique de la régression linéaire confirme les résultats obtenus numériquement car on voit bien que la majorité des points sont loin la première bissectrice.

- **Validation (données non sélectionnées)**



**Figure 6 : Régression linéaire ((Energie observée brute / Energie simulée brute) validation**

On voit d'après la représentation graphique que pratiquement tous les points sont loin de la première bissectrice ceci montent que l'écart entre l'énergie simulée et l'énergie observée est très grand.

### 3.3.2. La deuxième sélection (Sans NH4)

Dans le cas d'un intervalle de sélection de  $\pm 0,1$  du rendement optimal.

	DR	T	D	AB	IP	E
10/10/2011	5,45	6,09	3,6	0,47	0,11	0,31
13/11/2011	4,73	5,33	3,99	0,62	0,13	0,3
16/08/2009	3,94	7,01	4,43	0,68	0,15	0,18
15/11/2015	4,87	5,63	5,48	0,87	0,16	0,24
01/08/2012	4,49	6,77	3,97	0,69	0,17	0,18
10/09/2012	4,69	6,37	4,89	0,78	0,18	0,2
01/02/2009	3,74	3,97	4,33	0,84	0,19	0,16
12/11/2012	4,37	5,12	4,55	1,08	0,2	0,18
12/07/2009	3,96	6,69	3,85	0,79	0,21	0,19
15/07/2012	5,53	6,66	4,27	0,92	0,22	0,19
07/06/2009	3,92	5,89	4,2	0,87	0,23	0,18

27/03/2011	2,96	4,72	4,17	0,89	0,24	0,15
06/06/2012	5,17	6,32	4,07	1,09	0,25	0,17
06/07/2009	3,91	6,72	3,7	0,9	0,26	0,2
10/04/2011	3,12	5	3,92	0,94	0,27	0,17
28/08/2011	6,06	6,89	4,48	1,07	0,28	0,21
24/08/2010	4,84	6,91	4,18	1,36	0,29	0,23
24/06/2012	5,83	6,31	4,42	1,08	0,3	0,19
26/04/2015	5,91	5,31	5,51	1,48	0,31	0,23
10/05/2009	3,93	5,34	4,38	1,39	0,32	0,19
20/06/2012	5,03	6,38	3,72	1,07	0,33	0,24
13/12/2015	4,87	4,35	5,08	1,47	0,34	0,15
20/03/2011	3,3	4,46	4,27	1,27	0,35	0,16
16/02/2009	3,4	4,04	3,47	1,05	0,36	0,16
24/04/2011	1,86	5,1	4,76	1,76	0,37	0,23
25/08/2014	5,54	6,38	5,36	1,86	0,38	0,29
16/01/2013	3,48	3,75	3,54	1,21	0,39	0,28
24/05/2009	2,82	5,7	3,92	1,35	0,4	0,19
01/03/2009	3,71	4,1	3,83	1,28	0,41	0,19
05/08/2015	5,58	5,81	5,87	2,43	0,42	0,23
17/05/2011	5,48	5,54	2,43	1,02	0,45	0,33
25/01/2009	3,61	4,07	3,49	1,45	0,46	0,22
01/12/2015	4,99	4,74	5,32	1,99	0,47	0,26
29/03/2009	2,11	4,69	3,74	2,24	0,62	0,2
09/09/2015	5,67	6,37	5,78	3,75	0,84	0,33

Au total nous avons obtenus 35 journées sélectionnées jugées des bonnes expériences où on a un rendement proches du rendement optimal et une consommation en énergie électrique minimale.

- **Elaboration du modèle d'optimisation**

Dans cette partie nous allons exploiter les matrices résultantes de la sélection des données afin d'établir notre modèle de couplage ACP Neuro-flou le moins contraignant permettant de simuler la consommation en énergie électrique des aérateurs avec la plus grande facilité.

En appliquant une ACP sur les données centrés réduits de la partie calage (les expériences jugées bonnes), on a pu voir que l'énergie est bien représentée sur l'axe F3. Donc on prendra l'axe F3 comme variable à expliquer et les autres axes (F1, F2, F4, F5 et F6) comme variables explicatives.

Afin d'obtenir notre matrice de données (calage et validations) on a procédé comme suit :

- ✓ Les données de la parties calage sont ceux obtenues par ACP (matrice des coordonnées des observations)
- ✓ Les données de la partie validation ont été calculées à partir de la matrice de passage :

$$F1= 0,234*DR+0,028*T+0,379*D+0,612*AB+0,531*IP$$

$$F2= 0,61*DR+0,611*T+0,195*D-0,206*AB-0,358*IP$$

$$F4= -0,316*DR+0,764*T-0,3*D+0,016*AB+0,315*IP$$

$$F5= 0,683*DR-0,194*T-0,356*D+0,016*AB+0,315*IP$$

$$F6=0,059*DR-0,065*T-0,303*D+0,723*AB-0,613*IP$$

Pour les (5) axes on a pris  $E=0$  et qui est égale à l'énergie moyenne car on travaille avec des données centrées réduites.

$$F2= -0,057*DR+0,015*T+0,714*D+0,133*AB-0,161*IP-0,665*E$$

Comme l'énergie est représentée par l'axe (2) donc on l'a pris en considération dans l'équation du F2.

En remplaçant DR, T, D, AB, IP et l'énergie par leurs valeurs on obtiendra la matrice des données de la partie validation.

Ensuite on passe à la simulation de l'axe F2 par le modèle Neuro-floue.

On génère un système d'inférence flou à l'aide de MATLAB, c'est-à-dire notre modèle, On fait usage de l'outil de génération de systèmes neuro-flou « The Neuro-Fuzzy Design Toolbox ».

On fait les jeux de données de calage et de validation, et on choisit une fonction d'appartenance et le nombre de règles floues.

Après avoir complété l'apprentissage du modèle, on procède à son utilisation dans un programme qui a comme sorties les deux colonnes de F2 simulée par Neuro-floue de la partie calage ainsi que celle de la partie validation.



A partir de la matrice de passage, on obtient le modèle suivant qui nous permettra de simuler l'énergie à partir d'un couplage ACP Neuro-floue pour la partie calage ainsi que pour la partie validation :

$$E_{\text{sim}} = 0,381 \cdot F1 + 0,217 \cdot F2 - 0,665 \cdot F3_{\text{sim}} - 0,308 \cdot F4 - 0,518 \cdot F5 - 0,039 \cdot F6$$

Le calcul des critères de validation pour la partie calage ainsi que pour la validation sont données par le tableau suivant :

	Calage	Validation
$R^2$	89,6%	29%
$\overline{R^2}$	89,3%	28,7%

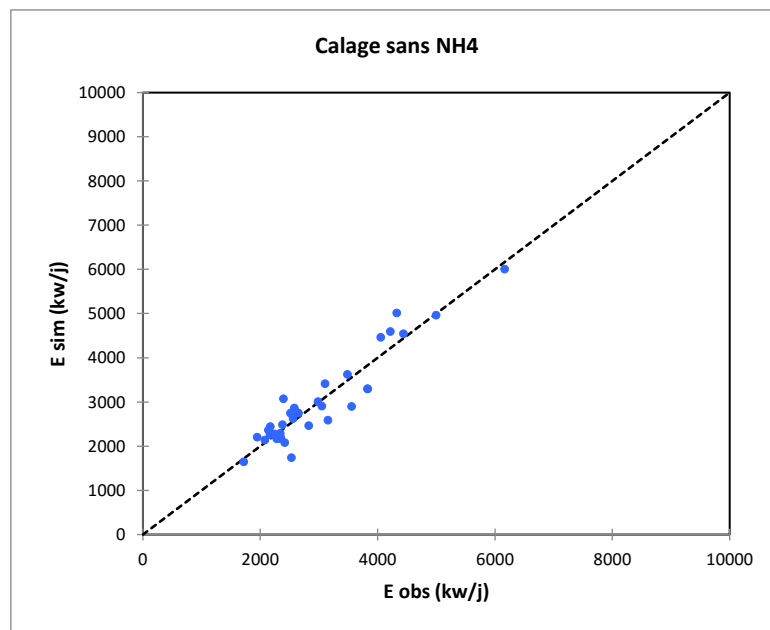
**Tableau 5 : Critères de validation de la deuxième sélection**

La valeur du  $R^2$  et du  $\overline{R^2}$  ( $R^2$  ajusté) sont satisfaisants et donc on conclure que le modèle couplage ACP Neuro-floue établi sur cette sélection a donné les meilleurs résultats par rapport à la première sélection.

Une fois qu'on a calculé l'énergie simulée par couplage ACP Neuro-floue des expériences jugées bonnes ainsi que l'énergie simulée des expériences non sélectionnées, on a fait une régression linéaire entre les énergies observées brutes et les énergies simulées brutes.

Les résultats de la régression sont représentés dans les figures suivantes :

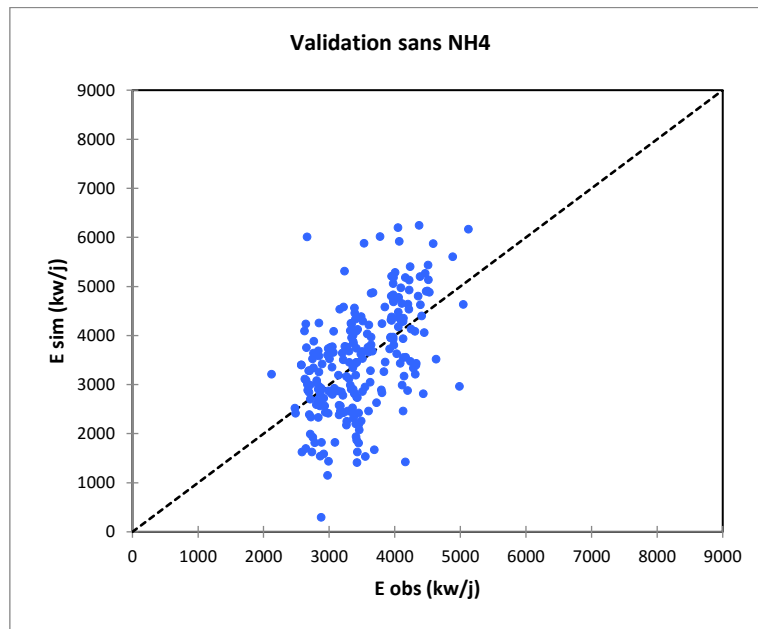
- **Calage (données sélectionnées)**



**Figure 7 : Régression linéaire ((Energie observée brute / Energie simulée brute) calage**

La représentation graphique de l'énergie simulée brute en fonction de l'énergie observée brute montre que le nuage de points se trouve presque superposé sur la première bissectrice. Ceci confirme les résultats obtenus numériquement.

- **Validation (données non sélectionnées)**



**Figure 8 : Régression linéaire ((Energie observée brute / Energie simulée brute) validation**

On voit d'après la représentation graphique que pratiquement tous les points sont loin de la première bissectrice ceci montre que l'écart entre l'énergie simulée et l'énergie observée est très grand.

- **Conclusion**

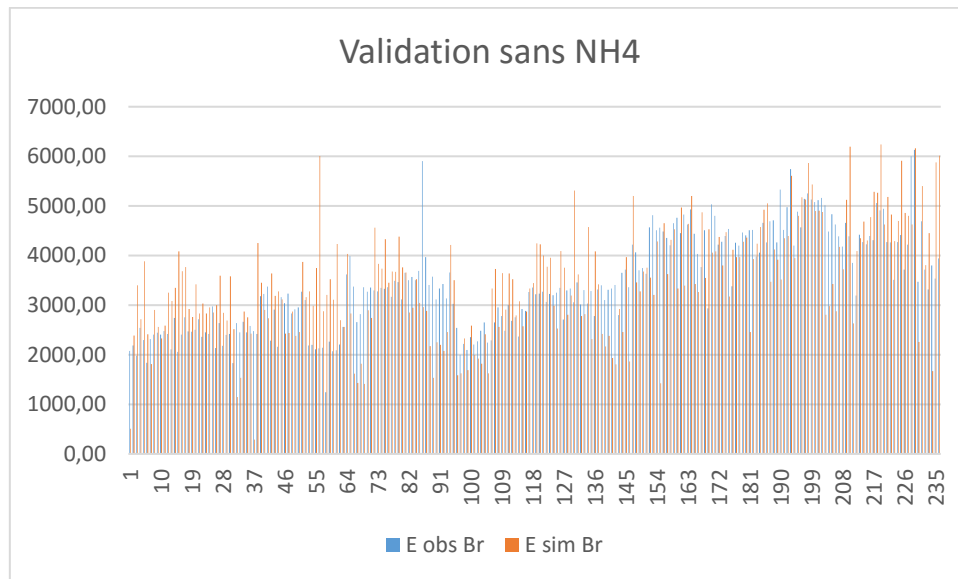
D'après les résultats obtenus, et une fois qu'on a établi les deux modèles de couplage ACP Neuro-flou pour les deux sélections des données jugées bonnes, on peut dire que le modèle de la deuxième sélection a donné les meilleurs résultats et donc c'est le modèle qui représente le mieux notre phénomène.

### **3.4. Le gain en énergie électrique à partir du modèle couplage ACP**

#### **Neuro flou de la deuxième sélection**

Dans cette partie nous allons exploiter le modèle choisi afin de déterminer le gain en matière d'énergie électrique durant les années « 2008-2015 ».

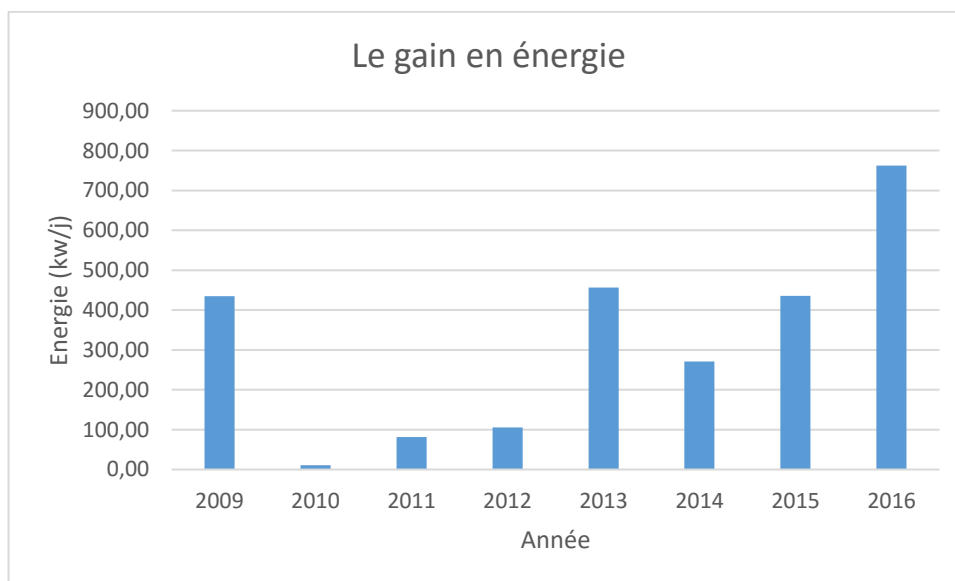
Les résultats de la simulation de l'ensemble des journées non sélectionnées sont illustrés dans l'histogramme suivant :



**Figure 9 : Comparaison entre l'énergie simulée et l'énergie observée**

L'histogramme ci-dessus représente l'énergie électrique réelle consommée par les aérateurs et l'énergie simulée par notre modèle d'optimisation.

L'écart entre l'énergie réelle et l'énergie simulée représente le gain en matière d'énergie électrique, les résultats du gain moyen annuel journalier d'énergie électrique sont représentés dans l'histogramme suivant :



**Figure 10 : Gain moyen annuel journalier en matière d'énergie**

On remarque à partir de l'histogramme précédent que les valeurs les plus élevées en matière de gain énergétique se trouvent à partir de l'année 2013, ceci est expliqué par le dépassement de la capacité nominale en matière de débit de la STEP.

## **Conclusion générale**

Le but de ce travail est l'optimisation de la consommation de l'énergie dans la station d'épuration de Boumerdes. La bonne gestion des procédés épuratoires nous entraîne à une bonne estimation des coûts et un respect des objectifs de qualité à la sortie de la station.

Afin d'atteindre ces objectifs, on a fait appel à la modélisation par le modèle couplage ACP Neuro-flou qui a été établi sur les données des expériences jugées bonnes.

Tout d'abord, nous avons commencé par établir deux modèles de couplage ACP Neuro-flou sur deux sélections de bonne expériences, ensuite les résultats des critères de validation nous ont permis de choisir le meilleur modèle.

Afin d'établir notre modèle, nous avons pris en considération (5) variables explicatives : IP, DR, D, T et AB ainsi qu'une variable à expliquer et qui est l'énergie électrique.

Ce modèle nous a permis de calculer le gain en énergie qui est assez grand à partir de l'année 2013 et ceci est dû au fait que la station de Boumerdes a commencé à travailler en surcharge à partir de l'année 2013 et on a vu que cela a pu influencer certains de nos résultats car l'énergie a beaucoup augmenté à partir de cette année.

## **Bibliographie**

[1] : C. Duby, S. Robin « Analyse en Composantes Principales » AgroParisTech

[2] : HAMRICHE, « Contribution à l'étude et à la simulation des paramètres hydrométriques par l'analyse en composantes principales (ACP) », Mémoire de Projet de Fin d'Etude, ENP Alger, 1993.

[3] : A. BERMAD, « Elaboration d'une méthodologie de simulation et de prévision des variables climatologiques aléatoires et cycliques par l'analyse en composantes principales». Thèse de doctorat d'Etat, ENP 2004.

[4] : T.DJAOUI, « Optimisation de l'énergie électrique de la STEP de Boumerdes ». Projet de fin d'études, 2016.