République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

École Nationale Polytechnique

Laboratoire de Commande des Processus

Département d'Automatique

End of studies' project

In view of obtaining the State Engineering Diploma in Automatic Control

---

Performance Prediction of a Reverse Osmosis System Using Machine Learning

---

BRITAH Adem and **BELALA Haithem Abderrahmane**

Under the supervision of **Pr. TADJINE Mohamed**
and **Dr. CHAKIR Messaoud**

Publicly presented and defended on 03/07/2024 in front of the jury composed of :

President:	Dr. ACHOUR Hakim	ENP
Examinator:	Pr. BOUDANA Djamel	ENP
Promoter :	Dr. CHAKIR Messoud and Pr. TADJINE Mohamed	ENP

ENP 2024

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

École Nationale Polytechnique

Laboratoire de Commande des Processus

Département d'Automatique

End of studies' project

In view of obtaining the State Engineering Diploma in Automatic Control

---

Performance Prediction of a Reverse Osmosis System Using Machine Learning

---

BRITAH Adem and **BELALA Haithem Abderrahmane**

Under the supervision of **Pr. TADJINE Mohamed**
and **Dr. CHAKIR Messaoud**

Publicly presented and defended on 03/07/2024 in front of the jury composed of :

President:	Dr. ACHOUR Hakim	ENP
Examinator:	Pr. BOUDANA Djamel	ENP
Promoter :	Dr. CHAKIR Messoud and Pr. TADJINE Mohamed	ENP

ENP 2024

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

École Nationale Polytechnique

Laboratoire de Commande des Processus

Département d'Automatique

Mémoire de projet de fin d'études

En vue de l'obtention du diplôme d'Ingénieur d'État en Automatique

---

Prédiction des performances d'un système d'osmose inverse à l'aide de l'apprentissage automatique

---

BRITAH Adem et **BELALA Haithem Abderrahmane**

Sous la supervision de **Pr. TADJINE Mohamed**
et **Dr. CHAKIR Messaoud**

Présenté et soutenue publiquement le 03/07/2024 devant le jury composé de :

President:	Dr. ACHOUR Hakim	ENP
Examinateur:	Pr. BOUDANA Djamel	ENP
Promoteurs :	Dr. CHAKIR Messoud et Pr. TADJINE Mohamed	ENP

ENP 2024

ملخص

تحظى عملية التناضح العكسي بأهمية كبيرة في صناعة معالجة المياه. على الرغم من استخدامها الشائع، تعاني هذه العملية من انسداد الغشاء، مما يؤثر على جودة المياه المنتجة وأداء الغشاء نفسه. حتى الآن، تعتمد عمليات تشغيل أنظمة التناضح العكسي على خبرة المشغلين، مع إجراء أنشطة الصيانة وفقًا لجداول زمنية أو معايير محددة مسبقًا. يتضمن هذا العمل تطوير تقدير للانسداد باستخدام مراقب الوضع الانزلاقي، واستخدام تقنيات التعلم الآلي المختلفة لتوفير تنبؤات في الوقت الفعلي وتوصيات الصيانة. توفر النتائج رؤى قيمة حول أداء وملاءمة هذه الأساليب التقديرية.

الكلمات المفتاحية : تحلية المياه، التناضح العكسي، النمذجة، انسداد الأغشية، تنبؤ الانسداد، التعلم الآلي، الذاكرة طويلة وقصيرة الأمد، المحول، مراقب الوضع الانزلاقي.

Résumé

Le processus d'osmose inverse revêt une grande importance dans l'industrie du traitement d'eau. Malgré son utilisation courante, ce processus souffre de l'encrassement des membranes, ce qui affecte la qualité de l'eau produite et les performances de la membrane elle-même. Jusqu'à présent, le fonctionnement des systèmes d'osmose inverse repose sur l'expérience des opérateurs, les activités de maintenance étant réalisées selon des calendriers ou des critères prédéfinis. Ce travail consiste à développer une estimation de l'encrassement basée sur un observateur en mode glissant et à utiliser diverses techniques d'apprentissage automatique pour fournir des prédictions en temps réel et des recommandations de maintenance. Les résultats offrent des perspectives précieuses sur les performances et la pertinence de ces approches d'estimation.

Mots clés : Dessalement, Osmose Inverse, Modélisation, Encrassement des Membranes, Prédiction de l'Encrassement, Apprentissage Automatique, Mémoire à Long et Court Terme, Transformateur, Observateur en Mode Glissant.

Abstract

The reverse osmosis process holds great importance in the water treatment industry. Despite its common use, this process suffers from membrane fouling, which affects the quality of the produced water and the performance of the membrane itself. So far, the operation of reverse osmosis systems relies on the operators' experience, with maintenance activities carried out according to predefined schedules or criteria. This work involves developing a sliding mode observer-based fouling estimation, and using various machine learning techniques to provide real-time predictions and maintenance recommendations. The results provide valuable insights into the performance and suitability of these estimation approaches.

Keywords: Desalination, Reverse Osmosis, Modeling, Membrane Fouling, Fouling Prediction, Machine Learning, Long Short-Term Memory, Transformer, Sliding Mode Observer.

Acknowledgements

First of all, we thank **GOD** Almighty **Allah** for giving us the courage and patience for developing this modest work.

We would also like to extend our deepest appreciation to **Our Parents**, whose unconditional love, encouragement, and sacrifices have made our education and this project possible. Their unwavering belief in us has been a constant motivation.

Our deepest thanks go to our supervisors, Doctor **Messaoud CHAKIR** and Professor **Mohamed TADJINE**. Their expertise, guidance, and mentorship have been invaluable throughout the entire project. Their commitment to excellence and their willingness to share their knowledge have inspired us and shaped our academic growth. We are truly fortunate to have had the opportunity to work under their guidance.

Finally, we would like to thank all the individuals who directly or indirectly contributed to the successful completion of our project. Their contributions, whether big or small, have played a significant role in shaping our understanding and achievements.

Dedication

To my dear mother,
To my dear father may Allah grant him jannah,
To my brothers Sohaib, Wail and Ishak,
To my dear grandparents,
To all my family and friends,
To my partner in this work Adem,

Haithem

To my dear parents,
To my dear paternal grandmother, may Allah grant her Jannah,
To my dear maternal grandparents,
To my brothers Abdessamed, Ayoub, Dayaa eddine and Mosaab,
To all my family and friends,
To my partner in this work Haithem,

Adem

Contents

List of Tables

List of Figures

List of Abbreviations

ML	Machine Learning
RO	Reverse Osmosis
ED	Electrodialysis
MSF	Multistage Flash Distillation
MED	Multi-Effect Distillation
MVC	Mechanical Vapor Compression
LSTM	Long Short-Term Memory
NF	Nanofiltration
MD	Membrane Distillation
CECP	Cake-Enhanced Concentration Polarization
MF	Microfiltration
UF	Ultrafiltration
SR	Solute Rejection
SWRO	Seawater Reverse Osmosis
BWRO	Brackish Water Reverse Osmosis
TMP	Transmembrane Pressure
TFC	Thin Film Composite
SMO	Sliding Mode Observer
RNN	Recurrent Neural Network
seq2seq	Sequence-to-Sequence model
NLP	Natural Language Processing
SD	Solution-Diffusion
ESD	Extended Solution-Diffusion

SDI Solution-Diffusion-Imperfection

CWP Concentration of Water in Permeate

TDS Total Dissolved Solids

List of Symbols

J_w	Permeat flow
ΔP	Applied pressure
π	Osmotic pressure
η	Permeat viscosity
R_m	Membrane resistance
J_c	Permeat flow with cake formation
R_c	Cake layer resistance
A_t	Total membrane area
J_s	Permeat flow with surface blockage
A_b	Membrane area occupied by surface crystals
$A_{\mathbf{free}}$	Free membrane surface, uncovered by surface crystals
J_t	Permeat flow combining cake filtration and surface blockage models
R_T	Total membrane resistance
$A_w(t,x)$	Membrane permeability coefficient with respect to time and space
A_0	Initial water permeability coefficient
FF	Fouling factor
A_n	Normalized water permeability coefficient
C_p	Permeate concentration
C_f	Feed concentration
$\hat{X}_i$	Estimated state variables
$\hat{f}_i$	Estimated nonlinear functions
$\hat{\theta}$	Estimated parameter

L_i	Observer gains
$\hat{y}_1$	Estimated output corresponding to y_1
$\hat{y}_2$	Estimated output corresponding to y_2
S	Sliding surface
e_i	Observation errors
Δf_i	Modeling uncertainties
$V(e)$	Lyapunov function
$\dot{V}$	Time derivative of the Lyapunov function
f_t	Forget gate layer
i_t	Input gate layer
C_t	Cell state
o_t	Output gate
h_t	Hidden state
σ	Sigmoid function
tanh	Hyperbolic tangent function
W_f, W_i, W_c, W_o	Weight matrices for the forget, input, cell candidate, and output gates
b_f, b_i, b_c, b_o	Bias terms for the forget, input, cell candidate, and output gates
Q	Query in attention mechanism
K	Key in attention mechanism
V	Value in attention mechanism
$Attentien(Q, K, V)$	Attention mechanism function

General Introduction

"**We made from water every living thing**"

Approximate translation of verse 30, Surah Al-Anbya, the Quran

The demand for clean water is a critical issue in many countries with limited water resources. According to the United Nations, by 2025, approximately 20% of countries will face severe water shortages, and 2.7 billion people will require access to clean drinking water. Clean water is essential for human survival and various industries such as agriculture, energy production, tourism, recreation, and construction. Additionally, rapid urbanization has led to an increase in annual water consumption, further exacerbating water pollution and scarcity issues [1].

To meet these essential needs, it is necessary to produce clean water to support human life and other activities that depend on water. Seawater, which constitutes about 97.5% of the Earth's accessible water, represents a significant potential source for potable water. However, only about 2% of the water exists as polar ice and glaciers, leaving less than 0.5% as freshwater available for human use [2].

In Algeria, water consumption is characterized by high demand across agricultural, domestic, and industrial sectors. The country consumes approximately 6 billion cubic meters of water annually, with the agricultural sector accounting for about 70-75% of total usage, domestic water use around 15-20%, and the industrial sector about 5-10%.

Efforts are underway to develop and expand alternative water production technologies to address the global water crisis. Scientists and government officials are also working to raise awareness about the importance of water management and conservation.

Desalination has emerged as a viable solution to address the global scarcity of drinkable freshwater by converting seawater into potable water through the removal of salt. Desalination methods can be classified into four categories: Thermal, Crystallization, Membrane, and Other.

Thermal-based and membrane-based separation methods are the most commonly used desalination techniques [3]. Multistage flash distillation (MSF), multi-effect distillation (MED), and vapor compression (VC) are the most utilized thermal desalination processes for large and medium scales. Except for MVC, all thermal desalination techniques require low-temperature heat energy to increase the temperature of the saline input and electrical energy to run the pumps [4].

Among membrane-based desalination techniques, reverse osmosis (RO) and electrodialysis (ED) are the most prevalent. RO and ED require electrical energy; for example, in RO, electricity is used for pumping, whereas in ED, a direct current between electrodes is necessary for ionic membrane separation [5].

RO technology is widely recognized and extensively researched, being used by some of the world's largest desalination plants [6]. In this pressure-driven process, a semipermeable membrane rejects dissolved components in the water through size exclusion and physical-chemical interactions between the solute, solvent, and membrane.

This work focuses on the reverse osmosis (RO) process for desalination due to its high efficiency in salt rejection, energy efficiency compared to methods like multi-stage flash (MSF) and multi-effect distillation (MED), scalability, flexibility for various settings, and enhanced reliability and lifespan from continuous technological improvements.

Despite the many benefits of membrane technology, membrane fouling remains a significant challenge during operation. Membrane fouling increases operational costs and shortens membrane life [7]. In extreme cases, uncontrolled fouling can cause membrane failure, severely impacting plant operations.

The need for effective prediction and management strategies for membrane fouling is critical to enhance the longevity and reliability of RO systems. This work aims to address this challenge by developing predictive models to estimate and manage membrane fouling effectively.

This work is organized into several chapters to provide a systematic and comprehensive exploration of the proposed membrane fouling prediction problem. The following outlines the organization of the thesis:

This chapter introduces the topic, providing an overview of the motivations, objectives, and contributions of this work. It also presents the scope of our work and outlines the organization of the subsequent chapters.

Chapter 1 provides an exploration of the background and state-of-the-art for reverse osmosis (RO) desalination. It discusses the principles and terminology, compares natural osmosis to reverse osmosis, and covers osmotic pressure, water recovery, solute rejection and passage, permeate and salt flux, and concentration polarization. It also examines different RO membrane modules and their characterization.

The chapter 2 focuses on the mathematical modeling of RO processes. It introduces various transport models, including phenomenological and mechanistic models, and details the solution-diffusion model. It also covers the spatial dependence in the solution-diffusion model and presents a nonlinear state-space model for the RO process. The chapter concludes with model verification through experimental validation and simulation analysis.

In chapter 3, the modeling macroscale fouling mechanisms in RO membranes takes center stage. It covers the formation mechanisms of cake layers and surface blockage, and combines different fouling mechanisms. The impact of fouling on operational variables in RO plants is analyzed, both with and without membrane discretization, providing insights into long-term performance.

The chapter 4 presents the proposed solution for predicting membrane fouling in RO systems. It includes the design of a sliding mode observer for online parameter estimation and discusses multi-step time series forecasting using machine learning techniques, such as LSTM and Transformer models.

In chapter 5, we focus on the methods used for data acquisition and preprocessing necessary for predictive modeling of membrane fouling. It discusses the collection and transformation of data, the application of sliding mode observers, and the preparation of the dataset for analysis.

The chapter 6 applies the predictive methods discussed earlier to realistic scenarios. It evaluates the performance of LSTM and Transformer models in predicting membrane fouling, analyzing

the results through residual analysis and autocorrelation of residuals. The chapter concludes with a summary of findings and the practical implications of the predictive models.

Finally, the general conclusion presents a summary of the research findings. It provides a concise overview of the key outcomes obtained throughout this work and explores potential future directions, paving the way for further investigations and advancements in the field.

Chapter 1

Background & State-of-the-art for Reverse Osmosis Desalination

1.1 Introduction

Water desalination is a primary method for producing clean water from various sources[8]. With advancements in membrane science, membrane-based processes have emerged as the most promising and practical desalination options due to their high energy efficiency[9].Various membrane-based processes, such as reverse osmosis (RO), nanofiltration (NF), and membrane distillation (MD), are widely used for treating seawater and brackish water from different sources[10].

Reverse osmosis (RO) is currently the most reliable and advanced state-of-the-art technique for desalinating seawater and brackish water. It has become a crucial alternative for producing clean water while minimizing associated costs. The use of RO for desalination has seen a substantial increase since the 1950s[11]. Currently, around 50% of the desalinated water available globally is produced by using RO.

This chapter aims to provide an overview on the state of the art of various aspects of the RO desalination process. We will focus on its fundamental principles, components, and challenges. We explore the theoretical underpinnings of osmosis and reverse osmosis, and the various membrane configurations used in RO systems.

1.2 Principles and Terminology

Osmosis, a well-established physical phenomenon, has been understood by humanity for many years. It can be succinctly described as the spontaneous movement of water molecules across a semipermeable membrane from a region of low solute concentration (resulting in low osmotic pressure) to one of high solute concentration (resulting in high osmotic pressure), as depicted in Figure 1.1 (a). The semipermeable membrane selectively allows the passage of water molecules while rejecting solutes[12].

1.2.1 Natural Osmosis Vs Reverse Osmosis

Osmosis persists until osmotic equilibrium is attained, where the chemical potentials across the membrane become equal, as shown in Figure 1.1 (b). The flow of water molecules can be altered or reversed by applying external pressure to the solution with higher concentration (known as the feed solution)[13]. When the applied pressure exceeds the osmotic pressure difference across the membrane, water molecules are compelled to move in the opposite direction to the natural osmosis phenomenon. This process, termed reverse osmosis (RO), is illustrated in Figure 1.1(c).

1.2.2 Osmotic Pressure

Osmotic pressure is closely related to the colligative properties of a solution such as freezing point depression and boiling point elevation. For ideal dilute solutions, the osmotic pressure (π) can be estimated using the van't Hoff equation given below[14]

$$\pi = CRT \tag{1.1}$$

where, C is the molar concentration of a non-permeable solute in the solution (mol/L), R is the universal gas constant (0.08206 L atm/mol K).

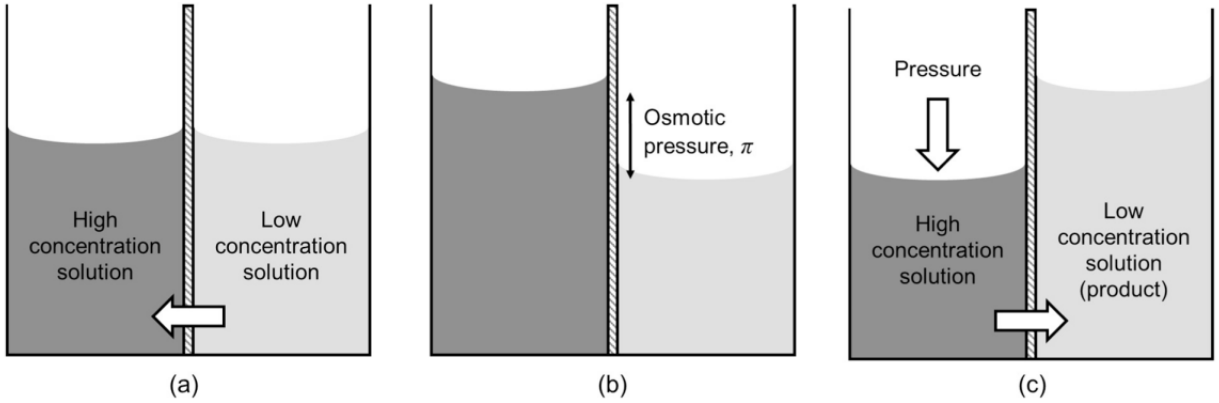


Figure 1.1: Schematic of (a) osmosis (b) osmotic equilibrium (c) RO

For non-ideal concentrated solutions, the following logarithmic equation is valid for the estimation of osmotic pressure [29]:

$$\pi = -\frac{RT}{V_1^0} ln(a_1) \tag{1.2}$$

where, a_1 is the activity of the solvent and V_1^0 is the molar volume of pure solvent (L/mol).

1.2.3 Water Recovery

In a continuous RO process, the feed water stream splits into two streams (Figure 1.2). The first stream, called permeate or product water, consists of water molecules that have passed through the membrane and has a low solute concentration. The second stream, termed brine, concentrate, or reject, contains the remaining water molecules and rejected solutes, resulting in a higher solute concentration than the feed.

Recovery or conversion of an RO process is defined as the volume fraction of feed water that is recovered as permeate or product water. The percentage recovery (r) can be calculated as follows [15]:

$$r = \frac{Q_p}{Q_f} \times 100\% \tag{1.3}$$

where, Qp and QF are the flow rates of the permeate and feed streams, respectively.

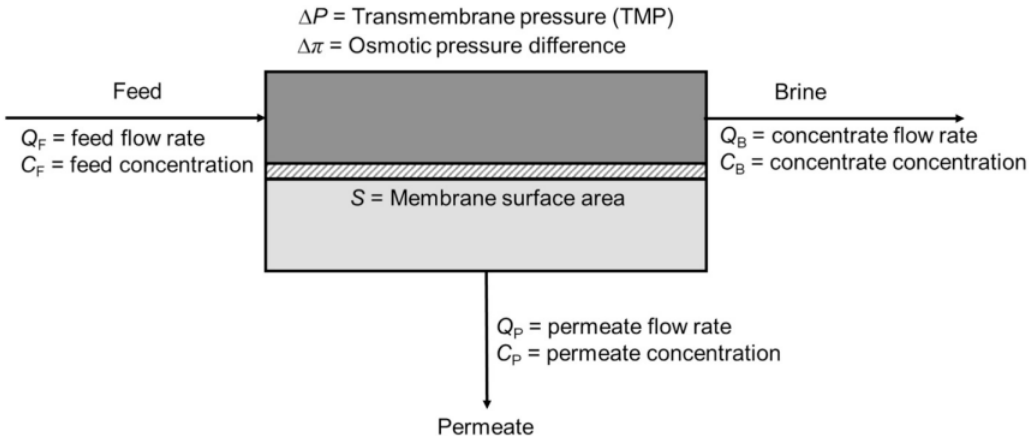


Figure 1.2: Schematic of a continuous RO system

1.2.4 Solute rejection and passage

Solute or salt rejection (SR) is defined as the percentage of a particular incoming solute that is rejected by the RO membrane. The apparent (observed) SR is given as follows:

$$SR = (1 - \frac{C_p}{C_f}) \times 100\% \tag{1.4}$$

where, C_p and C_f represent permeate and feed solute concentrations (mg/L), respectively. Rejection depends on the type of feed constituents, their characteristics, and the type of RO membrane.

1.2.5 Permeate and salt flux

Permeate or water flux J_W refers to the volumetric flow rate of permeate per unit surface area of the RO membrane and is generally proportional to the net pressure driving force across the membrane.
Conversely, salt flux J_s represents the amount of salt passing through the membrane per unit surface area per unit time, and is proportional to the salt concentration difference across the membrane.

1.2.6 Concentration polarization

During RO desalination, solutes flow from the bulk feed towards the membrane, creating a high concentration on the membrane surface and forming a boundary layer with elevated solute concentration [16]. This leads to diffusional back-transport of solutes away from the membrane. However, due to convection dominating over diffusion, solutes accumulate in the boundary layer and on the membrane surface, resulting in a higher concentration on the membrane surface C_m compared to the bulk feed water C_F. This buildup of rejected solutes near the membrane is known as concentration polarization C_P [11][13].

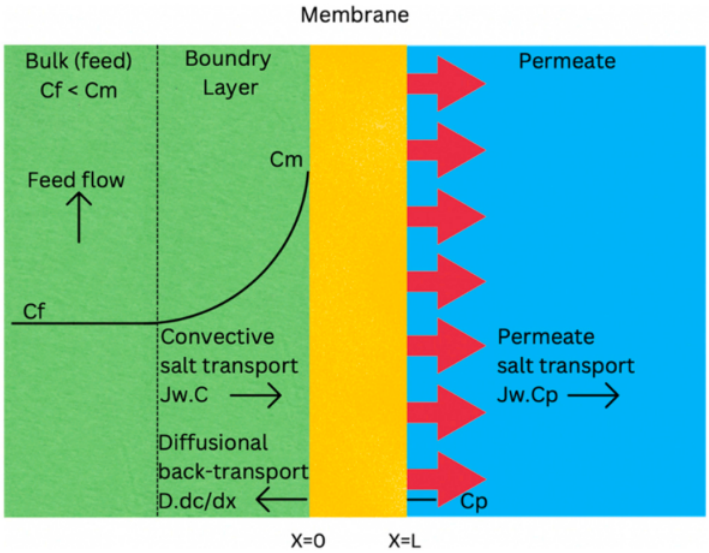


Figure 1.3: Concentration polarization and concentration gradient profile across the membrane surface

Effects of concentration polarization

Concentration polarization (CP) in RO desalination reduces permeate flux and membrane solute rejection while increasing fouling and scaling risks[17]. CP effects are intensified by colloidal particle presence, leading to cake-enhanced concentration polarization (CECP).

1.3 RO membranes modules

There exist four different types of membrane modules for RO desalination,: (1) plate and frame, (2) tubular, (3) spiral wound, and (4) hollow fiber.
RO membranes were initially designed in tubular and plate-and-frame configurations. However, due to their inherently low packing density, these designs were phased out and replaced by hollow fiber and spiral wound modules.[18]

1.3.1 Plate and frame modules

One of the earliest designs of RO membranes is plate and frame modules. In these modules, a flat sheet RO membrane is attached to both sides of a rigid plate made from materials such as solid plastic, porous fiberglass, or reinforced porous paper. Multiple plates are stacked within a pressurized support framework, with grooved channels in the plates providing paths for permeate flow. As the feed solution enters from one end, water molecules permeate the membrane and are collected in a central permeate collection manifold, while the brine or concentrate exits from the other end as shown in figure 1.4.
Despite their low packing density and high cost due to complex design and construction, plate and frame modules are easy to clean, making them suitable for feed streams with high suspended solids content. However, they are prone to fouling because of dead zones within the modules[19][15].

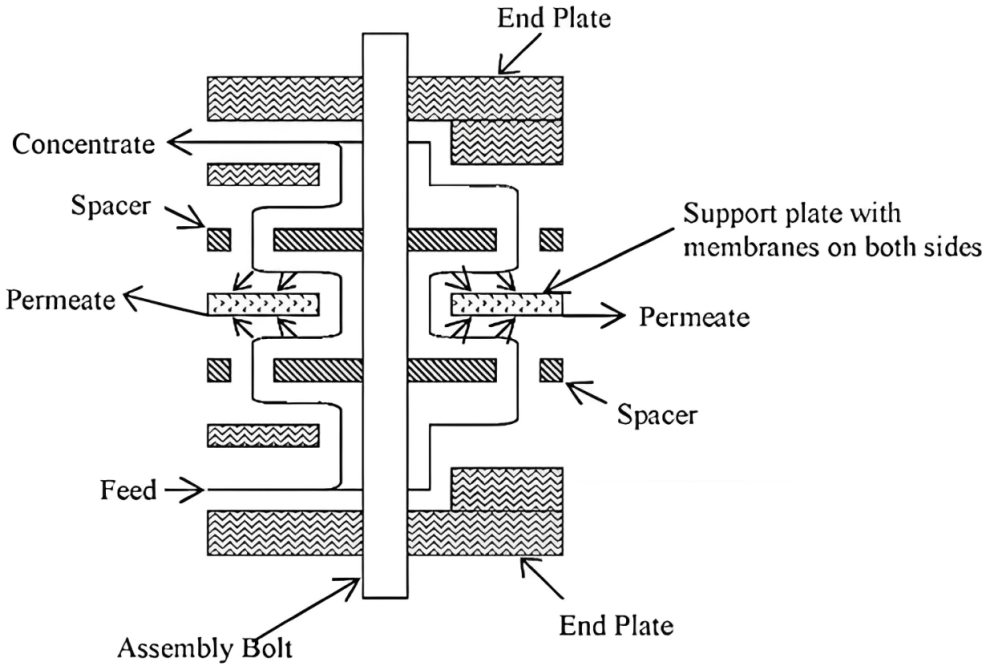


Figure 1.4: Plate and frame membrane module

1.3.2 Tubular modules

Tubular modules are straightforward in construction, typically consisting of a porous tube with an inserted or surface-coated RO membrane, as shown in Figure 1.5. These tubes, made from materials like ceramic, carbon, paper, plastic, or fiberglass, allow pressurized feed water to enter from one end, permeate radially through the membrane to produce product water, and discharge the concentrate from the other end[20]. Multiple tubes can be arranged in series or parallel to enhance system capacity[19] [21].
Despite being easy to clean, tubular membrane modules are considered uneconomical and have low packing density, making them less common in RO desalination but more frequently used in microfiltration (MF) and ultrafiltration (UF) applications[22].

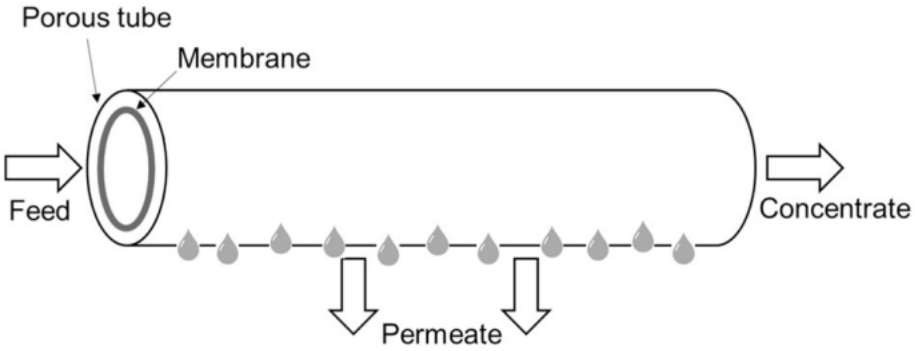


Figure 1.5: Tubular module

1.3.3 Hollow fiber modules

A hollow fiber module comprises numerous small-diameter (hair-like) fibers housed within a pressure vessel. One end of the module features an epoxy tube sheet where the fiber ends are potted in epoxy but remain open for permeate flow. The other end has an epoxy nub that seals the fibers to prevent feed from bypassing to the concentrate outlet [23]. The module also includes a porous feed distributor (core tube) running the length of the module [24]. As pressurized feed water enters through the core tube, water molecules permeate radially into the fibers and exit through the open ends in the epoxy tube sheet, while the concentrate exits at the same end as the feed inlet (Figure 1.6).
Hollow fiber modules are economical and have high packing density and recovery rates. However, they are difficult to clean and highly susceptible to fouling due to the small spacing between fibers [25].

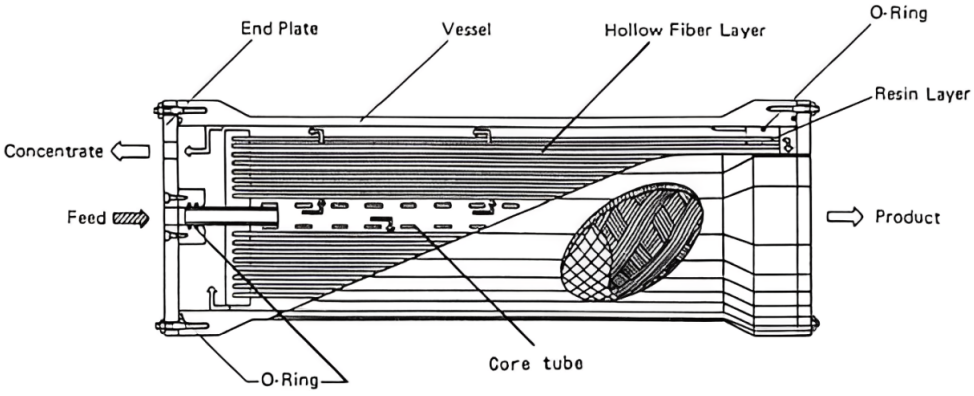


Figure 1.6: Hollow fiber modules

1.3.4 Spiral wound modules

The most common type of module used for RO desalination are currently spiral wound modules. In a spiral wound module, shown in Figure 1.7, two membrane sheets are placed together with a permeate spacer (made of nylon or dacron) in between to form a leaf[24]. The membrane sheets are glued from three sides with the fourth side left open and connected to a central perforated permeate collector tube [26]. The leaves are then placed together with a feed/concentrate mesh spacer to induce turbulence and minimize the CP effects. The combination of leaves and feed/concentrate spacers is wrapped around the permeate collector tube to create a spiral configuration and finally placed inside a pressure vessel (also known as housing).[15]

Feed water enters the module from one end and travels axially along its length. Water molecules permeate through the membrane and are collected as permeate in the perforated permeate collector tube, while the concentrate exits at the opposite end from the feed inlet.

Spiral wound modules are cost-effective, have high packing density, and enable high mass transfer rates due to the presence of feed spacers. However, they are difficult to clean and prone to fouling if pre-treatment is inadequate. Additionally, they can result in a high feed side pressure drop[27].

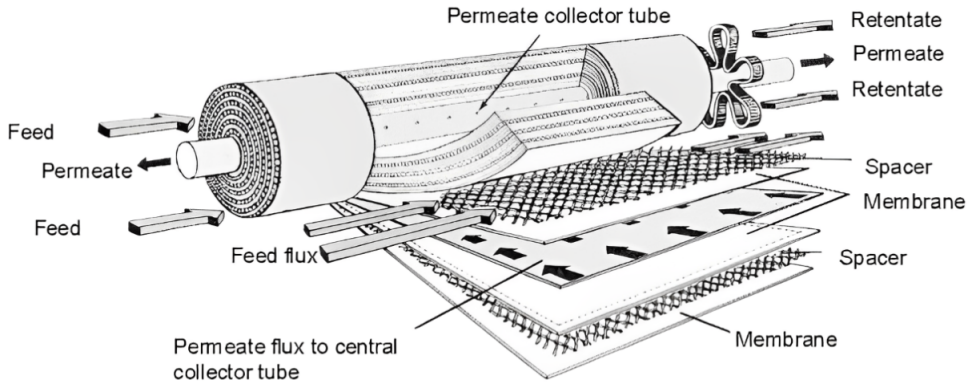


Figure 1.7: Spilar wound module

1.3.5 Comparison of Module types

Table 1.1 compares the four RO membrane modules in terms of packing density, fouling propensity, ease of cleaning, and manufacturing cost[28].

Module type	**Packing density** (ft^2/ft^3)	**Fouling propensity**	**Ease of cleaning**	**Manufacturing cost**
Plate and frame	45-150	Moderate	Good	High
Tubular	6-120	Low	Excellent	Very high
Spiral wound	150-380	High	Poor	Moderate
Hollow fiber	150-1500	Very high	Poor	Low

Table 1.1: Comparison of RO membrane module types

1.4 Membrane characterization

Characterization of RO membranes is important in understanding the membrane structure, morphology, chemical composition, and physio-chemical properties.
This process involves measuring pore size, examining surface features, and analyzing chemical composition. Techniques like microscopy and spectroscopy are used to visualize and study these aspects. Additionally, methods to assess thermal stability, surface wetting properties, and electrical characteristics are important. Overall, these characterizations help in selecting and designing effective RO membranes for various applications.

1.5 Conclusion

In conclusion, this chapter delved into the state-of-the-art research on RO desalination, its basics and principles, and an overview on membrane technology.

Throughout the chapter, several key aspects were discussed, including the basics of a reverse osmosis process, the difference between natural osmosis and reverse osmosis, as well as the performance parameters such as recovery, rejection and flux. An analysis of the different types of membrane modules was conducted, in addition to the characterization

In the next chapter, the mathematical modeling of a reverse osmosis process is carried out, establishing a solid understanding of the underlying dynamics of the transport models.

Chapter 2

Mathematical modeling of a reverse osmosis process

2.1 Introduction

The choice of a mathematical model for a given process largely depends on the intended application. Regarding reverse osmosis, several modeling approaches are found in the literature. In this chapter, we will justify our choice of the solution-diffusion (SD) dynamic model, develop the corresponding equations, and discuss their implementation.

We also present the simulations conducted on the model developed along with the experimental validation. Also, we will analyze the performance of the RO system and the effect of various parameters on the system's behavior.

2.2 Transport models

Membrane transport models are crucial for understanding the movement of solutes and water through RO membranes. These models play a significant role in predicting membrane performance and aid in the design of advanced membranes with enhanced properties. Over the years, various transport models have been developed, which can broadly be categorized into two main types: phenomenological and mechanistic models [29].

2.2.1 Phenomenological models

Phenomenological models, grounded in the principles of irreversible thermodynamics, do not rely on the specifics of the transport mechanism or membrane structure. They treat the membrane as a "black box" near equilibrium, dividing the system into smaller subsystems with local equilibrium [29].

These models describe membrane performance and flux using measurable quantities such as water flux and salt rejection. While convenient due to their minimal data requirements and applicability when membrane structure is unknown, phenomenological models do not offer insights into the actual transport mechanisms involved [30].

2.2.2 Mechanistic models

The performance of a membrane is related to the physicochemical parameters of both the membrane and the solutes. Mechanistic models assume a certain type of transport mechanism and relate the membrane performance to to the physicochemical parameters of both the membrane and the solutes[29].

These models can be classified into non-porous and porous transport models. The non-porous models assume that the membrane is non-porous or homogeneous. Common non-porous transport models include the solution-diffusion (SD) model, the extended solution-diffusion (ESD) model, and the solution-diffusion-imperfection (SDI) model. Fig. 2.1 shows the classification of the membrane transport model.

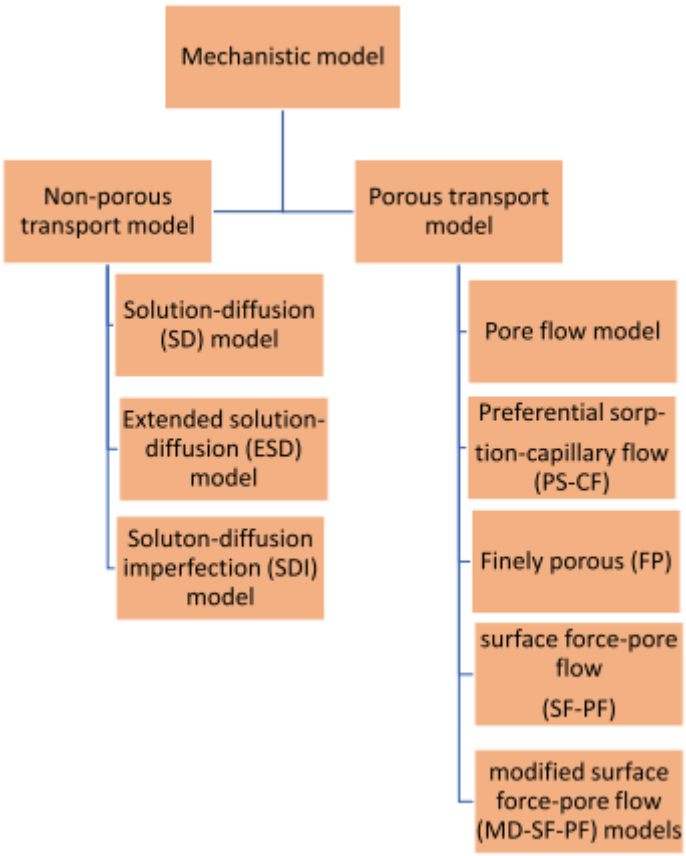


Figure 2.1: Classification of mechanistic model in RO.d

2.3 Solution-Diffusion model

The solution-diffusion model is the most used among all models. It is based on the diffusion of the solvent and the solute trough the membrane. This model assumes that both solute and solvent dissolve at the membrane surface and then diffuse across it. The solute and solvent diffusion are separate processes resulting from concentration and pressure differences across the membrane.[31]

We opted to choose this model due to its ability to describe solute and solvent transport through membranes, which is crucial for predicting fouling behavior.

2.3.1 Transport equations

The solvent flux of the permeate depends on the hydraulic pressure applied across the membrane, minus the difference in the osmotic pressure of the solutions of the feed and permeates side of the membrane [32] :

$$J_w = A_w(\Delta P - \Delta\pi) \tag{2.1}$$

A_w is the water permeability constant, $\Delta\pi$ represents the difference in osmotic pressure on both sides of the membrane. It is expressed as:

$$\Delta\pi = \pi_F - \pi_P \tag{2.2}$$

where the subscripts F refers to the feed side and the subscript P to the permeate side.

For moderate solute concentration, the osmotic pressure is approximately a linear function of solute concentrations:[31]

$$\pi = kC \tag{2.3}$$

where k is a proportionality coefficient. By substituting Eq. 2.3 into Eq. 2.1 we obtain:

$$J_w = A_w(\Delta P - k\Delta C) \tag{2.4}$$

where ΔC is the difference in solute concentration across the membrane, expressed as:

$$\Delta C = C_F - C_P \tag{2.5}$$

where C_F and C_P are the concentration in the feed and the permeate side, respectively.

The volumetric flow rate can be expressed as:

$$Q_w = \frac{J_w S_a}{\rho_w} \tag{2.6}$$

where S_a is the membrane surface area and ρ_w is the water density.

For the solute flux, it is assumed that the chemical potential difference due to pressure is negligible and so the driving force is almost entirely due to concentration differences. From Fick's law, the solute mass flux is:

$$J_s = B_s(\Delta C) \tag{2.7}$$

where B_s is the solute permeability coefficient which is a function of the solute composition and membrane structure.

The solute mass flow rate is expressed as:

$$\dot{Q}_s = J_s S_a = B_s S_a (C_F - C_p) \tag{2.8}$$

2.4 Spatial Dependence

The spatial dynamics in the solution-diffusion model provide a more comprehensive and detailed understanding of the system. This enhanced perspective allows for the precise localization of fouling within the membrane, which is crucial for optimizing membrane performance and maintenance.
The primary objective of this model is to distinguish each elemental section from the others through the use of differential dependency equations. By incorporating spatial variables, the model accounts for variations and dependencies that occur at different points within the membrane.
This differentiation is achieved by solving a set of differential equations that describe the behavior and interactions of the system's components across the spatial domain [33]. The following schematic represents the change in flow rate across a section ΔX

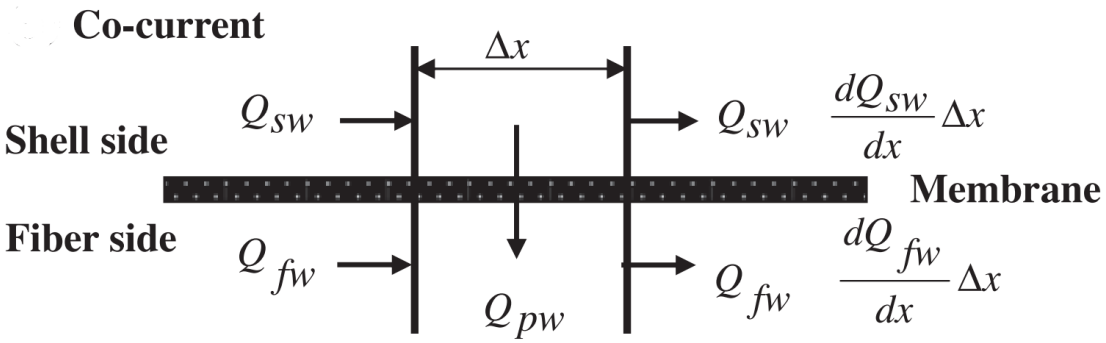


Figure 2.2: Schematic of the change in flow rate across a section ΔX.

In this flow pattern, the permeate and the feed in the fiber side and the shell, respectively, flow co-currently [33]. According to the solution-diffusion model, the rate of permeate in an elemental section length ΔX for water is Q_{pw}:

$$Q_{pw} = \frac{A_w}{\rho_w}\left(\pi D_m \Delta x\right)\left(\Delta P - \kappa(C_F - C_p)\right) \tag{2.9}$$

where D_m is the membrane mean diameter.

The material balance for water in the shell side can be written as:

$$Q_{sw} = \left(Q_{sw} + \frac{dQ_{sw}}{dx}\Delta x\right) + Q_{pw} \tag{2.10}$$

where Q_{sw} is the water volumetric flow rate in the shell side.
The first subscript indicates the shell or the fiber side, and the second subscript indicates the solute or the water.

This equation can be written in the following form:

$$\frac{dQ_{sw}}{dx} = -\frac{Q_{pw}}{\Delta x} \tag{2.11}$$

Substituting Q_{pw} for its expression in Eq. 2.9 gives:

$$\frac{dQ_{sw}}{dx} = -\frac{A_w}{\rho_w}(\pi D_m)(\Delta P - \kappa(C_f - C_p)) \tag{2.12}$$

C_F can be expressed as:

$$C_F = \frac{\dot{Q_{ss}}}{Q_{sw}} \tag{2.13}$$

where $\dot{Q_{ss}}$ is the solute mass flow rate in the shell side and Q_{sw} is the volumetric flow rate.

A similar equation is obtained for the solute concentration C_p in the permeate flow rate (fiber side):

$$C_p = \frac{\dot{Q}_{fs}}{Q_{fw}} \tag{2.14}$$

where $\dot{Q_{fs}}$ and Q_{fw} are, respectively, the solute mass flow rate and the water volumetric flow rate in the fiber side.

Substituting C_F and C_p for their expressions in Eq. 2.12 gives:

$$\frac{dQ_{sw}}{dx} = -\frac{A_w}{\rho_w}(\pi D_m)\left(\Delta P - \kappa\left(\frac{\dot{Q_{ss}}}{Q_{sw}} - \frac{\dot{Q_{fs}}}{Q_{fw}}\right)\right) \tag{2.15}$$

The material balance for water on the fiber side is obtained in the same manner:

$$\frac{dQ_{fw}}{dx} = \frac{A_w}{\rho_w}(\pi D_m)\left(\Delta P - \kappa\left(\frac{\dot{Q_{ss}}}{Q_{sw}} - \frac{\dot{Q_{fs}}}{Q_{fw}}\right)\right) \tag{2.16}$$

Similarly, the material balance for the solute in the shell side Fig. 2.2 can be written as:

$$\dot{Q}_{ss} = \left(\dot{Q}_{ss} + \frac{d\dot{Q}_{ss}}{dx}\Delta x\right) + \dot{Q}_{ps} \tag{2.17}$$

where $\dot{Q_{ss}}$ is the solute mass flow rate in the shell side, and $\dot{Q_{ps}}$ the solute mass flow rate across the membrane.

Arranging this equation and substituting $\dot{Q_{ps}}$ for its expression Eq. 2.8 gives:

$$\frac{d\dot{Q_{ss}}}{dx} = -B_s(\pi D_m)\left(\frac{\dot{Q_{ss}}}{Q_{sw}} - \frac{\dot{Q_{fs}}}{Q_{fw}}\right) \tag{2.18}$$

A similar equation is obtained for the fiber side:

$$\frac{d\dot{Q_{fs}}}{dx} = B_s(\pi D_m)\left(\frac{\dot{Q_{ss}}}{Q_{sw}} - \frac{\dot{Q_{fs}}}{Q_{fw}}\right) \tag{2.19}$$

Finally, the mathematical model obtained is composed of a set of four ordinary differential equations. To determine the permeate flow rate at the end of the module, equations 2.15, 2.16, 2.18 and 2.19 must be integrated simultaneously.

$$\begin{cases} \dfrac{dQ_{sw}}{dx} = -\dfrac{\pi D_m A_w}{\rho_w}\left(\Delta P - \kappa\left(\dfrac{\dot{Q}_{ss}}{Q_{sw}} - \dfrac{\dot{Q}_{fs}}{Q_{fw}}\right)\right) \\ \dfrac{dQ_{fw}}{dx} = \dfrac{\pi D_m A_w}{\rho_w}\left(\Delta P - \kappa\left(\dfrac{\dot{Q}_{ss}}{Q_{sw}} - \dfrac{\dot{Q}_{fs}}{Q_{fw}}\right)\right) \\ \dfrac{d\dot{Q}_{ss}}{dx} = -\pi D_m B_s\left(\dfrac{\dot{Q}_{ss}}{Q_{sw}} - \dfrac{\dot{Q}_{fs}}{Q_{fw}}\right) \\ \dfrac{d\dot{Q_{fs}}}{dx} = \pi D_m B_s\left(\dfrac{\dot{Q}_{ss}}{Q_{sw}} - \dfrac{\dot{Q}_{fs}}{Q_{fw}}\right) \end{cases} \tag{2.20}$$

2.5 Modeling of a closed-loop reverse osmosis installation

There are two main modes of an RO system operation:

Continuous mode : the feed characteristics remain the same, and the retentate or concentrate is collected separately from the permeate. If an initial feed volume is used, the feed is run to exhaustion. This mode assumes no mass transfer inhibition, and rejection, flux, and stream concentrations ideally remain constant over time.

Semi-Batch Mode : We're adopting this mode which operates in an unsteady state where the retentate is recycled to the feed tank, and permeate is collected separately as shown in Fig 2.3 . As the operation progresses, the feed volume decreases, and the concentration of the feed increases over time. This creates a closed-loop concentrating system where the feed becomes more concentrated, affecting the flux and recovery rates[32].

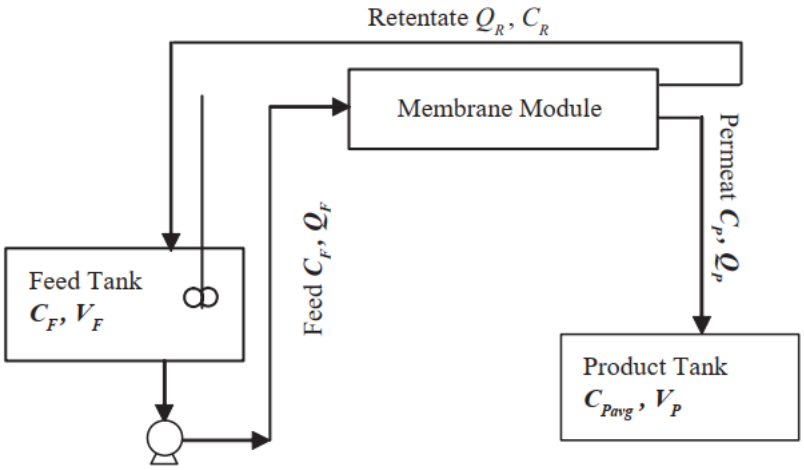


Figure 2.3: Schematic of reverse osmosis system with semi-batch mode.

The system material balances, together with the mass transfer models, were used to simulate system operation. Correlation of flux, solute concentrations and rejection with operating time and overall system recovery are functions of the model.

This model also predicts operational performance characteristics of the system at various times and recoveries. The effects of pressure, feed concentration, volume and membrane characteristics on separation efficiency can also be described.

A material balance made on the product tank yields :

$$Q_P C_P = \frac{d(V_P C_{Pavg})}{dt} \tag{2.21}$$

$$Q_P C_P = \frac{d(V_P)}{dt} C_{Pavg} + \frac{d(C_{Pavg})}{dt} V_P \tag{2.22}$$

The change in the volume of the permeate with time is the production rate of the membrane.

$$\frac{d(V_P)}{dt} = Q_P \tag{2.23}$$

By substitution of this in Eq. (2.7) we get

$$\frac{d(C_{Pavg})}{dt} = \frac{Q_P(C_P - C_{Pavg})}{V_P} \tag{2.24}$$

The material balance around the membrane module is

$$Q_P C_P = Q_F C_F - Q_R C_R \tag{2.25}$$

In this balance an assumption is made that in this system the concentration with in the membrane does not change greatly with spatial distribution. A mean permeate concentration from the membrane module was used.
Similarly the balance around the feed tank becomes

$$Q_R C_R - Q_F C_F = \frac{d(V_{Ft} C_{Ft})}{dt} \tag{2.26}$$

In the model it was assumed that the feed tank was well mixed. Thus the concentration of the feed to the membrane equals the concentration in the feed tank. Therefore,at any instant in time, t, $C_{Ft} = C_F$
The combination of Eqs. (2.25) and (2.26) with substitution of C_{Ft} as C_F, gives

$$-Q_P C_P = \frac{d(V_{Ft})}{dt} C_F + \frac{d(C_F)}{dt} V_{Ft} \tag{2.27}$$

The change in the feed volume with time can be take as the production rate, so

$$-\frac{d(V_{Ft})}{dt} = Q_P \tag{2.28}$$

Integrating with boundary condition at t = 0 with $V_F = V_F^0$

$$V_F = V_F^0 + Q_P t \tag{2.29}$$

the substitution of these expressions in Eq 2.12 gives :

$$\frac{d(C_F)}{dt} = \frac{Q_P(C_F - C_P)}{V_F^0 + Q_P t} \tag{2.30}$$

To get the solution of Eq. (2.30) we need the relationship between Q_P, and C_F with the expression for C_P, in terms of C_F. While the solute flux depends on the concentration gradient:

$$J_s = B_s(C_F - C_P) \tag{2.31}$$

the permeate concentration is equal to material balance around the membrane:

$$J_s = J_W \frac{C_P}{C_{WP}} \tag{2.32}$$

By combining Eq (2.31), Eq (2.32) and Eq (2.4), we get :

$$C_f = \frac{A_w \Delta P C_P}{B_s C_{WP}} - \frac{A_w k C_F C_P}{B_s C_{WP}} + \frac{A_w k C_P^2}{B_s C_{WP}} + C_P \tag{2.33}$$

Approximating the equation through $\frac{A_w \Delta P}{B_s} \gg \frac{A_w k}{B_s}$ for high rejection :

$$C_P = \frac{C_f}{a_3 - a_4 C_F} \tag{2.34}$$

By substitution of the expression for C_P, in Eq (2.4) we get the expression for flux in terms of C_F :

$$J_w = A_w(\Delta P - kC_F + \frac{C_f}{a_3 - a_4 C_F}) \tag{2.35}$$

Knowing that The flow of solvent through the membrane is defined in terms of flux

$$J_W = \frac{Q_p}{S_a} C_{WP} \tag{2.36}$$

substitution of Eq. (2.20) into Eq. (2.21):

$$Q_p = a_1 - a_2 C_F + \frac{a_2 C_F}{a_3 - a_4 C_F} \tag{2.37}$$

Putting the expression for C_P and Q_P, into the expression for the concentration change with time into Eq. 2.30, we get:

$$\frac{d(C_F)}{dt} = \frac{[a_1 - a_2 C_F + \frac{a_2 C_F}{a_3 - a_4 C_F}][C_F - \frac{C_f}{a_3 - a_4 C_F}]}{a_5 - a_1 t + a_2 C_{FT} - a_2 C_F + \frac{a_2 C_F}{a_3 - a_4 C_F}} \tag{2.38}$$

where the model constants are : $a_1 = \frac{S_a A_w \Delta P}{C_{WP}}$, $a_2 = \frac{S_a A_w k}{C_{WP}}$, $a_3 = 1 + \frac{A_w \Delta P}{B_s C_{WP}}$, $a_4 = \frac{A_w \phi}{B_s C_w}$, $a_5 = V_{F0}$, $a_6 = C_{f0}$, $a_7 = \frac{S_a A_w}{C_w}$.

The overall mass balance is :

$$V_{F0} C_{F0} = (V_{F0} - V_P) C_F + V_P C_{Pav} \tag{2.39}$$

$$C_{F0} = [C_F(1 - \frac{V_P}{V_{F0}} + \frac{V_P}{V_{F0}} C_{Pav}] \tag{2.40}$$

The equation for total dissolved solid (TDS) concentration in the product tank can be obtained by substituting Eq. (27) into Eq. (14) :

$$\frac{d(C_{Pav})}{dt} = \frac{[a_1 - a_2 C_F + \frac{a_2 C_F}{a_3 - a_4 C_F}][\frac{C_f}{a_3 - a_4 C_F} - C_{Pav}]}{a_5 - a_1 t + a_2 C_{FT} - a_2 C_F + \frac{a_2 C_F}{a_3 - a_4 C_F}} \tag{2.41}$$

2.6 Nonlinear State-Space Model for Reverse Osmosis Process

We can represent the nonlinear system expressed in equations Eq. (2.23) and Eq. (2.26). Let's define the state variables as follows:

- $x_1 = C_f$: feed concentration (kg/m^3)
- $x_2 = C_{pavg}$: permeate concentration (kg/m^3)

The outputs are defined as follows:

- $y_1 = C_f$
- $y_2 = J_w$

2.6.1 Differential Equations

The nonlinear state-space representation can be formulated using the given parameters. The system can be described by the following differential equations:

$$\begin{cases} \frac{dX_1}{dt} = \frac{(a5-\frac{Sa\cdot Aw\cdot u}{Cw}\cdot t+\frac{Sa\cdot Aw\cdot \phi}{Cw}\cdot X_1\cdot t-\frac{Sa\cdot Aw\cdot \phi}{Cw}\cdot X_1\cdot t)}{(Cw\cdot\frac{Sa\cdot Aw\cdot u}{Cw}\cdot X_1-Cw\cdot\frac{Sa\cdot Aw\cdot \phi}{Cw}\cdot X_1^2+1+\frac{Bs\cdot Cw}{Aw}\cdot u-\frac{Bs\cdot Cw}{Aw}\cdot\phi\cdot X_1)\cdot(1-(1+\frac{Bs\cdot Cw}{Aw}\cdot u-\frac{Bs\cdot Cw}{Aw}\cdot\phi\cdot X_1)^{-1})} \\ \frac{dX_2}{dt} = \frac{(a5-\frac{Sa\cdot Aw\cdot u}{Cw}\cdot t+\frac{Sa\cdot Aw\cdot \phi}{Cw}\cdot X_1\cdot t-\frac{Sa\cdot Aw\cdot \phi}{Cw}\cdot X_2\cdot t)}{(Cw\cdot\frac{Sa\cdot Aw\cdot u}{Cw}\cdot X_1-Cw\cdot\frac{Sa\cdot Aw\cdot \phi}{Cw}\cdot X_2^2+1+\frac{Bs\cdot Cw}{Aw}\cdot u-\frac{Bs\cdot Cw}{Aw}\cdot\phi\cdot X_1)\cdot(1-(1+\frac{Bs\cdot Cw}{Aw}\cdot u-\frac{Bs\cdot Cw}{Aw}\cdot\phi\cdot X_1)^{-1})} \end{cases}$$

2.7 Model Verification

Experimental data from several studies were used to verify the model, with a simple aqueous salt (NaCl) solution[34]. The experimental setup comprised three tubular cellulose acetate membranes with a combined surface area of 0.181 m^2. The system's pH was maintained at 4.5, and the temperature was kept at 20°C. The feed rate to the membrane modules was 0.063 kg/s. An operating pressure gradient of $4.02 \times 10^{13} kg/mh^2$ was applied.

2.7.1 Experimental Validation

The first experimental run to which the model was tested utilized a 2.00 kg/m^3 NaCl aqueous solution. The reverse osmosis unit was operated for 30 h to obtain data for comparison to the model. Experimental results of solute concentration were plotted as functions of time and recovery. The constants and initial conditions for the simulation are given in Table 2.1. The variation of recovery with time was linear during the early stages of the experiment, but became nonlinear toward the end of the experiment.[35]

$C_{\mathrm{f}_0} = 2.00\ \mathrm{kg/m^3}$	$\Delta P = 4.02 \times 10^{13}\ \mathrm{kg/mh^2}$
$V_{\mathrm{f}_0} = 1.50\ \mathrm{m^3}$	$A_w = 4.20 \times 10^{-13}\ \mathrm{h/m}$
$C_{\mathrm{w}} = 1.0 \times 10^3\ \mathrm{kg/m^3}$	$B_s = 1.12 \times 10^{-4}\ \mathrm{m/h}$
$S_a = 0.181\ \mathrm{m^2}$	$\psi = 1.02 \times 10^{12}\ \mathrm{m^2/h^2}$

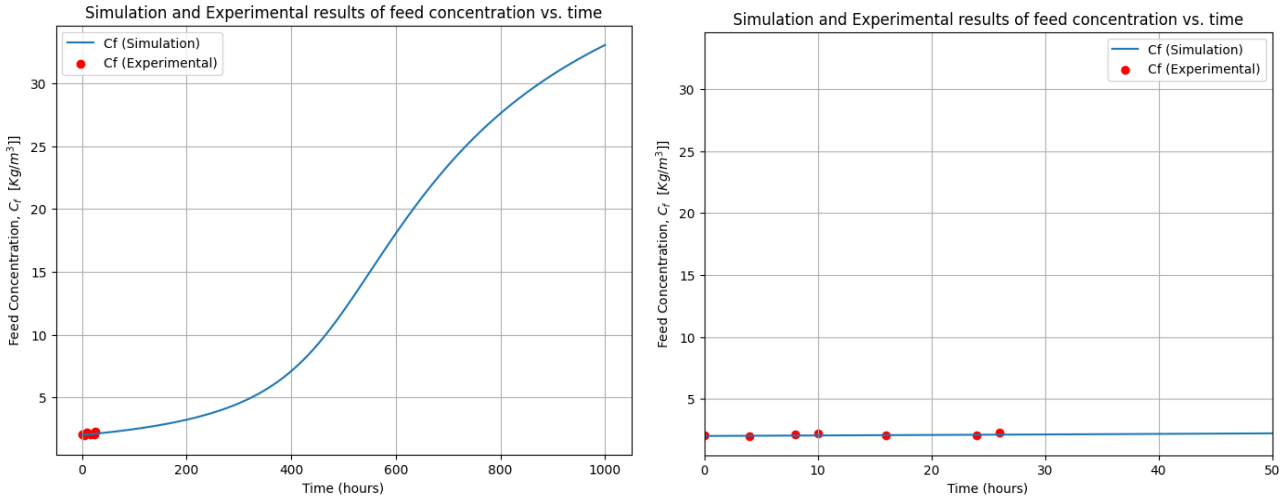
Table 2.1: The constants and initial conditions for the simulation

The experimental data for solute concentration and flux over time were plotted and compared with the model's predictions.

The results demonstrated a strong correlation between the experimental data and the model, with an accuracy of 95%. Specifically, the model accurately predicted the exponential increase in feed concentration over time and the behavior of flux under the given operating conditions as shown in Fig. 2.4.

Furthermore, the model's simulation of flux showed a good fit with the experimental data, as illustrated in Fig. 2.5. This agreement is critical because it confirms that the model can reliably predict the system's performance under different conditions.

This high degree of agreement between the experimental data and the model supports the validity of the model and its underlying assumptions. Consequently, the model can be used confidently for predicting the performance of similar systems under various operational scenarios.



(a) The whole time interval.

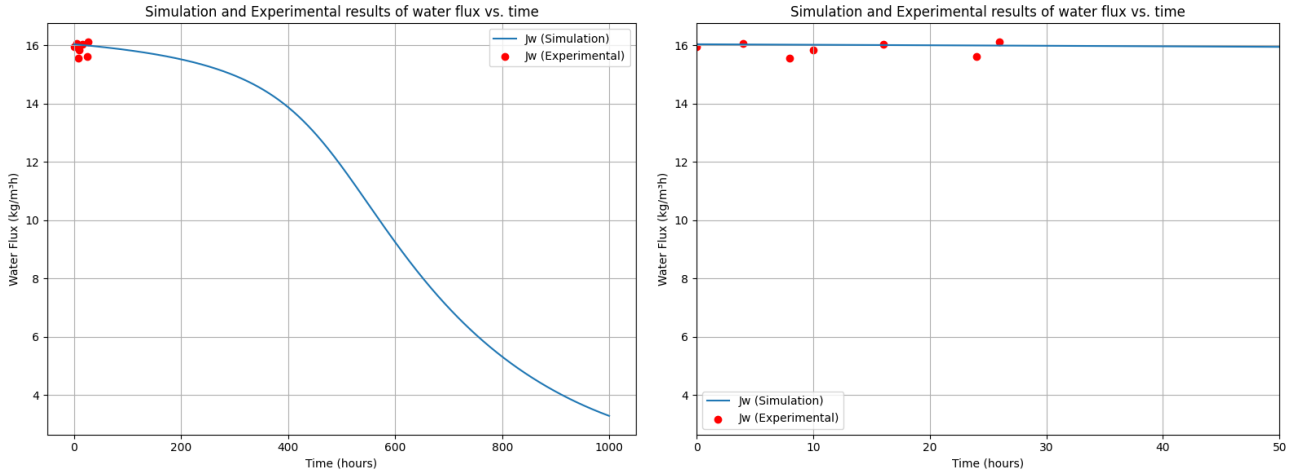
(b) The experimental results zoomed in.

Figure 2.4: Simulation and Experimental results of feed concentration vs. time

2.7.2 Process Analysis

Fig. 2.4a illustrates the simulation results of solute concentration in the feed over time. According to the model, the feed concentration increases linearly but slowly during the initial hours (up to 400 hours) of operation. After this period, the feed concentration begins to increase exponentially.

Initially, the system operates with a set feed volume, and the retentate concentration is con-

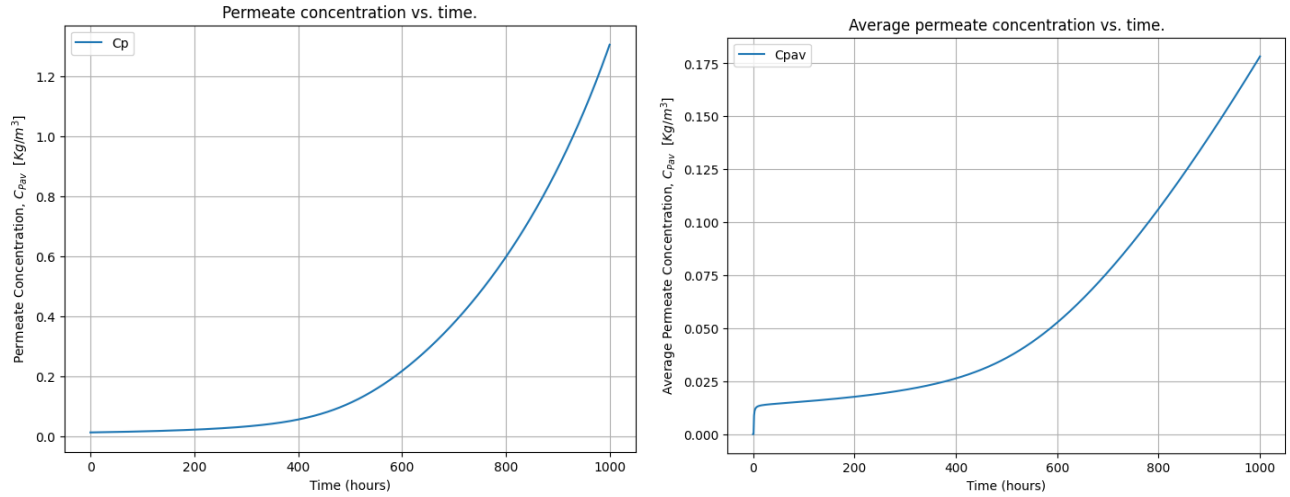


(a) The whole time interval. (b) The experimental results zoomed in.

Figure 2.5: Simulation and Experimental results of water flux vs. time.

sistently slightly higher than the feed concentration. The retentate is recycled back into the feed tank (as depicted in Fig. 2.3). The feed volume decreases over time due to permeate production, resulting in an increase in feed concentration. If the initial feed volume was larger, the linear increase in concentration would persist for a longer period.

Fig. 2.5a shows the variation of permeate flux with time. Initially, as operating time increases, the permeate flux decreases slowly. After around 400 hours, the flux decreases sharply. The simulation results suggest that while the permeate flux declines gradually at first, it falls more sharply over time due to the increasing concentration gradient.



(a) (b)

Figure 2.6: Simulation results of permeate concentration and average permeate concentration vs. time.

Fig. 2.6 shows the simulation results for both permeate and average product concentrations over time. According to the model, the permeate concentration increases gradually during the first 400 hours of operation. After this period, it rises rapidly, reaching 1.27 kg/m^3 by the end of the 1000-hour simulation. The average permeate concentration also increases but at a more gradual pace, reaching 0.178 kg/m^3 due to dilution effects in the product tank.

2.7.3 Effect of Initial Feed Concentration (V_{f0})

To understand the influence of different system parameters on the performance of the reverse osmosis system, we varied two parameters: initial feed volume V_{f0} and operating pressure gradient ΔP.

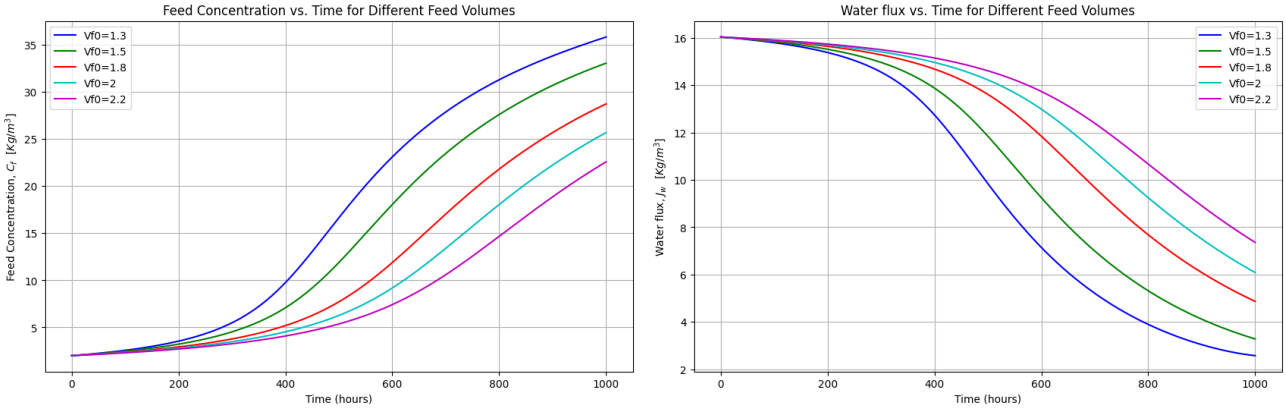
The influence of varying the initial feed volume (V_{f0}) on the system's performance was analyzed by simulating different V_{f0} values while keeping other parameters constant. The selected values of V_{f0} were 1.3, 1.5, 1.8, 2.0, and 2.2 m^3.

Feed Concentration (C_f) vs. Time

The rate of increase in C_f over time decreases with higher V_{f0}. This behavior is attributed to the dilution effect provided by the larger feed volume, which reduces the concentration of solutes in the feed stream over time, achieving lower final feed concentrations.

Water Flux (J_w) vs. Time

The water flux decreases more gradually at higher initial feed volumes. This is due to the higher dilution effect resisting the buildup of osmotic pressure, thus maintaining the driving force for water flux for a longer period.



(a) Feed Concentration vs. Time

(b) Water flux vs. Time

Figure 2.7: Effect of Initial Feed Volume (V_{f0})

2.7.4 Effect of Operating Pressure Gradient (dP)

The operating pressure gradient (dP) was varied to observe its effect on the system's performance. The chosen dP values were 3.7×10^{13}, 3.9×10^{13}, 4.0×10^{13}, 4.3×10^{13}, and 4.7×10^{13} kg/m/h^2.

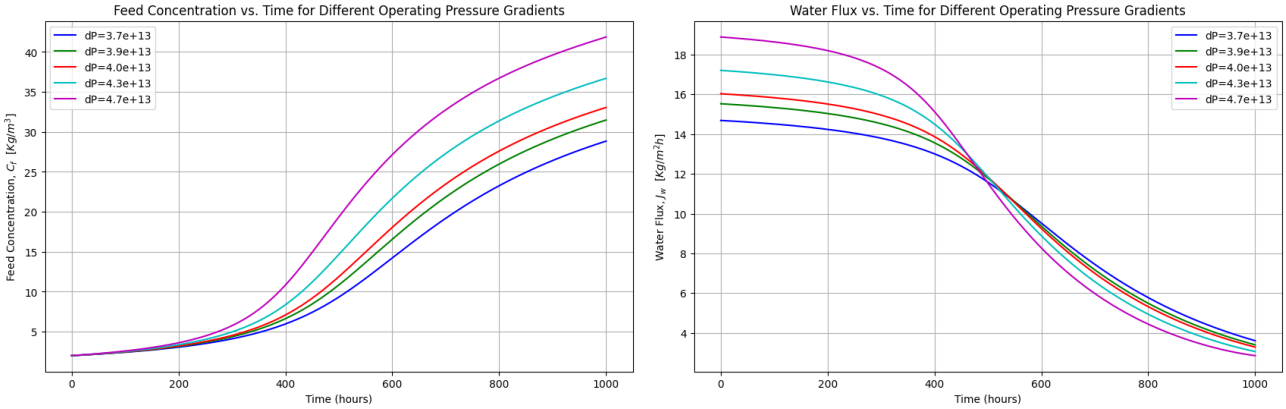
Feed Concentration (C_f) vs. Time

Higher operating pressure gradients result in faster increases in feed concentration. This is because the increased pressure gradient enhances the driving force for water permeation, leading to quicker diminution of the feed volume.

Water Flux (J_w) vs. Time

Water flux increases significantly with higher operating pressure gradients. This is expected as the pressure gradient directly enhances the permeation rate of water through the membrane.

Operating at higher pressure gradients can improve overall system efficiency by increasing water flux, though it must be balanced against potential increases in permeate concentration and energy costs.



(a) Feed Concentration vs. Time

(b) Water flux vs. Time

Figure 2.8: Effect of Operating Pressure Gradient (dP)

2.7.5 Simulation of Multivariable Distribution through Membrane Discretization

In this simulation, the membrane is discretized into six interconnected cells, governed by a system of four differential equations with respect to space (Eq: 2.20). As a result of this discretization, each cell is analyzed independently. Consequently, the scalar variables C_f and C_{pav} are transformed into vector variables. This multivariable model provides a clearer perspective and enhanced monitoring capabilities of the membrane, enabling not only the detection and quantification of fouling but also the localization of fouling within each individual cell and the determination of the fouling percentage in each cell.

The following figure presents the graphs of the vectorial operational variables over time.

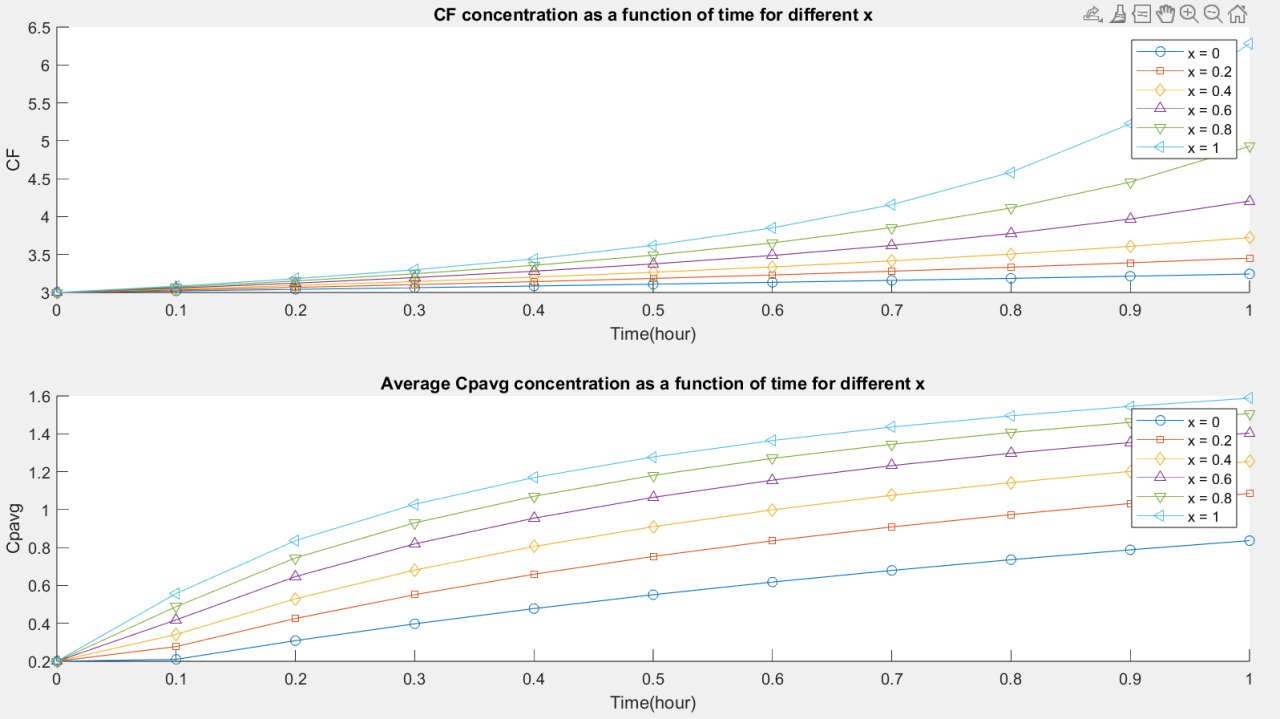


Figure 2.9: Multivariable Feed Concentration and Permeate Average Concentration vs. Time

Based on the results obtained from the model, we can infer that the variation of the feed concentration and the average permeate concentration is proportional to the spatial position. Over time, the average permeate concentration converges to a defined value, while the feed concentration varies exponentially. These results are consistent with observations from real plant operations, validating the accuracy of the model.

2.8 Conclusion

In this chapter, we have demonstrated our motivations for the choice of the solutions diffusion model. We derived the fundamental transport equations for solvent and solute fluxes through the RO membrane, which are crucial for understanding the system's behavior under different operating conditions.

Then, we extended the solution-diffusion model to include spatial dependence provided a more detailed view of the membrane. We implemented also the model in MATLAB and conducted simulations to validate it against experimental data, demonstrating the model's accuracy in predicting the system's performance. Finally, we examined the impact of different operational parameters on the system's performance, providing insights for optimizing RO efficiency.

Chapter 3

Modeling macroscale fouling in reverse osmosis membranes

3.1 Introduction to Membrane Fouling

Membrane fouling is a process that occurs when contaminants, such as minerals, organic matter, and microorganisms, accumulate on the surface of a reverse osmosis (RO) membrane, decreasing its ability to effectively purify the water. This can occur because of different factors such as scaling, biofouling, particulate fouling, and chemical fouling. The accumulation of these impurities leads to a decrease in water flow, increased pressure drops, and decreased water quality.

It is important to regularly clean and maintain the RO membrane to prevent fouling and maintain its efficiency in filtering water. Fouling occurs when soluble salts precipitate, adsorb, form a cake or gel, or clog pores of the membrane[36]. Fig. 3.1 shows that fouling can be of two types: external fouling on the surface of the membrane and internal fouling inside the membrane's pores.

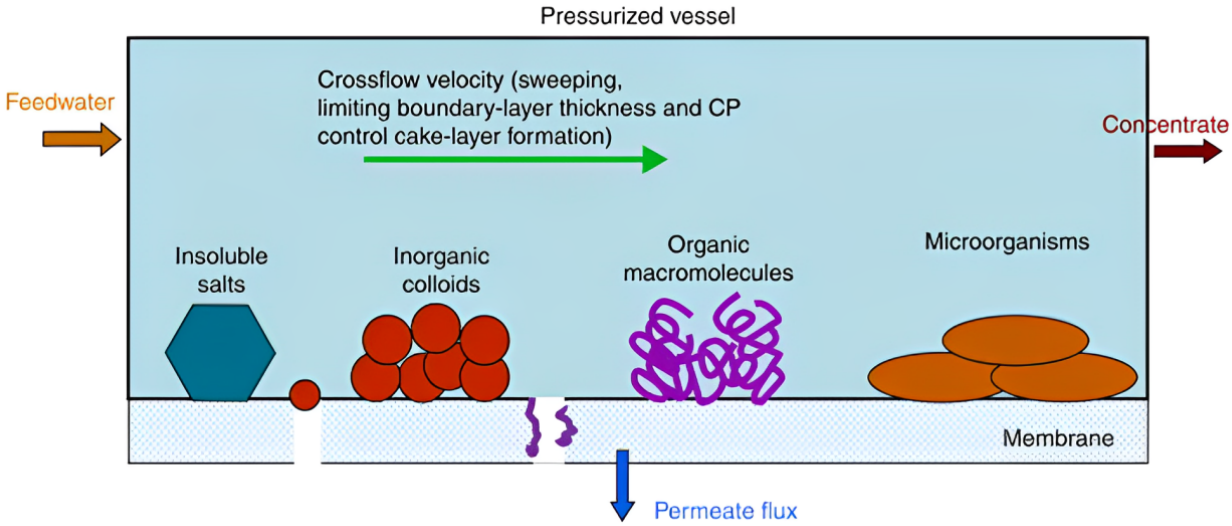


Figure 3.1: Representation of several fouling types

3.2 Modeling Cake Formation Mechanisms

Fouling of membranes has been a constant problem in membrane operations, lowering plant efficiency and increasing maintenance costs since the membrane business was first developed [35]. Modeling and simulation for anticipating fouling in RO systems have been the subject of several kinds of research since this would enable diagnostic and preventative steps to be implemented, thereby reducing the impacts of fouling [37, 38].

The scale formation mechanism in nanofiltration (NF) reverse osmosis (RO) systems can be theoretically modeled based on crystallization kinetics and the resistance-in-series model. If no fouling occurs, the initial water flux of the nanofiltration membrane will be given as:

$$J_w = \frac{\Delta P - \pi}{\eta R_m} \tag{3.1}$$

where J_w is estimated from the resistance-in-series model, ΔP is the applied pressure, η is the permeate viscosity, and R_m is the membrane resistance. Scale formation is a complex

process involving both crystallization and hydrodynamic transport mechanisms. Two pathways for crystallization have been identified: surface (heterogeneous) crystallization and bulk (homogeneous) crystallization [39, 40].

In the cake layer formation mechanism, crystal particles formed in the bulk phase through either bulk (homogeneous) or secondary crystallization deposit to form a precipitate layer on the membrane [40]. This accumulation leads to flux decline as described below:

$$J_c = \frac{\Delta P - \pi}{\eta(R_m + R_c)} \tag{3.2}$$

where J_c is the permeate flux estimated from the cake filtration model and R_c is the resistance due to the cake layer formation. R_c can be calculated using filtration theory and is given by:

$$R_c = \alpha \frac{A_c}{A_t} \tag{3.3}$$

where α is the specific cake resistance per unit weight of cake, A_c is the total accumulated weight of the precipitated scale, and A_t is the total membrane area.

3.3 Surface Blockage Mechanism

On the other hand, scale formation may also occur due to surface crystallization on the membrane surface through the surface blockage mechanism. Thus, the membrane surface would be blocked by the lateral growth of crystals. Assuming the areas occupied by crystals are completely impermeable, the flux in the absence of cake formation could be expressed as follows:

$$J_s = \frac{\Delta p - \pi}{\eta(R_m)} \times \frac{A_{free}}{A_t} = \frac{\Delta p - \pi}{\eta(R_m)} \times \frac{A_t - A_b}{A_t} \tag{3.4}$$

where J_s is estimated from the surface blockage model, A_t is the total membrane surface, A_b is the membrane area occupied by surface crystals, and $(A_t - A_b)$, A_{free} is the free membrane surface, uncovered by surface crystals. If the thickness of the crystal layer formed on the membrane surface is constant, the A_b could be defined as [41]:

$$A_b = \beta m_s \tag{3.5}$$

where β is the area occupied per unit mass, and m_s is the weight of scale formed directly on the membrane surface.

3.4 Combined Mechanisms of Fouling

At high operating recoveries, as the bulk phase becomes supersaturated, both cake formation and surface blockage may occur simultaneously [37]. The resulting permeate flux can be represented as given below:

$$J_t = \frac{\Delta p - \pi}{\eta(R_m + R_c)} \times \frac{A_t - A_b}{A_t} \tag{3.6}$$

where J_t is the permeate flux estimated by combining the cake filtration and the surface blockage models.

One way to interpret the fouling impact on an RO membrane is as an increase in total membrane resistance (R_T) or a decrease in the water permeability coefficient (A). One approach to quantifying RO membrane fouling involves figuring out the fouling factor that reduces A, as indicated in Eq.:

$$A = A_0 \times FF \tag{3.7}$$

The above equation represents A_0 as the membrane's initial water permeability, and FF is the fouling factor. Notably, concentration polarization (CP) events are connected to the RO membrane's feed spacer structure and are often studied using film theory [42, 43]. When estimating the coefficient A_0, membrane makers should consider this essential phenomenon for RO desalination plants. This coefficient would account for membrane compaction caused by transmembrane pressure (TMP).

3.5 Long-Term Performance Models

Wilf and Klinko [44] developed a model to predict the decline in permeate flow during extended operation, validated by experimental data from multiple SWRO desalination plants over three years, showing a 20–25% decrease in permeate flow. The model is:

$$A_n = t^m \tag{3.8}$$

where A_n is the normalized water permeability coefficient ($A_n = A/A_0$), t is time, and m is a parameter between -0.035 and -0.041.

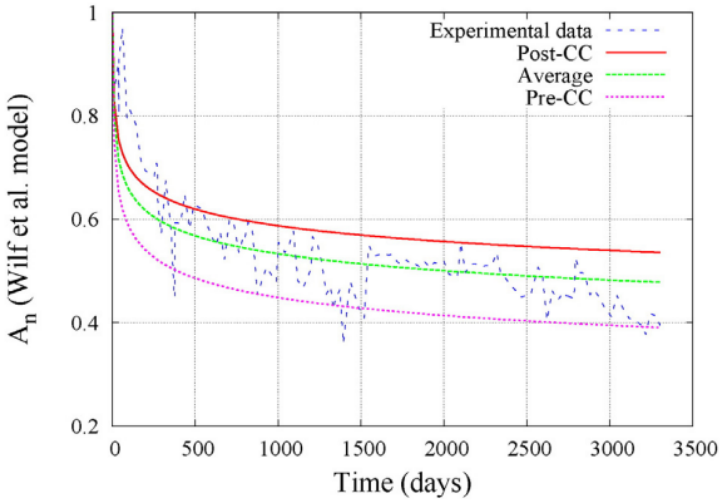


Figure 3.2: Wilf and al. model

Tolba and Mohamed [45] used this model to study the TFC 2822 Fluid system's efficiency in an SWRO plant, achieving 26–33% flux recovery with 6.70 MPa input pressure.

Abbas and Al-Bastaki [46] created an exponential model based on three factors and time, using five years of data from a BWRO plant with a BW30-400 FilmtecTM membrane. The model is:

$$A_n = \alpha e^{\frac{\beta}{t+\gamma}} \tag{3.9}$$

with parameters $\alpha = 0.68$, $\beta = 79$, and $\gamma = 201.1$.

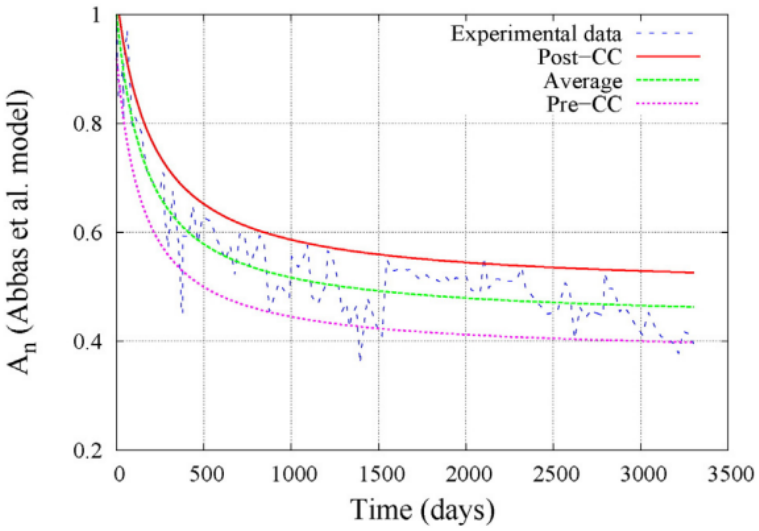


Figure 3.3: Abbas and al. model

An exponential equation-based approach for predicting coefficient 'A' was also presented [47], verified with a Dupont™B-1 hollow fiber membrane. The feed water concentration was constant at 35,000 mg/L and temperature at 27°C, with feed pressure varying from 6.28 to 7.09 MPa.

Belkacem et al. [48] used this model, focusing on membrane resistance rise rather than permeate flow decline, in a BWRO plant with a BW30LE-440 Filmtec™ membrane.

Lee et al. [49] proposed a model for quantifying membrane fouling, initially developed for NF membranes but applicable to RO membranes. The model is:

$$J = A_0 \times FF \times TMP \times \frac{S_t - S_b}{S_t} \tag{3.10}$$

where S_t is the total membrane area, S_b is the area occupied by scaling, and J is the permeate flow per unit area.

Ruiz-García and Nuez [50] proposed a model distinguishing two phases of coefficient 'A' decrease, explained by:

$$A_n = \delta_1 \cdot e^{\left(\frac{-t}{\tau_1} k_{fp}\right)} + \delta_2 \cdot e^{\left(\frac{-t}{\tau_2} k_{fp}\right)} \tag{3.11}$$

where δ_1 and τ_1 relate to the first phase, and δ_2 and τ_2 to the second.

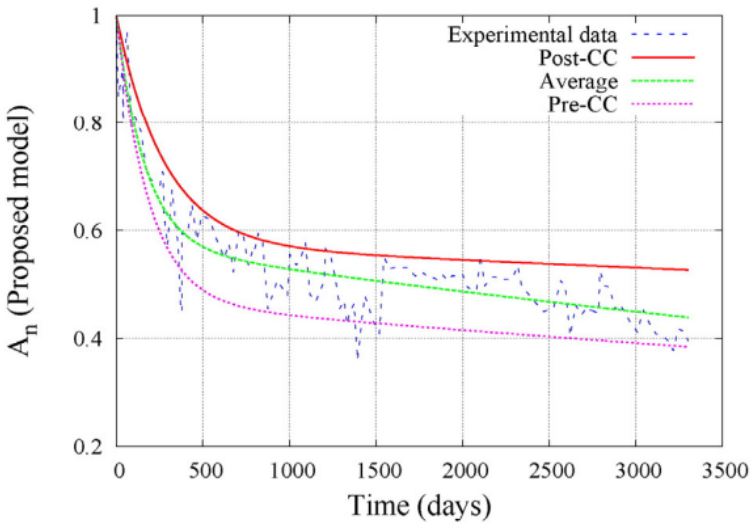


Figure 3.4: Ruiz and al. model

3.6 Impact of Fouling on Operational Variables in Reverse Osmosis Plant

3.6.1 Analysis Without Membrane Discretization

In reverse osmosis (RO) membrane systems, the water permeability coefficient, A_w, is a crucial parameter representing the membrane's ability to transmit water under a given pressure gradient. Fouling, including scaling, biofouling, and particulate fouling, directly affects A_w by adding resistance to water flow, thereby decreasing the membrane's permeability over time.

The simulation results for the fouling in membrane filtration systems demonstrate the effect of different forms of A_w (membrane permeability) on the system's performance. The three exponential forms for A_w are given as follows:

$$A_{w1}(t) = 4.2 \times 10^{-13} \exp(-0.001t)$$
$$A_{w2}(t) = 4.2 \times 10^{-13} \exp(-0.002t)$$
$$A_{w3}(t) = 4.2 \times 10^{-13} \exp(-0.003t)$$

The variation of A_w over time for the three different exponential forms is shown in Figure 3.5.

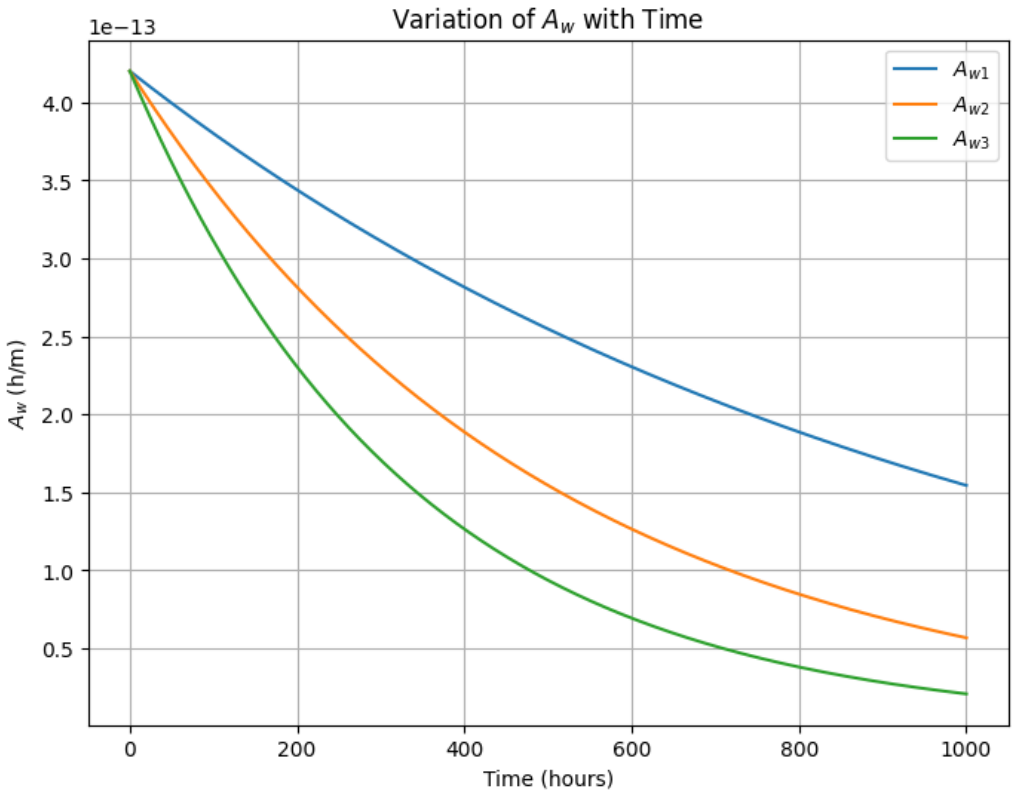


Figure 3.5: Variation of A_w with Time

Figure 3.6 shows the comparison of water flux J_w over time for the three different A_w functions.

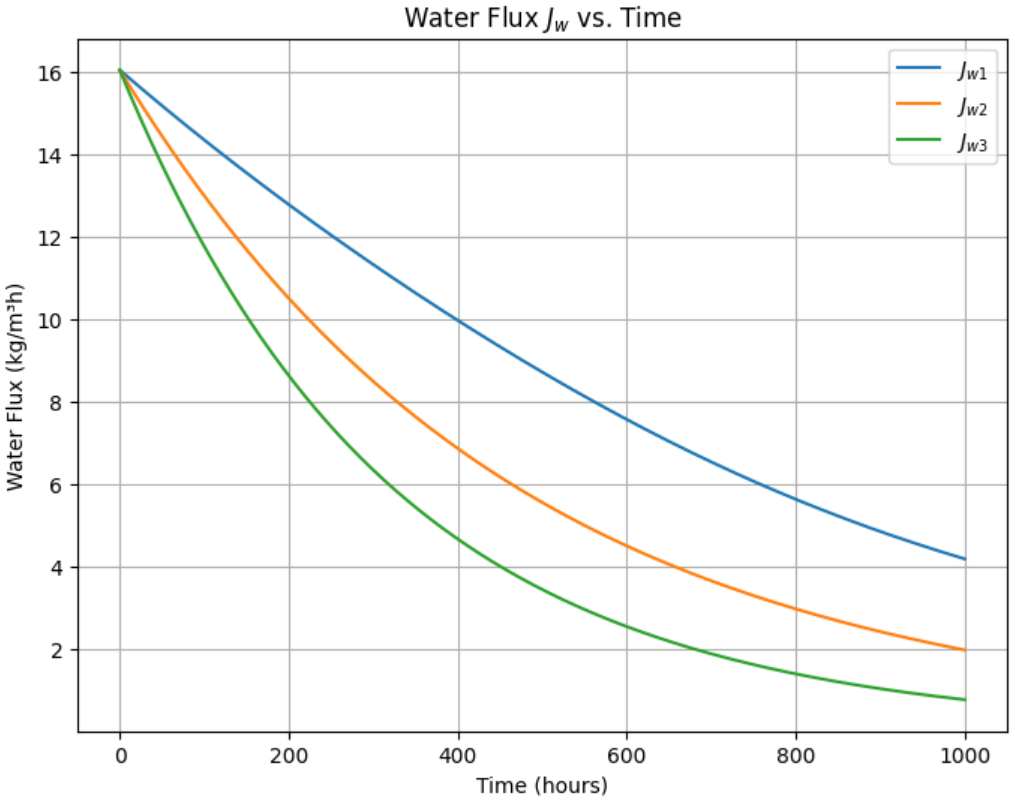


Figure 3.6: Comparison of Water Flux J_w vs. Time

Variation of A_w

The simulation results provide insights into how different rates of fouling, represented by the different exponential decays of A_w, impact the performance of the membrane filtration system.

As shown in Figure 3.5, all three exponential forms of A_w indicate a decrease in membrane permeability over time. The rate of decrease varies, with A_{w1} representing Standard pore blockage, A_{w2} Intermediate pore blockage, and A_{w3} Complete pore blockage. This suggests that A_{w1} represents a scenario with slower fouling, while A_{w3} represents a scenario with more rapid fouling.

Impact on Water Flux J_w

Figure 3.6 demonstrates the corresponding water flux J_w for the different forms of A_w. The results show that:

- For A_{w1}, the water flux decreases gradually over time. This indicates that the system maintains higher performance for a longer period under slower fouling conditions.
- For A_{w2} and A_{w3}, the water flux decreases more rapidly. Especially in the case of A_{w3}, the rapid fouling leads to a significant drop in performance early in the process.

These results highlight the importance of managing fouling in membrane filtration systems. Slower fouling rates (as in A_{w1}) allow for more sustained performance, while faster fouling rates (as in A_{w3}) can drastically reduce the system's efficiency.

3.6.2 Analysis With Membrane Discretization

The permeability coefficient A_w represents the mirror of fouling in our model. The localization of fouling and its level in each cell of the membrane will help us monitor the membrane more effectively.

After integrating fouling into our model for each cell, the following equations represent the permeability coefficient with respect to time and space:

$$
\begin{aligned}
A_w(t, x = 0) &= 4.2 \times 10^{-13} \exp(-0.001t) \\
A_w(t, x = 0.2) &= 4.2 \times 10^{-13} \exp(-0.002t) \\
A_w(t, x = 0.4) &= 4.2 \times 10^{-13} \exp(-0.003t) \\
A_w(t, x = 0.6) &= 4.2 \times 10^{-13} \exp(-0.004t) \\
A_w(t, x = 0.8) &= 4.2 \times 10^{-13} \exp(-0.005t) \\
A_w(t, x = 1) &= 4.2 \times 10^{-13} \exp(-0.006t)
\end{aligned}
$$

The following figure represents the simulation of the permeability coefficient with respect to time and space in the membrane:

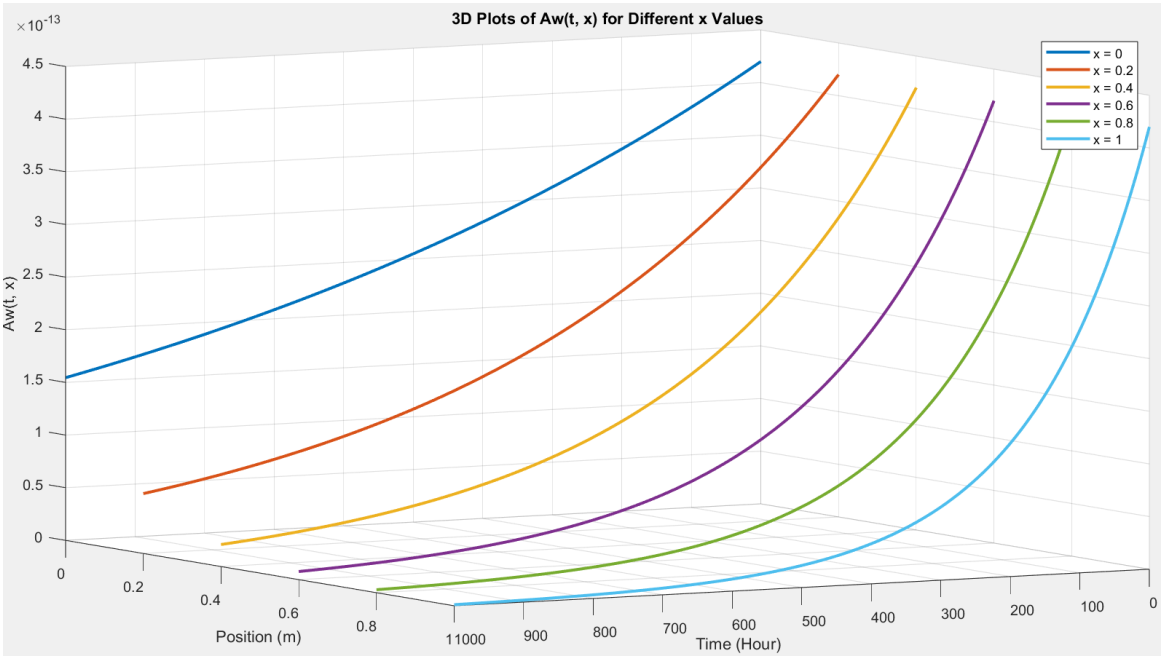


Figure 3.7: Variation of Permeability Coefficient A_w with Time and Space

The following figure represents the simulation of the permeate concentration with respect to time and space in the membrane:

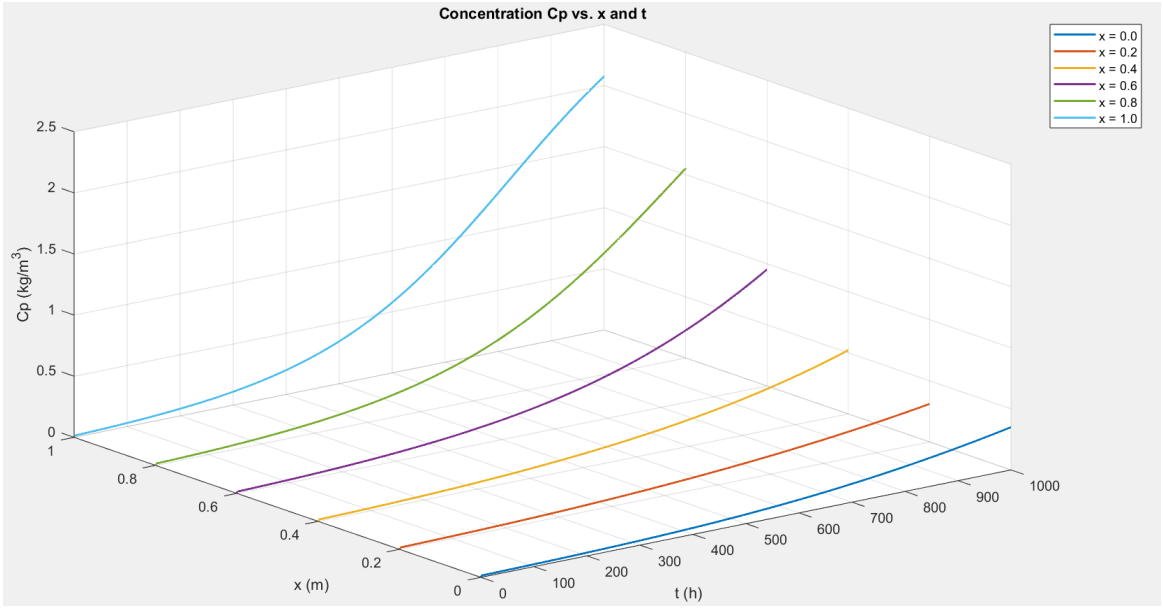


Figure 3.8: Variation of Permeate Concentration C_p with Time and Space

The following figure represents the simulation of the feed concentration with respect to time and space in the membrane:

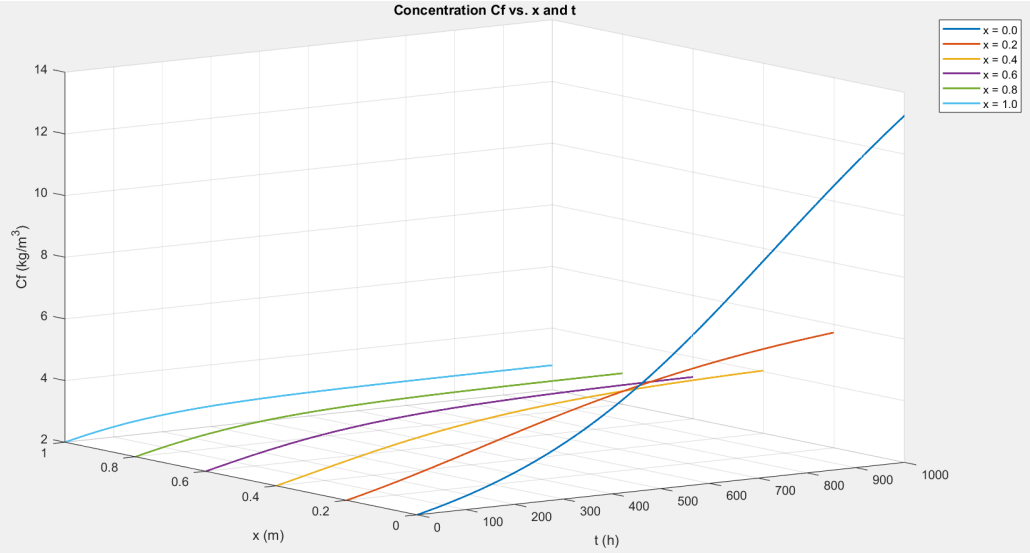


Figure 3.9: Variation of Feed Concentration C_f with Time and Space

From these two variables, we can observe that membrane fouling directly affects the variation of the operational variables with respect to space, entirely changing the system's behavior. Based on the time-space model that provides the state of the variables, we can localize and estimate the level of fouling in each cell.

This space-time model is a powerful tool that can help us develop a more advanced predictor for space and time, and determine solutions to mitigate the membrane fouling problem. Enhanced visualization leads to better solutions, predictions, and monitoring.

3.7 Conclusion

In conclusion, the simulation findings underscore the profound influence of fouling rate on critical parameters such as feed concentration, permeate concentration, and water flux in membrane filtration systems. This understanding underscores the importance of managing fouling rates effectively to optimize the design and operation of these systems, thereby improving their efficiency and durability. Leveraging a time-space model enables more accurate prediction and control of fouling, ensuring the reliability and sustainability of reverse osmosis plants.

Moreover, the detailed analysis provided by membrane fouling discretization offers valuable insights into localized fouling processes, enabling targeted maintenance and cleaning strategies. This approach enhances overall system efficiency and longevity by addressing fouling issues at specific problem areas which will surely reduce downtime and operational costs.

Chapter 4

Proposed Solution for Prediction of Membrane Fouling in Reverse Osmosis Plant

4.1 Introduction

Membrane fouling in RO systems significantly impacts efficiency, operational costs, and longevity, leading to performance decline, increased energy consumption, and higher maintenance expenses. Predicting long-term fouling is crucial for implementing proactive measures to mitigate its effects before they become irreversible.

To effectively tackle the issue of predicting membrane fouling, the problem can be systematically decomposed into two interrelated sub-problems:

0. **Online Membrane Fouling Parameter Estimation for RO Systems**: This involves developing a real-time estimation methodology to continuously monitor and evaluate the fouling parameters of the membrane. Accurate online estimation allows for immediate detection of changes in the fouling state, enabling timely intervention to prevent severe degradation and provide input time series data for prediction.

0. **Multi-step Ahead Prediction of Fouling**: Building on the real-time estimation data, this sub-problem focuses on forecasting the future state of membrane fouling over multiple time steps. The objective is to predict the progression of fouling, providing insights that facilitate advanced planning and maintenance scheduling to avoid critical performance losses.

The following figure illustrates the permeability coefficient forecasting architecture for a reverse osmosis plant using machine learning techniques:

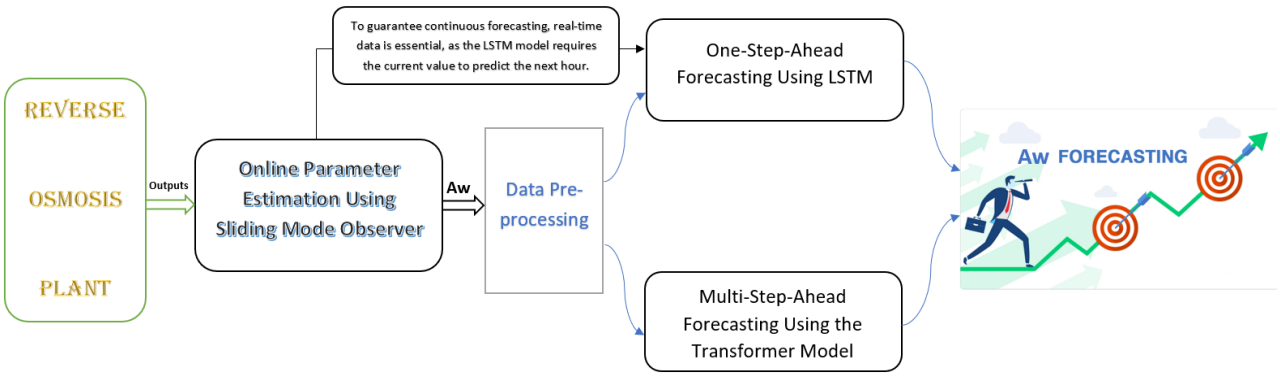


Figure 4.1: Machine Learning-Based Permeability Coefficient Forecasting Architecture for Reverse Osmosis Plant

4.2 Online Parameter Estimation Using Sliding Mode Observer

Accurate parameter variation estimation is critical for maintaining control and ensuring the performance of the RO system without degradation. The Sliding Mode Observer-Based Online Parameter Estimation is proposed due to its efficiency in control calculation and finite-time convergence.
Addressing the multifaceted challenges of membrane fouling prediction in reverse osmosis (RO)

systems necessitates a robust and efficient approach capable of accommodating the nonlinear, time-, and parameter-variant characteristics inherent to such systems. In this regard, we advocate for the utilization of the Sliding Mode Observer (SMO) for online parameter estimation, owing to its array of inherent advantages:

- **Reduced Observation Error Dynamics**: The SMO is crafted to operate within minimized observation error dynamics, ensuring the precise estimation of the system's states with unparalleled accuracy.
- **Step-by-step Design**: Offering a systematic and methodical design process, the SMO enables practitioners to navigate implementation and fine-tuning procedures with utmost ease and efficacy.
- **Finite Time Convergence**: Crucially, the SMO guarantees finite time convergence for all observable states, facilitating the swift attainment of reliable parameter estimates without compromise.
- **Applicability to Nonsmooth Systems**: Under specific conditions, the SMO can be tailored to accommodate systems characterized by nonsmooth dynamics, thereby enhancing its versatility and applicability spectrum.
- **Robustness Under Parameter Variations**: With adherence to the dual of the well-known matching condition, the SMO exhibits robust performance even in the face of parameter variations, thereby ensuring operational reliability amidst system uncertainties.

4.2.1 Sliding Mode Observer Design

We consider a nonlinear time- and parameter-variant system described by the following equations:

$$X_1 = f_1(x, t, u, \theta) \tag{4.1}$$
$$X_2 = f_2(x, t, u, \theta) \tag{4.2}$$
$$\vdots \tag{4.3}$$
$$X_n = f_n(x, t, u, \theta) \tag{4.4}$$
$$y_1 = C_1 x \tag{4.5}$$
$$y_2 = C_2(\theta) x \tag{4.6}$$

Here, X_i represents the state variables, f_i are the nonlinear functions, t is time, u denotes the control input, θ is the parameter to be estimated, and y is the output.

The sliding mode observer is designed to estimate the state variables and the parameter θ, as depicted in the following schematic:

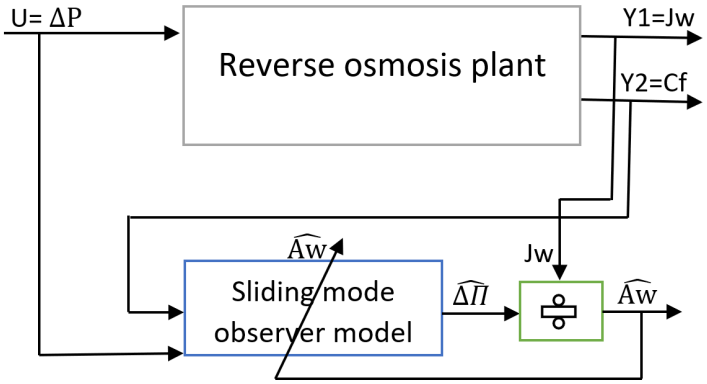


Figure 4.2: Sliding Mode Observer-Based Online Parameter Estimation

The observer equations are:

$$\hat{X}_1 = \hat{f}_1(\hat{x}, t, u, \hat{\theta}) + L_1 \operatorname{sign}(y_1 - \hat{y}_1) \tag{4.7}$$
$$\hat{X}_2 = \hat{f}_2(\hat{x}, t, u, \hat{\theta}) + L_2 \operatorname{sign}(y_1 - \hat{y}_1) \tag{4.8}$$
$$\vdots \tag{4.9}$$
$$\hat{X}_n = \hat{f}_n(\hat{x}, t, u, \hat{\theta}) + L_n \operatorname{sign}(y_1 - \hat{y}_1) \tag{4.10}$$

where $\hat{X}_i$ are the estimated state variables, $\hat{f}_i$ are the estimated nonlinear functions, $\hat{\theta}$ is the estimated parameter, and L_i are the observer gains.

The output equations of the observer are given by:

$$\hat{y}_1 = C_1 \hat{X} \tag{4.11}$$
$$\hat{y}_2 = \hat{\theta} C_2 \hat{X} \tag{4.12}$$

Sliding Surface and Error Dynamics

To avoid the accumulation of observation errors in the estimated parameter, the sliding surface is defined based on y_1, the output that is independent of θ: :

$$S = y_1 - \hat{y}_1 \tag{4.13}$$

Our objective is to drive S to zero, ensuring that the estimated output matches the actual output.

The parameter $\hat{\theta}$ is determined using the relationship:

$$\hat{\theta} = \frac{y_2}{C_2 \hat{X}} \tag{4.14}$$

The observation errors e_i are defined as:

$$e_1 = X_1 - \hat{X}_1 \tag{4.15}$$
$$e_2 = X_2 - \hat{X}_2 \tag{4.16}$$
$$\vdots \tag{4.17}$$
$$e_n = X_n - \hat{X}_n \tag{4.18}$$

The dynamics of the observation errors are described by:

$$\dot{e}_1 = \Delta f_1 - L_1 \operatorname{sign}(S) \tag{4.19}$$
$$\dot{e}_2 = \Delta f_2 - L_2 \operatorname{sign}(S) \tag{4.20}$$
$$\vdots \tag{4.21}$$
$$\dot{e}_n = \Delta f_n - L_n \operatorname{sign}(S) \tag{4.22}$$

where Δf_i represents the modeling uncertainties.

4.2.2 Stability Analysis Using Lyapunov's Second Theorem

To ensure the stability of the observer, we employ Lyapunov's second theorem. We define the Lyapunov function as:

$$V(e) = \frac{1}{2} e^T e \tag{4.23}$$

where e is the vector of observation errors.

The Lyapunov function is positive definite (i.e., $V(e) > 0$ for all $x \neq 0$ and $V(0) = 0$ for $x = 0$), indicating that it is always positive except at the origin where it is zero.

The time derivative of V is defined by:

$$\dot{V} = \sum_{k=1}^{n} e_k \dot{e}_k = \sum_{k=1}^{n} e_k (\Delta f_k - L_k \operatorname{sign}(S)) \tag{4.24}$$

For the observer to be finite-time stable, the following conditions must be satisfied:

$$L_1 > \Delta f_1, \quad L_2 > \Delta f_2, \quad \ldots, \quad L_n > \Delta f_n \tag{4.25}$$

By ensuring that the observer gains L_i are greater than the respective modeling uncertainties Δf_i , The time derivative of V is negative definite, indicating that it is always negative except at the origin where it is zero.

The observer is guaranteed to converge in finite time, thereby providing accurate online estimation of the parameters' variation.

4.2.3 Tuning the Parameters of a Sliding Mode Observer

Understanding the System Dynamics

Before tuning the parameters, it's crucial to understand the dynamics of the system you're working with. This includes the state variables, the nonlinear functions, and any parameters involved. You should also identify the modeling uncertainties Δf_i in your system.

Initializing the Observer Gains

The observer gains L_i are crucial parameters in a sliding mode observer. They are used to adjust the estimated state variables $\hat{X}_i$ to drive the sliding surface $S = y_1 - \hat{y}_1$ to zero. In practice, we might start with high values for the observer gains.

Reducing the Observer Gains

Once the observer gains have been initialized, you can start the process of tuning. This typically involves gradually decreasing the observer gains from their initial high values. The goal is to reduce the gains to a point where the estimated output matches the actual output, while preventing excessive chattering or high-frequency oscillations in the estimated states.

Evaluating the Performance

After each adjustment of the observer gains, evaluate the performance of the observer. This could involve checking the error between the estimated and actual states, observing the behavior of the sliding surface S, and the behavior of the permeability coefficient estimation error.

The process of tuning the observer gains is typically iterative. After evaluating the performance, you may need to adjust the gains further and re-evaluate. This process continues until you achieve the desired performance.

Considerations for Stability

To ensure the stability of the observer, the observer gains L_i should be greater than the respective modeling uncertainties Δf_i. This ensures that the dynamics of the observation errors $\dot{e}_i$ are stable.

The data obtained using the Sliding Mode Observer will be used for training, validation, and prediction with machine learning models. The following section will explain the theoretical aspects of the machine learning methods employed for prediction: Long Short-Term Memory (LSTM) and Transformer models.

LSTM is utilized for One-Step-Ahead Forecasting. To ensure continuous forecasting, real-time data is crucial, as the LSTM model requires the current value to predict the next hour. Conversely, the Transformer model is used for Multi-Step-Ahead Forecasting.

4.3 Time Series Prediction

4.3.1 Introduction

Time series forecasting is a critical problem in numerous domains. With the rapid advancement of computational power and the growing popularity of machine learning in recent years, there has been an increasing application of machine learning models, particularly deep learning models, to time series forecasting. These models include recurrent neural networks (RNNs) as the long short-term memory (LSTM) networks. More complex neural network structures are being used, notably the sequence-to-sequence (seq2seq) models, such as the transformer model. Seq2seq models are designed to convert sequences from one domain into sequences in another domain, thereby providing a more robust framework for handling intricate sequence data.

We will explore the application of the aforementioned methods to the problem of predicting fouling. We will evaluate and compare the performance of all these methods to determine the most effective approach for multi-step time series forecasting in this context.

4.3.2 Long Short Term Memory (LSTM)

Introduction

Long Short Term Memory networks (LSTMs) are a specialized type of RNN, created to overcome the issue of long-term dependencies. They were introduced by Sepp Hochreiter and Jurgen Schmidhuber in their 1997 paper titled "LONG SHORT-TERM MEMORY" [41]. The LSTM unit incorporates three gates: an input gate, a forget gate, and an output gate. Figure 4.3 illustrates the LSTM cell.

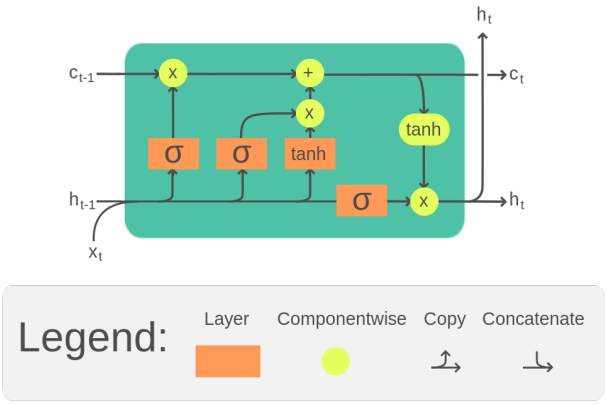


Figure 4.3: LSTM Cell by Guillaume Chevalier, CC BY-SA 4.0

LSTM Architecture

The state updates in LSTMs follow these operations: The first step in the LSTM is to determine what information should be discarded from the cell state. This decision is made by a sigmoid layer called the Forget Gate Layer f_t. It examines the previous hidden state of the LSTM cell h_{t-1} and the input x_t at the current time step t and outputs a number between 0 and 1 for each component in the cell state C_{t-1} [42, 51].

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{4.26}$$

where:

- b_f: represents the bias term associated with the forget gate.

Next, the LSTM decides what new information to store in the cell state. This consists of two parts: first, a sigmoid layer called the input gate layer i_t determines which values will be updated. Then, a tanh layer creates a vector C'_t of new candidate values that can be added to the state. These two parts are then combined to update the state [42, 51].

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{4.27}$$
$$C'_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \tag{4.28}$$

The old cell state C_{t-1} is then updated to the new cell state C_t. The old state is multiplied by f_t, forgetting the previous information as decided. Then, $i_t C'_t$ is added to the state [42, 51].

$$C_t = f_t C_{t-1} + i_t C'_t \tag{4.29}$$

The next step is to determine the output. This output will be taken from our cell state but in a more refined manner. Initially, it is passed through a sigmoid layer, which decides which

parts of the cell state to output. Then, a tanh transformation is applied to scale the cell state to the range of -1 to 1, and it is multiplied by the result of the sigmoid gate. This ensures that the output is based only on the selected components [42, 51].

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{4.30}$$
$$h_t = o_t \tanh(C_t) \tag{4.31}$$

4.3.3 Multi-step time series forcasting using transformer

Introduction

Transformers are a class of machine learning models that use self-attention or the scaled dot-product operation as their primary learning mechanism. They were initially proposed for neural machine translation, one of the most challenging natural language processing (NLP) tasks. Recently, Transformers have been successfully employed to tackle various problems in machine learning and achieve state-of-the-art performance.

Apart from classical NLP tasks, Transformers have been used in areas like image classification, object detection and segmentation, image and language generation, sequential decision-making in reinforcement learning, multi-modal (text, speech, and image) data processing, and analysis of tabular and time-series forecasting [52]. This section focuses on time series forecasting using Transformers.

Transformer Architecture

The Transformer architecture discards the need for RNNs by employing attention mechanisms, which facilitate the alignment of input sequences with output sequences of varying lengths. Like sequence-to-sequence (seq2seq) architectures, The Transformer follows an encoder-decoder structure architecture. In brief, the model uses an encoder to process the input data to generate a sequence of information and a decoder to decode the mapped sequence and output the desired result [53]. In both the encoder and decoder, using stacked self-attention and point-wise, fully connected layers, which will be shown in the left and right halves of the figure 4.4. The expression of the attention function used is as follows:

$$\mathit{Attention}(Q, K, V) = \text{softmax}(f(Q, K)) \cdot V \tag{4.32}$$

where Q, K, and V represent the query, key, and value, respectively. This analogy, introduced by Vaswani et al.[54], mirrors information retrieval systems, where a query searches for the most similar key to retrieve its corresponding value. Various differentiable similarity functions (f) can be utilized, with the Scaled Dot Product being a prominent choice:

$$f(Q, K) = \frac{Q \cdot K^T}{\sqrt{d_K}} \tag{4.33}$$

In this equation, d_K denotes the length of the key vector K. This scaled version is effective for comparing sequences of different lengths, as demonstrated in the initial Transformer publication.

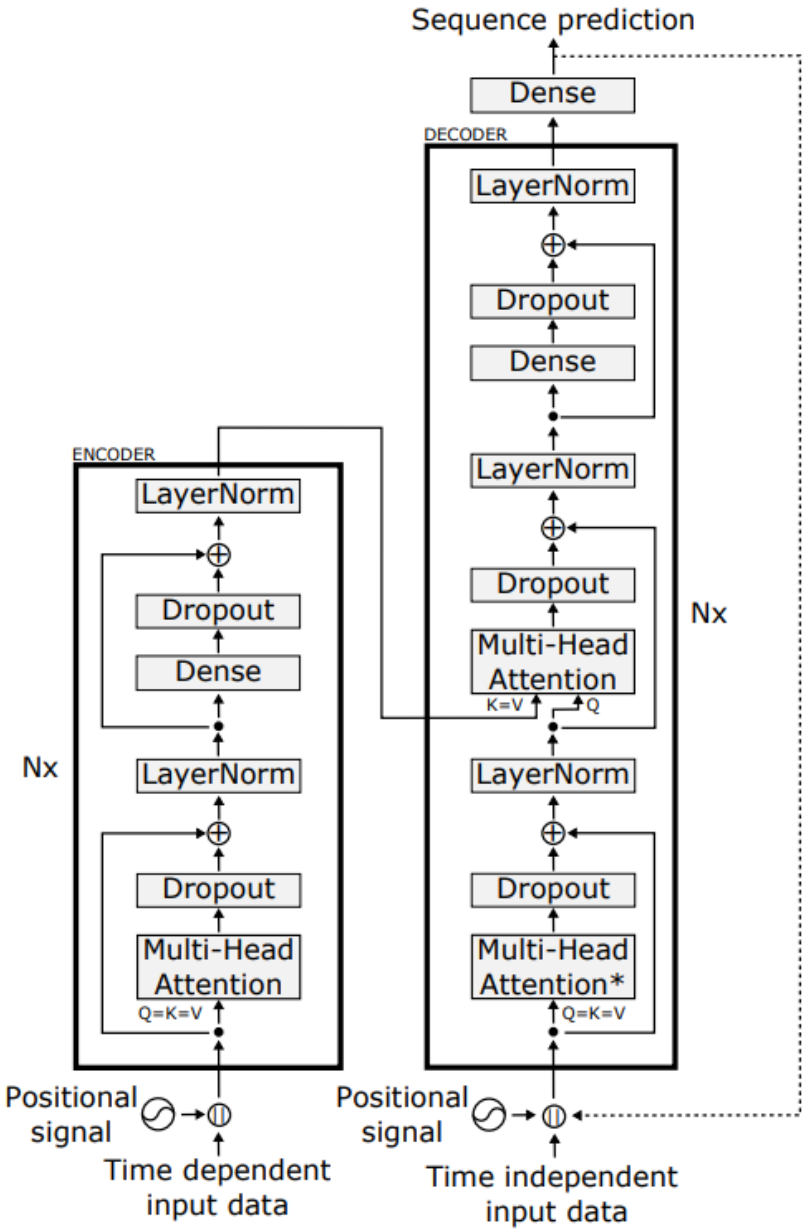


Figure 4.4: Transformer architecture used to perform time series forecasting. In the diagram, the || symbol stands for the concatenation operation, and similarly as in Vaswani et al. (2017), NX represents the number of repetitions of the encoder and decoder blocks.

There are at least three places in the Transformer that use the multi-head attention: in "encoder-decoder attention" layers; in the self-attentive layer in the encoder; in the decoder to prevent leftward information flowGoing a little deeper, a fully connected feedforward network in the Transformer is used in each layer of the encoder and decoder. This network will be the same at every position. The activation function here is chosen to be ReLU.

Attention Mechanisms

The Transformer employs three types of attention mechanisms, each serving a distinct purpose within the architecture:

0. **Encoder-Encoder Attention:** This form of self-attention is used in the encoder module,

where the query, key, and value originate from the same time series.

0. **Decoder-Decoder Masked Attention:** Another form of self-attention, this mechanism is constrained to be causal, utilizing only past time steps and masking future ones. Here, the query, key, and value also come from the same time series.

0. **Encoder-Decoder Attention:** Unlike self-attention, this mechanism compares the decoder's information with that of the encoder. The query is derived from the decoder, while the key and value are taken from the encoder output.

Positional Encoding in Transformer Models

Positional encoding is a crucial technique employed in Transformer models to incorporate an understanding of sequence order. Unlike conventional recurrent neural networks (RNNs) or convolutional neural networks (CNNs), Transformers lack inherent awareness of token order. Positional encodings provide the model with information about the position of tokens in the sequence, ensuring that the model can differentiate between tokens based on their position [55, 56].

The positional encoding for a given position *pos* in the sequence and a dimension i is computed as follows:

- For even indices: $PE_{(pos,2i)} = sin\left(\frac{pos}{10000^{2i/d_{model}}}\right)$
- For odd indices: $PE_{(pos,2i+1)} = cos\left(\frac{pos}{10000^{2i/d_{model}}}\right)$

These formulas help the model to capture the order of data points in the time series. The output of the positional encoding layer is a matrix, where each row of the matrix represents an encoded object of the sequence summed with its positional information [55, 56].

Conclusion

In this chapter, we tackled the prediction of membrane fouling in reverse osmosis (RO) plants by addressing two key areas: online parameter estimation and multi-step ahead prediction. We proposed the use of a Sliding Mode Observer (SMO) for real-time estimation.

For forecasting, we explored Long Short Term Memory (LSTM) networks and Transformer models. Comparing these models helps identify the best approach for predicting membrane fouling progression.

By combining SMO for real-time estimation with advanced deep learning models for forecasting, we offer a comprehensive solution for managing membrane fouling in RO systems.

Chapter 5

Data Acquisition and Preprocessing

5.1 Introduction

In this chapter, we focus on the preparation and transformation of the dataset needed for predictive modeling of membrane fouling.

We discuss methods for data generation, collection, and preprocessing to ensure the data accurately represents real-world conditions. By incorporating random fluctuations, we create a robust dataset for training and validating predictive models.

5.2 Acquiring Fouling Parameters through Sliding Mode Observer: A Data Acquisition Approach

The efficiency of our system relies heavily on the real-time monitoring of fouling and the precise acquisition of relevant data. Accurate data collection is crucial for enhancing the reliability and consistency of predictive outcomes. This section details the results obtained using the Sliding Mode Observer (SMO), evaluating its suitability and effectiveness for parameter estimation by examining its performance in terms of precision, robustness, and reliability.

5.2.1 Results and Discussion

Actual Permeability Coefficient

Figure 5.1 illustrates the actual permeability coefficient of the reverse osmosis membrane. This parameter is essential for understanding the membrane's performance and detecting fouling.

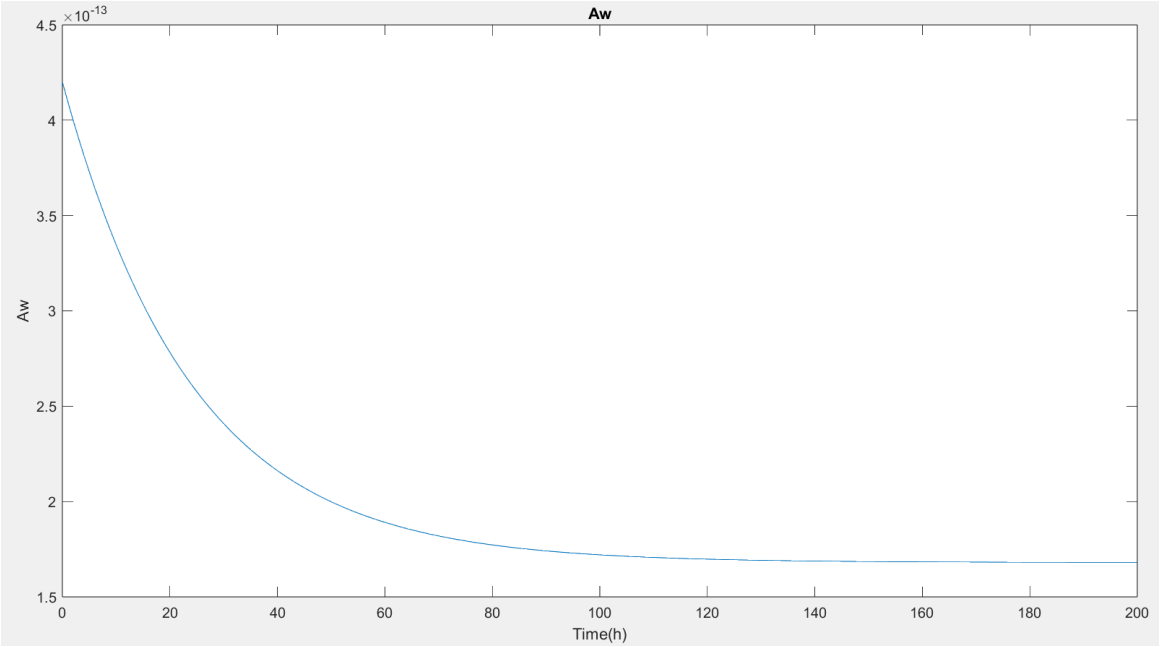


Figure 5.1: Actual Permeability Coefficient in Reverse Osmosis Membrane

Estimated Permeability Coefficient

Figure 5.2 shows the permeability coefficient estimated using the Sliding Mode Observer. The estimated values closely match the actual values, demonstrating the observer's accuracy.

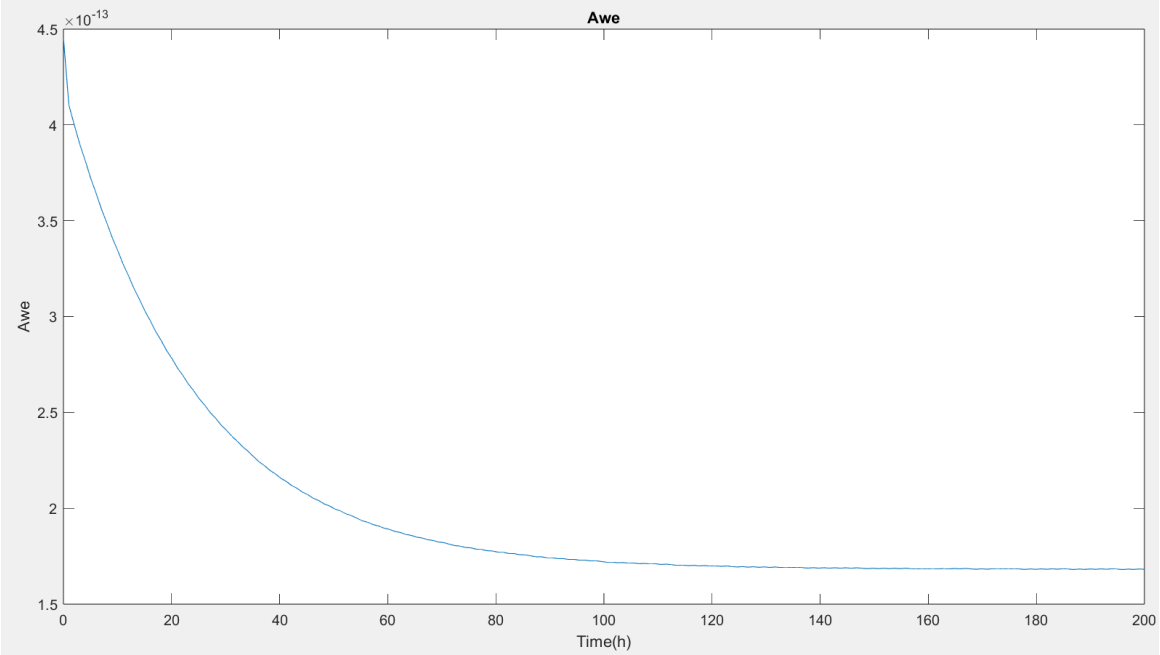


Figure 5.2: Estimated Permeability Coefficient using Sliding Mode Observer

Estimation Error Analysis

Figure 5.3 presents the estimation error of the permeability coefficient using the SMO. The maximum estimation error is $\pm 5 \times 10^{-16}$, indicating high precision.

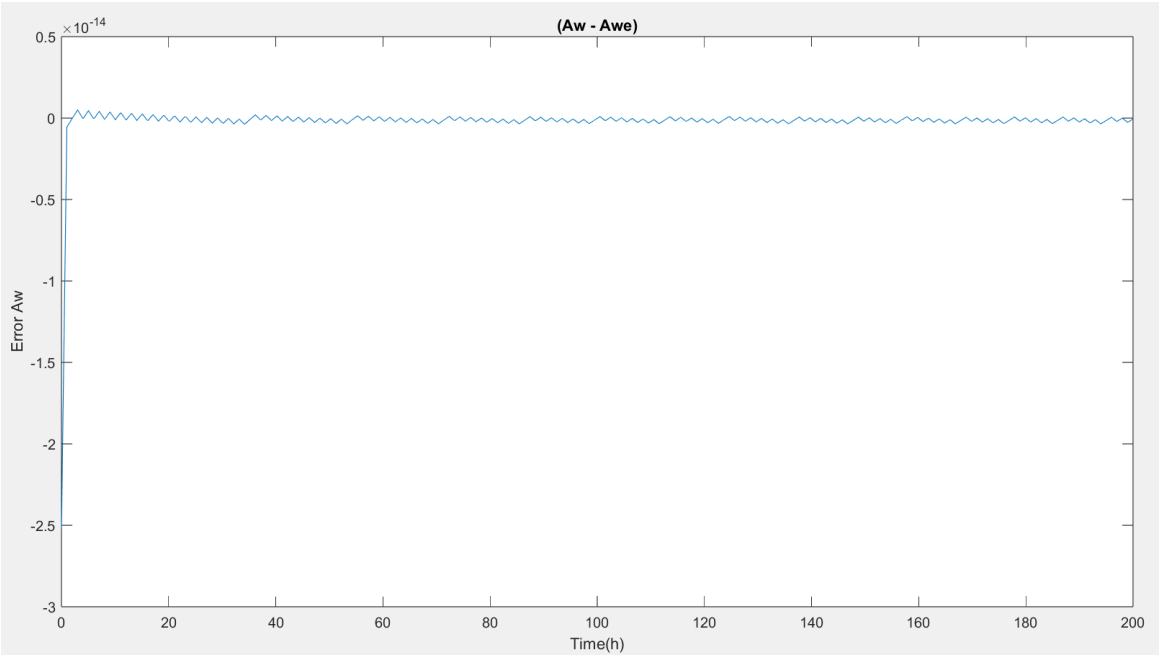


Figure 5.3: Estimation Error of the Permeability Coefficient using Sliding Mode Observer

To evaluate this error, we calculate the maximum possible relative error using equation 5.1:

$$\text{Maximum Relative Error} = \frac{\text{Absolute Maximum Error}}{\text{Minimum True Value}} = \frac{\max |V_{\text{estimated}} - V_{\text{true}}|}{\min(V_{\text{true}})} \tag{5.1}$$

Our calculations reveal that the Maximum Relative Error is:

$$\text{MRE} = \frac{4.9369 \times 10^{-16}}{1.683 \times 10^{-13}} = 2.9292 \times 10^{-3} \tag{5.2}$$

This result indicates that the parameter estimation using the SMO is precise and reliable for predictive purposes.

Actual Feed Concentration

Figure 5.4 illustrates the actual Feed Concentration of the reverse osmosis plant.

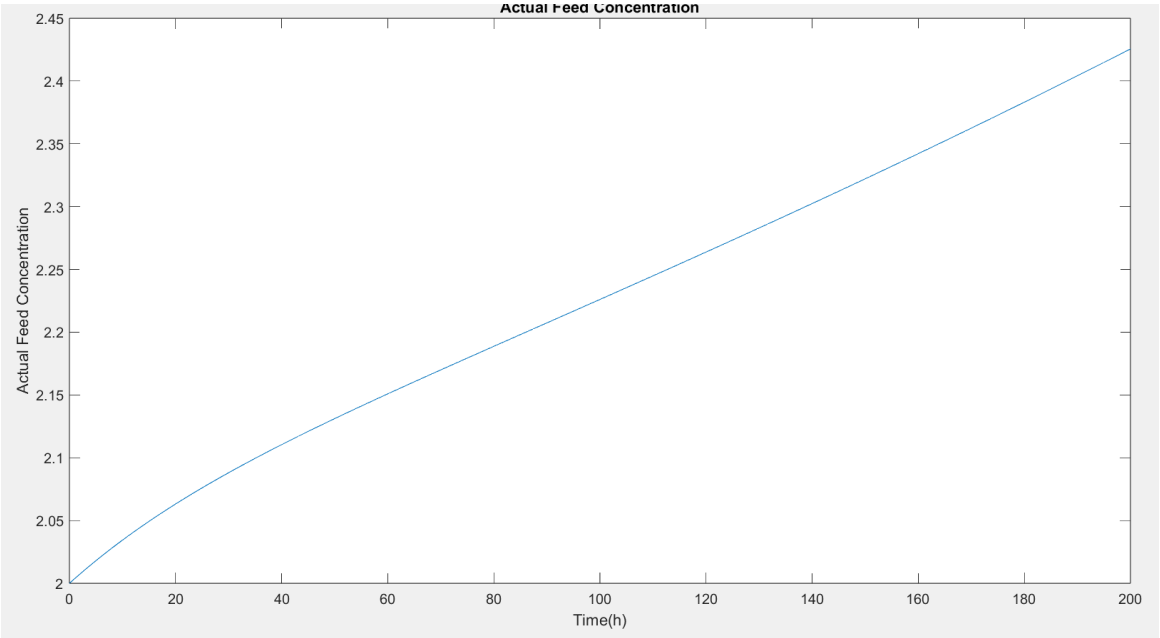


Figure 5.4: Actual Feed Concentration in Reverse Osmosis Plant

Observed Feed Concentration

Figure 5.5 shows the Feed Concentration observed using the Sliding Mode Observer.

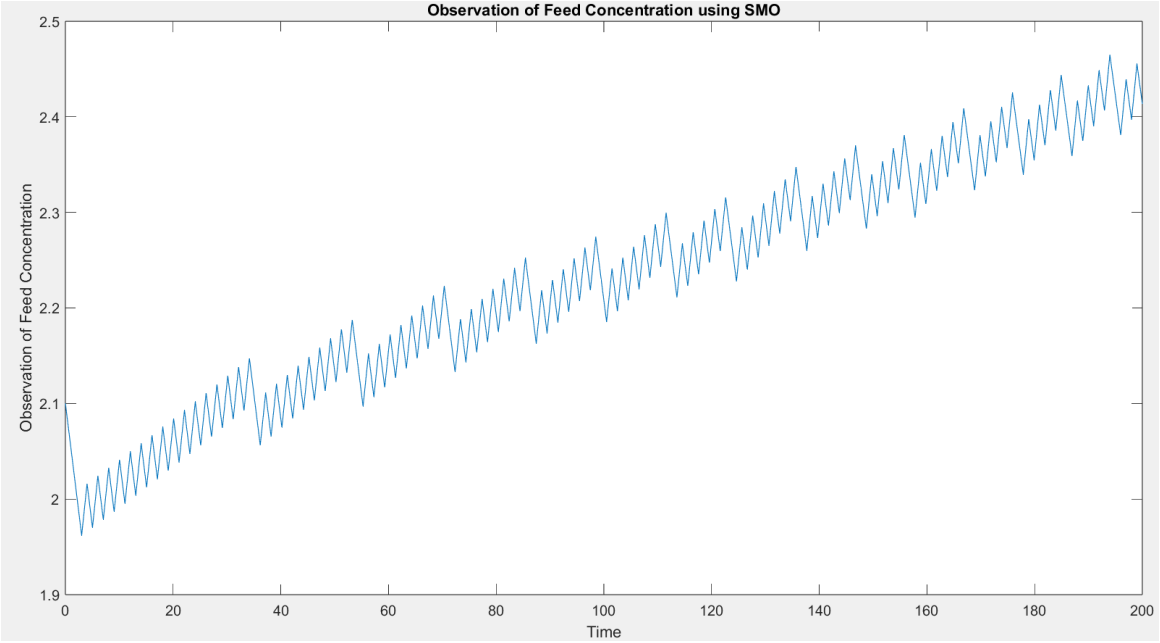


Figure 5.5: Observed Feed Concentration using Sliding Mode Observer

Feed Concentration Error Analysis

Figure 5.6 shows the error between the actual and observed feed concentration, which is a critical output of our system. The observed output error ranges between ± 0.05, with the maximum relative error calculated as:

$$\text{MRE} = \frac{5 \times 10^{-2}}{2} = 2.5 \times 10^{-2} \tag{5.3}$$

which is within acceptable limits for practical applications.

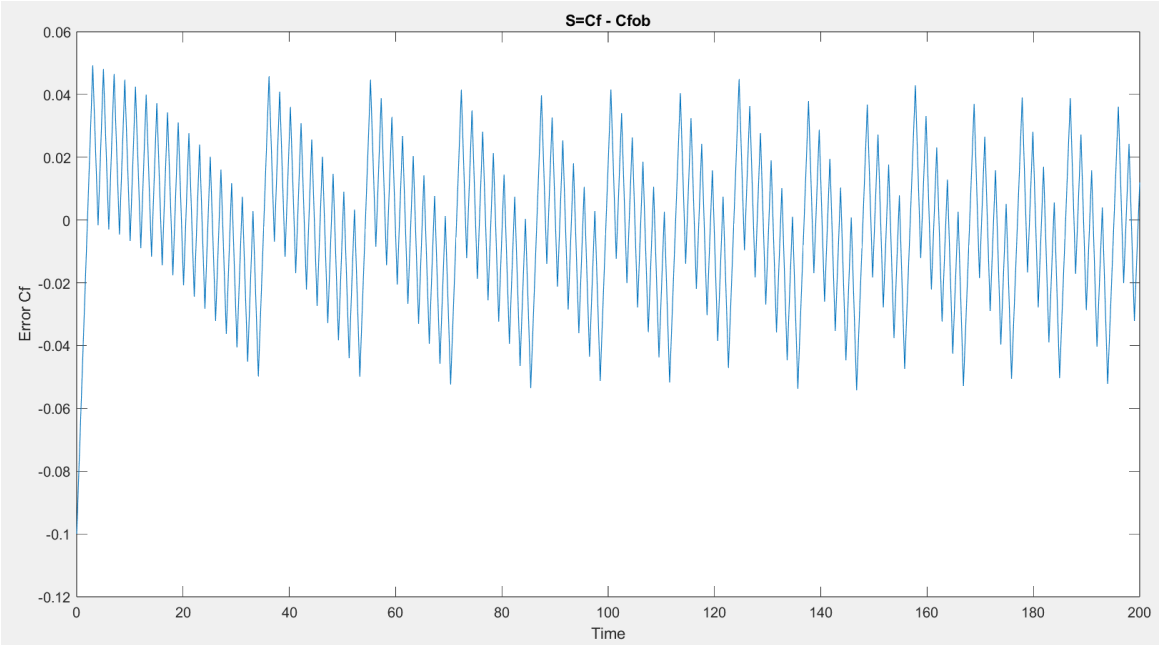


Figure 5.6: Error between Actual and Observed Feed Concentration

5.2.2 Generalization for Discretized Membrane Model

To accurately estimate fouling within each of the N cells of the membrane, the method requires the deployment of N sensors instead of a single sensor for output measurement. The output variables, represented as $Cf_1(t), Cf_2(t), \ldots, Cf_N(t)$, are used in the analysis. The sliding surface for the sliding mode observer is defined as $S(x,t) = Cf(x,t) - Cf_{predicted}(x,t)$. By employing this methodology consistently across each cell, we achieve a localized and detailed estimation of fouling.

This approach ensures a more granular and precise estimation of fouling, significantly enhancing the overall accuracy and reliability of the membrane model. The use of multiple sensors allows for real-time monitoring of spatial variations in fouling.

5.2.3 Data Collection and Transformation

To emulate the dynamic behavior of the permeability coefficient (Aw) in a RO membrane, we generated data that incorporates both the base permeability and deterministic fluctuations. It reflects how permeability changes over time due to factors such as fouling and operational conditions.

The Random Fluctuations were incorporated in order to represent random noise and variability found in real experimental data, challenging the robustness of predictive models under more stochastic conditions.

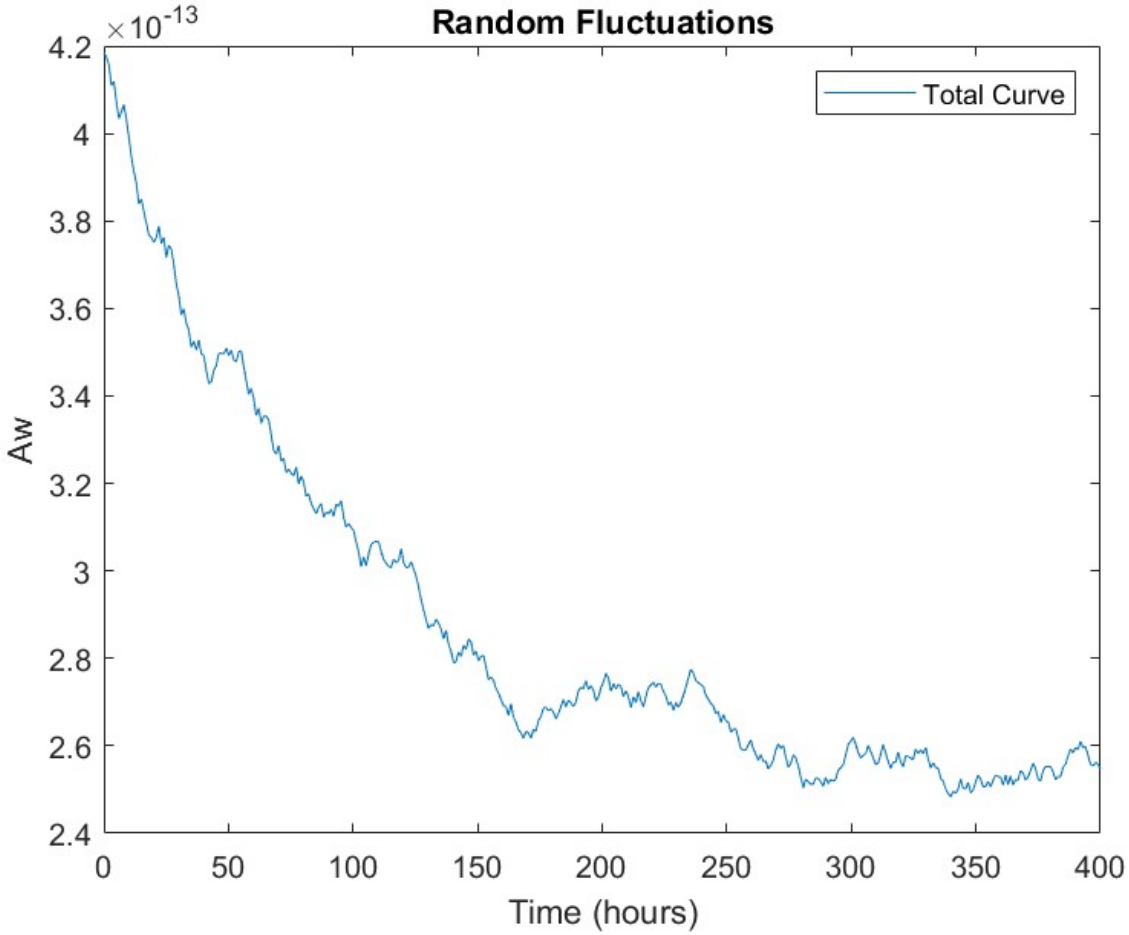


Figure 5.7: Aw vs. time with Random Fluctuations

Data Transformation and Export

The simulated data was then exported to an Excel file, which was subsequently converted into a CSV format. This transformation ensures that the data is in a convenient format for further analysis and predictive modeling. The CSV file contains time-series data of the permeability coefficient. A sample of this univariate time series is shown in Table 5.1.

Table 5.1: Sample of Permeability Coefficient Data

Hour	**Aw (h/m)**
0.000000	4.200000e-13
1.002506	4.197739e-13
2.005013	4.195020e-13
3.007519	4.191393e-13
4.010025	4.186460e-13

5.2.4 Data Analysis and Testing

To ensure the robustness and reliability of our predictive models, we performed a series of tests and analyses on the permeability coefficient data. These tests help us understand the underlying properties of the data and guide the selection of appropriate modeling techniques.

Stationarity Test

Stationarity is a crucial assumption for many time series forecasting methods. We conducted the Augmented Dickey-Fuller (ADF) test to check for the presence of a unit root in the time series.

Table 5.2: ADF Test Results for Random Fluctuations Dataset

Metric	**Random Fluctuations**
ADF Statistic	-3.802972620274889
p-value	0.00287704721079618
Critical Value (1%)	-3.447585
Critical Value (5%)	-2.869136
Critical Value (10%)	-2.570816

The ADF statistic is lower (more negative) than the critical values at the 1%, 5%, and 10% significance levels. Additionally, the p-value is significantly less than 0.05. Therefore, we reject the null hypothesis and conclude that the time series for the random fluctuations dataset is stationary.

Autocorrelation Analysis

To analyze the dependencies in the time series data, we examined the ACF. This analysis helps identify significant lags in the data, which are crucial for building autoregressive models.

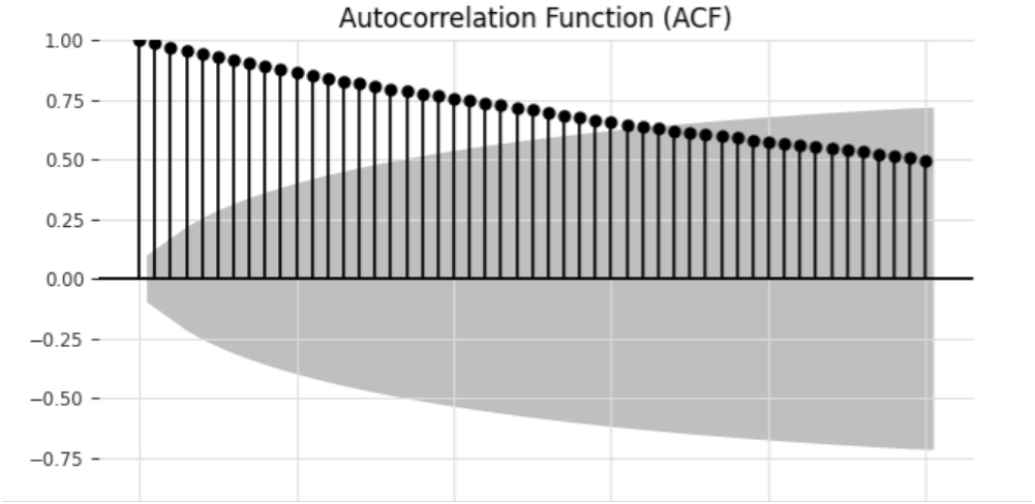


Figure 5.8: Autocorrelation Function for Random Fluctuations Dataset

The ACF plot for the random fluctuations dataset shows high initial autocorrelation, with values gradually decreasing as the lag increases. Significant autocorrelation at higher lags indicates strong temporal dependencies.

5.2.5 Data Splitting and Preparation

The dataset was strategically divided into three distinct subsets: training, validation, and test sets as detailed above :

Training Set : It comprises the portion of data used to train the model. It allows the model to learn and understand patterns within the data.

Validation Set : It includes the remaining part of data which is used to evaluate the model's performance on unseen data, thereby assisting in the fine-tuning of the model parameters for optimal performance.

Test Set : After the model has been trained and validated, the test set is used. This set is used to assess the model's predictive performance on new data that it has never seen before. It provides the most realistic estimate of how the model will perform when making predictions in real-world applications.

To ensure a fair comparison among different prediction models, the same dataset was used for training and evaluating each model. This approach guarantees that any observed differences in performance are attributable to the models themselves rather than variations in the data.

5.3 Conclusion

This chapter detailed the preparation and transformation of the dataset for predicting membrane fouling in reverse osmosis systems. We emphasized the importance of accurate data collection and preprocessing, simulating realistic conditions with random fluctuations.

The findings presented underscore the effectiveness of the Sliding Mode Observer (SMO) in providing high-resolution parameter estimation within complex models that exhibit parameter variations.

The resulting dataset will be foundational for the predictive techniques explored in subsequent chapters.

Chapter 6

Application of Predictive Methods for Membrane Fouling in Reverse Osmosis Plant

6.1 Introduction

In this chapter, we explore various prediction methods, focusing on Long Short-Term Memory (LSTM) networks, and Transformer model. We present the simulation results for each model, evaluating their performance based on specific metrics.

6.2 Evaluation Criteria

To ensure a comprehensive assessment of the prediction models, we will evaluate their performance using the following metrics:

- **Mean Absolute Error**: Measures the average magnitude of the errors in a set of predictions, without considering their direction.
- **Mean Squared Error**: Measures the average of the squares of the errors, providing a sense of how large the errors are.
- **R-squared**: Indicates the proportion of the variance in the dependent variable that is predictable from the independent variables.
- **Residual Analysis**: Examines the differences between the actual and predicted values to ensure the model has captured the underlying patterns in the data.
- **Autocorrelation of Residuals**: Analyzes the residuals to ensure there is no significant autocorrelation, indicating a well-fitted model.

6.3 Long Short-Term Memory (LSTM) Networks

6.3.1 Simulation Setup

The LSTM network was trained and validated using the preprocessed dataset of permeability coefficient data with random fluctuations. The dataset was divided into training and testing sets, with 80% of the data used for training and 20% for testing, the dataset consists of 400 sequences of time series data. These sequences are split into a training set of 320 sequences and a validation set of 80 sequences. The model's performance was evaluated based on several metrics, including Mean Absolute Error, Mean Squared Error, and R-squared score.

6.3.2 LSTM Model Implementation

Data Structure and Preprocessing

The model is designed to process time series data with a single feature. The input shape is '(32, 400, 1)', where '32' is the batch size and '400' is the sequence length. The sequence length represents the number of time steps in our single input sequence, which consists of 400 consecutive data points. The time series data is normalized using Min-Max scaling. This preprocessing step is crucial for improving the training efficiency and performance of the LSTM model.

Network Architecture

The input layer receives time series data organized into sequences with a specified window size of 3 time steps, where each sequence is a feature vector of past observations. This data is processed by one or more LSTM layers, which can be stacked to increase the model's capacity for learning complex patterns. Each LSTM layer consists of a certain number of units (LSTM cells), with the option to return either the full sequence of outputs or only the output for the last time step. In a typical architecture, an initial LSTM layer might return the full sequence to facilitate stacking another LSTM layer, while subsequent layers may only return the last output. Following the LSTM layers, a Dense (fully connected) layer is added to map the LSTM outputs to the desired prediction shape, usually a single value for time series forecasting. This final output layer provides the prediction for the next time step. The model predicts the next value in the sequence based on the input sequences, and the predictions are rescaled back to the original scale using the same scaler applied during training. This architecture is well-suited for capturing temporal dependencies in the data, enabling accurate one-step-ahead predictions.

Training Process

The model uses the Adam optimizer for training. Adam is an optimization algorithm that can be used instead of the classical stochastic gradient descent procedure to update network weights iteratively based on the training data. The learning rate for the Adam optimizer is set to 0.001. The loss function used is the Mean Squared Error (MSE).

6.3.3 Results

The LSTM model was trained on the training set and tested on the testing set. The following results were obtained:

Table 6.1: Performance Metrics for LSTM Model

Dataset	**MAE**	**MS**	R^2
Training	1.435e-2	3.000e-4	0.688
Testing	2.347e-2	8.917e-4	0.996

These results indicate that this model exhibits varying performance for the two different sets. Here is a context-specific interpretation:

- Both the training and testing datasets show extremely low MAE values, indicating that the average difference between the predicted and actual values is minimal for both sets.
- The MSE values for both datasets are extremely low, indicating that the model's predictions are very close to the actual values for both the training and testing sets. The testing MSE is higher than the training MSE, which further supports the abscence of overfitting.
- The R-squared value for the training data is approximately 0.69, indicating that around 69% of the variance in the permeability coefficient is explained by the model. For the testing data, the R-squared value is much higher at approximately 0.997, suggesting that the model explains almost all the variance in the testing dataset.

The following figures illustrate the comparison between the actual and predicted permeability coefficients for both the training and testing sets.

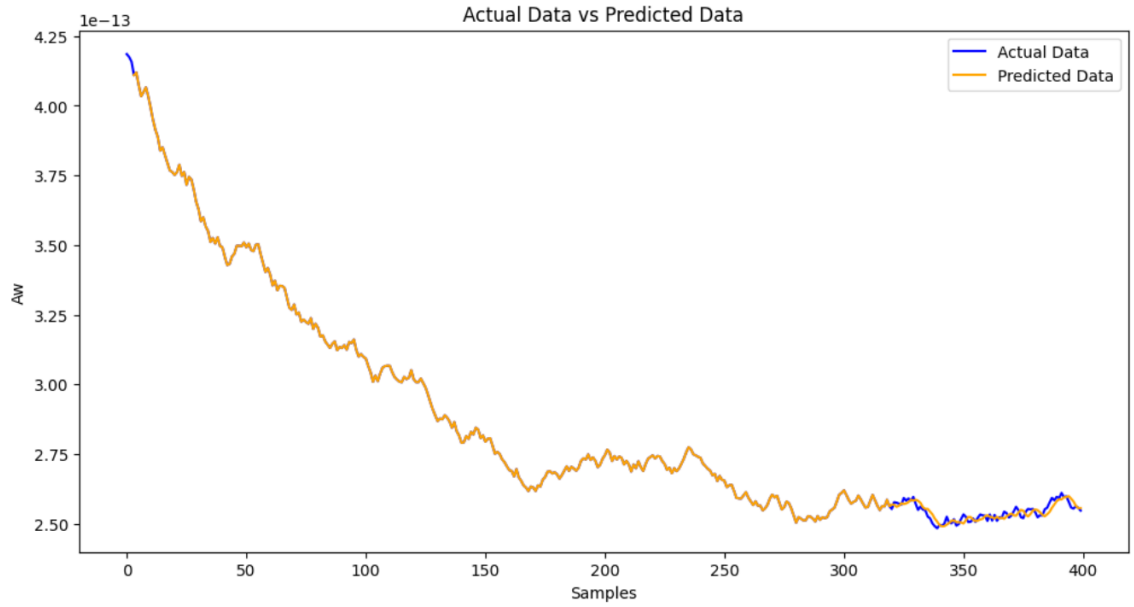


Figure 6.1: Actual vs Predicted Permeability Coefficient (Training Set)

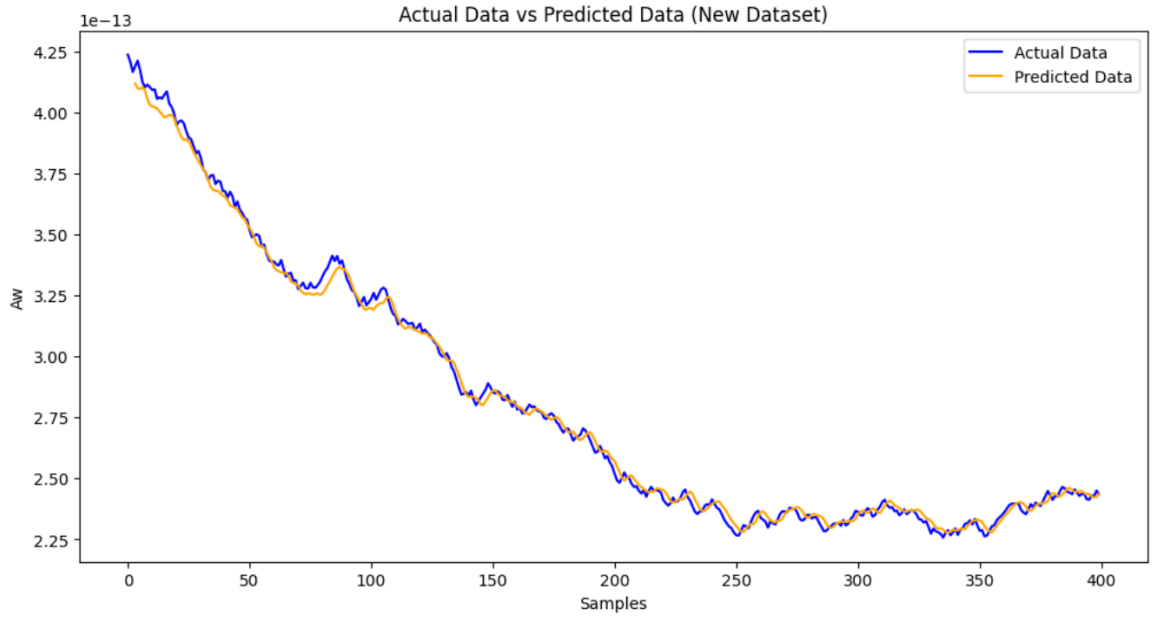


Figure 6.2: Actual vs Predicted Permeability Coefficient (Testing Set)

6.3.4 Residual Analysis

Analyzing the residuals (the differences between the actual and predicted values) is crucial for understanding the model's performance. The residuals for the testing set are shown in the following figure:

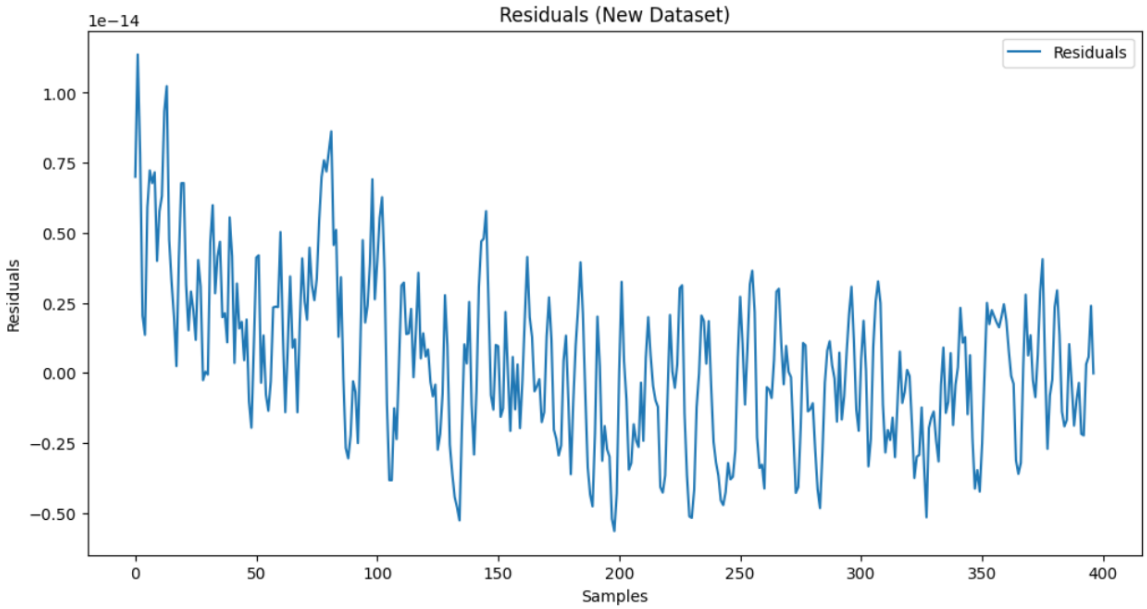


Figure 6.3: Residuals for Testing Set

The residual plot indicates that the errors are randomly distributed in the range of approximately zero, suggesting that the LSTM model has captured the underlying patterns in the data effectively.

6.3.5 Autocorrelation of Residuals

To further validate the model, the ACF of the residuals was analyzed, the plot is shown below:

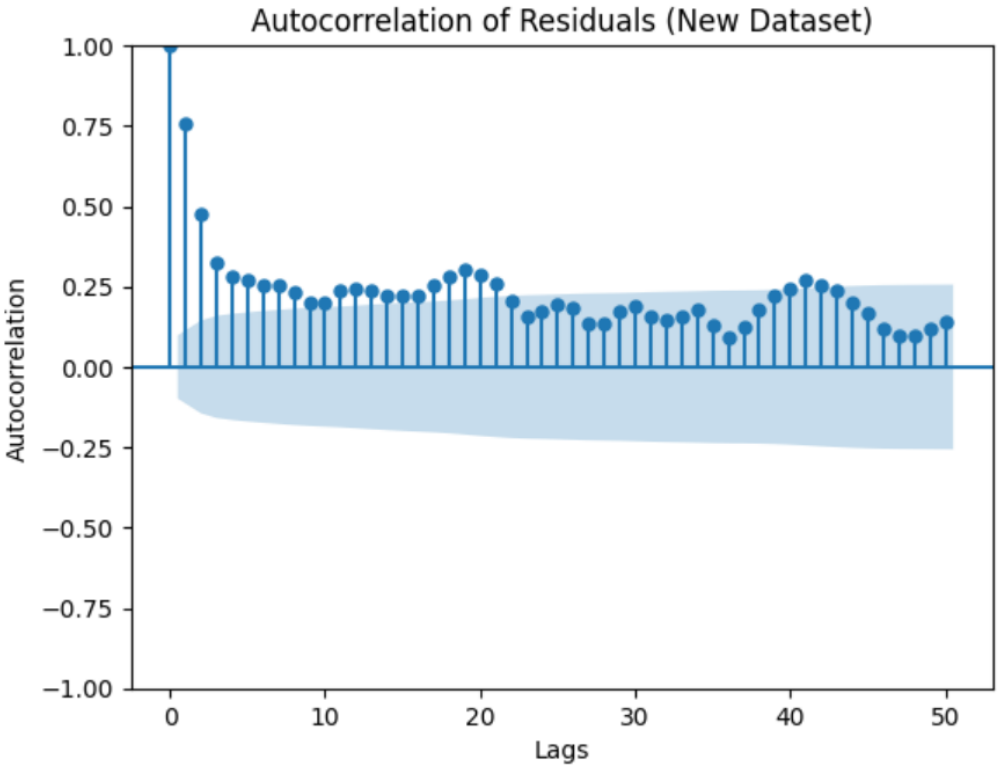


Figure 6.4: Autocorrelation Function (ACF) of Residuals

The ACF plot demonstrates that there is no strong autocorrelation in the residuals, indicating that the LSTM model has adequately modeled the time series data.

6.3.6 Conclusion

The LSTM network demonstrated strong performance in predicting the permeability coefficient of the reverse osmosis membrane. The model achieved high accuracy with low MAE and MSE values and a high R-squared score, indicating a good fit to the both training and testing datasets. The residual analysis and autocorrelation function further confirmed the model's effectiveness.

These results highlight the potential of LSTM networks for accurate and reliable prediction of membrane fouling in RO systems.

6.4 Fouling Prediction via Transformer Model

6.4.1 Transformer Model Implementation

Data Structure and Preprocessing

The model processes time series data with a single feature. The input shape is $(40, 250, 1)$, where 40 is the batch size, 250 is the sequence length, and 1 is the number of features. The sequence length represents the number of time steps in our input sequence. The time series data is normalized using Min-Max scaling, which is crucial for improving training efficiency and model performance. The data is divided into sequences using a sliding window approach, creating input-output pairs for the model to learn from.

Network Architecture

The core of the model is a Transformer architecture, effective for capturing long-range dependencies in sequential data. The Transformer model includes several key components:

- **Positional Encoding**:
  - The Transformer model does not inherently capture the order of the sequence. Therefore, positional encoding is added to the input embeddings to incorporate information about the position of each time step. This is done using sine and cosine functions of different frequencies.
- **Transformer Encoder**:
  - The model comprises a stack of Transformer encoder layers. Each encoder layer includes:
    - **Multi-Head Self-Attention Mechanism**: This mechanism allows the model to focus on different parts of the sequence simultaneously. It helps the model to learn intricate temporal patterns by attending to various positions in the sequence.

    * **Feed-Forward Neural Network**: Following the self-attention mechanism, a feed-forward neural network is applied to each position in the sequence independently. This consists of two linear transformations with a ReLU activation in between.
    * **Layer Normalization and Residual Connections**: These are used to stabilize and enhance the training process. Layer normalization is applied before the multi-head attention and feed-forward network, and residual connections are added to help gradients flow through the network.

- **Decoder Layers**:
  - The decoder is responsible for generating the output sequence. Similar to the encoder, the decoder layers include multi-head self-attention mechanisms and feed-forward neural networks. Additionally, the decoder incorporates an attention mechanism to attend to the outputs of the encoder, allowing it to generate contextually relevant predictions based on the encoded input sequence.
- **Output Layer**:
  - The final layer is a fully connected dense layer with a linear activation function, producing a single output as it is a regression problem. This layer maps the high-dimensional representation back to the original feature space.

Training Process

The model uses the AdamW optimizer with a learning rate of 0.005. AdamW is an optimization algorithm that modifies the typical Adam algorithm by including weight decay to improve generalization. The learning rate is adjusted using a StepLR scheduler, which decreases the learning rate by a factor of 0.98 every epoch to allow the model to converge smoothly. The loss function used is the Mean Squared Error (MSE), which measures the average squared difference between the predicted and actual values.

Dataset and Evaluation

The dataset consists of time series data split into training and validation sets. The sequences are divided as follows:

- **Training Set**: 80% of the data
- **Validation Set**: 20% of the data

Furthermore, the model's performance is evaluated by predicting future time steps on the test data set:

- **Prediction for 50 Steps Ahead**: This evaluates the model's ability to predict the next 50 time steps after the last observed data points.
- **Prediction for 200 Steps Ahead**: This evaluates the model's performance in predicting further into the future, 200 time steps ahead.

The training process involves multiple epochs, where the model's parameters are updated iteratively. During each epoch, the model is trained on the training set, and its performance is evaluated on the validation set. The model's effectiveness is measured using various metrics, including MAE, MSE, and R^2, which assess the accuracy and reliability of the predictions. The predictions are rescaled back to the original scale using the same scaler applied during training. The best model parameters are selected based on the lowest validation loss achieved during training.

6.4.2 Univariate Time Series Forecasting

In this analysis, we focus on univariate time series forecasting, focusing on both short-term and long-term predictive capabilities. Specifically, we will test the model's efficacy over a 50-step horizon to gauge its short-term forecasting accuracy, as well as extend our examination to a 200-step horizon to assess its long-term prediction performance.

Short-Term Forecasting (50-Step Horizon)

Analysis of Experimental Results The Transformer model was trained and tested for short-term forecasting, producing the following results:

Table 6.2: Performance Metrics for Transformer Model (50-Step Horizon)

Dataset	**MAE**	**MSE**	R^2
Testing	3.2110e-2	1.5406e-3	0.8653

The results suggest:

- Extremely low MAE value for testing dataset, indicating minimal average prediction error.
- Very low MSE values **for testing dataset**, indicating predictions are very close to actual values.
- The R^2 value for testing, this metric assesses how well the model explains the variance in the data. An R^2 value of 0.8653 means that the model captures about 86.53% of the variance in the testing dataset. Higher R2 values indicate better explanatory power.

The following figure depict the comparison between actual and predicted permeability coefficients for testing set.

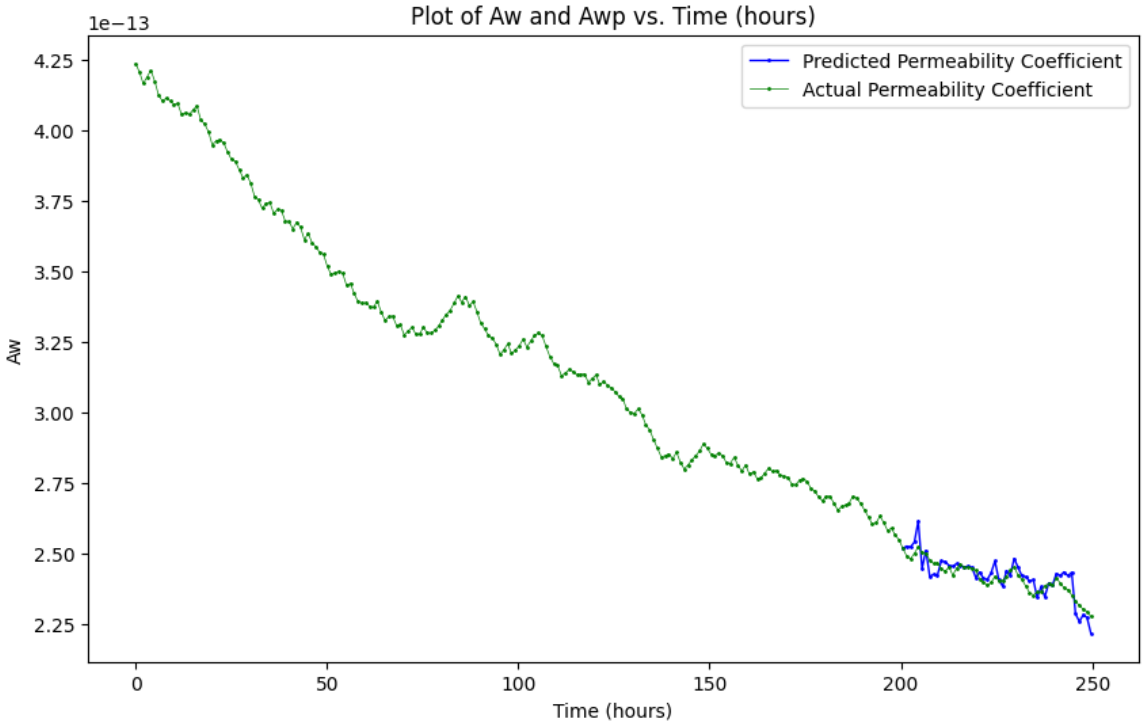


Figure 6.5: Actual vs Predicted Permeability Coefficient (Testing Set, 50-Step Horizon)

Residual Analysis for Short-Term Forecasting

The residuals for the testing set are shown below:

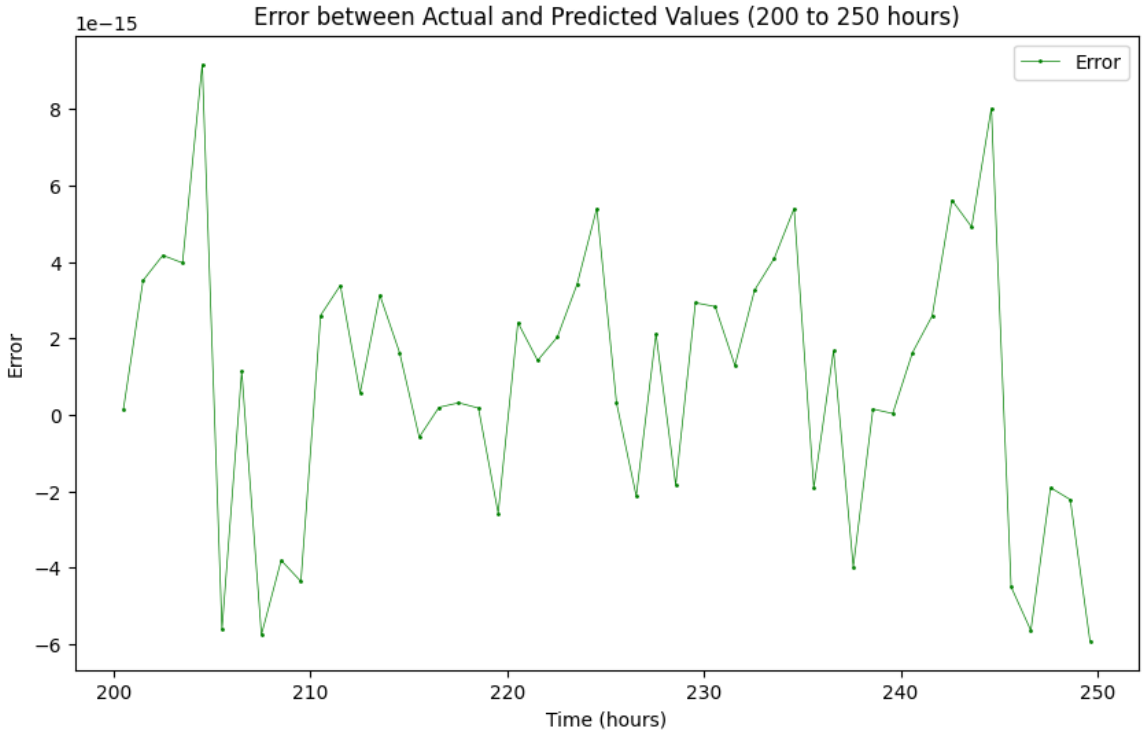


Figure 6.6: Residuals for Testing Set (50-Step Horizon)

The residual plot indicates that the errors are randomly distributed around zero, suggesting the Transformer model effectively captures the underlying data patterns in the short term.

Autocorrelation of Residuals for Short-Term Forecasting

The ACF of the residuals was also analyzed, as shown below:

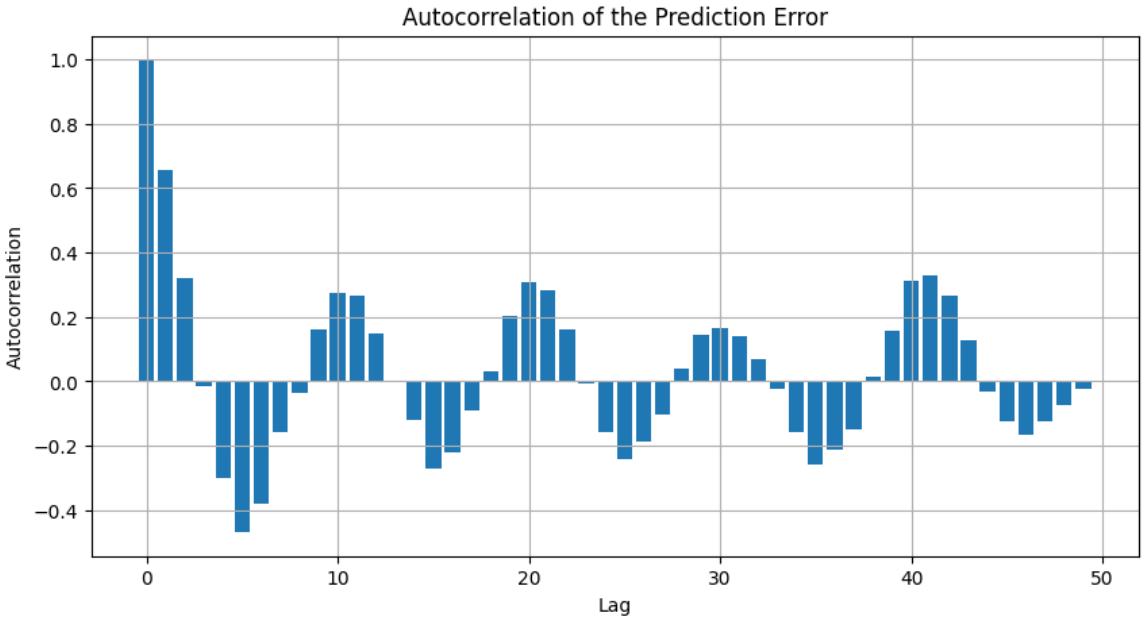


Figure 6.7: Autocorrelation Function (ACF) of Residuals (50-Step Horizon)

The ACF plot reveals that there is no significant autocorrelation in the residuals. As the lag (time gap between residuals) increases, the autocorrelation diminishes—a typical behavior when modeling intricate time series data. These findings provide strong evidence for the effectiveness of the Transformer model in capturing dependencies, indicating that it has successfully represented the underlying dynamics of the time series.

Long-Term Prediction (200-Step Horizon)

Analysis of Experimental Results The Transformer model was also trained and tested for long-term prediction, producing the following results:

Table 6.3: Performance Metrics for Transformer Model (50-Step Horizon)

Dataset	**MAE**	**MSE**	R^2
Testing	4.1361e-2	2.6712e-3	0.3028

The results suggest:

- Slightly higher MAE values compared to short-term forecasting, indicating more significant average prediction errors.
- Higher MSE values, suggesting the long-term predictions are less accurate but still close to actual values.
- The R^2 value for the testing data is approximately 0.3028, indicating that the model explains about 30.28% of the variance in the permeability coefficient. This suggests that

while the model captures some of the variance in the testing dataset, there is still a substantial portion of the variance that remains unexplained.

The following figure depict the comparison between actual and predicted permeability coefficient for testing set in the long term.

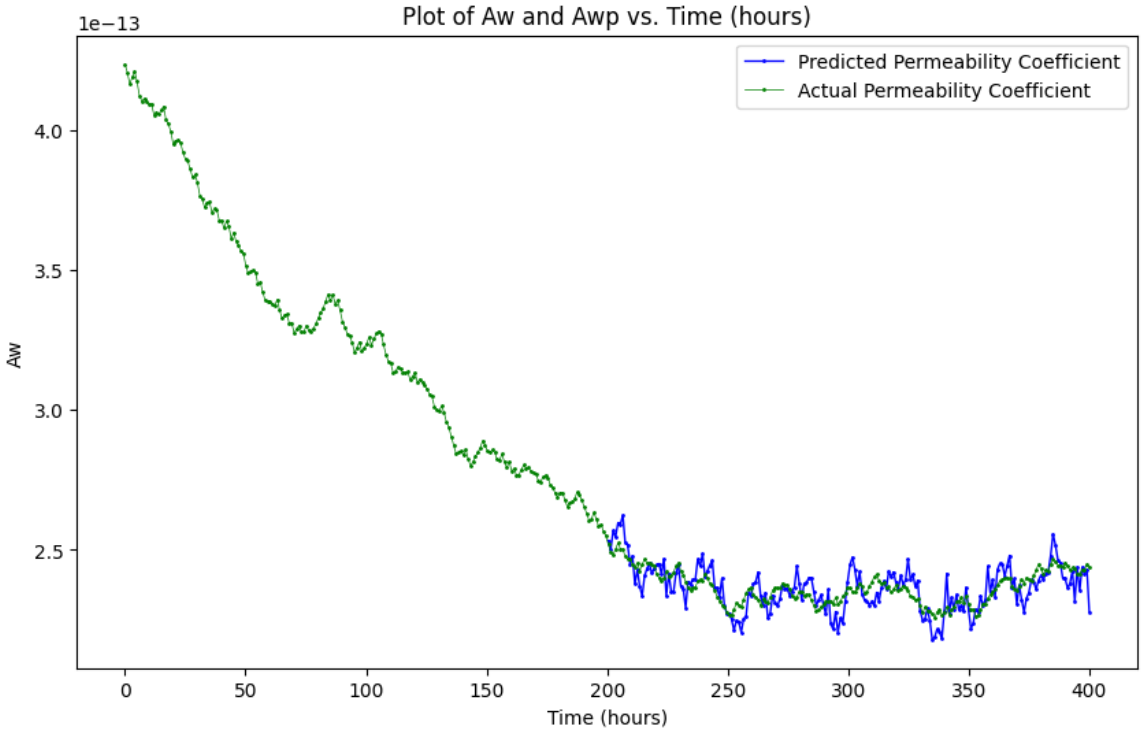


Figure 6.8: Actual vs Predicted Permeability Coefficient (Testing Set, 200-Step Horizon)

Residual Analysis for Long-Term Prediction

The residuals for the testing set are shown below:

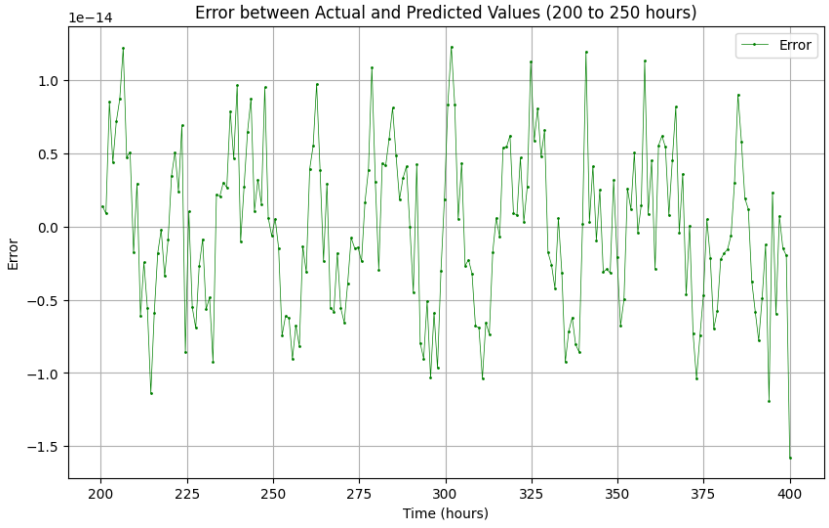


Figure 6.9: Residuals for Testing Set (200-Step Horizon)

The residual plot indicates that the errors are randomly distributed around zero, suggesting the Transformer model effectively captures the underlying data patterns in the long term.

Autocorrelation of Residuals for Long-Term Prediction

The ACF of the residuals was also analyzed, as shown below:

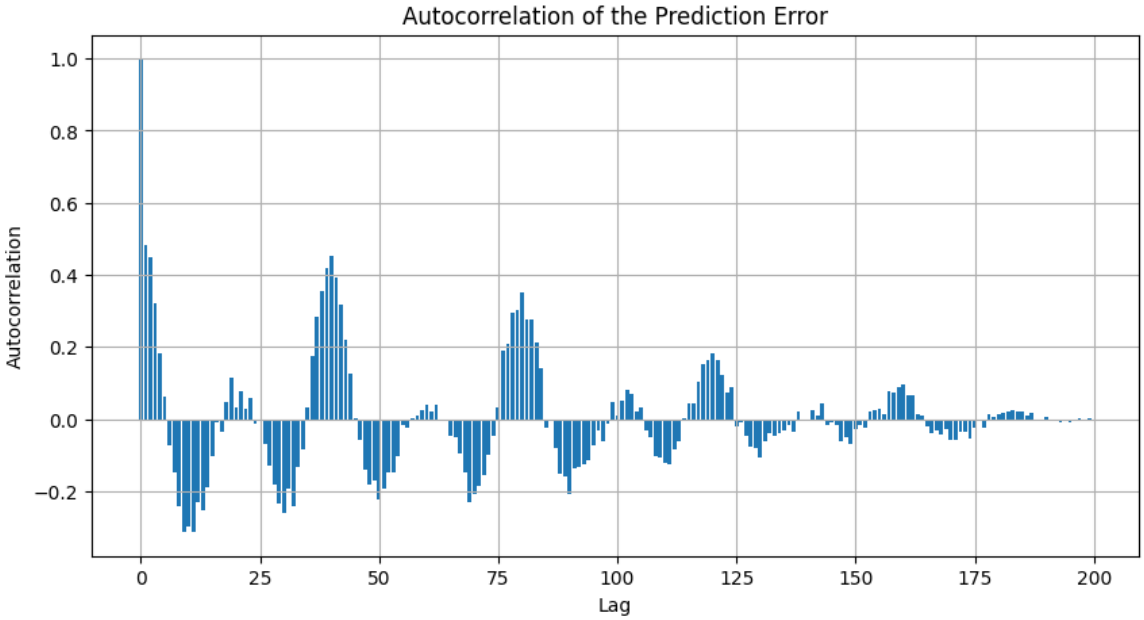


Figure 6.10: Autocorrelation Function of Residuals (200-Step Horizon)

The ACF plot demonstrates that there is no strong autocorrelation in the residuals. As you increase the lag (time gap between observations), the autocorrelation decreases. These insights validate the Transformer model's effectiveness in capturing both dependencies and modeling the time series data.

Simulation Analysis and Conclusion

The Transformer model demonstrates strong performance in both short-term and long-term forecasting. In the short term, the model achieves extremely low error metrics, reflecting its ability to accurately predict the permeability coefficient over a 50-step horizon. For long-term predictions, while the errors are slightly higher, the model still maintains a high degree of accuracy, effectively capturing the overall trends and patterns over a 200-step horizon. These results underscore the model's robustness and versatility in different forecasting scenarios.

6.4.3 Multivariate Time Series Forecasting

For multivariate prediction tasks, we analyzed the permeability coefficient's variation over time across multiple cells. The study focused on six cells with short-term prediction horizon (50-step).

The following figure shows the comparison between actual and predicted permeability coefficients testing set.

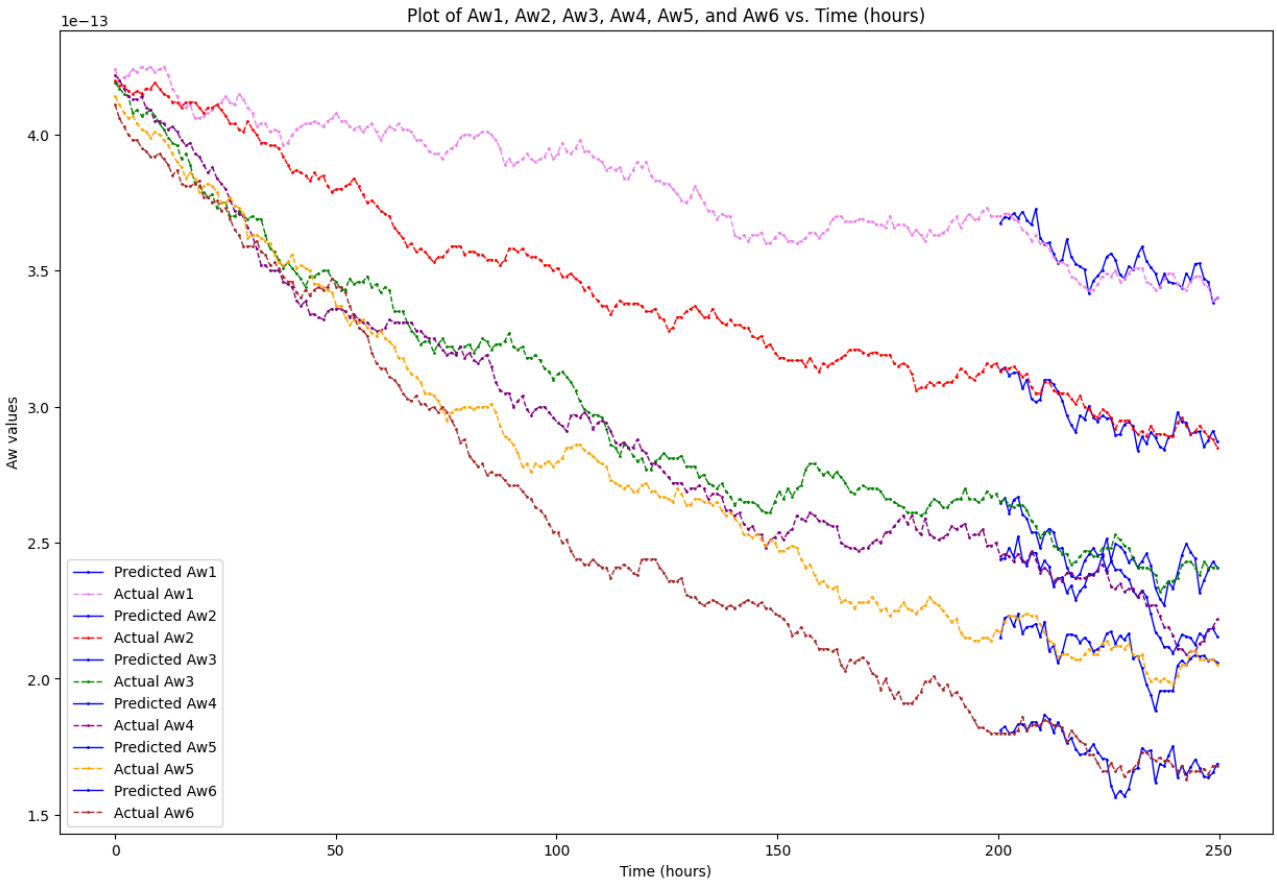


Figure 6.11: Actual vs Predicted Permeability Coefficient (Testing Set) for Multiple Cells (Short-Term)

Analysis of Experimental Results

The Transformer model was trained on the training dataset and evaluated on the testing dataset, yielding the following performance metrics across six cells:

Table 6.4: Performance Metrics for Transformer Model Across Six Cells

Metric	Cell 1	Cell 2	Cell 3	Cell 4	Cell 5	Cell 6
Testing MAE	3.4797e-2	2.7082e-2	3.0200e-2	3.4336e-2	3.2344e-2	2.8725e-2
Testing MSE	1.9464e-3	1.3112e-3	1.2797e-3	1.9273e-3	1.6576e-3	1.4289e-3
Testing R^2	0.7368	0.8162	0.8296	0.8504	0.6768	0.7015

From the results in Table 6.4, we can make the following observations:

1. **Mean Squared Error**:
    - The MSE values for the cells span an order of magnitude from approximately 1.28×10^{-29} to 1.95×10^{-29}.
    - Cell 3 exhibits the lowest MSE, signifying superior predictive accuracy.
    - Conversely, Cell 1 manifests the highest MSE.
2. **Mean Absolute Error**:
    - The MAE values, ranging from approximately 2.71×10^{-15} to 3.48×10^{-15}, provide further insights.
    - Cell 2 consistently outperforms others, boasting the lowest MAE and indicating more accurate predictions.
    - In contrast, Cell 1 displays the highest MAE.
3. R^2 **Score**:
    - The R^2 values, spanning from 0.68 to 0.85, elucidate the model's explanatory power.
    - Cell 4 achieves the highest R^2 value, implying that the model explains most of the variance in the testing dataset.
    - Conversely, Cell 5 exhibits the lowest R^2 value, indicating less variance explained by the model.

In conclusion, the R^2 value for Cell 4 is approximately 0.85, indicating that the model explains about 85% of the variance in the permeability coefficient for this particular cell. This high R^2 value suggests that the model effectively captures a significant portion of the variance in the testing dataset for Cell 4.

However, the performance of the Transformer model varies significantly across different cells. While Cells 2 and 3 also show strong performance, with lower MAE and MSE values, Cells 1, 5, and 6 exhibit relatively lower R^2 values, indicating less explained variance in these datasets. Therefore, while the Transformer model demonstrates considerable promise, its efficacy is highly dependent on the specific dataset used for training and testing. This highlights the need for further refinement and potential customization of the model to ensure consistent performance across diverse datasets.

Residual Analysis

Residuals, representing the differences between actual and predicted values, were analyzed to assess the model's performance. The residuals for the testing set are illustrated below:

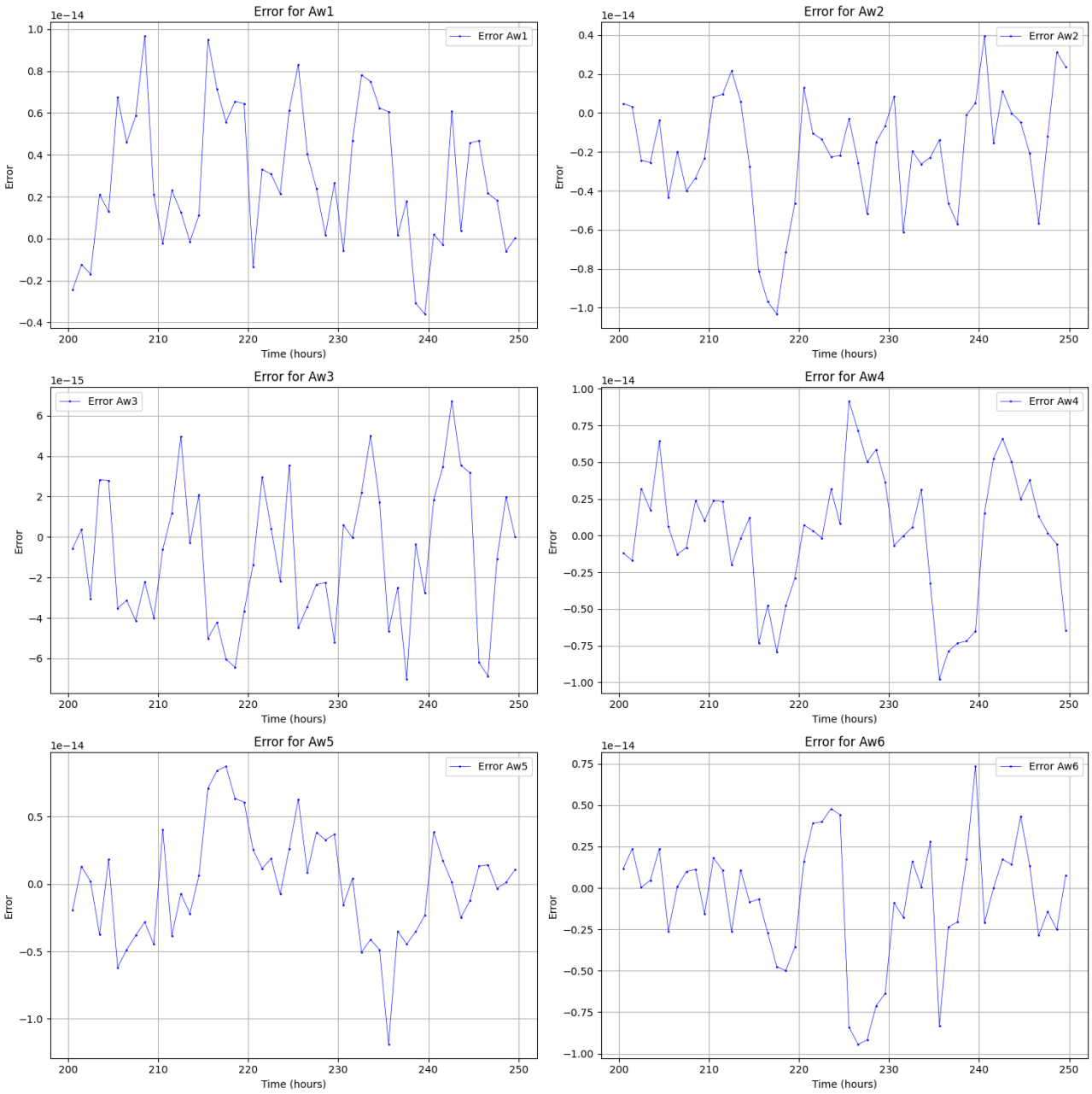


Figure 6.12: Residuals for Multiple Cells (Short-Term)

In our analysis, we observe that the residuals across all prediction cells exhibit random variation. However, the absolute maximum error varies from one cell to another. This behavior is closely tied to the input data, specifically the permeability coefficient. When there is significant variability in the training and testing sets, the predictive performance of the Transformer model may experience a slight degradation. Nevertheless, the overall error remains consistently low, and the results from all cells remain both useful and reliable.

These findings suggest that the Transformer model effectively captures the underlying multivariate patterns in the data. Consequently, it provides valuable insights into the future development of fouling in reverse osmosis membranes. Leveraging this model can enhance predictive maintenance strategies, optimizing the performance of the RO plant.

Autocorrelation of Residuals

To further validate the model, the ACF of the residuals was analyzed, as shown below:

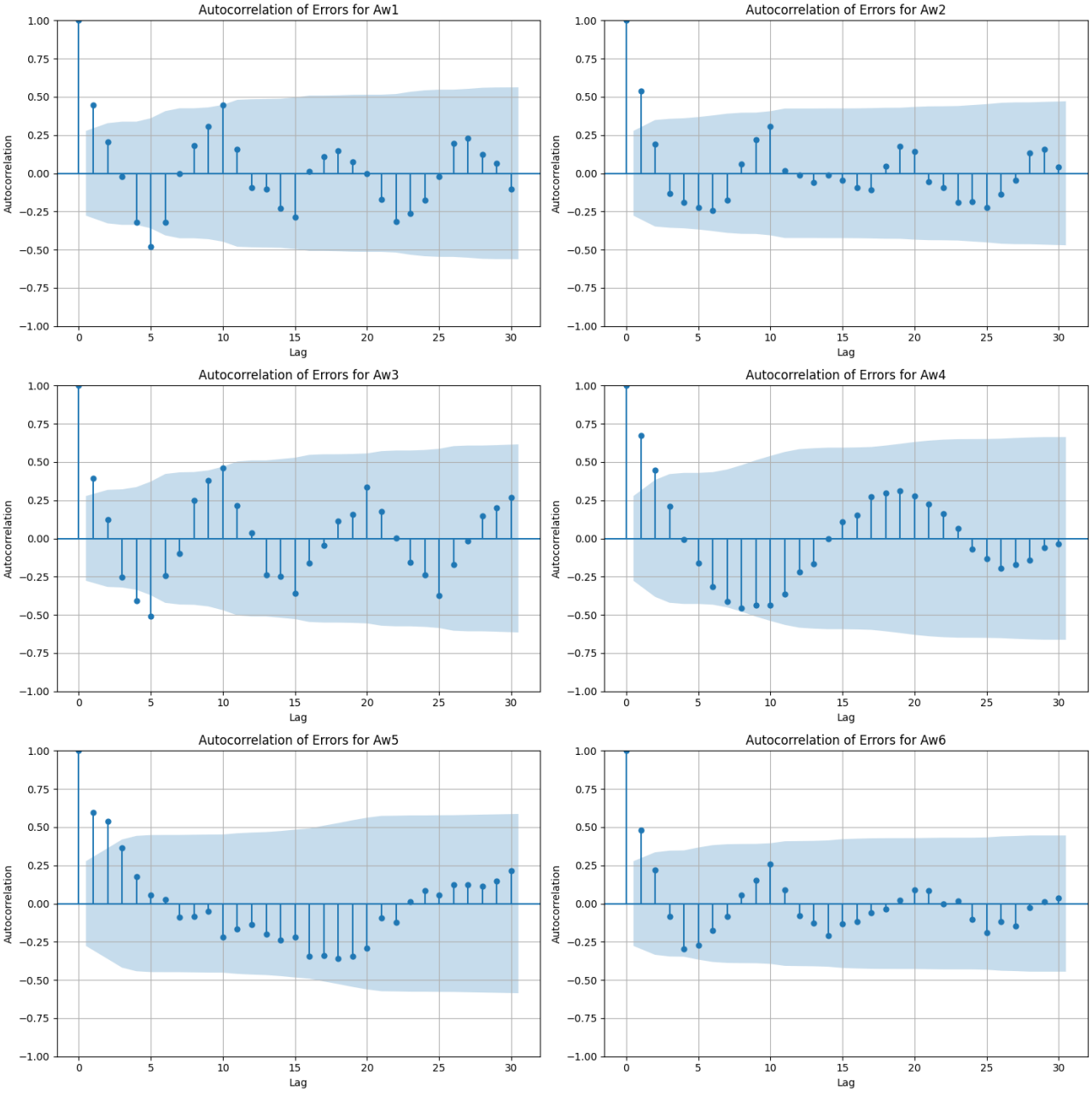


Figure 6.13: Autocorrelation Function of Residuals for Multiple Cells (Short-Term)

When examining the autocorrelation plots for all membrane cells, we find no significant autocorrelation in the residuals. These observations validate the Transformer model's ability to capture the dependencies.

6.5 Conclusion

This study comprehensively explored the application of predictive methods, particularly focusing LSTM networks and Transformer model, for forecasting membrane fouling in reverse osmosis systems. Through detailed simulations and evaluations, several key findings emerged.

The LSTM network demonstrated robust performance in predicting the permeability coefficient of the RO membrane. With low MAE and MSE values, alongside high R^2 scores, the model effectively captured the underlying data patterns, validating its efficacy for both short-term and long-term predictions. Residual analysis and the ACF confirmed the absence of significant errors and autocorrelations, further underscoring the model's accuracy.

The Transformer model also exhibited strong predictive capabilities in both univariate and multivariate contexts. For univariate time series forecasting, the model achieved minimal prediction errors and high R^2 values, maintaining its accuracy over different forecast horizons. In multivariate forecasting, the model efficiently handled multiple variables, providing a comprehensive view of permeability variations across different cells. The low prediction errors and high R^2 values across multiple cells indicated the model's robustness and versatility.

In conclusion, both LSTM and Transformer model are highly effective for predicting membrane fouling in RO systems. Their ability to provide accurate forecasts based on historical data suggests significant potential for improving operational efficiency and reducing maintenance costs in such systems. Future research could enhance these models by incorporating additional features, optimizing model architectures, and applying them to diverse real-world datasets and extended prediction horizons. Such advancements would further solidify the role of these predictive methods in the proactive maintenance and management of membrane filtration systems.

General conclusion

In this work, we have explored the performance prediction of a desalination system using machine learning, demonstrating the efficacy of various models and methodologies in optimizing reverse osmosis processes. The investigation began with a comprehensive analysis of the solution-diffusion model, which provided a fundamental understanding of solvent and solute transport through RO membranes under varying operational conditions. The implementation and validation of this model in MATLAB showcased its accuracy and offered insights for enhancing RO efficiency.

Subsequently, the significant impact of fouling rates on critical parameters such as feed concentration, permeate concentration, and water flux was examined. The findings highlighted the importance of managing fouling to improve system efficiency and longevity, with a time-space model offering precise prediction capabilities. Detailed analysis of membrane fouling through discretization enabled targeted maintenance strategies, enhancing operational decision-making and reducing downtime.

The preparation and transformation of datasets for predicting membrane fouling were also detailed, emphasizing the role of accurate data collection and preprocessing. The application of the Sliding Mode Observer demonstrated high-resolution parameter estimation, forming the foundation for the predictive techniques employed in this study.

Finally, advanced machine learning models, including LSTM networks and Transformer models, were evaluated for their predictive performance. Both models exhibited strong capabilities in forecasting membrane permeability and fouling, with low prediction errors. These models proved effective in both univariate and multivariate contexts, suggesting their potential for improving operational efficiency and reducing maintenance costs in RO systems.

Overall, this report underscores the significant potential of machine learning in enhancing the performance prediction and management of desalination systems. Future research should focus on further refining these models, incorporating additional features, and applying them to diverse real-world datasets to solidify their role in the proactive maintenance and optimization of membrane filtration systems.

Bibliography

[1] E.T. Sayed B. Soudan M.A. Abdelkareem, M. el Haj Assad. Recent progress in the use of renewable energy sources to power water desalination plants. *Desalination*, 435:97–113, 2018.

[2] A.E. Kabeel. Performance of solar still with a concave wick evaporation surface. *Energy*, 34(10):1504–1509.

[3] R. Hashaikeh N. Hilal H. Nassrullah, S.F. Anis. Energy for desalination: a stateof-the-art review. *Desalination*, 491, 2020.

[4] L.L. Kazmerski A. Al-Karaghouli. Energy consumption and water production cost of conventional and renewable-energy-powered desalination processes. *Renew.*

[5] D. Prats D. Zarzo. Desalination and energy consumption. what can we expect in the near future? *Desalination*, 427:1–9, 2018.

[6] Baltasar Peñate and Lourdes García-Rodríguez. Current trends and future prospects in the design of seawater reverse osmosis desalination technology. *Desalination*, 284, 01 2011.

[7] F. del Vigo S.P N. Pena, ~ S. Gallego. Evaluating impact of fouling on reverse osmosis membranes performance. *Desalin. Water Treat*, 51:4–6, 2013.

[8] S.H. Kim N. Prihasto, Q.F. Liu. Pre-treatment strategies for seawater desalination by reverse osmosis system. *Desalination*, 409:308–316, 2009.

[9] M. Elimelech J.R. Werber, A. Deshmukh. The critical need for increased selectivity, not increased water permeability, for desalination membranes. *Environ. Sci. Technol. Lett*, 3:112–120, 2016.

[10] H.R. van Tonningen A. Altaee, G. Zaragoza. Comparison between forward osmosis-reverse osmosis and reverse osmosis processes for seawater desalination. *Desalination*, 336:50–57, 2014.

[11] S. Gray S. Burn. Efficient desalination by reverse osmosis: A guide to ro practice. *IWA Publishing, London, UNITED KINGDOM*, 2015.

[12] George M. Ayoub Lilian Malaeb. Reverse osmosis technology for water treatment: State of the art review. *Desalination*, 267:1–8, 2011.

[13] P.-X. Sheng K.G.N. Nanayakkara L.K. Wang Y.-P. Ting J.P. Chen, E.S.K. Chian. Desalination of seawater by reverse osmosis. *Membr. Desalin. Technol, Humana Press, Totowa, NJ, 2011.*

[14] A. Yokozeki. Osmotic pressures studied using a simple equation-of-state and its applications. *Appl. Energy*, 83:15–41, 2006.

[15] Noora N. Darwishb Naif A. Darwisha Nidal Hilal Muhammad Qasima, Mohamed Badrelzamana. Reverse osmosis desalination: A state-of-the-art review. *Desalination*, 459:59 – 104, 2019.

[16] S. Srinivasan and Tien Chi. A simplified method for the prediction of concentration polarization in reverse osmosis operation for multi-component systems. *Desalination*, 7(2):133–145, 1970.

[17] A.P. Reverberi, B. Fabiano, Cristiano Cerrato, and V.G. Dovì. Concentration polarization in reverse osmosis membranes: Effect of membrane splitting. *Chemical Engineering Transactions*, 39:763–768, 01 2014.

[18] J. Kucera. *Desalination: Water From Water*. Wiley-Scrivener, Salem, 2014.

[19] J. Bundschuh A. Figoli, J. Hoinkis. Membrane technologies for water treatment: Removal of toxic trace elements with emphasis on arsenic, fluoride and uranium. *CRC Press, London*, 2016.

[20] Xue Yun Long, Rui Zhang, Bing Cao, and Pei Li. *Tubular membranes and modules*, pages 431–448. 01 2021.

[21] Abdellah Shafieian, Mehdi Khiadani, and Masoumeh Zargar. Performance analysis of tubular membrane distillation modules: An experimental and cfd analysis. *Chemical Engineering Research and Design*, 183:478–493, 2022.

[22] R.W. Baker. Membrane technology and applications. *2nd ed., John Wiley Sons Ltd., New York*, 2014.

[23] Marian G. Marcovecchio, Nicolás J. Scenna, and Pío A. Aguirre. Improvements of a hollow fiber reverse osmosis desalination model: Analysis of numerical results. *Chemical Engineering Research and Design*, 88(7):789–802, 2010.

[24] J. Corson MacNeil. Membrane separation technologies for treatment of hazardous wastes. *Crit. Rev. Environ. Control*, 18:91–131, 1988.

[25] Dongfei li, Rong Wang, and Tai-Shung Chung. Fabrication of lab-scale hollow fiber membrane modules with high packing density. *Separation and Purification Technology*, 40:15–30, 11 2004.

[26] G. Schock and A. Miquel. Mass transfer and pressure loss in spiral wound modules. *Desalination*, 64:339–352, 1987.

[27] T. Wintgens T. Melin C. Fritzmann, J. Löwenberg. State-of-the-art of reverse osmosis desalination. *Desalination*, 18:1–76, 2007.

[28] J. Kucera. *Reverse Osmosis: Design, Processes, and Applications for Engineers*. Wiley, 2010.

[29] A.K. Mishra M.T.M. Pendergast M.C.Y. Wong B.B. Mamba-V. Freger A.R.D. Verliefde E.M.V. Hoek J. Wang, D.S. Dlamini. A critical review of transport through osmotic membranes. *J. Memb. Sci*, 454:516–537, 2014.

[30] H. Mehdizadeh. Modelling of transport phenomena in reverse osmosis membranes. *McMaster Unversity*, 1990.

[31] Sid El Mahi Lamine Kadi Omar Belhamiti Belkacem Absar. Reverse osmosis modeling with the orthogonal collocation on finite element method. *Desalination and Water Treatment*, 21(1-3):23–32, 1990.

[32] M. Kamil K. JAMAL, M.A. Khan. Mathematical modeling of reverse osmosis systems. *Desalination*, pages 20 – 42, 2004.

[33] Belkacem Absar, Sid El Mahi Lamine Kadi, and Omar Belhamiti. Reverse osmosis modeling with the orthogonal collocation on finite element method. *Desalination and Water Treatment*, 21, 2010.

[34] Zielinski J. M. Wendel R. G. Uchrin C. G Slater, C. S. Modeling of small scale reverse osmosis systems. *Desalination*, 53:267 – 284, 1885.

[35] F. Suja M.Z. Meor Talib C.H. Koo, A.W. Mohammad. Use and development of fouling index in predicting membrane fouling. *Sep. Purif. Rev.*, 42(4):296–339, 2013.

[36] Haleema Saleem S. Javaid Zaidi. Reverse osmosis systems: Design, optimization and troubleshooting guide. page 489, 2021.

[37] S. Prasad Das A. Uppu, A. Chaudhuri. Numerical modeling of particulate fouling and cake-enhanced concentration polarization in roto-dynamic reverse osmosis filtration systems. *Desalination*, 468, 2019.

[38] L.N. Sim , et al. A review of fouling indices and monitoring techniques for reverse osmosis. *Desalination*, 434:169–188, 2018.

[39] T. Wagener and D. Solomatine. *Hydrological Modeling*, volume 2 of *Treatise on Water Science*. 2011. Accessed: Nov. 28, 2022.

[40] A.G. Pervov. Scale formation prognosis and cleaning procedure schedules in reverse osmosis systems operation. *Desalination*, 83(1–3):77–118, 1991.

[41] D. Hasson J. Gilron. Calcium sulphate fouling of reverse osmosis membranes: flux decline mechanism. *Chem. Eng. Sci.*, 42(10):2351–2360, 1987.

[42] I. Nuez A. Ruiz-García. Long-term intermittent operation of a full-scale bwro desalination plant. *Desalination*, 489, 2020.

[43] X. Mao Wang X. Huang W. Chen Lin, R. Peng Shao. Impacts of non-uniform filament feed spacers characteristics on the hydraulic and anti-fouling performances in the spacer-filled membrane channels: experiment and numerical simulation. *Water Res.*, 185, 2020.

[44] K. Klinko M. Wilf. Performance of commercial seawater membranes. *Desalination*, 96(1–3):465–478, 1994.

[45] R.A. Mohamed A.M. Tolba. *Performance and Characteristics of Reverse Osmosis Membranes*. 2019.

[46] N. Al-Bastaki A. Abbas. Performance decline in brackish water film tec spiral wound ro membranes. *Desalination*, 136(1–3):281–286, 2001.

[47] M. Al-Ahmad M. Zhu, M.M. El-Halwagi. Optimal design and scheduling of flexible reverse osmosis networks. *J. Membr. Sci.*, 129(2):161–174, 1997.

[48] K. Bensadok M. Belkacem, S. Bekhti. Groundwater treatment by reverse osmosis. *Desalination*, 206(1–3):100–106, 2007.

[49] C.H. Lee S. Lee, J. Kim. Analysis of caso4 scale formation mechanism in various nanofiltration modules. *J. Membr. Sci.*, 163(1):63–74, 1999.

[50] I. Nuez A. Ruiz-García. Long-term performance decline in a brackish water reverse osmosis desalination plant. predictive model for the water permeability coefficient. *Desalination*, 397:101–107, 2016.

[51] I. de la N. Pestana A. Ruiz-García. Feed spacer geometries and permeability coefficients. effect on the performance in bwro spriral-wound membrane modules. *Water*, 11(1):152, 2019.

[52] Iván Vallés-Pérez, Emilio Soria-Olivas, Marcelino Martínez-Sober, Antonio J. Serrano-López, Juan Gómez-Sanchís, and Fernando Mateo. Approaching sales forecasting using recurrent neural networks and transformers, 2022.

[53] Sining Ang. Precipitation forecasting using transformer: A comparative study with unet. *Highlights in Science, Engineering and Technology*, 39:627–632, 04 2023.

[54] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 6000–6010. NIPS, 2017.

[55] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.

[56] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[57] Haoyi Cui Fan Zhang Jinjiang Li Xiaofeng Zhang Mingli Zhang Caiming Zhang Siyuan Huang, Yepeng Liu. Meaformer: An all-mlp transformer with temporal external attention for long-term time series forecasting. *Information Sciences*, 669, 2024.

[58] et al. N. Najid. Fouling control and modeling in reverse osmosis for seawater desalination: a review. *Comput. Chem. Eng.*, 162, 2022.

[59] S.L. Ong W.J. Ng K.L. Chen, L. Song. The development of membrane fouling in full-scale ro processes. *J. Membr. Sci.*, 232(1–2):63–72, 2004.

[60] D.Y. Kim M. Park D.R. Yang J.H. Kim Y.G. Lee, Y.S. Lee. A fouling model for simulating long-term performance of swro desalination process. *J. Membr. Sci.*, 401–402:282–291, 2012.

[61] Wouter Kool, Herke van Wijk, and Max de Hoog. The transformer network for the traveling salesman problem. *arXiv preprint arXiv:2103.03012*, 2021.

[62] Youshan Li, Nan Wang, Jianping Shi, Jiachen Liu, Junxiong Hou, Shilin Zhang, and Qi Liu. itransformer: Transformer time series forecasting model. *arXiv preprint arXiv:2104.14257*, 2021.

[63] Ziyang Liu, Stephen Merity, Alec Radford, Yutong Xie, and Pengcheng Ma. Cats: Cross-attention time series transformer. *arXiv preprint arXiv:2106.03806*, 2021.