

Serial N° :/2026

University of Kasdi Merbah Ouargla

Faculty of Hydrocarbons, Renewable Energies, Earth and the Universe Sciences

Department of Production



Master Dissertation

Specialization: Academic Production

By Abdelhamid CHIKH AISSA

-THEME-

Machine Learning-Based Waterflood Optimization System for Improved Oil Recovery

Jury Members:

Jury Chair	ZIARI	Saber	Univ. Ouargla
Supervisor	MERABTI	Hocine	Univ. Ouargla
Co-Supervisor	MEDJANI	Fethi	Univ. Ouargla
Examiner	GHALI	Ahmed	Univ. Ouargla

Academic year 2025/2026

Acknowledgments

First and foremost, I thank ALLAH the Merciful for giving me the strength, opportunity, and patience to complete this modest work.

I would like to express my deepest gratitude to my dissertation supervisor, **Dr. Hocine MERABTI**, for the trust he placed in me by accepting to supervise this work, as well as for guidance, which greatly contributed to the development of this dissertation.

I would also like to express my sincere gratitude to my co-supervisor, **Pr. Fethi MEDJANI**, for his continuous support, guidance, encouragement, and invaluable assistance throughout the realization of this work. His advice, availability, and involvement were of great importance and highly appreciated.

My special gratitude also to my project mentors in Groupement Berkine for their practical guidance, support, and assistance during my internship.

Finally, I thank all the individuals who contributed to the success of this work and assisted me in completing this work.

Dedication

To my dear father, **Youcef**, and my dear mother, **Nanna ADDAOUD**, and to my dear brothers, and beloved sister. To my esteemed friend **Ziyad MESSAOUDI**. Your unwavering support and love have been my guiding light throughout this journey.

Thank you for everything.

Abdelhamid CHIKH AISSA

Abstract

Although waterflooding remains the primary method of secondary oil recovery, its efficiency is often limited by reservoir heterogeneity, inefficient injection allocation, non-uniform sweep efficiency, and uncertainty in injector-producer connectivity. Traditional full-physics reservoir simulation models are computationally intensive and complex to use, which makes optimization and decision-making challenging. This dissertation therefore proposes a hybrid physics-informed machine learning optimization system for waterflooding management and injection allocation optimization, developed and validated using data from the EMN sub-field of the El Merk oil field in the Berkine Basin in Algeria. The system integrates a Capacitance-Resistance Model (CRM) for connectivity, a Long Short-Term Memory (LSTM) model for forecasting fractional flow, a Physics-Informed Neural Network (PINN) for solving the heterogeneous Buckley–Leverett equation, and a constrained Sequential Least Squares Programming (SLSQP) optimizer for injection allocation. This system was developed using 8460 well-test observations for the machine learning models and 27 574 allocated observations for the CRM model, drawn from six producers and two injectors over 156 months. The LSTM model achieved a coefficient of determination (R^2) of 0.944, while the PINN delivered an overall R^2 of 0.978. The optimizer produced the best possible injection rates, leading to an estimated incremental oil volume uplift of 222 285 STB per year. This hybrid system is more efficient, faster, and simpler than traditional methods, demonstrating its potential to optimize waterflooding and improve oil recovery.

Keywords: Waterflooding optimization, Capacitance-Resistance Model (CRM), Physics-Informed Neural Networks (PINNs), Long Short-Term Memory (LSTM), Buckley-Leverett equation, Sequential Least Squares Programming (SLSQP), El Merk field

Résumé

Bien que le waterflooding reste la principale méthode de récupération secondaire du pétrole, son efficacité est souvent limitée par l'hétérogénéité du réservoir, une allocation inefficace de l'injection, un balayage non uniforme et l'incertitude sur la connectivité injecteur-producteur. Les modèles traditionnels de simulation de réservoir basés sur la physique sont très coûteux en calcul, chronophages et complexes à utiliser, ce qui rend l'optimisation et la prise de décision particulièrement difficiles. Cette dissertation propose donc un système hybride d'optimisation basé sur l'apprentissage automatique informé par la physique pour la gestion du waterflooding et l'optimisation de l'allocation des injections, développé et validé à partir de données issues du sous-champ EMN du champ pétrolier d'El Merk, situé dans le bassin de Berkine en Algérie. Ce système intègre un modèle de Capacitance-Résistance (CRM) pour la connectivité injecteur-producteur, un modèle Long Short-Term Memory (LSTM) pour la prévision du flux fractionnaire, un réseau de neurones informé par la physique (PINN) pour la résolution de l'équation hétérogène de Buckley–Leverett, ainsi qu'un algorithme d'optimisation sous contraintes basé sur la méthode Sequential Least Squares Programming (SLSQP) pour l'allocation des débits d'injection. Le système a été développé à partir de 8460 observations de tests de puits utilisées pour les modèles d'apprentissage automatique et de 27 574 observations d'allocation utilisées pour le modèle CRM, provenant de six producteurs et deux injecteurs sur une période de 156 mois. Le modèle LSTM a atteint un coefficient de détermination (R^2) de 0,944, tandis que le PINN a obtenu un R^2 global de 0,978. L'algorithme d'optimisation a fourni les débits d'injection optimaux, conduisant à une augmentation estimée de la production pétrolière de 222 285 STB par an. Ce système hybride est plus efficace, plus rapide et plus simple que les méthodes traditionnelles, démontrant son potentiel pour optimiser le waterflooding et améliorer la récupération du pétrole.

Mots-clés: Optimisation du waterflooding, Capacitance-Résistance Model (CRM), Physics-Informed Neural Networks (PINNs), Long Short-Term Memory (LSTM), Équation de Buckley–Leverett, Sequential Least Squares Programming (SLSQP), Champ d'El Merk.

الملخص

على الرغم من أن عملية حقن المياه في المكنن (Waterflooding) تُعد الطريقة الأساسية للاسترجاع الثانوي للنفط، فإن فعاليتها غالبًا ما تكون محدودة بسبب عدم تجانس المكنن، عدم كفاءة توزيع الحقن، وعدم انتظام كفاءة كنس النفط، إضافة إلى عدم اليقين في طبيعة الاتصال بين آبار الحقن والإنتاج. كما أن نماذج المحاكاة التقليدية للمكانن المعتمدة كليًا على القوانين الفيزيائية تتسم بأنها مكلفة حسابيًا، تستغرق وقتًا طويلاً، ومعقدة الاستخدام، مما يجعل اتخاذ قرارات التحسين في الوقت الانى أمرًا صعبًا. لمعالجة هذه التحديات، تطوّر هذه الأطروحة نظامًا هجينًا للتحسين قائمًا على التعلم الآلي المدعوم بالفيزياء (Physics-Informed Machine Learning) لإدارة عمليات حقن المياه وتحسين توزيع معدلات الحقن، وقد تم تطويره والتحقق من صحته باستخدام بيانات من الحقل الفرعي EMN التابع لحقل المرق (El Merk) الواقع في حوض بركين في الجزائر. يدمج هذا الإطار بين نموذج السعة والمقاومة (CRM) لتحديد الترابط بين الآبار، ونموذج LSTM للتنبؤ بالتدفق المائي، وشبكة عصبية مدعومة بالفيزياء (PINN) لحل معادلة Buckley-Leverett في وسط غير متجانس، بالإضافة إلى خوارزمية تحسين مقيدة تعتمد على أسلوب البرمجة التتابعية للمربعات الصغرى (Sequential Least Squares Programming – SLSQP) لتوزيع معدلات الحقن. تم تطوير النظام باستخدام 8460 خلية من العمليات على الآبار استخدمت لتطوير نماذج التعلم الآلي، بالإضافة إلى 27 574 خلية مُخصّصة استُخدمت لنموذج CRM اخذت من ستة آبار إنتاج واثنين من آبار الحقن على مدى 156 شهرًا. حقق نموذج LSTM معامل تحديد (R^2) قدره 0.944، بينما حققت شبكة PINN قيمة إجمالية لـ R^2 بلغت 0.978. كما أدى المُحسن إلى تحديد معدلات حقن مثالية، مما نتج عنه زيادة تقديرية في إنتاج النفط بمقدار 222 285 برميل قياسي في السنة. يتميز هذا النظام الهجين بأنه أقل استهلاكًا للموارد الحسابية، أسرع، وأبسط مقارنة بالطرق التقليدية، مما يُظهر إمكانيات قوية لتحسين فعال عمليات حقن المياه وزيادة معدل الاسترجاع النفطي.

الكلمات المفتاحية: تحسين حقن المياه، نموذج السعة والمقاومة (CRM)، الشبكات العصبية المدعومة بالفيزياء (PINNs)، الذاكرة طويلة وقصيرة المدى (LSTM)، معادلة Buckley-Leverett، البرمجة التتابعية للمربعات الصغرى (SLSQP)، حقل المرق (El Merk).

Table of Contents

Acknowledgments	I
Dedications	II
Abstract	III
Table of contents.....	IV
List of figures.....	V
List of tables	VI
List of abbreviations	VII
General introduction	1

Chapter I: Presentation of El Merk field

Introduction.....	03
I.1 Historical Background of Groupement Berkine	04
I.2 General Overview of El Merk Field	04
I.3 Geographical Location of the Berkine Basin	04
I.4 Geological Context of the Berkine Basin	05
I.4.1 Structural Framework	05
I.4.2 Stratigraphic Framework	06
I.5 Reservoir and Petrophysical Properties of El Merk Field	06
I.5.1 TAGI Reservoir (Trias Argilo-Gréseux Inférieur)	08
I.5.2 RKF Reservoir	08
I.5.3 Strunian Reservoir	08

I.5.4 Petrophysical properties	09
I.6 The subdivision of El Merk field	10
Conclusion	10

Chapter II: Waterflooding Engineering

Introduction.....	12
II.1 Waterflooding overview	13
II.2 Criteria for Selecting Reservoirs for Waterflooding	14
II.2.1 Reservoir Geometry	14
II.2.2 Fluid Properties	14
II.2.3 Reservoir Depth	15
II.2.4 Lithology and Rock Properties	15
II.2.5 Fluid Saturations	16
II.2.6 Reservoir Uniformity and Pay Continuity	16
II.2.7 Primary Reservoir Driving Mechanisms	17
II.3 Waterflooding Patterns	20
II.3.1 Irregular Injection Patterns	20
II.3.2 Peripheral Injection Patterns	21
II.3.3 Regular Injection Patterns	22
II.4 Overall Recovery Efficiency of Waterflooding	22
II.5 Displacement Efficiency	23
II.6 Frontal Displacement Theory	24
II.6.1 Fractional flow equation	24
II.6.2 Frontal Advance Equation	29

II.7 Areal Sweep Efficiency	32
II.7.1 Mobility Ratio “M”	33
II.7.2 Cumulative Water Injected.....	36
II.8 Vertical Sweep Efficiency	36
II.9 Waterflood Diagnostic Plots and Surveillance Methods	38
Conclusion	39

Chapter III: Overview of Machine Learning Concepts in Petroleum Engineering

Introduction	41
III.1 Machine Learning Definition	41
III.2 Machine Learning Categories	42
III.2.1 Supervised Learning	42
III.2.2 Unsupervised Learning	44
III.2.3 Deep Learning	45
III.2.4 Reinforcement Learning	49
III.3 Machine Learning Libraries	49
III.4 Model Evaluation Metrics	49
III.4.1 Regression Metrics	50
III.4.2 Classification Metrics	50
Conclusion	51

Chapter IV: Waterflooding Optimization System

Introduction.....	53
IV.1 Materials Used.....	53
IV.2 Methodology	54
IV.2.1 Data collection	54

IV.2.2 Data preprocessing	56
IV.2.3 Model development	58
IV.2.4 Optimization Unit	74
Conclusion	78
General Conclusion	80

REFERENCES

APPENDIX

List of figures

Figure I.1 Groupement Berkine Logo	03
Figure I.2 Geographical location of the Berkine Basin	05
Figure I.3 Stratigraphic Cross-Section of El Merk field	07
Figure I.4 Distribution of reservoirs in stratigraphic cross-section	09
Figure I.5 Subdivision map of El Merk field	10
Figure II.1 Oil recovery mechanisms	13
Figure II.2 Peripheral injection pattern	21
Figure II.3 Regular injection patterns	23
Figure II.4 Linear waterflood scheme and fractional flow curves	26
Figure II.5 Effect of oil viscosity on fractional flow	27
Figure II.6 Effect of injection well location on fractional flow	29
Figure II.7 Water flow through a differential element	30
Figure II.8 Fractional flow curve and saturation derivative	31
Figure II.9 Effect of the capillary pressure on the shape of the fractional flow curve	32
Figure II.10 Water saturation profile as a function of time t and distance x	33
Figure II.11 Mobility ratio variation	34
Figure II.12 X-ray shadowgraphs of waterflood progress	35
Figure II.13 Idealized layered pattern system	37
Figure III.1 Machine Learning Categories	42

Figure III.2 Schematic representation of an MLP	47
Figure III.3 Schematic depiction of long short-term memory (LSTM) cell architecture	48
Figure III.4 An illustration of confusion matrix	51
Figure IV.1 Methodological framework of the project	54
Figure IV.2 EMN field water-injection history of the two water injectors (2013–2026)	55
Figure IV.3 EMN field production history for the six oil producers (2013–2026)	55
Figure IV.4 EMN field oil production history (monthly average) for all six producers	56
Figure IV.5 Waterflood Optimization System Architecture	59
Figure IV.6 Illustrative decline-and-response decomposition for N_P17X and N_P6X producers.	60
Figure IV.7 The fitted CRM connectivity matrix and the producer time constants	63
Figure IV.8 Corey kr curves with the EMN SCAL data	65
Figure IV.9 FFM training and validation results	68
Figure IV.10 All-Pair PINN performance on EMN field observations	71
Figure IV.11 Saturation profiles $S_w(x,t)$ from the PINN for four representative pairs	72
Figure IV.12 Flood front advance position and the volumetric sweep efficiency of PINN	73
Figure IV.13 The optimum injection-allocation and oil rate gain	78

List of tables

Table II.1 Primary driving mechanisms oil recovery range	20
Table IV.1 Data before outliers' removal	57
Table IV.2 Data after outliers' removal	58
Table IV.3 CRM detailed input specification (per producer, per monthly refit)	61
Table IV.4 CRM detailed output specification (per producer)	62
Table IV.5 Fitted CRM parameters for the EMN field	62
Table IV.6 Physical constants of the EMN field	64
Table IV.7 Detailed input features of the FFM model	66
Table IV.8 FFM network architecture	66
Table IV.9 Training configuration of the FFM	67
Table IV.10 EMN_PINN architecture and specification	69
Table IV.11 PINN outputs summary	74
Table IV.12 Detailed input specification of the injection optimizer	77
Table IV.13 Detailed output specification of the injection optimizer	77

List of Abbreviations

API: American Petroleum Institute

BL: Buckley-Leverett

CPF: Central Processing Facility

CRM: Capacitance-Resistance Model

EDA: Exploratory Data Analysis

EMK: El Merk

EME: El Merk East

EMN: El Merk North

EKT: El Merk (specific sector)

FBHP: Flowing Bottom Hole Pressure

FFM: Fractional Flow Forecasting Model

GB: Groupement Berkine

GOR: Gas-Oil Ratio

HBNS: Hassi Berkine North South

IQR: Interquartile Range

kNN: k-Nearest Neighbors

MAD: Median Absolute Deviation

ML: Machine Learning

MLP: Multilayer Perceptron

OOIP: Original Oil In Place

Oxy: Occidental Petroleum

PINN: Physics-Informed Neural Network

PVT: Pressure-Volume-Temperature

RBF: Radial Basis Function

RF: Recovery Factor / Random Forest

RMSE: Root Mean Squared Error

SCAL: Special Core Analysis

STB: Stock Tank Barrel

SVR: Support Vector Regression

TAGI: Trias Argilo-Gréseux Inférieur

VRR: Voidage Replacement Ratio

WOR: Water-Oil Ratio

XGBoost: Extreme Gradient Boosting

General Introduction

The global oil and gas industry is perpetually seeking novel methods to optimize hydrocarbon recovery from mature fields, while concurrently concentrating on reducing operational costs and enhancing efficiency. In Algeria, where the oil and gas industry is pivotal to the national economy, optimizing the performance of existing mature fields (including waterflooded fields) is imperative to ensure continued production and enhance the overall hydrocarbon recovery.

El Merk field, a significant oil and gas development in the Berkine Basin of southeastern Algeria, exhibits characteristics that are not uncommon in such fields. These include water-flooded subfields facing challenges such as reservoir heterogeneity, non-uniform sweep efficiency, premature water breakthrough, and inefficient injection allocation. Conventional full-physics reservoir simulation models, despite their intricate nature, frequently prove to be excessively computationally intensive, time-consuming, and complex for regular operational decision-making processes.

To overcome these limitations, this dissertation proposes a hybrid physics-informed machine learning optimization system designed specifically for waterflooding management and injection allocation optimization, developed and validated using data from the EMN sub-field of El Merk. The developed framework combines traditional reservoir engineering tools with modern artificial intelligence techniques to create a practical, fast, and efficient solution for waterflood management and optimization.

The primary objective of this work is to design, implement, and validate an integrated optimization system that includes a Capacitance-Resistance Model (CRM) for injector-producer connectivity, an LSTM model for fractional flow forecasting, a Physics-Informed Neural Network (PINN) for solving the heterogeneous Buckley-Leverett frontal advance equation, and a constrained optimizer to determine the optimal monthly water injection rates.

This dissertation is structured as follows:

- Chapter I presents a detailed overview of El Merk field, its geological setting, reservoir properties, and production history.
- Chapter II covers the core physics of waterflooding engineering.
- Chapter III introduces key concepts and techniques in machine learning and deep learning.

- Chapter IV describes the development, implementation, and results of the proposed hybrid optimization system using real field data.

The dissertation concludes with a summary of the main findings, contributions, and recommendations for future research.

Chapter I: Presentation of El Merk field

Introduction

This chapter presents an overall overview of El Merk field, which belongs to Groupement Berkine one of the major hydrocarbon developments located in southeastern Algeria within the Berkine Basin. The purpose of this chapter is to describe the key characteristics of the field, including its geographical location, geological framework, reservoir properties, fluid characteristics, production history, and the challenges it faces.

I.1 Historical Background of Groupement Berkine

The Groupement Berkine was established in 1998 as an operating entity formed by Sonatrach and Anadarko Petroleum and now Occidental Petroleum (Oxy), one of the major independent oil and gas exploration and production companies worldwide. (Occidental Petroleum Corporation, 2025)

The Groupement is represented by the visual representation shown in **Figure I.1**, and is responsible for managing the joint operations between Sonatrach and its international partners. By May 2002, it had already achieved a cumulative production of approximately 100 million barrels of oil, reflecting the importance of the Berkine assets in Algeria's hydrocarbon production portfolio.

The Groupement Berkine oversees several major production assets, including the Hassi Berkine and El Merk fields, which are among the key contributors to national oil and gas output. (Groupement Berkine, 2025)



Figure I.1 Groupement Berkine Logo

I.2 General Overview of El Merk Field

El Merk field is one of the most recent large-scale developments in southeastern Algeria. The field is situated in a remote Saharan region approximately 300 to 350 km southeast of Hassi Messaoud and about 80 km from the HBNS field, in an environment characterized by extensive sand dunes and challenging operating conditions.

The project was developed through a partnership led by Sonatrach in association with several international companies, including Anadarko Petroleum, ENI, Maersk, ConocoPhillips, and Talisman. With an estimated investment of approximately 4 billion USD, El Merk project is designed to sustain hydrocarbon production for 30 to 40 years.

The exploration and development of El Merk field progressed through several key stages:

- 1993 – 1998: Discovery of hydrocarbon accumulations by Anadarko Petroleum
- 1998 – 1999: Acquisition and interpretation of 2D seismic data
- 1999: Execution of 3D seismic studies and correction of 2D maps
- 2001: Drilling of more than 13 exploratory wells and coring operations
- 2002 – 2003: Conducting PVT studies, petrophysical analysis, and reserve estimation
- 2006: Recognition of El Merk as a field with significant hydrocarbon reserves
- 2006 – 2008: Reservoir studies, field modeling, well placement, and development planning
- 2008: Start of development phase, including drilling operations and construction of CPF and infrastructure
- March 2013: Start of production. (Groupement Berkine, 2025)

I.3 Geographical Location of the Berkine Basin

The Berkine Basin is located in the Algerian Sahara, approximately 350 km southeast of Hassi Messaoud. It extends over a large intracratonic area covering parts of eastern Algeria and bordering regions of Tunisia and Libya (see **Figure I.2**).

The basin is bounded by:

- The Tunisian border to the east (~114 km)
- The Libyan border (~160 km)
- The Illizi Basin to the south
- The Dahar dome to the north
- Hassi Messaoud region to the west

This strategic location makes the Berkine Basin one of the most important hydrocarbon provinces in Algeria. (Groupement Berkine, 2025; Saoud, 2013)

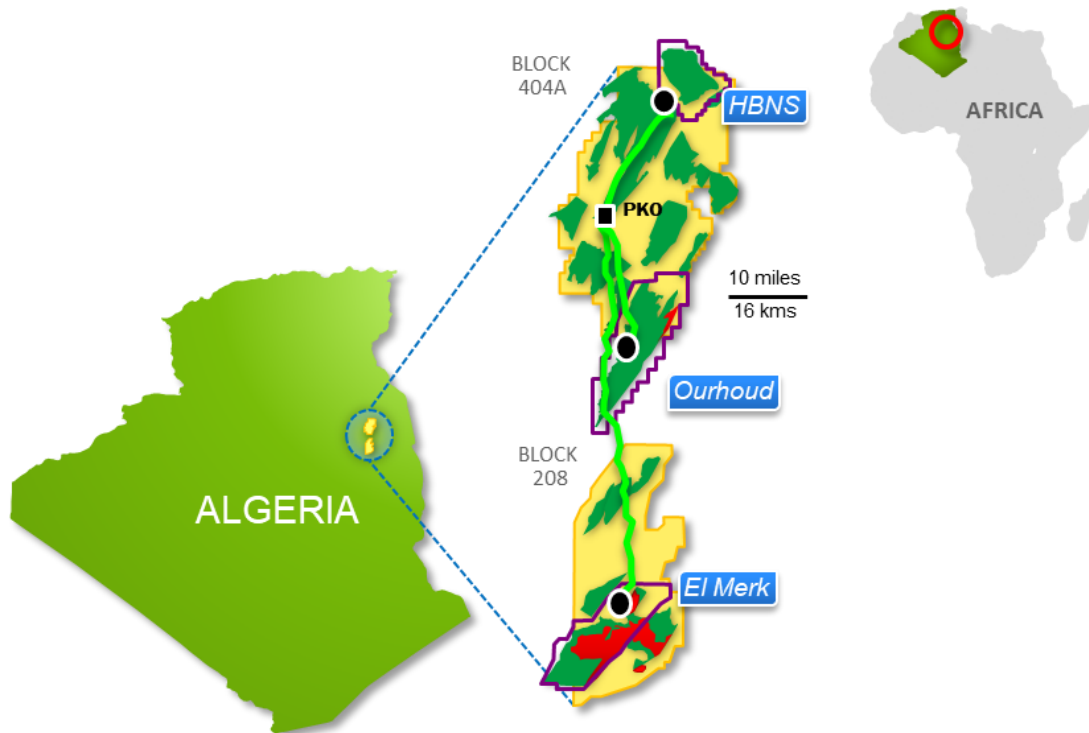


Figure I.2 Geographical location of the Berkine Basin

I.4 Geological Context of the Berkine Basin

The geological framework of the Berkine Basin is characterized by distinct structural and stratigraphic features that control hydrocarbon accumulation and reservoir behavior.

I.4.1 Structural Framework

From a structural perspective, the Berkine Basin is marked by a clear contrast between two main domains, both influenced by basement tectonics:

a) Western domain

This area is strongly structured, with major fault systems exhibiting significant vertical displacements that can exceed 500 meters. Structural closures are highly pronounced, reaching up to 900 meters in certain areas such as Rhourde Nouss. These features create favorable conditions for hydrocarbon trapping.

b) Central domain

In contrast, the central part of the basin is less deformed and characterized by relatively gentle structuring. It presents a monoclinal uplift toward the south, with lower fault intensity compared to the western domain.

This structural configuration plays a crucial role in trap formation, reservoir compartmentalization, and the distribution of fluids within the basin. (Saoud, 2013)

I.4.2 Stratigraphic Framework

The stratigraphic succession of the Berkine Basin, as represented in **Figure I.3**, is composed of several sedimentary sequences deposited over geological time. The basin is predominantly filled with Mesozoic deposits, which form the main sedimentary sequence. These deposits lie in major unconformity over partially known Paleozoic formations, which include important hydrocarbon-bearing reservoirs. Above the Mesozoic sequence, Tertiary detrital deposits are present, forming a discontinuous cover. The total thickness of the sedimentary column reaches approximately 7000 meters, reflecting the long and complex geological evolution of the basin.

I.5 Reservoir and Petrophysical Properties of El Merk Field

El Merk field produces hydrocarbons from three main reservoir formations, namely TAGI, RKF, and Strunian (**Figure I.4**).

These reservoirs differ in terms of lithology, fluid type, and production behavior, which adds complexity to field development and reservoir management. (Groupement Berkine, 2025)

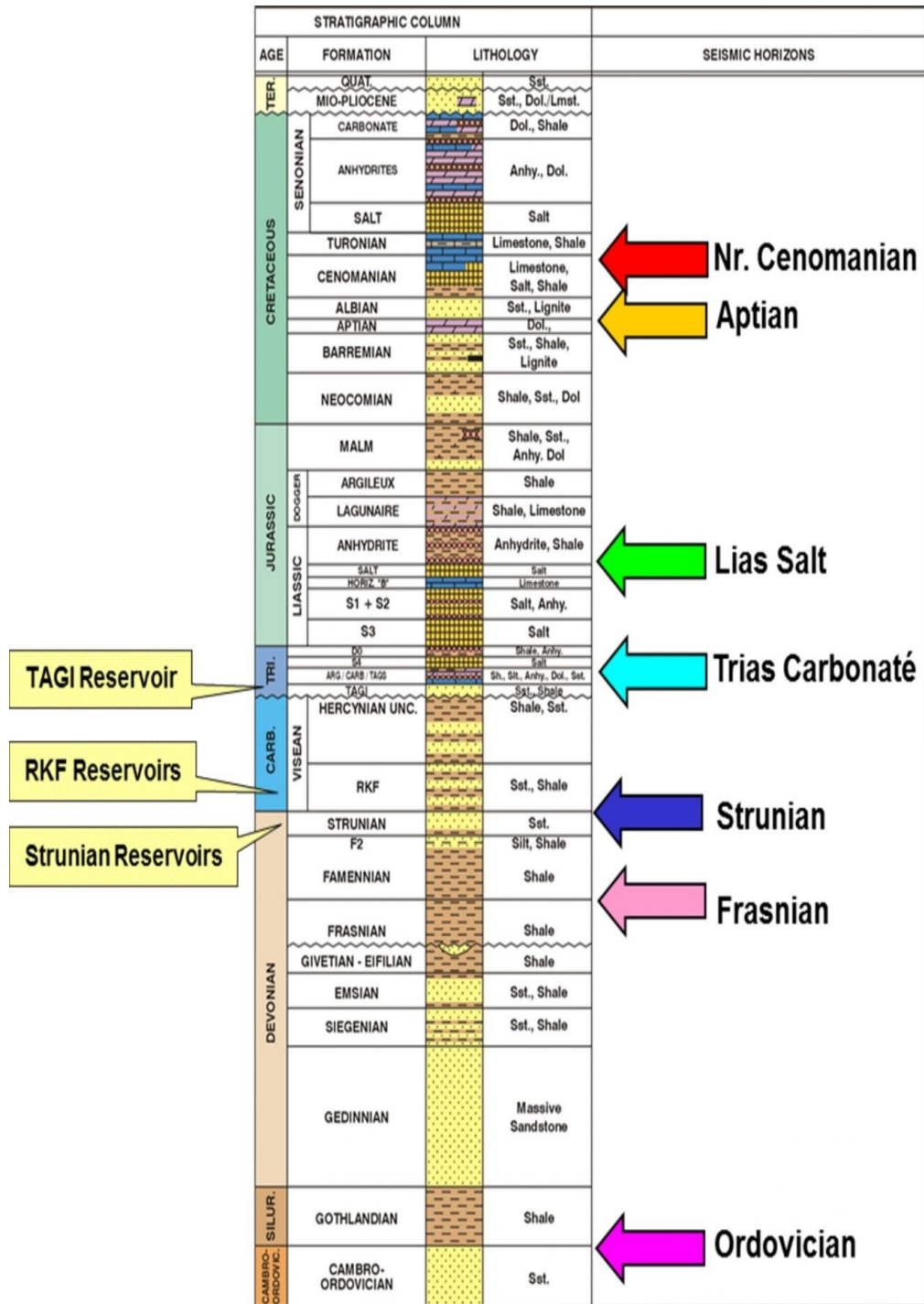


Figure I.3 Stratigraphic Cross-Section of El Merk field

I.5.1 TAGI Reservoir (Trias Argilo-Gréseux Inférieur)

The TAGI reservoir constitutes the primary oil-producing formation across all sub-fields of El Merk.

a) Geological origin

Predominantly fluvial depositional environment, leading to sandstone-dominated facies

b) Fluid type

Mainly crude oil, with limited quantities of gas condensate

c) Production strategy

Oil is exported, the produced gas is processed, and dry gas is re-injected for pressure maintenance

The TAGI reservoir is subdivided into three main units:

- 1) Upper TAGI (U)
- 2) Middle TAGI (M)
- 3) Lower TAGI (L)

These units are further divided into approximately 10 sub-units, reflecting vertical heterogeneity and variations in reservoir quality.

I.5.2 RKF Reservoir

The RKF reservoir is primarily a gas-condensate producing formation, present across all fields. It is subdivided into 7 sub-units, where certain sub-units, such as RKFB and Tournaisian shale, are considered non-reservoir zones due to their shale composition. This reservoir contributes significantly to gas and condensate production within the field.

I.5.3 Strunian Reservoir

The Strunian reservoir is also a gas-condensate producing formation, present in all sub-fields except EKT. It is subdivided into 7 sub-units, where the Strunian H unit is identified as a non-reservoir shale layer. (Groupement Berkine, 2025; Saoud, 2013)

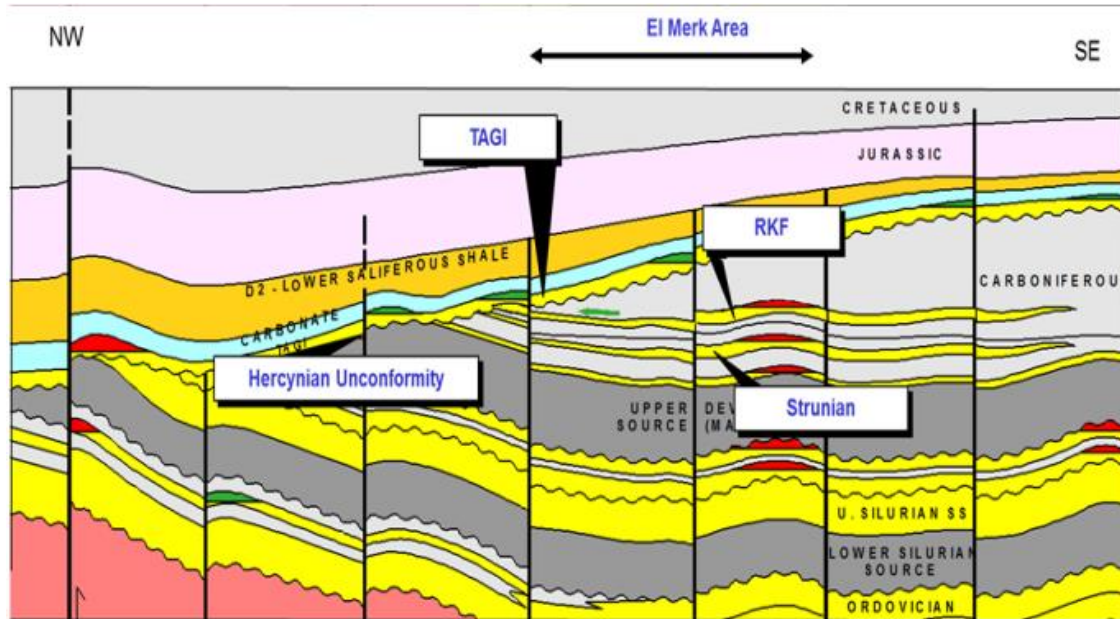


Figure I.4 Distribution of the three reservoirs in the stratigraphic cross-section

I.5.4 Petrophysical properties

a) TAGI Reservoir

- Clay volume cut-off: 0.40 v/v
- Average thickness: 37m
- Average porosity: 16%
- Average permeability: 385 mD

b) RKF and Strunian Reservoirs

- Clay volume cut-off: 0.40 v/v
- Average thickness: 18m
- Average porosity: 10%
- Average permeability: 210 mD

I.6 The subdivision of El Merk field

The El Merk oil field is subdivided into four main sectors: EMK, EMN, EKT, and EME as shown in **Figure I.5**.

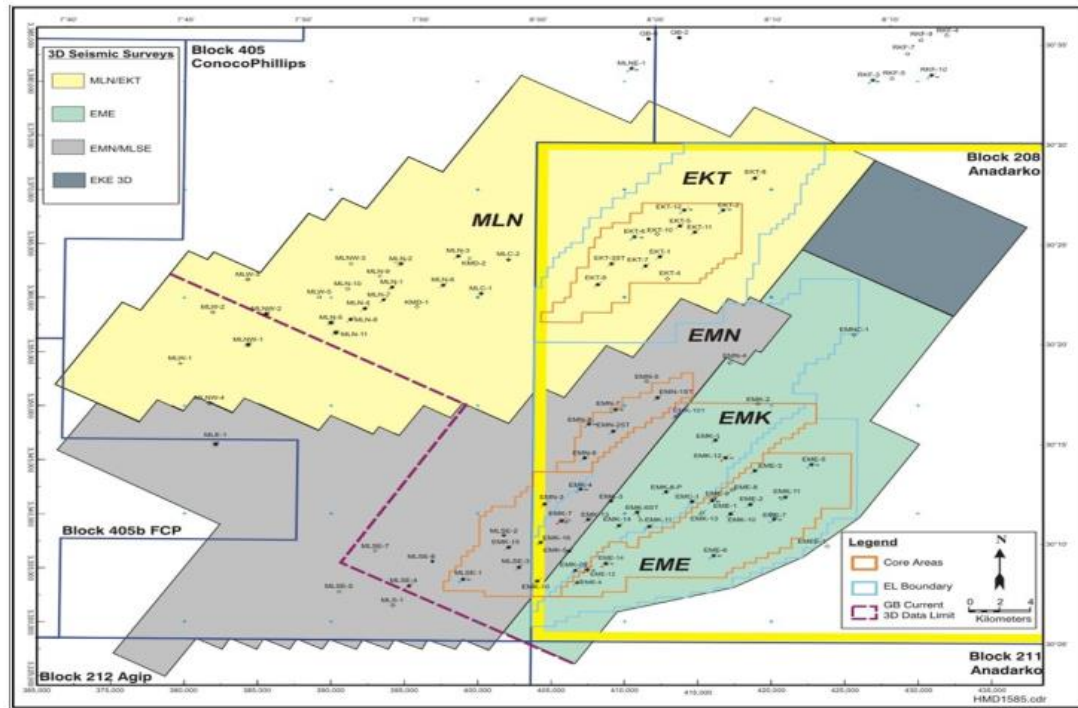


Figure I.5 Subdivision map of El Merk field

The field development includes the following well inventory:

- 38 oil production wells
- 10 gas-condensate production wells
- 18 gas injection wells
- 23 water injection wells
- 9 water production wells (Groupement Berkine, 2025)

Conclusion

This chapter presented an overview of El Merk field within the Berkine Basin, covering its location, geological framework, reservoir characteristics, and development strategy. The field is

characterized by a multi-reservoir system (TAGI, RKF, Strunian) with generally good petrophysical properties.

The well configuration and use of injection processes reflect an integrated approach to reservoir management. These characteristics provide a solid basis for the analysis and optimization work presented in the following chapters.

Chapter II: Waterflooding Engineering

Introduction

In most oil reservoirs, the natural energy initially present is insufficient to extract a significant portion of the original oil in place (OOIP) during the first stage of production. That is due to the fact that primary recovery mechanisms are relatively inefficient and result in low overall oil recovery, as a large fraction of oil remains trapped in the reservoir once the natural reservoir energy declines and becomes insufficient to sustain production.

To overcome these limitations and enhance recovery, secondary recovery techniques are widely implemented. Among these techniques, waterflooding is the most commonly applied method due to its operational simplicity, relatively low cost, and effectiveness in maintaining reservoir pressure while displacing oil toward production wells (Ahmed, 2019; Lake et al., 2014). In a waterflooding process, water is typically injected into the reservoir through injection wells to provide pressure support and improve the displacement of oil toward producing wells.

Despite its widespread adoption, the success of the waterflooding technique highly depends on reservoir rock properties such as heterogeneity, permeability distribution, wettability, and fluid properties. Reservoir heterogeneity often leads to uneven fluid displacement, causing the injected water to preferentially flow through high-permeability zones or fractures. This phenomenon may result in early water breakthrough and poor sweep efficiency, leaving large portions of the reservoir unswept and reducing the overall recovery factor (Green & Willhite, 2018; Lake et al., 2014).

To address these challenges, various conformance control techniques have been developed. Conformance control refers to a set of engineering strategies aimed at improving the distribution of injected fluids within the reservoir by reducing excessive water flow through high-permeability channels and redirecting it toward poorly swept regions of the reservoir. These techniques may involve chemical treatments such as polymer gels and relative permeability modifiers, as well as mechanical methods designed to regulate fluid flow paths and improve sweep efficiency (Seright, 2017; Sydansk & Romero-Zerón, 2011).

This chapter presents the fundamental principles of waterflooding and conformance control. It discusses the mechanisms governing water displacement in porous media, the factors affecting sweep efficiency, and the main operational challenges encountered during waterflooding projects.

II.1 Waterflooding overview

Waterflooding is the most widely used secondary oil recovery technique in petroleum reservoirs. The method involves injecting water into the reservoir through injection wells to maintain reservoir pressure and displace oil toward production wells. The primary objective of waterflooding is to improve the recovery of hydrocarbons that cannot be produced during the primary recovery stage due to insufficient reservoir energy (Ahmed, 2019; Green & Willhite, 2018).

During primary recovery, oil production is driven mainly by natural reservoir energy mechanisms such as solution gas drive, gas cap expansion, water drive, or rock and fluid expansion. However, these natural mechanisms typically recover only a limited fraction of the original oil in place (OOIP), often between 5% and 20%, depending on reservoir conditions. As reservoir pressure declines, the driving force required to move hydrocarbons toward production wells becomes insufficient. At this stage, secondary recovery methods such as water injection are implemented to provide artificial pressure support and enhance oil displacement (see **Figure II.1**). (Lake, Johns, Rossen, & Pope, 2014).

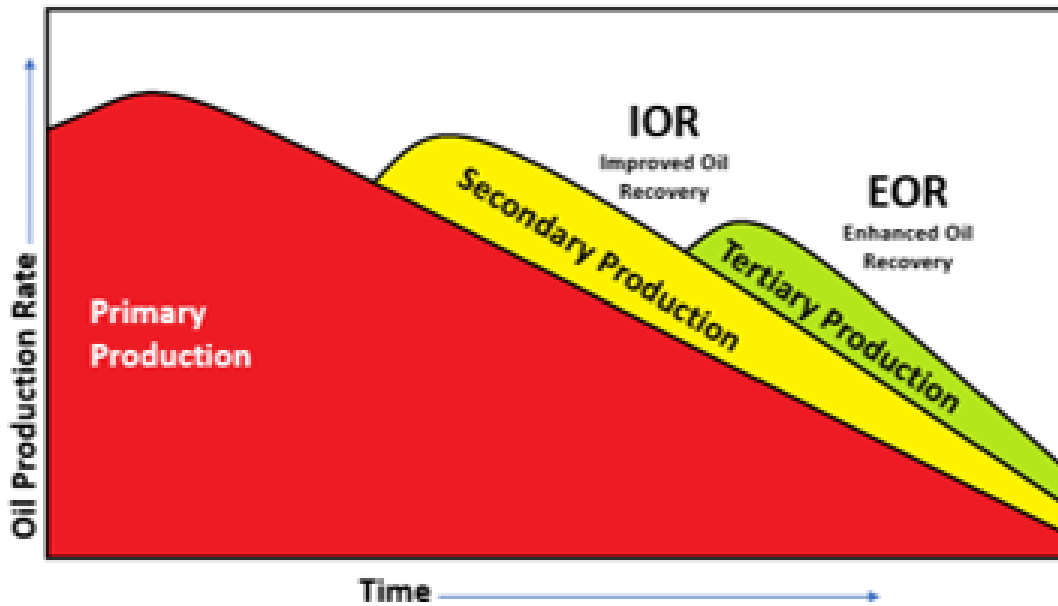


Figure II.1 Oil recovery mechanisms

II.2 Criteria for Selecting Reservoirs for Waterflooding

In determining whether a reservoir is a suitable candidate for waterflooding, several key characteristics must be evaluated (Thomas, Mahoney & Winter, 1989).

These parameters are outlined as follows:

- Reservoir geometry
- Fluid properties
- Reservoir depth
- Lithology and rock properties
- Fluid saturations
- Reservoir uniformity and pay continuity
- Primary reservoir driving mechanisms

Each one of these reservoir characteristics plays a critical role in assessing the feasibility and expected performance of a waterflooding project.

II.2.1 Reservoir Geometry

The areal geometry of a reservoir plays a key role in determining the optimal placement of production and injection wells. Overall, reservoir geometry governs the strategies that can be applied for water injection and production.

Analyzing reservoir geometry, together with past reservoir performance, which can be evaluated through historical production data is essential for identifying the existence and intensity of a natural water drive. This evaluation helps determine whether artificial water injection is necessary. In cases where the reservoir exhibits a strong or active natural water drive, additional water injection may not be required. (Ahmed, 2019)

II.2.2 Fluid Properties

The physical characteristics of reservoir fluids strongly impact the suitability of a reservoir to be considered for waterflooding. Among these properties, crude oil viscosity (μ_o), which is considered the most critical factor affecting the success of a waterflooding operation.

Oil viscosity directly impacts oil mobility (λ_o), which describes the ability of the oil to flow through the porous medium. This, in turn, affects the mobility ratio (M), which is given by the Formula (2.1) as the ratio of the displacing fluid mobility to that of the displaced fluid, and it presents a key parameter controlling displacement efficiency during water injection. (Willhite, 1986)

$$M = \frac{\lambda_w}{\lambda_o} = \left(\frac{k_w}{k_o} \right) \left(\frac{\mu_o}{\mu_w} \right) \quad (2.1)$$

II.2.3 Reservoir Depth

Reservoir depth significantly affects both the technical feasibility and economic viability of secondary and tertiary recovery processes such as waterflooding. As depth increases, the allowable injection pressure also rises. However, deeper reservoirs involve higher lifting costs, which limit the economically acceptable water–oil ratio and may reduce the ultimate recovery while increasing overall operating expenses.

In contrast, shallow reservoirs impose limitations on injection pressure, as it must remain below the fracture pressure of the formation. Exceeding this limit can lead to the initiation or propagation of fractures, causing injected water to channel through preferential paths and bypass significant portions of oil in the reservoir.

In waterflooding operations, a critical pressure gradient of approximately 1 psi/ft represents the threshold beyond which fracturing may occur. To ensure safe operation and avoid formation damage, a lower operational pressure gradient is typically maintained, commonly around 0.75 psi/ft. This safety margin helps prevent pressure-induced fracturing and ensures a more uniform displacement of fluids within the reservoir. (Ahmed, 2019)

II.2.4 Lithology and Rock Properties

Lithology plays a critical role in determining the efficiency of waterflooding systems. Several rock properties directly influence the performance and success of waterflooding operations, including:

- Porosity
- Permeability
- Clay content
- Net reservoir thickness

In complex reservoirs, only a limited portion of the total porosity such as fracture porosity that may provide sufficient permeability to allow effective fluid flow. In such cases, water injection primarily affects the fracture network, while the matrix porosity remains largely unaffected.

The presence of clay minerals can also impact waterflooding performance. During water injection, certain clays may swell or disperse, potentially reducing pore space and permeability. However, the extent of this effect is not always well quantified.

Reservoirs that are characterized by low permeability or small net thickness present additional challenges. The attainment of the desired injection rate in such formations frequently necessitates elevated injection pressures, as demonstrated by Equation (2.2).

$$p_{inj} = \frac{i_w}{hk} \quad (2.2)$$

Moreover, reservoir heterogeneity significantly affects waterflooding efficiency. Structural features such as sealing faults and permeability barriers can limit pressure support and fluid distribution. Additionally, high-permeability streaks may alter flow patterns, leading to uneven displacement and early water breakthrough. (Thomas, Mahoney & Winter, 1989)

II.2.5 Fluid Saturations

Fluid saturation is a key parameter in evaluating the suitability of a reservoir for waterflooding operations. A high initial oil saturation is essential, as it ensures the availability of a sufficient volume of recoverable hydrocarbons, which is a primary condition for successful flooding.

Higher oil saturation at the start of injection also enhances oil mobility (λ_o), facilitating fluid flow and improving displacement efficiency. Consequently, reservoirs with elevated initial oil saturation generally exhibit better recovery performance during waterflooding processes.

II.2.6 Reservoir Uniformity and Pay Continuity

Reservoir uniformity is a key factor influencing the success of waterflooding operations. A relatively homogeneous reservoir promotes uniform fluid displacement, whereas heterogeneity can significantly reduce recovery efficiency. Several reservoir characteristics must therefore be carefully evaluated when assessing the feasibility of secondary recovery processes.

One major concern is the presence of high-permeability layers, commonly referred to as thief zones. These zones can lead to rapid water channeling and bypassing of oil, resulting in early water breakthrough and high water/oil ratios. If not properly identified and isolated, thief zones can severely compromise the economic viability of the flooding process.

In addition, highly permeable zones often experience lower depletion pressures, which further enhances water channeling effects. These zones typically contain lower residual oil saturation compared to tighter layers, meaning that water injection in such areas contributes less to overall oil recovery.

Another critical aspect is the areal continuity of the pay zone. For an effective waterflood, both injection and production wells must be connected within the same continuous reservoir body. In contrast, isolated lenses may be efficiently drained by individual wells but are not suitable for coordinated flooding operations.

Finally, discontinuities in the reservoir, such as those caused by depositional variations, fractures, or faulting, can create anisotropy and disrupt fluid flow patterns. Identifying and characterizing these features is essential for optimizing well placement and selecting an appropriate flooding pattern. (Willhite, 1986; Ahmed, 2019)

II.2.7 Primary Reservoir Driving Mechanisms

Primary recovery refers to the production of hydrocarbons using only the natural energy of the reservoir, without the application of external support such as water or gas injection. The efficiency of oil recovery during this stage depends strongly on the dominant reservoir drive mechanism. Therefore, identifying the primary driving mechanism is essential when evaluating a reservoir for potential secondary recovery methods, particularly waterflooding.

The main driving mechanisms (**Table II.1**) that provide natural energy for oil production are:

- Rock and fluid expansion
- Solution gas drive
- Gas cap drive
- Water drive
- Gravity drainage

➤ Combination drive

Each mechanism contributes differently to pressure maintenance and oil displacement, leading to varying recovery efficiencies.

a) Rock and fluid expansion (Undersaturated Reservoirs)

Undersaturated reservoirs are initially at pressures higher than the bubble point pressure of the oil. Under these conditions, no free gas is present in the reservoir, and the primary source of energy is the expansion of the reservoir rock, connate water, and crude oil. As production proceeds and pressure declines, this expansion provides limited energy for fluid flow.

This driving mechanism is characterized by a rapid decline in reservoir pressure, weak natural energy support, and limited oil recovery.

Due to their poor primary recovery efficiency, these reservoirs present strong opportunities for enhanced recovery techniques.

b) Solution Gas Drive Reservoirs

In solution gas drive reservoirs, the initial reservoir pressure is close to the bubble point pressure. As pressure drops below this value during production, dissolved gas is liberated from the oil and expands, providing the main driving force.

This driving mechanism is characterized by a rapid decline in reservoir pressure, an increasing gas–oil ratio (GOR) over time, and declining oil production rates.

These reservoirs are generally considered excellent candidates for waterflooding, as water injection acts as an artificial water-drive mechanism. In many cases, waterflooding can recover an additional volume of oil comparable to the primary recovery.

c) Gas Cap Drive Reservoirs

Gas cap drive reservoirs are characterized by the presence of a gas cap above the oil zone. As production proceeds, the gas cap expands and helps maintain reservoir pressure.

This driving mechanism is characterized by a slower pressure decline compared to solution gas drive, more stable production behavior, and improved recovery compared to depletion drive.

However, waterflooding in such reservoirs must be carefully controlled to avoid upward migration of oil into the gas cap and the formation of residual oil zones within the gas cap.

Waterflooding may be considered in these reservoirs under specific conditions such as poor vertical communication between the gas cap and oil zone, the presence of sealing faults or barriers, and controlled injection through selective well completions.

d) Water Drive Reservoirs

Water drive reservoirs are supported by an active aquifer that supplies water into the reservoir as pressure declines. This natural water influx helps maintain pressure and sustain production.

This driving mechanism is characterized by strong pressure support, slower production decline, and increasing water cut over time.

Although strong water-drive reservoirs are generally not ideal candidates for waterflooding, since natural water influx already provides pressure maintenance, water injection may still be applied in certain cases to increase production rates, improve areal sweep efficiency, and achieve better voidage replacement.

e) Gravity Drainage Mechanism

Gravity drainage occurs when density differences between fluids cause oil to move downward and gas or water to move upward within the reservoir.

This mechanism is particularly effective in thick reservoirs, reservoirs with good vertical permeability, and systems with strong gas cap or water support.

f) Combination Drive Reservoirs

In many reservoirs, more than one driving mechanism operates simultaneously. These are referred to as combination drive reservoirs.

For examples:

- Solution gas drive + water drive
- Gas cap drive + water drive

The overall recovery performance depends on the relative contribution of each mechanism. These reservoirs require careful analysis to design optimal recovery strategies. (Ahmed, 2019)

Table II.1 Primary driving mechanisms oil recovery range

Driving Mechanism	Oil Recovery Range, %
Rock and liquid expansion	3–7
Solution Gas-cap drive	20–35
Gas-cap drive	20–45
Water drive	35–75
Gravity drainage	<80
Combination drive	30–60

II.3 Waterflooding Patterns

One of the most important steps in designing a waterflooding project is the selection of an appropriate flooding pattern. The main objective is to ensure maximum contact between the injected fluid and the reservoir oil, thereby improving displacement efficiency and overall recovery.

This can be achieved either by converting existing production wells into injection wells or by drilling new injection wells. The selection of the most suitable pattern depends on several factors, including reservoir heterogeneity and directional permeability, the orientation of natural fractures, the availability of injection fluid (water or gas), the expected flood duration, targeted oil recovery, as well as well spacing, productivity, and injectivity. In practice, the chosen pattern is strongly influenced by the number and distribution of existing wells.

II.3.1 Irregular Injection Patterns

Irregular patterns arise when wells are not uniformly distributed due to surface constraints, subsurface conditions, or drilling limitations. In such cases, each injection well may influence a different area, resulting in non-uniform flow patterns that are difficult to predict.

Irregular patterns are often observed in small or marginal fields, in reservoirs affected by faults or strong heterogeneity, and in situations where only a few production wells are converted into injection wells.

II.3.2 Peripheral Injection Patterns

In peripheral flooding, injection wells are located along the boundaries of the reservoir, pushing oil toward centrally located production wells (**Figure II.2**).

This injection pattern is characterized by high potential oil recovery with relatively low water production, a delayed water breakthrough occurring at late production stages, and the requirement of relatively few injection wells.

However, this pattern has several limitations, including a slow reservoir response due to long displacement distances, the need for high permeability to allow effective fluid movement over large distances, and injection rate limitations caused by pressure losses. In addition, its performance is difficult to predict, and there is a risk of bypassing oil near central producers.

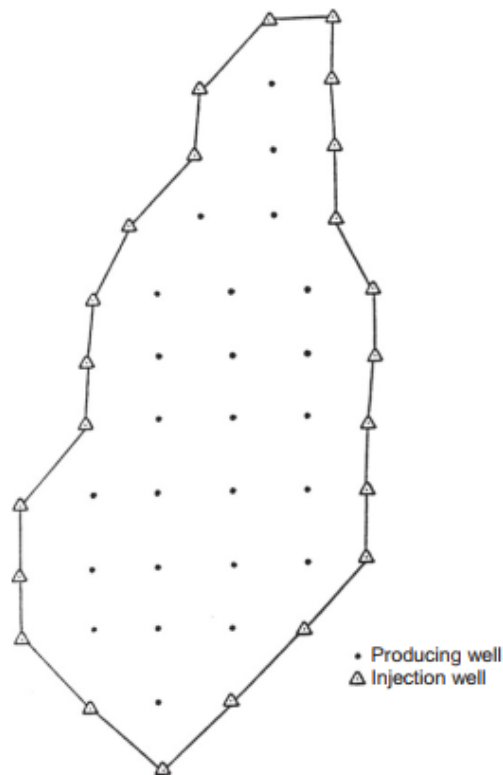


Figure II.2 Peripheral injection pattern

II.3.3 Regular Injection Patterns

Regular patterns are the most commonly used, particularly in fields developed on a grid system. These patterns provide better control of fluid movement and improved sweep efficiency.

Figure II.3 illustrates typical configurations.

- a) **Direct Line Drive:** Injectors and producers are aligned in parallel lines facing each other.
- b) **Staggered Line Drive:** Similar to direct line drive, but injectors are laterally offset to improve sweep.
- c) **Five-Spot Pattern:** Four injection wells form a square with a production well at the center.
- d) **Seven-Spot Pattern:** Six injectors arranged in a hexagon around a central producer.
- e) **Nine-Spot Pattern:** Eight injectors surrounding a central production well, providing dense injection coverage. (Willhite, 1986)

II.4 Overall Recovery Efficiency of Waterflooding

The overall recovery factor (RF) of any secondary or tertiary recovery process is represented by Equation (2.3), where it is defined as the product of three fundamental efficiency terms:

$$RF = E_D \times E_A \times E_V \quad (2.3)$$

In terms of cumulative production, this relationship can be expressed by the following equation:

$$N_p = N_s \times E_D \times E_A \times E_V \quad (2.4)$$

Where:

RF: Overall recovery factor

N_s : Initial oil in place at the start of the flood (STB)

N_p : Cumulative oil produced (STB)

E_D : Displacement efficiency

E_A : Areal sweep efficiency

E_V : Vertical sweep efficiency

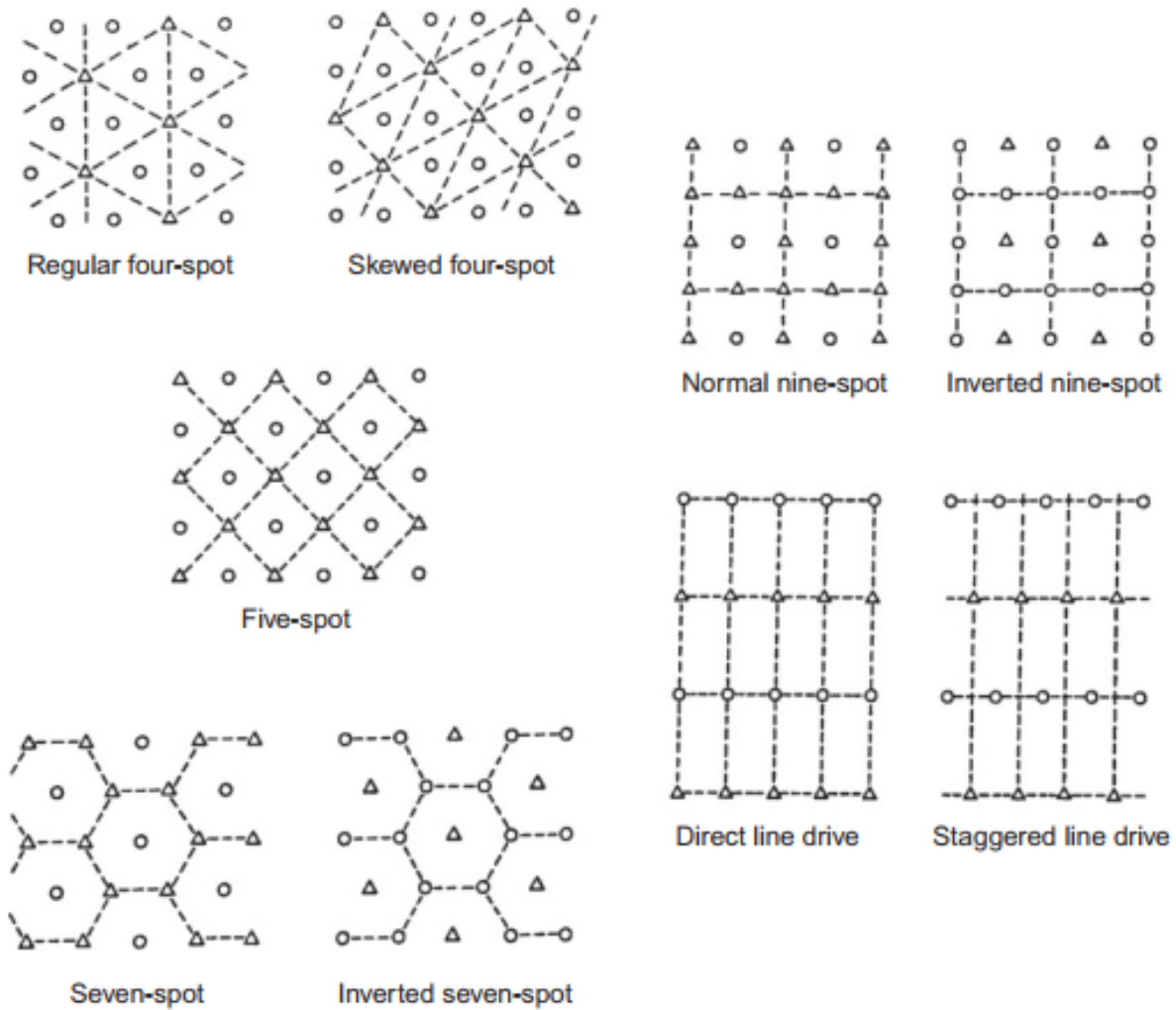


Figure II.3 Regular Injection Patterns

II.5 Displacement Efficiency

Displacement efficiency (E_D) represented by Equation (2.5) is defined as the fraction of movable oil that is recovered from the swept zone at a given time during a displacement process such as waterflooding. (Ahmed, 2019)

Mathematically, displacement efficiency can be expressed as:

$$E_D = \frac{\text{Volume of oil at start of flood} - \text{Remaining oil volume}}{\text{Volume of oil at start of flood}}$$

$$E_D = \frac{\overline{S_w} - S_{w_i}}{1 - S_{w_i}} \quad (2.5)$$

Where:

$\overline{S_w}$: Average water saturation in the swept area

S_{w_i} : Initial water saturation at the start of the flood

II.6 Frontal Displacement Theory

Frontal displacement theory is a theory that is developed by Buckley and Leverett and describes the movement of the injected water through a porous reservoir rock, where it displaces the resident oil in the form of a distinct advancing front separating the two fluids.

This classic theory consists of two equations:

- Fractional flow equation
- Frontal advance equation

II.6.1 Fractional flow equation

The development of the fractional flow equation is attributed to Buckley and Leverett, who established a fundamental analytical model to describe the displacement of one fluid by another in porous media. A schematic representation of the flood and its fractional flow curves is shown in **Figure II.4**. We start with Darcy's law, which is a key principle in fluid mechanics.

$$q_o = -\frac{kk_{ro}A}{\mu_o} \left(\frac{\partial P_o}{\partial x} + \rho_o g \sin \alpha \right)$$

$$q_o = -\frac{kk_{rw}A}{\mu_w} \left(\frac{\partial P_o}{\partial x} + \rho_w g \sin \alpha \right)$$

And we replace the water pressure by $P_w = P_o - P_{cow}$, so that

$$q_w = -\frac{kk_{rw}A}{\mu_o} \left(\frac{\partial(P_o - P_{cow})}{\partial x} + \rho_w g \sin \alpha \right)$$

After rearranging, the equation may be written as:

$$\begin{aligned} -q_o \frac{\mu_o}{kk_{ro}A} &= \frac{\partial P_o}{\partial x} + \rho_o g \sin \alpha \\ -q_w \frac{\mu_w}{kk_{rw}A} &= \frac{\partial P_o}{\partial x} - \frac{\partial P_{cow}}{\partial x} \rho_w g \sin \alpha \end{aligned}$$

By subtracting the first equation by the second one, we get:

$$-\frac{1}{kA} (q_w \frac{\mu_w}{k_{rw}} - q_o \frac{\mu_o}{k_{ro}}) = -\frac{\partial P_{cow}}{\partial x} + \Delta \rho g \sin \alpha$$

We substitute for $q = q_w + q_o$, and $f_w = q_w/q$, and solving for the fraction of water flowing, we obtain Equation (2.6) for the fractional of water flowing:

$$f_w = \frac{1 + \frac{kk_{ro}A}{q\mu_o} \left(\frac{\partial P_{cow}}{\partial x} - \Delta \rho g \sin \alpha \right)}{1 + \frac{k_{ro}}{\mu_o} \frac{\mu_w}{k_{rw}}} \quad (2.6)$$

For the simplest case of horizontal flow, with negligible capillary pressure, the Equation (2.6) reduces to Equation (2.7):

$$f_w = \frac{1}{1 + \frac{k_{ro}}{\mu_o} \frac{\mu_w}{k_{rw}}} \quad (2.7)$$

The above expression of the fractional flow indicates that during the displacement of oil by waterflooding, an increase in the water fractional flow (f_w) at any point in the reservoir will cause a proportional decrease in:

- Oil cut (f_o)
- Oil mobility

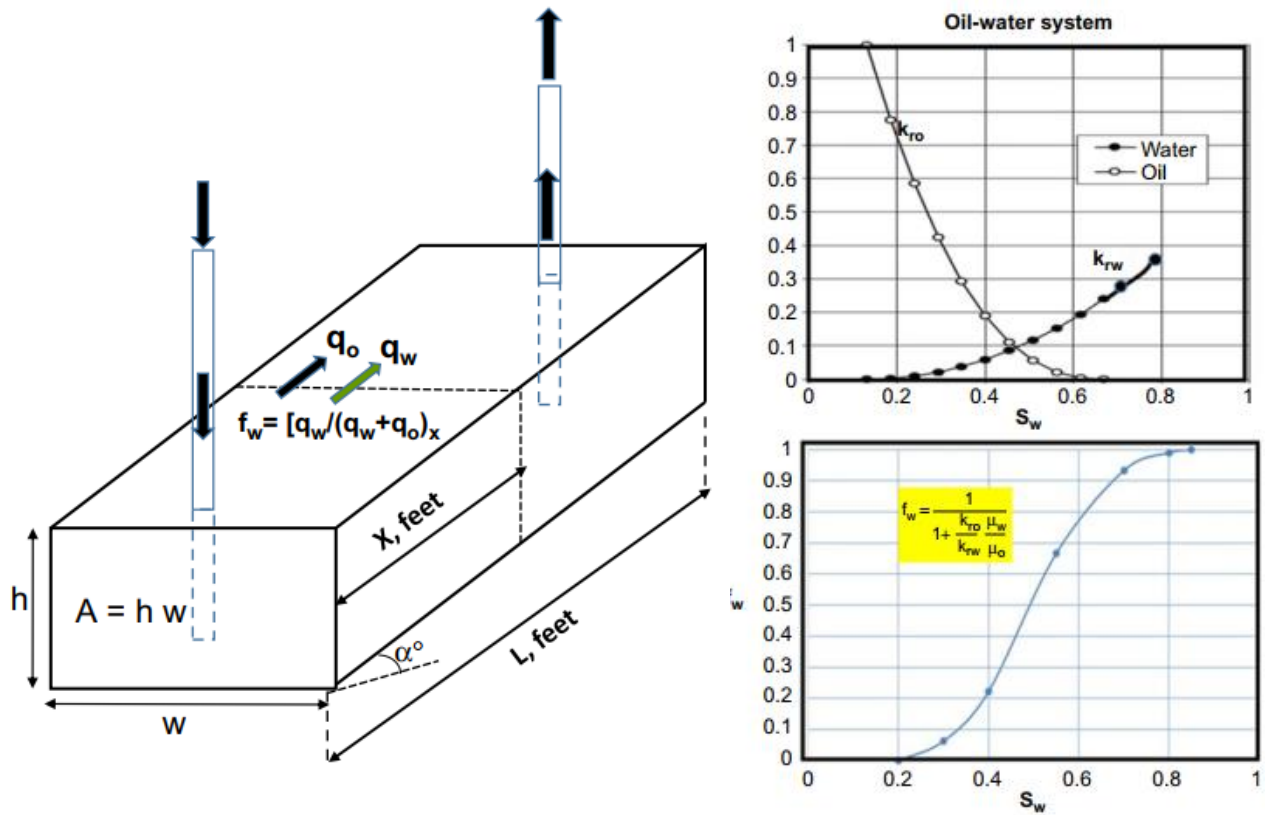


Figure II.4 Linear waterflood scheme and its fractional flow curves

Therefore, the objective is to select an appropriate injection scheme that can reduce the water fractional flow. This can be achieved by analyzing the effect of key parameters on water production behavior.

The overall effect on the fractional flow curve mainly depends on:

- Oil and water viscosities
- Water injection rate in relation to the formation dip angle (Ahmed, 2019)

II.6.1.1 Effect of Water and Oil Viscosities

The effect of fluid viscosity on the fractional flow curve is significant (see **Figure II.5**). Regardless of reservoir wettability (water-wet or oil-wet), it is observed that an increase in oil viscosity leads to an upward shift in the fractional flow curve, resulting in a higher water fractional flow (f_w).

On the other hand, the effect of water viscosity can be understood from the fractional flow equation. An increase in injected water viscosity leads to an increase in the denominator of the equation, which results in a reduction of f_w (i.e., a downward shift of the fractional flow curve).

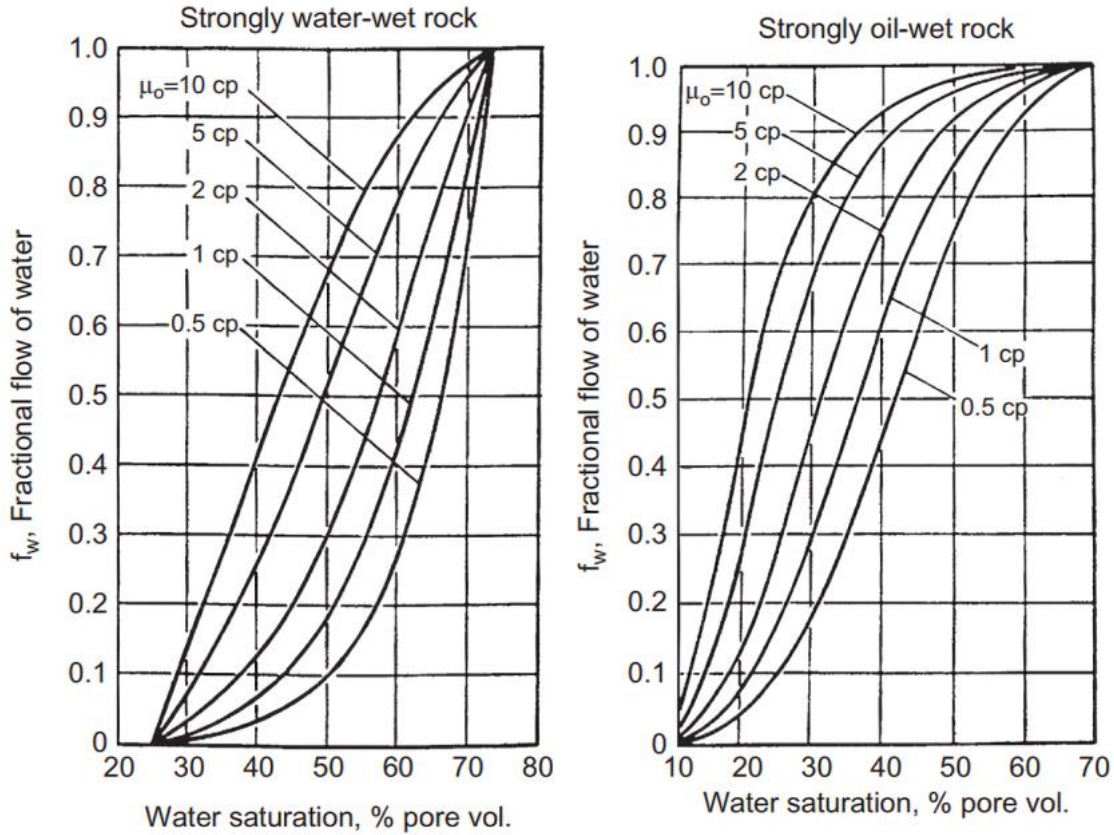


Figure II.5 Effect of oil viscosity on the fractional flow

II.6.1.2 Effect of Dip Angle and Injection Rate

To analyze the effect of the formation dip angle (α) and injection rate on displacement behavior (as seen in **Figure II.6**), the water fractional flow equation is considered. By assuming a constant injection rate and noting that the density difference ($\rho_w - \rho_o$) is always positive, Equation (2.6) can be simplified to Equation (2.8) as shown below to isolate the impact of these parameters on the water fractional flow (f_w).

$$f_w = \frac{1 - (X \frac{\sin(\alpha)}{i_w})}{1 + Y} \quad (2.8)$$

Two cases can be considered depending on the location of the injection well in a tilted reservoir:

a) Case 1: Injection Well Located Downdip

In this case, $\sin(\alpha)$ is positive, and the negative sign in the numerator of the above equation remains unchanged. When water is injected downdip and displaces oil updip, the term $X \sin(\alpha) / i_w$ remains positive.

This results in:

- A decrease in the numerator $[1 - (X \sin(\alpha) / i_w)]$
- A downward shift of the fractional flow curve (f_w)
- Improved displacement efficiency

Additionally, a lower injection rate (i_w) further decreases the numerator, which as a result enhances the downward shift of f_w and improves sweep performance.

b) Case 2: Injection Well Located Updip

In this case, $\sin(\alpha)$ is negative, which changes the sign in the numerator of the above equation to positive. When water is injected updip and displaces oil downdip, the term $X \sin(\alpha) / i_w$ becomes negative.

This results in:

- An increase in the numerator $[1 - (X \sin(\alpha) / i_w)]$
- An upward shift of the fractional flow curve (f_w)
- Reduced displacement efficiency

However, increasing the injection rate (i_w) helps reduce this effect and improves displacement performance in this configuration.

When the injection well is located updip and water displaces oil downdip, the fractional flow Equation (2.6) can be simplified by combining the term $X \sin(\alpha)$ into a constant C , we get Equation (2.9):

$$f_w = \frac{1 + \left(\frac{C}{i_w}\right)}{1 + Y} \quad (2.9)$$

This expression indicates that the water fractional flow (f_w) may exceed unity ($f_w > 1$) if:

$$\frac{C}{i_w} > Y$$

This condition typically occurs during downdip displacement at low injection rates (i_w).

Under this condition, a phenomenon known as counterflow may occur, where oil moves upward (updip), while water flows downward. This results in inefficient displacement and poor sweep performance. (Ahmed, 2019)

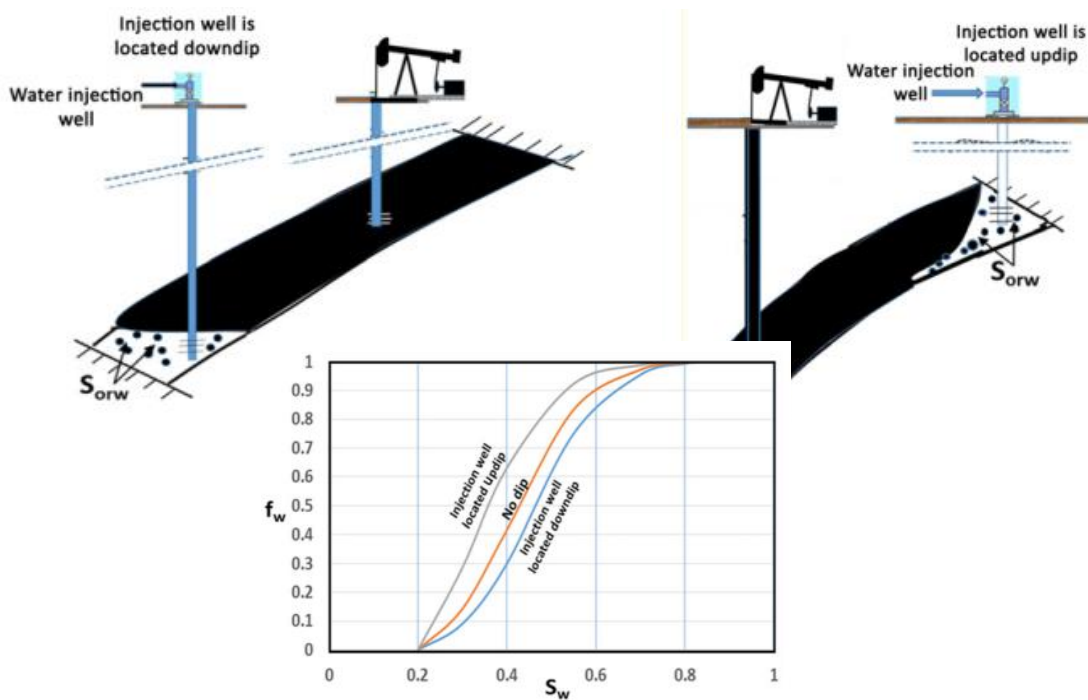


Figure II.6 Effect of injection well location on the fractional flow

II.6.2 Frontal Advance Equation

Buckley and Leverett (1942) developed the fundamental equation describing two-phase immiscible displacement in porous media. This equation is based on a material balance of the displacing fluid flowing through a differential element (**Figure II.7**): (Buckley & Leverett, 1942)

$$\text{Volume entering} - \text{Volume leaving} = \text{Accumulation}$$

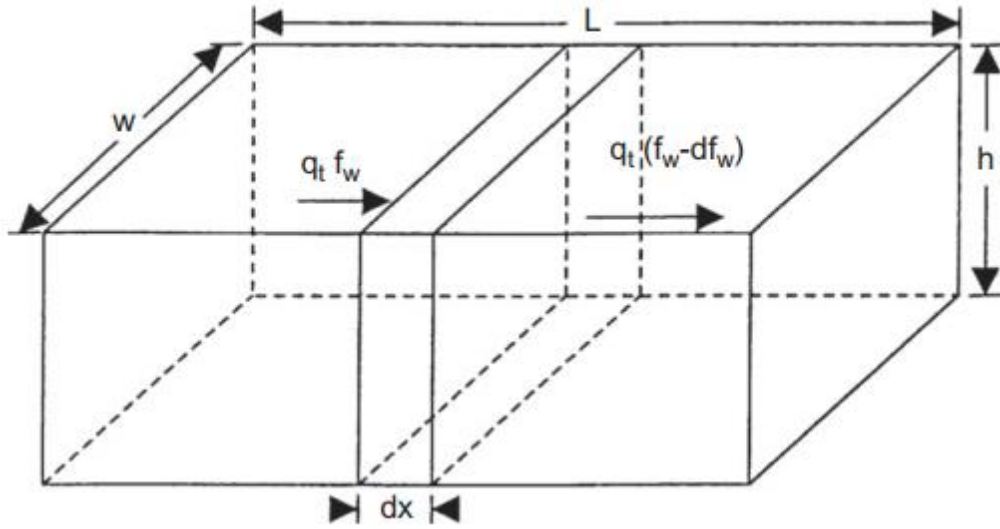


Figure II.7 Water flow through a linear differential element

Consider a differential element of length dx , cross-sectional area A , and porosity ϕ . Over a time interval dt :

Volume of water entering: $q_t f_w dt$

Volume of water leaving: $q_t (f_w - df_w) dt$

The accumulation of water within the element is obtained by subtracting the outlet from the inlet:

$$q_t f_w dt - q_t (f_w - df_w) dt = \frac{A\phi(dx)(dS_w)}{5.615}$$

By simplifying and separating variables we obtain Equation (2.10):

$$\left(\frac{dx}{dt}\right)_{S_w} = (v)_{S_w} = \left(\frac{5.615 q_t}{\phi A}\right) \left(\frac{df_w}{dS_w}\right)_{S_w} \quad (2.10)$$

Where:

$(v)_{S_w}$: Velocity of any specified value of S_w , ft/day

A : Cross-sectional area, ft^2

q_t : Total flow rate (oil + water), bbl/day

$(df_w/dS_w)_{S_w}$: Slope of the f_w vs. S_w curve at S

$$\int_0^x dx = \left(\frac{5.615 i_w}{\phi A} \right) \left(\frac{df_w}{dS_w} \right) \int_0^t dt$$

$$(x)_{S_w} = \left(\frac{5.615 i_w t}{\phi A} \right) \left(\frac{df_w}{dS_w} \right)$$

The fractional flow curve with its saturation-derivative are presented in **Figure II.8**.

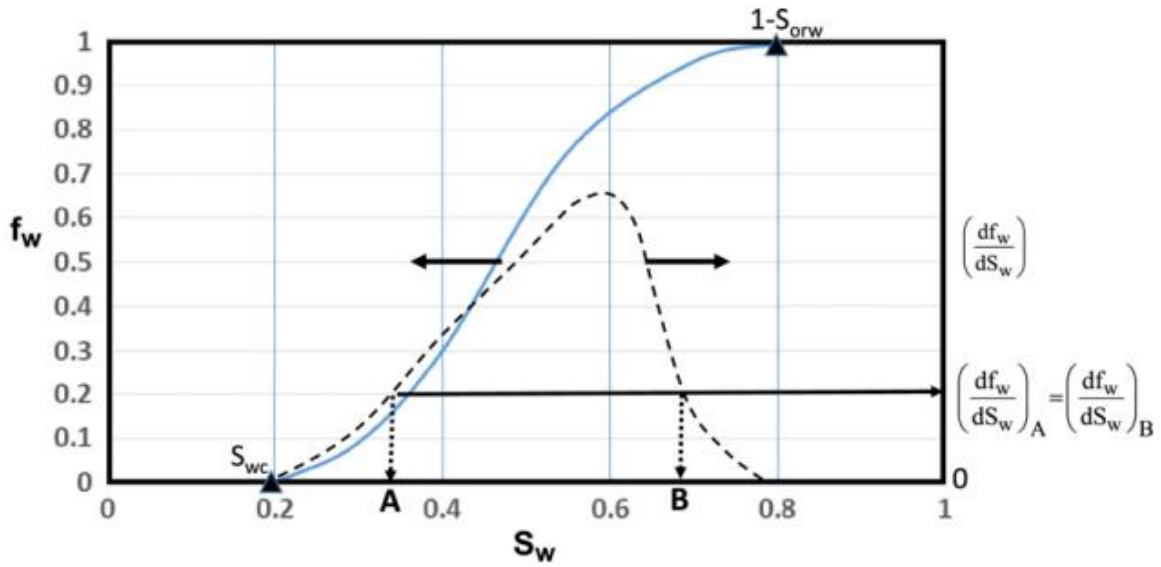


Figure II.8 Fractional flow curve with its saturation-derivative

However, **Figure II.8** shows that the derivatives at points A and B are identical. This implies that multiple water saturations could exist in the same location, which is physically impossible. Buckley and Leverett explained that this inconsistency arises from neglecting the capillary pressure gradient term in the fractional flow equation, which is given by:

$$\text{Capillary term} = \left(\frac{0.001127 k_o A}{\mu_o i_w} \right) \left(\frac{dP_c}{dx} \right)$$

When this capillary term is included, the fractional flow curve is modified and can be represented by two distinct segments (**Figure II.9**):

- A straight-line segment with constant slope $(df_w/dS_w)_{S_{wf}}$ from S_{wc} to S_{wf}

- A concave curve with decreasing slope, extending from S_{wf} to $(1 - S_{or})$

Terwilliger and his co-authors proposed that all water saturation values between S_{wc} and S_{wf} move at the same velocity as a function of time (t) and distance (x). As a result, the saturation profile within this interval maintains a constant shape during displacement forming a shock front. (Terwilliger et al., 1951)

In contrast, the saturation interval between S_{wf} and $(1 - S_{or})$ is termed the non-stabilized zone, which represents the saturation behind the front, where its velocities are variable, and the its profile changes shape with time as it is illustrated in **Figure II.10**.

In order to find the front saturation Welge proposed to achieve that by drawing a tangent from S_{wc} to the fractional flow curve, and the point of tangency defines the front saturation (S_{wf}). The corresponding ordinate gives the water cut at the front (f_{wf}). (Welge, 1952)

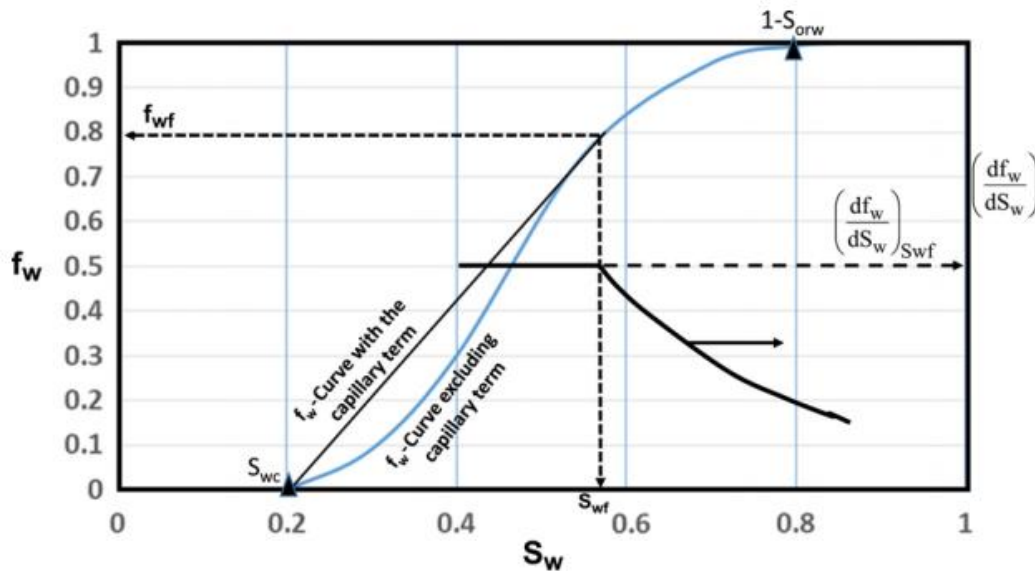


Figure II.9 Effect of the capillary pressure on the shape of the fractional flow curve

II.7 Areal Sweep Efficiency

The areal sweep efficiency (E_A) is defined as the fraction of the total flood pattern area that is contacted by the displacing fluid. It increases progressively from zero at the start of injection until breakthrough, after which it continues to increase at a slower rate.

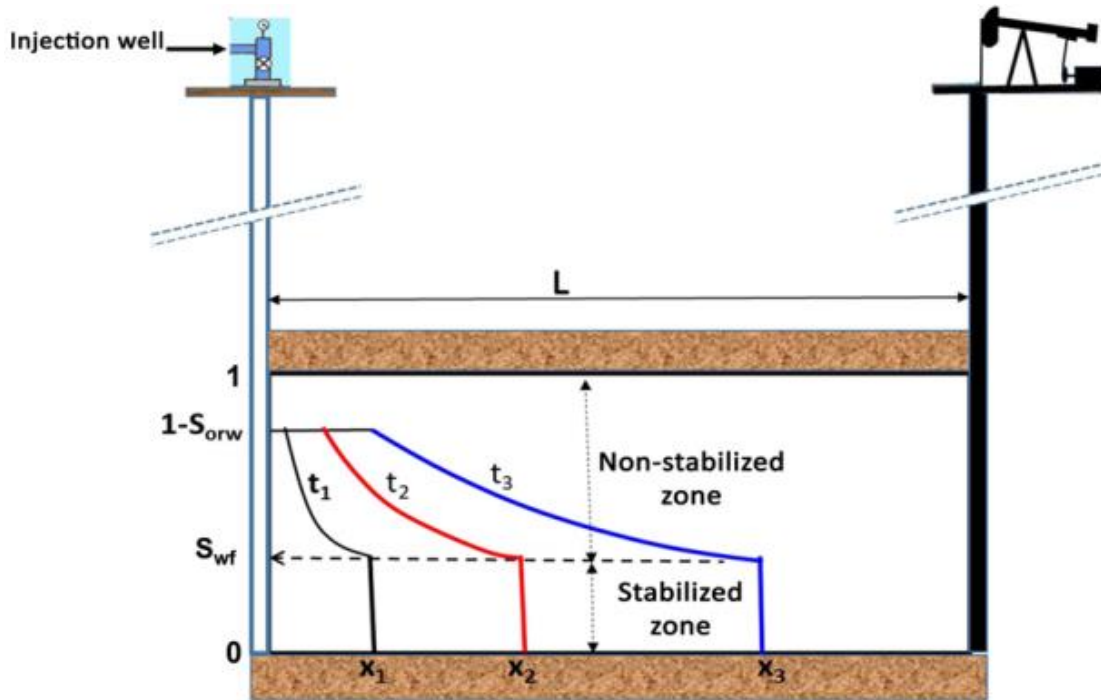


Figure II.10 Water saturation profile as a function of time t and distance x

The areal sweep efficiency mainly depends on the following factors:

- Mobility ratio (M)
- Flood pattern configuration
- Cumulative water injected (W_{inj})
- Pressure distribution between injectors and producers
- Directional permeability

If directional permeability trends are identified, well placement can be optimized to improve sweep efficiency. In addition, proper management of pressure distribution and careful selection of injection–production patterns can significantly enhance the overall areal sweep.

II.7.1 Mobility Ratio “ M ”

Muskat (1949) emphasized that correctly evaluating the mobility ratio (M) requires the careful selection of relative permeability values, as illustrated in **Figure II.11**. The following considerations must be applied:

- a) **Relative permeability of oil (k_{ro}):**

Since oil flows ahead of the water front in the non-invaded zone, k_{ro} should be evaluated at the initial water saturation (S_{wi}).

b) Relative permeability of water (k_{rw}):

The injected water forms a water bank characterized by an average water saturation (S_{wBT}) in the swept region. This saturation remains approximately constant until breakthrough, after which it increases (e.g., to S_{w2}). (Ahmed, 2019)

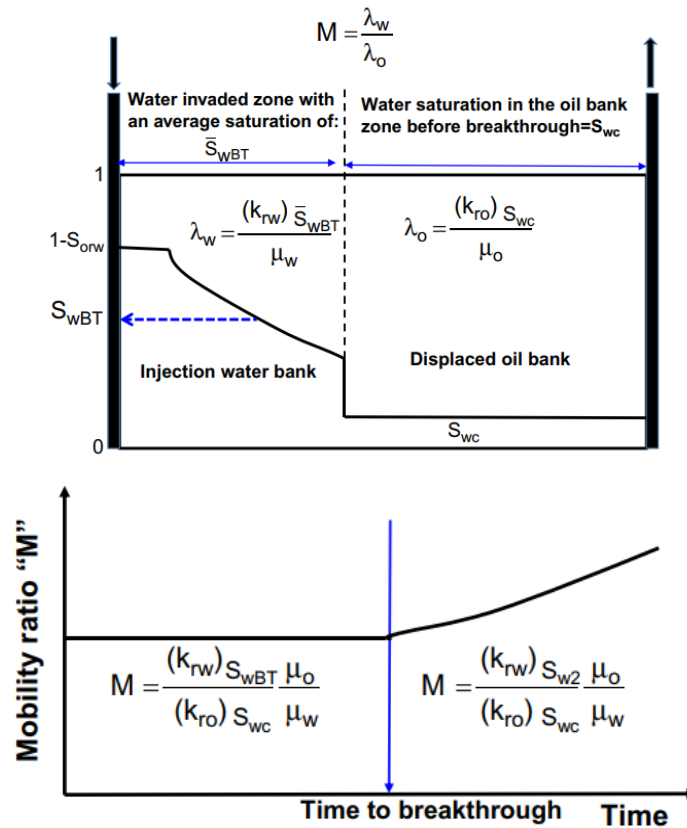


Figure II.11 Mobility ratio variation

In waterflood design, injection and production wells are commonly arranged in regular geometric patterns to ensure a symmetrical and interconnected flow system. The most common patterns include the direct line drive, staggered line drive, five-spot, seven-spot, and nine-spot configurations.

Among these configurations, the five-spot pattern is the most widely used and is generally considered as the reference pattern in waterflood studies.

Craig and his co-authors conducted experimental investigations using laboratory models representing a quadrant of a five-spot pattern to study the influence of fluid mobility on areal sweep efficiency as shown in **Figure II.12**. (Craig, 1971)

Their observations indicate that at the beginning of injection, the water front has a cylindrical shape around the injection well. As injection continues, pressure gradients and streamlines develop between injection and production wells. The shortest streamline, which directly connects the injector to the producer, exhibits the highest velocity, resulting in faster fluid movement along this path. As a result, the displacement front progressively deforms and develops a cusp toward the production well, leading to breakthrough.

At breakthrough, only a portion of the reservoir is swept. This fraction is defined as the areal sweep efficiency at breakthrough, denoted by E_{ABT} . For a mobility ratio of 1.43, E_{ABT} is approximately 65 %, whereas for a mobility ratio of 0.4, E_{ABT} is approximately 82.8 %.

These results demonstrate that lower mobility ratios improve areal sweep efficiency, while higher mobility ratios reduce sweep efficiency and accelerate breakthrough. After breakthrough, the areal sweep efficiency continues to increase gradually with continued injection until the reservoir is fully contacted.

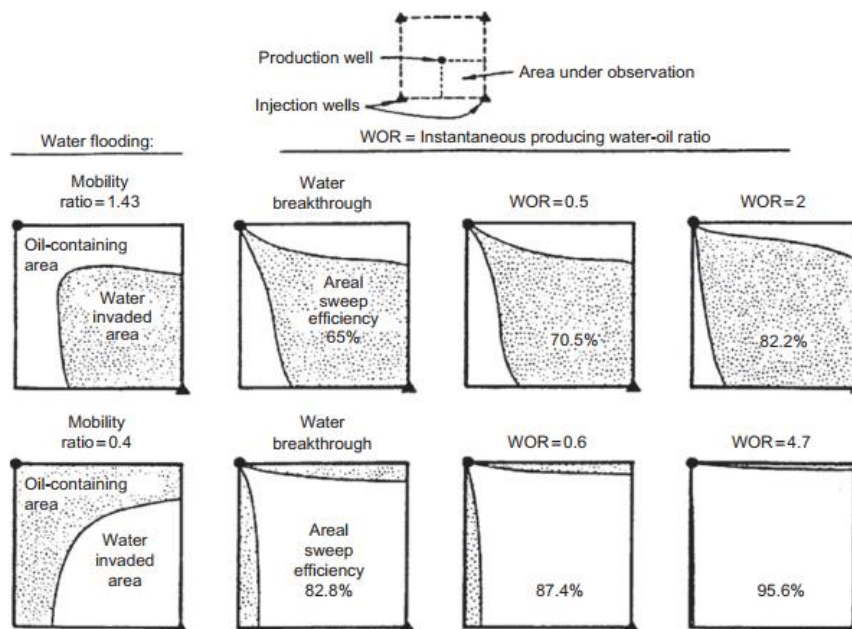


Figure II.12 X-ray shadowgraphs of waterflood progress

II.7.2 Cumulative Water Injected

Continued water injection after breakthrough can lead to a significant increase in oil recovery, particularly when the mobility ratio is unfavorable. Craig et al. demonstrated that important amounts of oil can still be displaced by water after breakthrough occurs.

This behavior becomes more pronounced as the mobility ratio increases, meaning that post-breakthrough production plays a more critical role in overall recovery efficiency for adverse mobility conditions. (Craig, 1971)

II.8 Vertical Sweep Efficiency

The vertical sweep efficiency (E_v) is defined as the fraction of the vertical section of the pay zone that is contacted by the injected fluid. It primarily depends on:

- The mobility ratio
- The total volume of injected fluid

Due to vertical heterogeneity in permeability, the injected fluid does not advance uniformly. Instead, it tends to flow preferentially through high-permeability layers, where it moves faster, while low-permeability zones remain poorly swept (**Figure II.13**). This results in an irregular displacement front and reduced sweep efficiency.

One of the major challenges in waterflood design is the accurate characterization of permeability variations, both vertically and areally. Among all factors, vertical heterogeneity is considered the most influential parameter affecting vertical sweep efficiency.

To properly evaluate E_v , the following aspects must be addressed:

- 1) Defining vertical permeability variations using appropriate mathematical models
- 2) Determining the minimum number of layers required to represent reservoir behavior accurately
- 3) Assigning representative average rock properties to each layer after upscaling (zonation process)

II.8.1 Calculation of Vertical Sweep Efficiency

Two classical methods are commonly used to estimate the vertical sweep efficiency (E_v): The Stiles method and the Dykstra–Parsons method.

Both approaches are based on representing the reservoir as an idealized layered system, where the layers are arranged in descending order of permeability. These methods rely on several simplifying assumptions, including the absence of cross-flow between layers, immiscible displacement, and linear flow conditions. In addition, they assume that the distance traveled by the injected water in each layer is proportional to its permeability and that the displacement process follows a piston-like behavior.

The fundamental concept behind both methods is to track the advance of the displacement front within each layer and determine its position at the time of breakthrough in successive layers. By considering the flow capacity of each layer, defined as the product of permeability and thickness ($k \cdot h$), it becomes possible to estimate the contribution of each layer to fluid production and to calculate the resulting water–oil ratio.

These methods provide a practical framework for evaluating the impact of vertical heterogeneity on sweep efficiency and overall reservoir performance. (Ahmed, 2019)

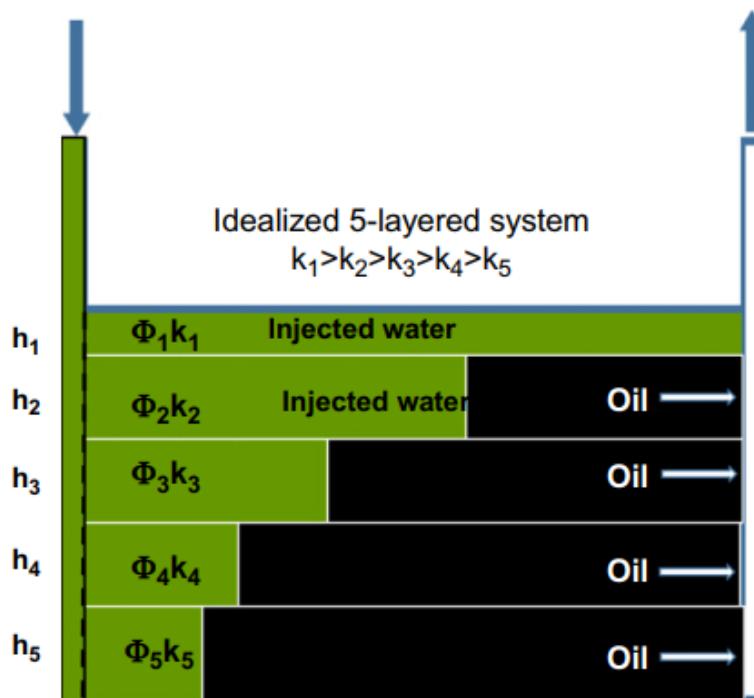


Figure II.13 Idealized pattern layered system

II.9 Waterflood Diagnostic Plots and Surveillance Methods

A successful waterflooding project relies on a well-designed and properly executed surveillance and monitoring program. Production data and performance curves are essential tools for evaluating reservoir behavior and detecting changes in well performance. Reliable interpretation of these curves requires accurate and regular well testing, since field production is allocated to individual wells based on these measurements.

According to Satter and Thakur (1998), an effective waterflood surveillance program requires accurate monitoring and recording of injector and producer performance data. This includes injection and production rates, bottom-hole flowing pressures and injection pressures, as well as oil cut, water–oil ratio (WOR), and gas–oil ratio (GOR).

In addition, actual reservoir performance should be periodically compared with predicted performance. Key parameters used in this comparison include cumulative oil and water production, oil and water production rates, injection rates, GOR and WOR trends, and water breakthrough time.

The surveillance program should also involve the evaluation of reservoir and waterflood performance indicators, such as volumetric sweep efficiency (E_{vol}), oil recovery factor at different stages of the flood, pore volumes injected (PVI), and the Voidage Replacement Ratio (VRR).

Furthermore, continuous monitoring of reservoir and operational conditions is essential. This includes assessing surface facility performance, updating reservoir characterization, monitoring injected water quality, and identifying potential flow assurance issues that may affect waterflood efficiency.

The main objective of a surveillance program is to maximize oil recovery while maintaining low water production, minimizing injected water volume, and reducing operational costs. This is achieved through effective control of the key waterflood efficiency parameters: areal sweep efficiency (E_A), vertical sweep efficiency (E_V), and displacement efficiency (E_D).

Several diagnostic and plotting techniques are commonly used to support performance analysis and optimize waterflood operations, including:

- Bubble maps
- Gas–oil ratio (GOR) plots

- Water–oil ratio (WOR) plots
- Hall plots
- X-plots
- Production decline curves
- ABC plots
- Heterogeneity index maps
- Pattern balancing analysis
- Tracer surveys
- Volumetric sweep evaluation
- Confinement factor analysis

These tools provide valuable insight into reservoir dynamics and help improve decision-making during the life of the waterflood project. (Ahmed, 2019)

Conclusion

Waterflooding remains one of the most effective and widely applied secondary recovery methods for enhancing oil production beyond the limitations of primary recovery. Its success depends on a thorough understanding of reservoir characteristics, including rock and fluid properties, saturation conditions, reservoir continuity, and the dominant drive mechanisms.

The efficiency of a waterflooding project is governed by the interplay of displacement efficiency, areal sweep efficiency, and vertical sweep efficiency, all of which are strongly influenced by reservoir heterogeneity and fluid mobility. Proper selection of injection patterns, control of mobility ratio, and optimization of operational parameters are essential to achieve favorable sweep and delay premature water breakthrough.

In addition, the application of frontal displacement theory and fractional flow analysis provides a fundamental framework for predicting fluid behavior during displacement, while surveillance and diagnostic tools play a critical role in monitoring performance and guiding field adjustments.

Ultimately, maximizing oil recovery through waterflooding requires an integrated approach that combines sound reservoir characterization, appropriate design strategies, and continuous

monitoring. When properly implemented, waterflooding can significantly improve recovery efficiency and extend the productive life of oil reservoirs.

Chapter III: Overview of Machine Learning Concepts in Petroleum Engineering

Introduction

In recent years, the oil and gas industry has increasingly adopted data-driven approaches to improve decision-making and optimize field performance. Among these approaches, machine learning has emerged as a powerful tool capable of extracting valuable insights from large and complex datasets generated during exploration, drilling, and production operations. Unlike traditional physics-based models, machine learning techniques can identify hidden patterns and complex relationships without requiring explicit formulation of governing equations.

In petroleum production and reservoir engineering, machine learning has shown strong potential in addressing challenges such as production forecasting, reservoir characterization, and fluids' flow modeling. By integrating historical field data with advanced algorithms, it becomes possible to enhance prediction accuracy, reduce uncertainty, and support real-time optimization of operations.

This chapter provides an overview of machine learning concepts and their relevance to petroleum engineering applications. It introduces key learning paradigms, commonly used algorithms, and highlights how these techniques can be applied to improve reservoir management and production performance.

III.1 Machine Learning Definition

Machine learning is a field of study in which computers are designed to learn from data and improve their performance on a specific task based on experience, rather than being explicitly programmed with fixed rules. The performance is evaluated using a defined metric, such as accuracy. For example, in handwritten digit recognition, the objective is to correctly classify digits, while the performance metric is the percentage of correctly identified images using a labeled dataset. The learning process relies on large volumes of data to capture underlying patterns and relationships.

With the rapid advancement in computational power and the development of more efficient algorithms, machine learning has significantly improved in recent years. It has been successfully applied in various domains, including image recognition, text analysis such as spam detection, sentiment analysis, information extraction, video games, and robotics. Its range of applications

continues to expand reaching domains such as the petroleum industry as new techniques and data sources become available.

III.2 Machine Learning Categories

Machine learning algorithms are generally classified into three main categories based on how they learn from data.

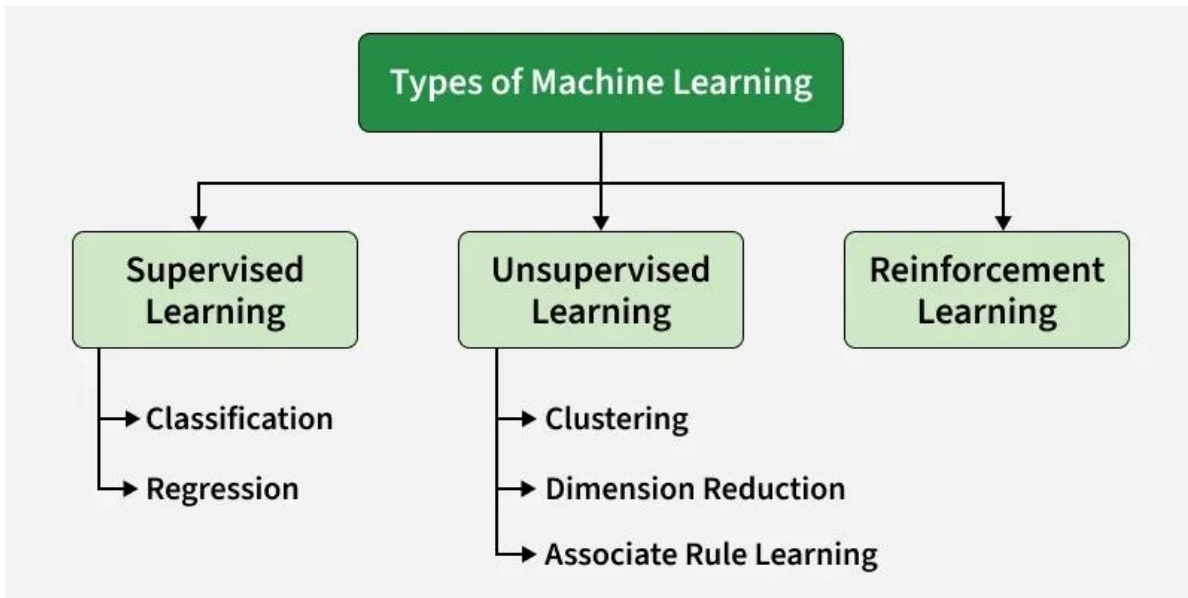


Figure III.1 Machine Learning Categories

III.2.1 Supervised Learning

Supervised learning relies on the availability of a known output variable. The algorithm learns the relationship between input features and the desired output. It includes regression, where the output is continuous, and classification, where the output consists of discrete classes. The model is trained by minimizing the error between predictions and actual values until acceptable accuracy is achieved.

III.2.1.1 Regression

Regression is a supervised learning technique used when the target variable is continuous or numerical. The objective of regression models is to establish a mathematical relationship between

input features and a continuous output variable. In petroleum engineering, regression algorithms are commonly used for predicting permeability, porosity, production rates, reservoir pressure, and enhanced oil recovery performance.

The following regression algorithms are commonly used in machine learning applications (Pandey et al., 2020):

a) Multiple Linear Regression

Multiple Linear Regression is one of the simplest regression techniques. It models the relationship between the dependent variable and multiple independent variables using a linear equation. Although simple and interpretable, linear regression may not accurately capture nonlinear relationships commonly encountered in petroleum reservoir data.

b) Decision Tree Regression

Decision Tree Regression uses a hierarchical tree structure to divide the dataset into smaller subsets based on decision rules. Predictions are generated from the average values of terminal nodes. Decision trees are highly interpretable and capable of modeling nonlinear relationships, but they may suffer from overfitting.

c) Random Forest Regression

Random Forest Regression is an ensemble learning method that combines multiple decision trees. Each tree is trained using random subsets of the data and input features. The final prediction is obtained by averaging the outputs of all trees. Random forests generally provide high prediction accuracy and improved generalization capability.

d) XGBoost Regression

Extreme Gradient Boosting (XGBoost) is a powerful boosting algorithm that sequentially trains decision trees to reduce prediction errors from previous models. XGBoost is known for its high computational efficiency, robustness, and strong predictive performance in complex engineering datasets.

III.2.1.2 Classification

Classification is a supervised learning technique used when the output variable is categorical or discrete. The objective is to assign observations into predefined classes based on input features. In

petroleum engineering, classification algorithms are applied in lithology identification, facies classification, drilling event prediction, and equipment fault detection.

Commonly used classification algorithms include:

a) Decision Tree Classifier

Decision Tree Classifier separates data into classes using hierarchical decision rules. The algorithm minimizes randomness and maximizes information gain during training, making it effective for interpretable classification tasks.

b) Random Forest Classifier

Random Forest Classifier combines multiple decision trees and predicts the final class using a voting mechanism. This approach improves classification accuracy and reduces overfitting compared to a single decision tree. (Hastie et al., 2001; Pandey et al. 2020)

III.2.2 Unsupervised Learning

Unsupervised learning does not use labeled outputs. Instead, it analyzes the input data to uncover hidden patterns or structures. It is commonly applied in clustering, dimensionality reduction, and association rule learning to better understand the data.

III.2.2.1 Clustering

Clustering is an unsupervised learning technique used to group data points with similar characteristics into distinct clusters. Observations within the same cluster are more similar to each other than to observations belonging to other clusters. Clustering methods are particularly useful in petroleum engineering for identifying reservoir zones, classifying lithofacies, analyzing production behavior, and detecting hidden geological patterns.

III.2.2.2 Dimensionality reduction

Dimensionality reduction is a key concept in data analysis used to handle datasets containing a large number of variables. Its main purpose is to transform high-dimensional data into a reduced set of meaningful features while preserving most of the important information. This process is particularly useful for improving the performance of machine learning models by enhancing computational efficiency and reducing complexity without significantly affecting predictive accuracy. Additionally, dimensionality reduction helps eliminate redundancy and multicollinearity

among input variables. One of the most widely used techniques in this category is Principal Component Analysis (PCA), which projects the original features into a lower-dimensional space while retaining maximum variance. (Pandey et al, 2020)

III.2.3 Deep Learning

Deep learning is a specialized branch of machine learning based on artificial neural networks with multiple hidden layers. Unlike traditional machine learning algorithms, deep learning models are capable of automatically learning complex and hierarchical representations from large datasets. These models are particularly effective in solving highly nonlinear problems and handling large volumes of structured and unstructured data.

Deep learning has gained significant importance in petroleum production and reservoir engineering due to the increasing availability of field data, sensor measurements, well logs, seismic data, and production histories. Deep neural networks can capture complex relationships between reservoir parameters and production behavior, making them highly suitable for prediction, classification, optimization, and pattern recognition tasks.

Some of the most widely used deep learning architectures include:

- Multilayer Perceptrons (MLP)
- Convolutional Neural Networks (CNN)
- Recurrent Neural Networks (RNN)
- Long Short-Term Memory Networks (LSTM)
- Multilayer Perceptron (MLP)

The Multilayer Perceptron (MLP) is one of the simplest and most fundamental deep learning architectures. It consists of interconnected artificial neurons organized into multiple layers. MLP models are capable of learning nonlinear relationships between inputs and outputs through the use of activation functions and iterative training processes.

Artificial neurons are the basic computational units of neural networks. Each neuron receives input values, multiplies them by corresponding weights, adds a bias term, and applies a nonlinear activation function to produce an output. The nonlinear transformations performed by neurons enable neural networks to model complex physical and engineering systems.

III.2.3.1 Components of an MLP Network

An MLP architecture as shown in **Figure III.2** generally consists of the following components:

1) Input Layer

The input layer receives the input features from the dataset and transfers them to the hidden layers for processing. Each node in the input layer typically represents one feature or variable from the dataset.

2) Hidden Layers

Hidden layers contain neurons responsible for learning complex relationships within the data. Each neuron performs mathematical operations using weights, biases, and activation functions. Increasing the number of hidden layers and neurons enables the network to learn more sophisticated patterns.

3) Output Layer

The output layer produces the final prediction of the neural network. The structure of the output layer depends on the problem type. For regression problems, the output layer generally contains a single neuron representing a continuous value, while classification problems may contain multiple neurons corresponding to different classes.

4) Optimizer

The optimizer updates the neural network parameters during training to minimize prediction errors. Optimization algorithms adjust weights and biases iteratively using gradient-based methods such as backpropagation. (Pandey et al, 2020)

III.2.3.2 Activation Functions

Activation functions introduce nonlinearity into neural networks, enabling them to model complex relationships. Several activation functions are commonly used in deep learning models:

a) Sigmoid Function

The sigmoid activation function transforms input values into outputs ranging between 0 and 1. It is widely used in binary classification problems.

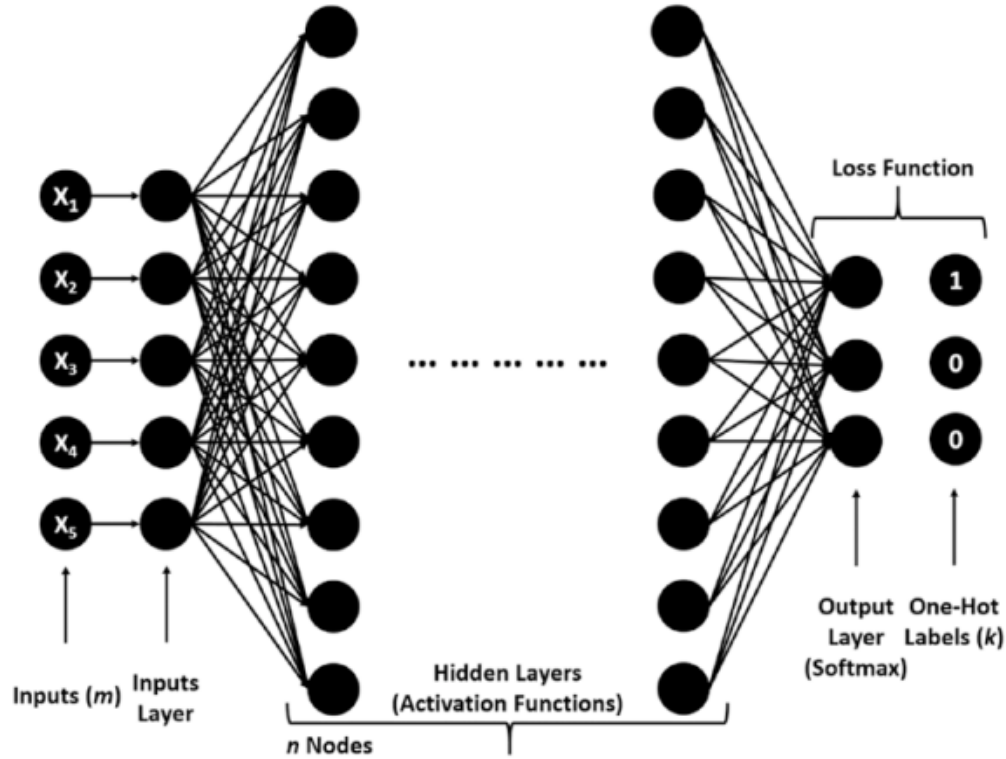


Figure III.2 Schematic representation of an MLP

b) Hyperbolic Tangent (Tanh)

The hyperbolic tangent function produces outputs between -1 and 1 and is often preferred over sigmoid because it is zero-centered. (Pandey et al, 2020)

III.2.3.3 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a specialized recurrent neural network architecture designed to overcome the limitations of standard RNNs, particularly the vanishing gradient problem. It achieves this by introducing a sophisticated gating mechanism that carefully regulates the flow of information across time steps.

At the core of an LSTM as shown in **Figure III.3** is a memory cell that maintains a cell state, which acts as a conveyor belt for long-term information. This cell state is modulated by three main gates:

1) Forget Gate

This gate decides what information from the previous cell state should be discarded. It takes the current input and the previous hidden state, processes them through a sigmoid activation function, and outputs values between 0 and 1. A value close to 0 means “forget this,” while a value close to 1 means “keep this.”

2) Input Gate

This gate controls which new information should be added to the cell state. It consists of two parts:

- A sigmoid layer that determines which values to update (0 to 1).
- A tanh layer that creates candidate values in the range of -1 to 1.
- The results of these two operations are combined to update the cell state.

3) Output Gate

This gate decides what information from the current cell state should be output as the hidden state. It uses a sigmoid function to select which parts of the cell state to output, and then applies a tanh function to the cell state. The filtered cell state (scaled by the output gate) becomes the final output for that time step. (Pandey et al, 2020)

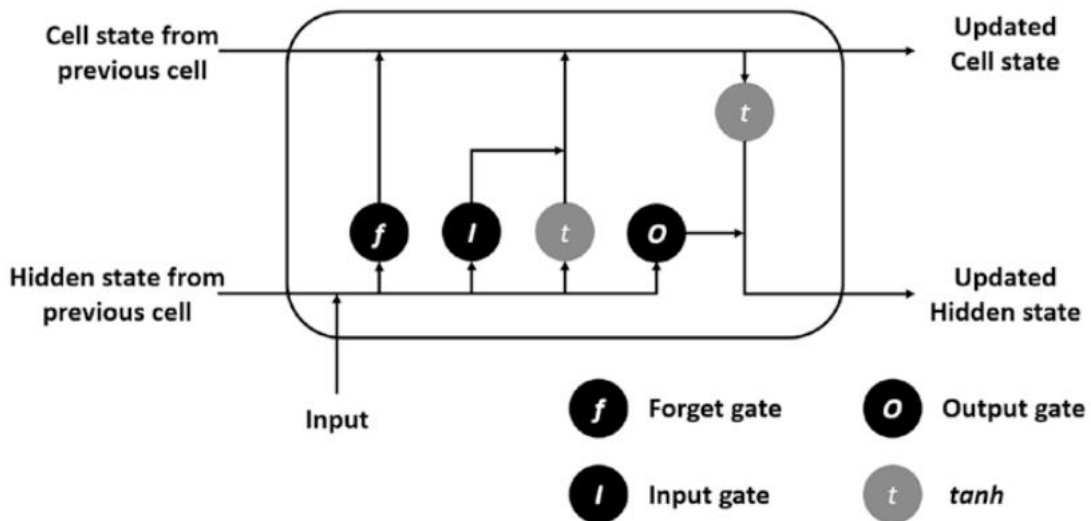


Figure III.3 Schematic depiction of long short-term memory (LSTM) cell architecture

III.2.4 Reinforcement Learning

Reinforcement learning is based on interaction with an environment where the algorithm makes decisions and receives rewards or penalties. Through this trial-and-error process, it learns the best sequence of actions to achieve a specific objective, often within a dynamic and uncertain system. (Pandey et al, 2020)

III.3 Machine Learning Libraries

Several open-source libraries have been developed to facilitate the implementation of machine learning algorithms, particularly in Python. These libraries provide efficient tools for data processing, model development, and performance evaluation.

- **Scikit-learn** is one of the most widely used libraries for machine learning in Python. Built on top of NumPy, SciPy, and Matplotlib, it offers a broad range of supervised and unsupervised learning algorithms through a simple and consistent interface. (Scikit-learn Developers, n.d.)
- **TensorFlow** is another powerful open-source library developed by Google for numerical computation and large-scale machine learning. It is particularly well-suited for neural networks and deep learning applications. TensorFlow uses dataflow graphs based on tensors, enabling efficient computation across multiple CPUs and GPUs, and supports several programming languages such as Python, C++, and Java. (TensorFlow Developers, n.d.)

Other libraries such as PyTorch, Theano, OpenCV, and Apache Spark MLlib are also widely used in machine learning applications. Among these, TensorFlow remains a preferred choice due to its scalability, flexibility, and strong community support. (Pandey et al, 2020)

III.4 Model Evaluation Metrics

To assess how well a machine learning model performs, especially on unseen data, it is essential to use quantitative evaluation metrics. These metrics provide a reliable way to measure prediction

accuracy and compare different models. The choice of metric depends on the type of problem being addressed.

III.4.1 Regression Metrics

For regression problems, where the output is continuous, common evaluation metrics include the coefficient of determination (R^2) and the root mean squared error (RMSE).

The R^2 metric measures how well the model explains the variability in the data. It is based on the comparison between the residual sum of squares (SS_{res}) and the total sum of squares (SS_{tot}), where y_i represents actual values, $\bar{y}$ is the mean value, and y_{reg} represents predicted values. A higher R^2 value indicates better model performance.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}, \text{ Where: } SS_{res} = \sum (y_i - y_{reg})^2$$

$$SS_{tot} = \sum (y_i - \bar{y})^2$$

The RMSE metric evaluates the average magnitude of prediction errors. It is calculated using the difference between predicted values and actual values, and N represents the total number of observations. Lower RMSE values indicate better accuracy.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

III.4.2 Classification Metrics

For classification problems, model performance is commonly evaluated using several metrics, including accuracy, precision, recall, and F1-score. Accuracy measures the proportion of correct predictions among all predictions, precision assesses the ability of the model to avoid false positives, recall evaluates the ability to identify all actual positive cases, and the F1-score provides a balanced measure of precision and recall.

A confusion matrix as shown in **Figure III.4** is also commonly used to evaluate classification performance. It is a square matrix where correct predictions appear along the diagonal, while misclassifications are represented in the off-diagonal elements. This matrix provides a clear visualization of model performance. (Pandey et al, 2020)

		Predicted class	
		P	N
Actual class	P	TP	FN
	N	FP	TN

Figure III.4 An illustration of confusion matrix

Conclusion

This chapter presented a comprehensive overview of machine learning and deep learning concepts and their relevance to petroleum engineering applications. The discussion began with the definition of machine learning and its major categories, including supervised, unsupervised, and reinforcement learning. Several supervised learning techniques for regression and classification were presented, including linear regression, support vector machines, decision trees, and random forests. In addition, unsupervised learning approaches such as clustering and dimensionality reduction were introduced as effective tools for discovering hidden structures and reducing dataset complexity. The fundamentals of deep learning were also highlighted, focusing on multilayer perceptrons, and LSTM. Deep learning models provide the capability to capture highly nonlinear relationships and process large-scale datasets, making them increasingly valuable in reservoir engineering and production optimization applications.

The chapter also introduced commonly used machine learning libraries such as Scikit-learn, TensorFlow, which provide efficient tools for data analysis, model development, and optimization.

Finally, the chapter explored the importance of data preprocessing through data cleaning, features scaling, and outliers' detection and treatment. These steps are essential for improving data quality, ensuring numerical stability, and enhancing the performance and generalization capability of machine learning models. The concepts of underfitting and overfitting were also discussed, emphasizing the need to balance model complexity and predictive performance.

Overall, machine learning and deep learning offer powerful data-driven approaches for solving complex petroleum engineering problems. With the increasing availability of field data and computational resources, these techniques are expected to play an increasingly important role in reservoir characterization, production forecasting, optimization, and intelligent decision-making within the oil and gas industry.

Chapter IV: Waterflooding Optimization System

Introduction

Waterflooding optimization and management remains challenging in the majority of fields, where inefficient water injection is allocated, premature water breakthrough, and inappropriate conformance practices lead to considerable recoverable oil left behind at abandonment. These challenges occur because of the absence of practical systems that monitor waterflooding processes and ensure optimized water injection that supports reservoir pressure and provides efficient oil sweep and displacement.

In recent years, the rapid advancement of machine learning (ML) techniques, coupled with the increasing availability of high-frequency historical production and injection data, has opened new avenues for data-driven reservoir management. This chapter presents the Machine Learning unit developed for the optimization of waterflooding in the EMN sub-field of El Merk field. The unit integrates the LSTM model for forecasting reservoir fractional flow and a hybrid Physics-Informed Neural Network for solving the Buckley-Leverett frontal advance equation, together with a physics-informed Capacitance-Resistance Model (CRM) for injector-producer connectivity.

IV.1 Materials Used

The following tools and environments were used in this study:

- **Python (Version 3.10):** The main programming language used for this project for data processing, model development, and optimization.
- **Jupyter Notebook:** The primary interactive development environment for coding, data exploration and visualization.
- **Visual Studio Code:** Used as a secondary code editor for script development, debugging, and project management.
- **Microsoft Excel:** Used for initial data storage, organization, statistical summaries, and exploratory data analysis.

IV.2 Methodology

The methodological framework comprises four sequential stages, as illustrated in **Figure IV.1**. They include: 1) Data collection; 2) data pre-processing and quality control; 3) model development, training and multi-level validation; and 4) integration into the optimization and decision-support system.

IV.2.1 Data collection

The dataset has been collected from EMN sub-field, and it contains historical production and injection data from March 2013 to January 2026, which is approximately 156 months period

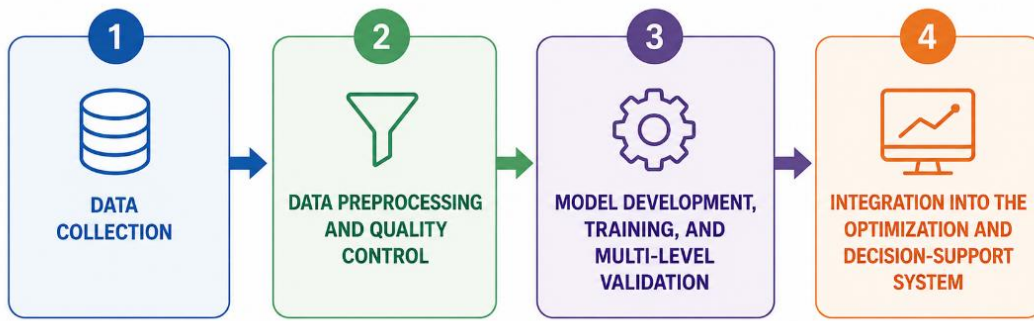


Figure IV.1 Methodological framework of the project

providing about 27574 allocated data and 8460 well test observations spread over six producers (N_P1X, N_P2X, N_P6X, N_P16X, N_P17X, N_P18X), and two water injectors (N_A7X, N_A12X). **Figures IV.2, IV.3, and IV.4** show the injection and production rates of all wells over 156 months.

The Well N_P1X was originally a producer but was later converted to a gas injector and subsequently shut-in. However, for the purpose of consistency in this study and to maintain a complete dataset for the waterflooding analysis, N_P1X is treated throughout this work as a producer. All production data, water-cut forecasts, and saturation profiles for this well are based on its historical performance as a producer before to the conversion.

The dataset also contains the following data:

- The flowing bottom hole pressure (FBHP) of all wells

- Well test data
- Petrophysical data (porosity and permeability)
- PVT data
- Relative permeability data
- Static data including wells coordinates and perforation intervals



Figure IV.2 EMN field water-injection history of the two water injectors (2013–2026)

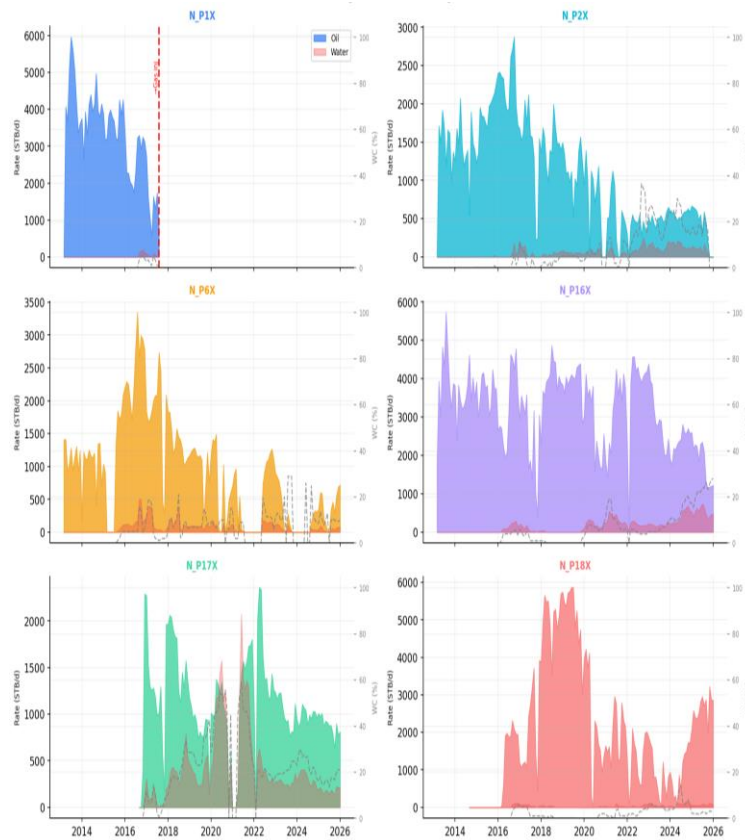


Figure IV.3 EMN field production history for the six oil producers (2013–2026)

IV.2.2 Data preprocessing

IV.2.2.1 Data quality control and cleaning

Data quality control followed a structured workflow designed to identify and mitigate anomalous observations within the dataset. Initially, physical plausibility judgment was applied to investigate the values statistically flagged as anomalies to determine whether it complies with physics. Then, apply the statistical method was applied to remove measurements that remain physically possible yet statistically anomalous, including meter malfunctions, allocation-reporting artefacts, and operational interventions.

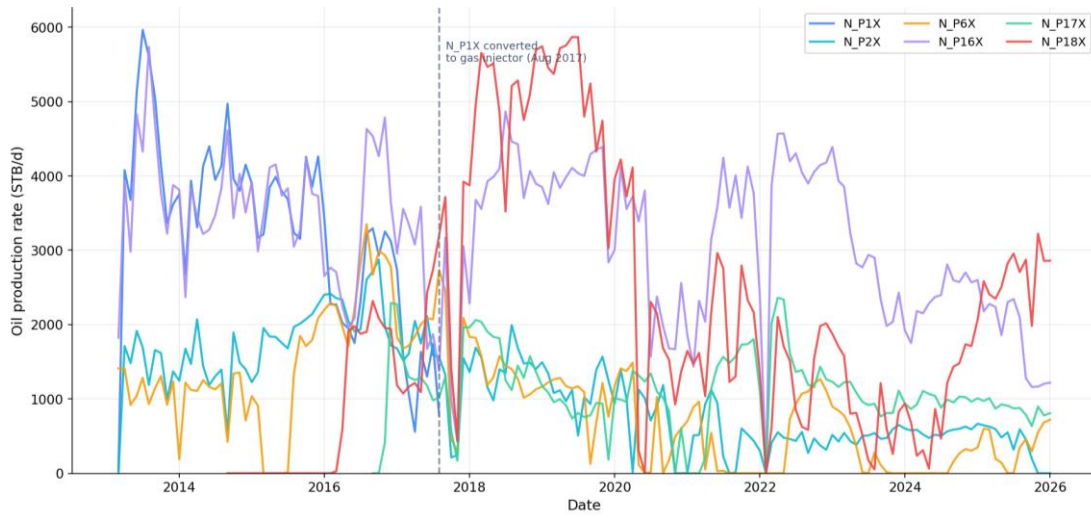


Figure IV.4 EMN field oil production history (monthly average) for all six producers

A diagnostic analysis of the EMN field dataset showed that approximately 6% of production and injection records exhibited these anomalies. Since these fluctuations introduce harmful noise into machine-learning and optimization workflows, they were explicitly treated rather than retained in the dataset.

IV.2.2.2 Hampel filter

The Hampel filter (window=5, $k=1.4826$, $n_sigma=3.0$, identical for all variables) is applied to remove outliers and it's preferred over IQR and Z-score for three specific reasons observable in the data:

- 1) **Non-stationary trends:** Water cut (WC) increases monotonically over time, while oil production rate (QOP) continuously declines. As a result, global statistical methods such as the Interquartile Range (IQR) and Z-score often misclassify valid late-life production measurements as outliers because they rely on overall dataset statistics. In contrast, the Hampel filter uses a local rolling median and median absolute deviation (MAD), allowing it to adapt to the evolving data level and distinguish genuine outliers from normal long-term production trends.
- 2) **Non-Gaussian distributions:** Seven variables exhibit skewness outside the range $[-1, 1]$, including QWP (3.56) and QGP (2.02), indicating strong departures from normality. Since the Z-score method assumes an approximately Gaussian distribution, its performance may be compromised. The Hampel filter, based on the median and median absolute deviation (MAD), is distribution-free and robust to skewed data.
- 3) **Operational anomalies:** Shut-ins, workovers, and meter faults often produce short-term spikes or drops lasting 1–3 months. The selected 11-point Hampel window ($w = 5$) is well matched to this timescale, enabling effective detection of operational anomalies without distorting the long-term production behavior.

The **Table IV.1** and **Table IV.2** shows a statistical summary of the data before and after removing outliers.

Table IV.1 Data before outliers' removal

Variable	Unit	N monthly	Corrected (Hamp.)	Min	Max	Mean	Std (σ)	Skew
Water Cut (WCUT_W)	%	681	78/681 (11.5%)	0	59.09	9.099	10.81	1.84
Oil Rate (QOP_W)	STB/d	681	31/681 (4.6%)	236.5	7091	2214	1409	0.89
Water Prod (QWP_W)	STB/d	681	47/681 (6.9%)	0	2002	162.5	243.1	3.56
Gas Prod (QGP_W)	MSCF/d	681	52/681 (7.6%)	0.546	43.58	8.229	9.473	2.02
GOR (GOR_W)	SCF/STB	681	77/681 (11.3%)	397.7	34.0k	5659	6801	1.54
WHP (WHP_W)	psia	681	54/681 (7.9%)	775.0	4093	1844	788.8	0.99
WHT (WHT_W)	°C	540	46/540 (8.5%)	59.00	186.9	161.5	16.59	-1.76
FBHP Gauge (BHP_W)	psia	399	37/399 (9.3%)	2402	5315	3899	608.7	-0.05
Injection (QWI_W)	STB/d	185	12/185 (6.5%)	848.7	11.3k	6716	1919	-1.13

Table IV.2 Data after outliers' removal

Variable	Unit	Min	Max	Mean	Std (σ)	Skew	Δ Mean (%)	Δ Std (%)	Δ Skew
Water Cut (WCUT_W)	%	0	59.09	8.519	10.39	1.94	-6.4%	-3.9%	+0.10
Oil Rate (QOP_W)	STB/d	236.5	7091	2210	1395	0.88	-0.2%	-1.0%	-0.01
Water Prod (QWP_W)	STB/d	0	2002	158.7	242.6	3.61	-2.3%	-0.2%	+0.05
Gas Prod (QGP_W)	MSCF/d	0.575	43.58	8.180	9.410	2.05	-0.6%	-0.7%	+0.02
GOR (GOR_W)	SCF/STB	397.7	28.7k	5582	6598	1.42	-1.4%	-3.0%	-0.11
WHP (WHP_W)	psia	775.0	4044	1847	782.8	0.97	+0.2%	-0.8%	-0.02
WHT (WHT_W)	°C	95.67	186.9	162.2	15.19	-1.27	+0.5%	-8.4%	+0.49
FBHP Gauge (BHP_W)	psia	2402	5315	3905	594.2	-0.04	+0.1%	-2.4%	+0.01
Injection (QWI_W)	STB/d	848.7	11.3k	6761	1803	-1.16	+0.7%	-6.0%	-0.03

IV.2.2.3 Normalisation

Min-Max scaling was applied to the feature matrix. Min-Max scaling was selected instead of z-score standardisation because it preserves the original distribution shape while maintaining relative feature relationships. In addition, constraining feature values to the range [0,1] improves the numerical stability and convergence behavior of the models.

IV.2.3 Model development

The proposed system is structured as three tightly coupled unites as shown in **Figure IV.5** rather than a single end-to-end black-box predictor. This modular architecture deliberately assigns each task to the modelling paradigm best suited to it:

- 1) **Unit 1 (Physics Engine):** A hybrid Capacitance-Resistance Model (CRM) that quantifies injector-producer connectivity and fluid distribution from first principles.
- 2) **Unit 2 (Machine Learning):** A regularized machine learning model dedicated exclusively to forecasting the water-cut trajectory.
- 3) **Unit 3 (Operational Optimizer):** A constrained non-linear programming solver that integrates outputs from the first two units to maximize injection efficiency.

The three unites operate within a monthly receding-horizon (model predictive control) framework. At the end of each month, the CRM and water-cut forecaster are re-fitted on the latest rolling window, the efficiency objective is updated, and the optimizer is re-solved. Only the first month of the resulting plan is implemented.

IV.2.3.1 The Capacitance-Resistance Model (Unit 1)

The Capacitance-Resistance Model is a reduced-physics and material-balance representation of waterflooding. It treats the reservoir between an injector and a producer as a system of connected control volumes (capacitance) and a flow resistance, and infers from rate history how strongly each injector communicates with each producer. It is retained as the structural backbone of the system

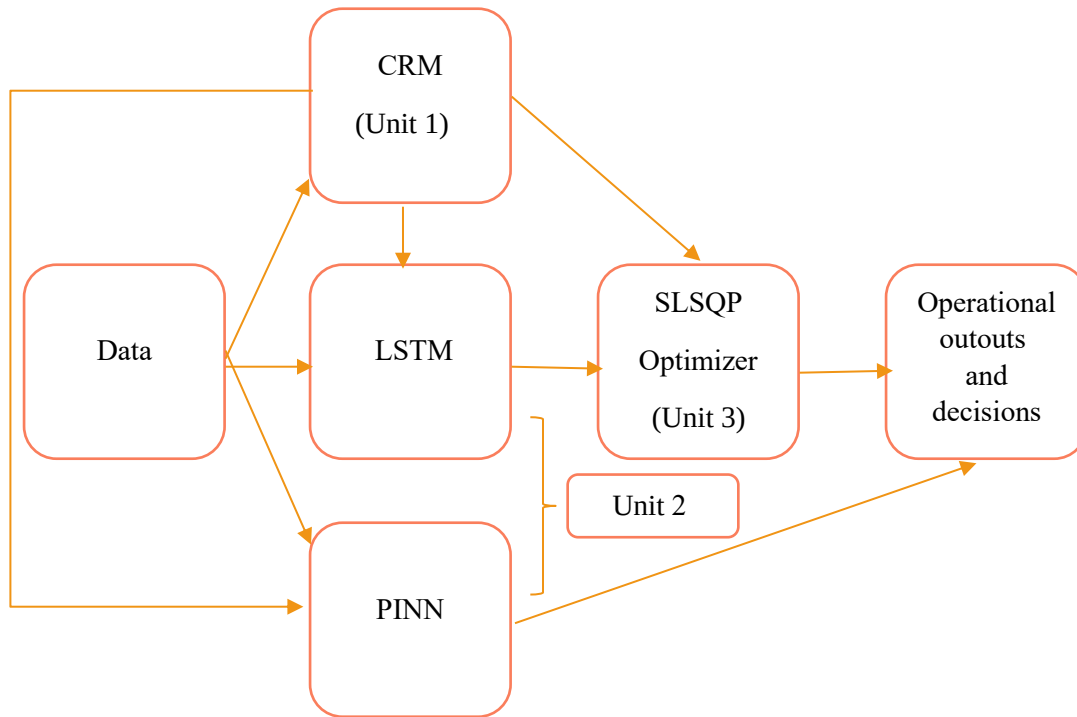


Figure IV.5 Waterflood Optimization System Architecture

each injector communicates with each producer. It is retained as the structural backbone of the system because for mature decline-dominated fields it is the most robust available estimator of inter-well connectivity (see **Figure IV.6**). (You et al., 2021)

For a single producer p influenced by injectors j , the continuity form of the producer-based CRM (CRMP) expresses the producing rate as a decaying memory of past production plus a weighted sum of injection signals (see Equation (4.1)):

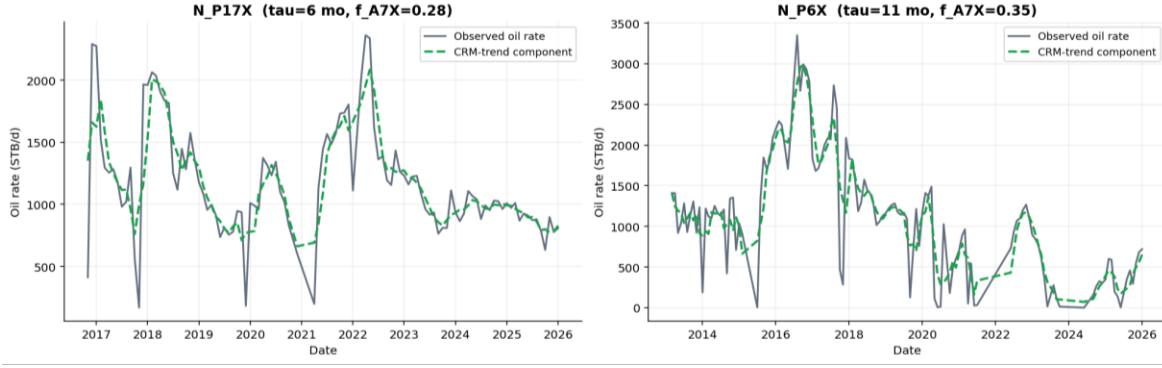


Figure IV.6 Illustrative decline-and-response decomposition for N_P17X and N_P6X producers

$$q_p(t) = q_p(t-1) \times e^{\left(-\frac{\Delta t}{\tau_p}\right)} + \sum_{j=1}^{N_i} \left[f_{pj} \times I_j(t) \times \left(1 - e^{\left(-\frac{\Delta t}{\tau_p}\right)} \right) \right] \quad (4.1)$$

where:

$q_p(t)$: liquid production rate of producer p at time t (STB/d).

$I_j(t)$: injection rate of injector j (STB/d).

$f_{pj} [0, 1]$: connectivity fraction (steady-state gain) between injector j and producer p .

τ_p : time constant (months) representing the drainage volume's capacitance and resistance.

A small τ means a fast, possibly channeled, connection, whereas a large τ means a diffuse and slow-responding drainage volume. The two parameters per injector-producer pair, f_{pj} and τ_p , fully characterize the connection.

Direct non-linear least-squares fitting of Equation (4.1) is unstable on the EMN data because the injection signals are strongly correlated and the production decline dominates the variance. A hybrid four-stage estimator is therefore used, which is both robust and physically interpretable:

- 1) **Decline detrending:** An exponential decline curve is fitted to the first twenty-four months of each producer's liquid rate and removed, isolating the fluctuation component that can respond to injection.
- 2) **Cross-correlation:** The detrended production fluctuation is cross-correlated with each normalized injector signal over lags of zero to twelve months; the peak positive correlation is taken as the raw connectivity strength and its lag informs the time constant.

- 3) **Distance prior:** A geological prior is formed from inter-well Euclidean distance d , where distance prior is equal to $1 / (1 + d/2000)$, expressing that nearer wells are more likely to be connected. The raw connectivity is combined as $0.70 \times \text{cross-correlation} + 0.30 \times \text{distance prior}$.
- 4) **Normalization:** Each producer's coefficients are scaled so that their sum does not exceed 0.9, enforcing the physical constraint that a producer cannot receive more than the total injected volume and reserving headroom for aquifer and gas-cap support.

The time constant is estimated from the autocorrelation decay of the detrended series taken as the lag at which autocorrelation falls below $1/e$. The estimator is re-run every month on a rolling sixty-month window so that the inferred connectivity adapts as the flood matures.

a) Inputs and outputs of the CRM

The CRM is fitted independently for each producer. Its inputs as shown in **Table IV.3** are entirely historical rate series and the inter-well coordinates used for the distance prior. Its outputs as shown in **Table IV.4** are the two physical parameters per injector-producer pair that the optimization consumes.

Table IV.3 CRM detailed input specification (per producer, per monthly refit)

Input	Description - dimension - units
Producer liquid rate	$q_p(t)$ = oil + water rate of the producer; one value per month over the rolling 60-month window, unit STB/d. Vector length ≈ 60 for each producer.
Injector rate series	$I_j(t)$ for N_A7X and N_A12X; one value per month over the same window, unit STB/d. Matrix 60×2 .
Well coordinates	UTM easting/northing of each producer and injector; used once to form the distance prior $1/(1+d/2000)$. Static, dimensionless after transform.
Fitting window	Rolling 60 months ending at the decision month; re-sliced every cycle so the estimate adapts.

In summary, the CRM maps a 60×2 injection matrix and a 60-vector of producer liquid rate to three scalars per injector-producer pair (f , τ and a diagnostic cross-correlation). For the six EMN

producers and two injectors this is a 6×2 connectivity matrix and a 6-vector of time constants, refreshed every month.

Table IV.4 CRM detailed output specification (per producer)

Output	Description, range and downstream use
Connectivity $f_{p,A7X}$	Fraction of N_{A7X} water reaching producer p . Dimensionless, range [0, 0.9]. Feeds the efficiency objective and the channeling penalty.
Connectivity $f_{p,A12X}$	Fraction of N_{A12X} water reaching producer p . Dimensionless, range [0, 0.9]. Same downstream use.
Time constant τ_p	Response memory in months, range [0.5, 24]. Used to flag channeling risk ($\tau < 4$) and to weight the receding-horizon projection.
Peak cross-correlation	Diagnostic strength of each connection, range [0, 1]. Reported for transparency; not used directly in optimization.

b) Results of the CRM Model

Table IV.5 reports the fitted CRM parameters. The connectivity matrix and time constants are shown graphically in **Figure IV.7**

Table IV.5 Fitted CRM parameters for the EMN field

Producer	τ (months)	$f(N_{A7X})$	$f(N_{A12X})$	$xcorr\ A7X$	Behaviour
N_P1X	21	0.4608	0.4392	0.6409	Slow; isolated drainage
N_P2X	13	0.4902	0.4098	0.4906	Moderate response
N_P6X	11	0.3547	0.2632	0.3831	Moderate response
N_P16X	3	0.346	0.3095	0.2732	Fast; channeling watch
N_P17X	6	0.2761	0.1281	0.148	Moderate response
N_P18X	12	0.5952	0.3048	0.806	Moderate response

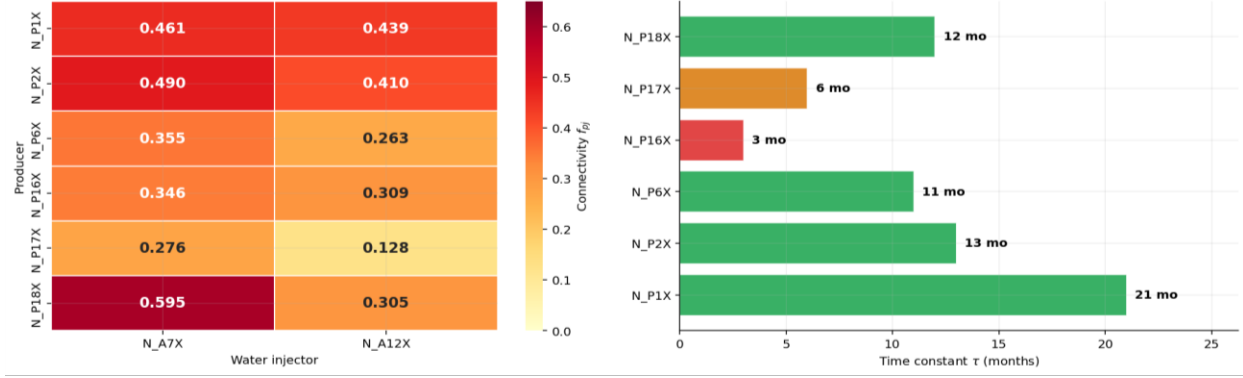


Figure IV.7 The fitted CRM connectivity matrix and the producer time constants

IV.2.3.2 The Machine-Learning Unit

This section presents the ML unit developed for the EMN waterflooding optimization system. It consists of two complementary models that address different aspects of the waterflooding displacement process. The models are developed on top of 8460 well test observations.

The first model uses LSTM to forecast the fractional flow using reservoir data and historical production data. The second model is a hybrid Physics-Informed Neural Network (PINN) designed to solve the frontal advance equation and deliver the full saturation profile $S_w(x,t)$ across the reservoir.

a) Physical correlations, constants and SCAL data

- The Corey fractional flow correlation (Equation (4.2)) connects reservoir saturation to reservoir fractional flow:

$$f_w(S_w) = \frac{k_{rw}(S_w)/\mu_w}{k_{rw}(S_w)/\mu_w + k_{ro}(S_w)/\mu_o} \quad (4.2)$$

$$k_{rw} = k_{rw,max} S_{wD}^{n_w}, \quad k_{ro} = k_{ro,max} (1 - S_{wD})^{n_o}$$

where $S_{wD} = (S_w - S_{wi}) / (1 - S_{or} - S_{wi})$. Fitting to the 30-point WOTABLE data gives $n_w = 2.200$, $n_o = 4.3439$. This constitutive law is not a prediction model; it is derived entirely from laboratory measurements and links the observable reservoir fractional flow to the pore-media saturation. The Welge construction applied to this curve yields the shock-front saturation $S_w^* = 0.5367$ and front velocity $v_{front} = 2.2022$ PV/t. (Corey, 1954)

- Heterogeneous Buckley-Leverett frontal advance PDE

Conservation of the water phase at reservoir conditions is given by Equation (4.3):

$$\phi(x) \frac{\partial S_w}{\partial t} + \frac{K(x)}{K} v_{eff} \frac{\partial f_w}{\partial S_w} \frac{\partial S_w}{\partial x} = 0 \quad (4.3)$$

where x in $[0,1]$ is the normalized streamline coordinate ($x=0$: injector, $x=1$: producer), t is injected pore volumes, and $K(x)$, $\phi(x)$ are the spatially-variable permeability and porosity extracted cell-by-cell from the 21,296-cell EMN history-matched grid. This PDE has no closed-form analytical solution when $K(x)$ and $\phi(x)$ are spatially variable. The Welge solution applies only when K and ϕ are constant (homogeneous).

- The physical constants of the EMN field data are shown in **Table IV.6** and **Figure IV.8**.

Table IV.6 Fitted CRM parameters for the EMN field

Parameter	Value
S_{wi} - connate water saturation	0.1500
S_{or} - residual oil saturation	0.0934
n_w - Corey water exponent	2.200
n_o - Corey oil exponent	4.3439
$k_{rw,end}$ - water kr at $S_w = 1 - S_{or}$	0.5611
$k_{ro,end}$ - oil kr at $S_w = S_{wi}$	0.800
μ_o / μ_w - oil and water viscosity	0.80 / 0.50 cP
B_o - oil formation volume factor	1.483 RB/STB
B_w - water formation volume factor	1.020 RB/STB

S_{w^*} - Welge shock-front saturation	0.5367
v_{front} - Welge front velocity	2.2022 PV/t
N_{cap} - capillary number	0.0431
$\varepsilon(S_{w^*})$ - diffusion coefficient at BT	0.376

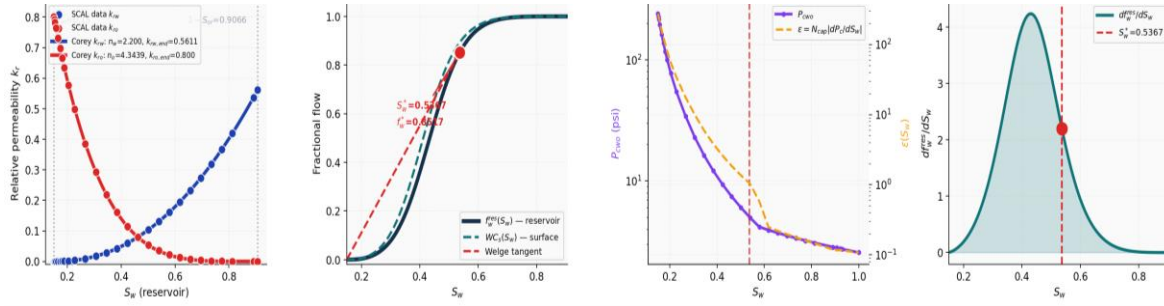


Figure IV.8 Corey kr curves with the EMN SCAL data

b) Fractional flow forecasting ML model

This ML model is used to forecast the fractional flow fw . This is the one quantity in the system that is genuinely pattern-rich and data-driven and it depends on saturation history, injection history, and breakthrough dynamics, and it is the quantity that the physics unit cannot project forward accurately on its own.

A model trained to predict oil rate obtains most of its apparent accuracy from a quantity irrelevant to the injection decision. The water-cut trajectory by contrast is smooth, physically driven, and forecastable, and it is exactly what the efficiency objective requires.

b-1) Algorithm selection

The LSTM-128 is chosen to be the water-cut forecasting component of the EMN waterflood ML system. Given a 12-month observation window per producer, it outputs $fw(t+1)$. The predicted water cut for the following month. This scalar feeds directly into the SLSQP constrained optimizer as the objective function input.

b-2) Input specification of the FFM

Each training example is a (12, 4) tensor: 12 consecutive monthly timesteps, 4 features each. The features are shown in the **Table IV.7**.

Table IV.7 Detailed input feature of the FFM model

Feature name	Source variable	Physical role in forecast
Water cut fw	WCUT_W	Primary target history. Provides current flood state and trend over the 12-month window.
Flowing BHP	BHP_W (gauge)	Encodes flood advance signal: BHP declines as waterfront approaches $x=1$.
Oil rate	QOP_W	Captures productivity decline and workover events that co-vary with water breakthrough.
Total injection	$\text{inj_tot} = \text{QWI_A7X} + \text{QWI_A12X}$	Voidage replacement signal. Increases in injection rate precede WC acceleration by 1–3 months (CRM lag).

b-3) FFM network architecture

The FFM network architecture used is shown in **Table IV.8**.

Table IV.8 FFM network architecture

Layer	Type / Config	Input shape	Output shape	Parameters	Key design choice
LSTM	nn.LSTM(input_size=4, hidden_size=128, num_layers=1, batch_first=True)	(N, 12, 4)	hn: (1, N, 128)	~268,800	Single layer avoids vanishing gradient risk; $h=128$ sufficient for 4 features x 12 steps.
Extract	hn[-1] (final hidden state)	(1, N, 128)	(N, 128)	0	Takes only the last hidden state — full sequence summary. No attention needed at this scale.

Layer	Type / Config	Input shape	Output shape	Parameters	Key design choice
BatchNorm1d	nn.BatchNorm1d(128)	(N, 128)	(N, 128)	256 (γ , β)	Normalises over the batch after the recurrent layer. Stabilises training, reduces internal covariate shift.
Dropout	nn.Dropout(p=0.15)	(N, 128)	(N, 128)	0	~15% neurons zeroed per forward pass during training. Disabled at inference. Complements weight decay.
Linear	nn.Linear(in_features=128, out_features=1)	(N, 128)	(N, 1)	129 (W+b)	No activation: fw(t+1) prediction is unbounded in training; physical range [0,1] enforced externally.

b-4) Training configuration

The training configuration of the FFM is shown in **Table IV.9**.

Table IV.9 Training configuration of the FFM

Parameter	Value	Rationale
Optimiser	AdamW (beta1=0.9, beta2=0.999)	Weight decay decoupled from gradient update; more stable than Adam for small datasets
Learning rate	OneCycleLR (max_lr=3e-3, pct_start=0.1)	Linear warmup for first 10% of steps, cosine decay thereafter; avoids divergence at start
Weight decay	5e-5	Mild L2 regularisation; complements Dropout; prevents memorisation of small training set
Gradient clipping	max_norm=1.0	Prevents exploding gradients in LSTM recurrence (unrolled across 12 timesteps)
Max epochs	350	Upper bound; EarlyStopping almost always terminates earlier (~87 epochs typical)

Parameter	Value	Rationale
Early stopping	Patience=40 on validation MSE	Prevents overfitting; best validation checkpoint saved and restored
Batch size	Full training set (N=361)	Small enough to load in memory; BatchNorm1d effective at this batch size
Loss function	MSE: $(1/N) * \sum((fw_hat - fw_obs)^2)$	Standard loss function; (physics enters via BHP feature)
Train set	361 sequences (60%)	Chronological split: earliest 60% of production history per producer
Validation set	90 sequences (15%)	Chronological: months 60-75% (used only for EarlyStopping)
Test set	152 sequences (25%)	Chronological: latest 25% — held out entirely until final evaluation
Adopted seed	55	Best $R^2=0.9440$ across 7 seeds evaluated; fixed for all downstream results

b-5) Results of the FFM

The FFM training and validation results are shown in the **Figure IV.9**.

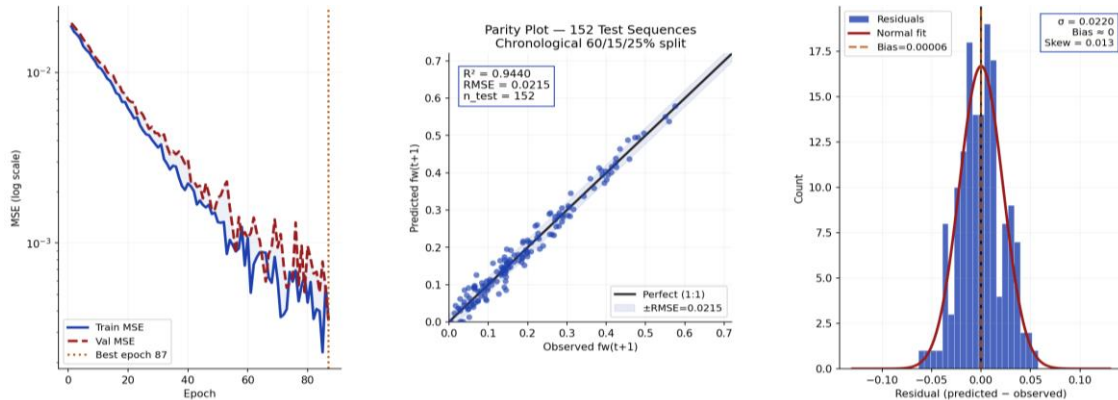


Figure IV.9 FFM training and validation results

b-6) The efficiency objective

The two unites are coupled through a single physically defined objective. The injection efficiency of producer p is the product of the CRM connectivity and the Buckley-Leverett oil fraction of the displaced fluid:

$$\eta_{p,j} = f_{pj} \times (1 - f_w(S_{w,p}))$$

where f_w is the water fractional flow evaluated at the producer's estimated saturation. The complement $1 - f_w$ is unity when injection displaces pure oil and approaches zero after water breakthrough, when injection merely recycles water.

The field objective is the total incremental oil rate attributable to injection and the headline performance metric is its ratio to the total injected volume:

- The field objective

$$\Phi = \sum_p (f_{p,7} \cdot I_7 + f_{p,12} \cdot I_{12}) \times (1 - f_w(S_{w,p}))$$

- The headline efficiency metric

$$E = \frac{\Phi}{I_7 + I_{12}}$$

This objective contains no decline term. It is constructed entirely from the physics unit (the connectivity coefficients and the fractional-flow curve) and the machine-learning unit (the forecast saturation that enters f_w). It is therefore fully decision-relevant: every unit of the objective is a direct consequence of the injection allocation being chosen.

c) Field-Trained Physics-Informed Neural Network model

c-1) PINN architecture and specification

The architecture and specification of the PINN model are shown in **Table IV.10**.

Table IV.10 EMN_PINN architecture

Component	Specification	Role and rationale
Input: position x	[0, 1]	0 = injector face (injected at $S_w=1-S_{or}$); 1 = producer face (observed fractional flow)

Input: time t	[0, 1.5] PV injected	Dimensionless injection time; covers full pre- and post-breakthrough range
Fourier encoding	12 freq: $\pi, 2\pi, \dots, 12\pi$ (sin+cos per dim)	Multi-scale representation; captures both near-injector plateau and near-front gradient.
Architecture	3 x 96 tanh (23,617 params)	Hard BC ansatz $D(x,t)=(1-e^{-25t})(1-e^{-20x})$ enforces IC/BC exactly; output clamped to $[S_{wi}, 1-S_{or}]$
PDE autodiff	dS_w/dt and dS_w/dx via torch.autograd	Residual of heterogeneous BL PDE evaluated at 60 collocation pts per step using real $K(x), \phi(x)$
$K(x), \phi(x)$ source	80-pt extraction from 21,296-cell grid	Inverse-distance weighted from EMN-HM26_KX.csv and EMN-HM26_POR.csv; separate profile per pair
Field data	277 S_w at $x=1$	Primary training signal; 755 monthly records after 5%-drop operational filter; per-pair CRM-normalised PV time
Training	1,200 epochs, Adam $lr=5e-3$, cosine annealing	One PINN per injector-producer pair (12 models); each trained independently on its own $K(x), \phi(x)$ and saturation history

c-2) Training and total loss function

The total loss function is:

$$\mathcal{L} = \underbrace{500 \left\| S_w(1, t_{obs}) - f_w^{-1}(f_w^{obs}) \right\|^2}_{\mathcal{L}_{data}: 277 \text{ EMN field points}} + \underbrace{50 \left\| \phi \partial_t S_w + K / \bar{K} \cdot v_{eff} \cdot f'_w \cdot \partial_x S_w \right\|^2}_{\mathcal{L}_{PDE}: \text{autodiff interior}} + 200(\mathcal{L}_{BC} + \mathcal{L}_{IC})$$

The primary training signal ($\lambda=500$) is 277 reservoir-condition S_w observations. These observations are at $x=1$ (producer face) and directly constrain the PINN to match the actual EMN reservoir behaviour. The PDE residual ($\lambda=50$) evaluates the heterogeneous BL equation at 60 interior collocation points per training step using automatic differentiation through the network; $K(x)$ and $\phi(x)$ are interpolated from the 80-point streamline extraction per pair. The PDE residual provides physics-consistent inductive bias for the unobservable interior $0 < x < 1$. The ($\lambda=200$) represents the boundary and initial condition losses, respectively.

The weights (500–50–200) were selected through systematic ablation experiments. Different weight combinations were tested, and the final values were chosen because they provided the best balance between minimizing the error with respect to the observed saturation data, ensuring good agreement with the analytical Welge solution, and maintaining stable and convergent model training.

12 independent PINNs are trained, one per injector-producer pair, each with its own $K(x)$, $\phi(x)$ profile and WC history. This is necessary because $K(x)$ and $\phi(x)$ differ substantially between pairs. (Raissi et al., 2019)

c-3) Results and output of the PINN

All-pair PINN performance on EMN field observations are shown in **Figure IV.10**.

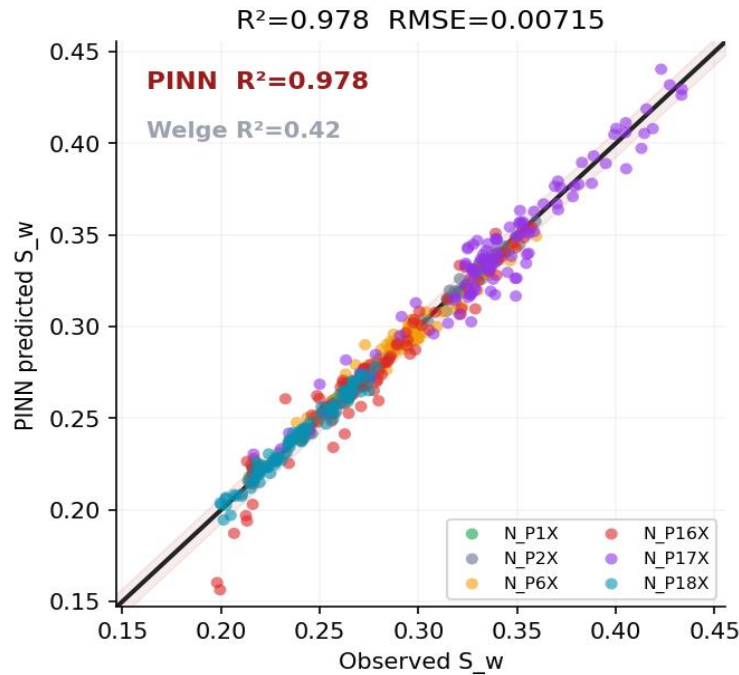


Figure IV.10 All-Pair PINN performance on EMN field observations

Figure IV.10 shows the overall performance of the PINN across all injector-producer pairs in the EMN field, comparing predicted and observed water saturations. The parity plot shows a high degree of agreement between the PINN predictions and the actual field measurements. Most points cluster tightly around the 1:1 line, indicating that the model successfully reconstructs the correct

water saturation. Overall, the error remains low, with an R^2 of 0.978 and an RMSE of 0.00715, and the majority of predictions fall within a narrow band around the observed values.

The PINN saturation profile output is shown in the **Figure IV.11**.

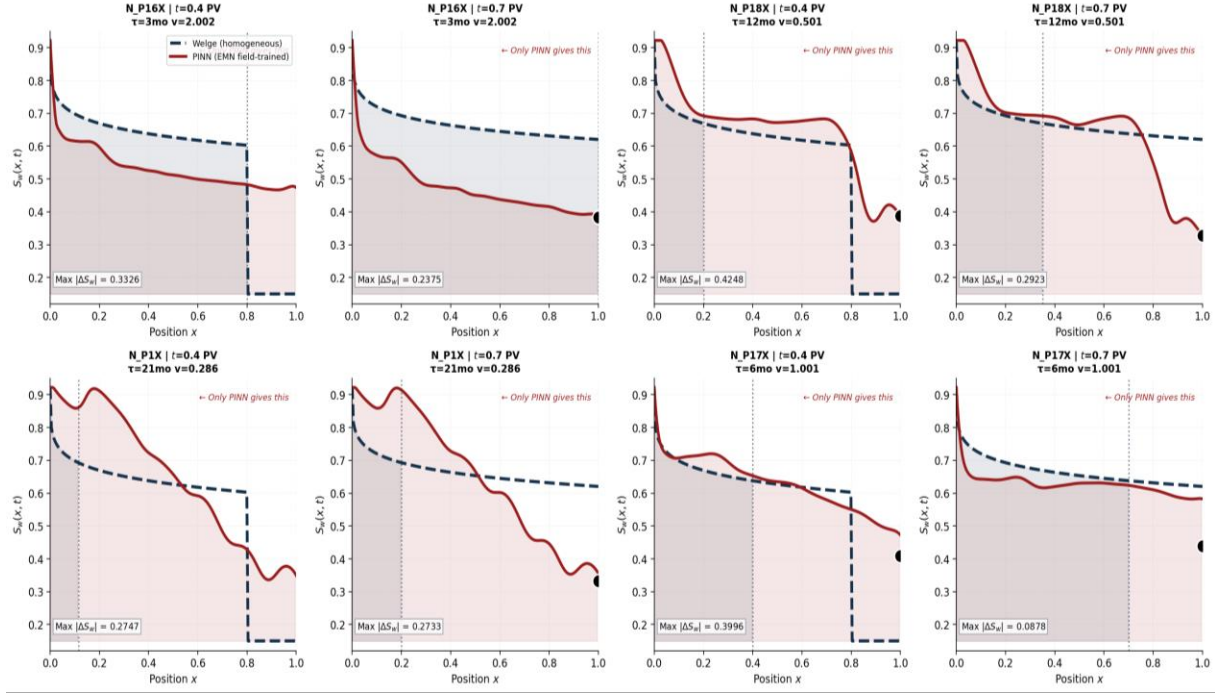


Figure IV.11 Saturation profiles $S_w(x,t)$ from the PINN for four representative pairs

Figure IV.11 displays the saturation profiles $S_w(x,t)$ predicted by the PINN for four representative injector-producer pairs at selected injection times, compared with the homogeneous Welge analytical solution.

The PINN profiles exhibit the characteristic Buckley-Leverett shape represented by a realistic sharp shock front, a smooth concave rarefaction behind the front, and correct boundary conditions at the injector. Importantly, the differences between the PINN and Welge curves reveal unswept zones that are not captured by the homogeneous analytical solution. These unswept regions, visible as areas where the PINN saturation remains lower than the Welge profile at the same time, highlight potential opportunities for infill drilling or targeted conformance treatments. The ability to quantitatively identify such unswept zones from the PINN output represents a key practical advantage of the hybrid model for field optimization.

The flood front advance position and the volumetric sweep efficiency delivered by PINN are shown in the **Figure IV.12**.

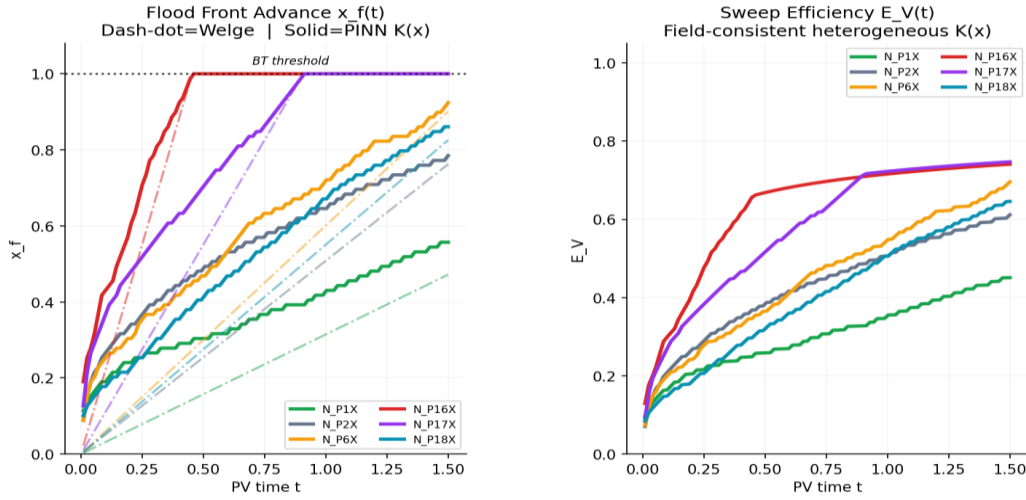


Figure IV.12 Flood front advance position and the volumetric sweep efficiency of PINN

Figure IV.12 shows the flood front advance position $x_f(t)$ predicted by the PINN along with the homogeneous Welge analytical solution flood front advance position.

The figure shows the evolution of the saturation front position over time for the injector-producer pairs. The PINN-derived front advances at different speeds depending on the pair-specific CRM time constants, whereas the Welge solution assumes a single homogeneous velocity.

This pair-specific front tracking is a major practical advantage. It allows for a more accurate estimation of breakthrough time for each producer, enabling better risk assessment and more timely implementation of mitigation strategies to prevent or delay early water breakthrough and be able to produce maximum oil. Unlike the idealized Welge front, the PINN provides a realistic representation of the flood front position that can be directly used for improved field exploitation and operational decision-making.

The figure also shows the volumetric sweep efficiency provided by the PINN, which is a key operational output. It offers a more realistic assessment of flood performance and enables better identification of poorly swept zones, supporting decisions on infill drilling and conformance control.

The summary of the PINN outputs is shown in **Table IV.11**.

Table IV.11 PINN outputs summary

Output	Format	Downstream use in EMN decision system
S_w(x,t) - full interior profile	Array: 200 x-points per t	Swept/unswept volume mapping; infill placement targeting; heterogeneity diagnostic
x_f(t) - flood front position	Scalar per (pair, t)	Breakthrough timing; identifies which producers are at risk; drives conformance plan
E_v(t) - sweep efficiency	Scalar: integral of normalised S _w	Flood performance monitoring
Breakthrough time t_{BT}	Time when x _f =1 or S _w (1)=S _w *	Risk status: N_P16X/N_P17X IMMINENT; N_P18X LOW (22 months); N_P1X LOW (24 months)

IV.2.4 The Optimization Unit

The optimization unit is the component that converts the physical understanding of the reservoir into an actionable monthly decision. The Capacitance-Resistance Model supplies inter-well connectivity, and the machine-learning forecaster supplies the projected water-cut state, and then the optimizer combines these into a single objective and answers one operational question that is what are the optimal water injection rates for the two controllable injectors (N_A7X and N_A12X) that maximize incremental oil recovery while respecting all physical and operational constraints of the field.

IV.2.4.1 Decision variables

The decision vector has exactly two continuous components: the water-injection rate of N_A7X, denoted I_7 , and that of N_A12X, denoted I_{12} , both in STB/d. These are the only quantities the operator can directly set on. The producer rates represent the outcome of this injection.

$$\mathbf{x} = (I_7, I_{12}), \quad I_7 \geq 0, \quad I_{12} \geq 0$$

IV.2.4.2 The objective function

The objective is the total incremental oil rate attributable to water injection. For each producer p , the water reaching it is the connectivity-weighted sum of the two injection rates, and the fraction of that displaced fluid which is oil is one minus the Buckley-Leverett water fractional flow

evaluated at the producer's forecast saturation. Summing the product over all producers gives the field incremental oil as shown in section **IV.3.1.2**

IV.2.4.3 The soft channeling penalty

A purely volumetric objective would happily push water toward a producer with very high connectivity even if that producer is a fast, watered-out thief well. To discourage this, a soft penalty is subtracted from the objective. The penalty activates only for a producer whose CRM time constant is short (τ below four months, indicating a fast, possibly fractured path) and whose water cut already exceeds 30%, and only for an injector whose connectivity to that producer exceeds 0.40. Its magnitude grows as the time constant shortens:

$$p_{chan} = \sum I_j \times f_{pj} \times \left(\frac{1 - \tau_p}{\tau_{crit}} \right), \quad \tau_{crit} = 4 \text{ months}$$

The penalty is multiplied by a small weight ($\lambda = 0.03$) and removed from the objective, so the optimizer trades a marginal amount of nominal incremental oil to avoid accelerating channeling. It is deliberately soft rather than a hard constraint, because channeling risk is a matter of degree and the engineer may still choose to accept some of it. The formulation surfaces the trade-off rather than forbidding it outright. The penalized objective actually maximized is:

$$Max \Phi(I_7, I_{12}) - \lambda \cdot p_{chan}(I_7, I_{12})$$

IV.2.4.4 The constraints

Four hard constraints bound the feasible region. Each expresses a physical limit that cannot be violated.

1) Injection budget

The combined water-injection rate cannot exceed the surface water-handling and pumping capacity allocated to the field, taken as $Q = 12,000$ STB/d. This is the constraint most operators expect to bind, and the result of this study is that, for the EMN field, it does not.

$$I_7 + I_{12} \leq 12,000 \text{ STB/d}$$

2) Pump capacity

Each injector has a mechanical upper rate set by its pump and tubing, taken as 11,000 STB/d, and a non-negativity lower bound. These enter the program as simple box bounds on the decision variables.

$$0 \leq I_7 \leq 11,000, \quad 0 \leq I_{12} \leq 11,000 \text{ STB/d}$$

3) Formation fracture pressure

Injecting water raises the bottomhole pressure at the injector, and if that pressure exceeds the formation parting pressure, the rock fractures, creating high-conductivity thief channels that short-circuit the flood and permanently degrade sweep. Thus, the injection pressure is held to a safety fraction of 85% of the fracture pressure. Therefore, with a fracture pressure of 6546 psia, the safe bottomhole-pressure ceiling is 5564 psia:

$$g_3 : \max(P_{wf,7}, P_{wf,12}) \leq 0.85 \cdot P_{\text{frac}} = 5564 \text{ psia}$$

4) Water voidage replacement

To maintain reservoir pressure without over-pressuring the formation, the water-only voidage-replacement ratio, which represents the reservoir-volume of water injected divided by the reservoir-volume of fluid produced is held within a band. The lower bound prevents pressure depletion and the upper bound prevents the pressure build-up that itself drives fracturing and channeling.

$$g_4 : 0.25 \leq \text{VRR}_w \leq 1.50$$

IV.2.4.5 The numerical optimization algorithm

1) Method

The program is a small, smooth, two-dimensional non-linear program with non-linear inequality constraints. It is solved by Sequential Quadratic Programming (SQP), which approximates the problem at each step by a quadratic model with linearized constraints and is well suited to smooth, low-dimensional, constrained problems. The objective and constraint functions are solved effectively instantaneous and can be repeated every month. (Nocedal & Wright, 2006)

2) Multi-start safeguard against local optima

Although the objective is well behaved, the channeling penalty introduces mild non-convexity, so a single starting point could in principle converge to a local optimum. To guard against this, the solver is run from four deliberately spread starting allocations, which are an even split, and three

skewed splits favouring each injector to differing degrees and the best converged result is retained. In practice all four starts converge to the same allocation, which is strong empirical evidence that the optimum is global for this problem.

3) Inputs and outputs of the optimizer

The optimizer is a deterministic mapping solved once per month. Its inputs as shown in **Table IV.12** are the current outputs of the physics and machine-learning unites together with the fixed physical limits. Its output as shown in **Table IV.13** is the recommended injection pair and a full diagnostic of every constraint, so that the recommendation can be audited before it is implemented.

Table IV.12 Detailed input specification of the injection optimizer

Input	Source and description
Connectivity matrix	From the CRM; 6×2 matrix of f_{pj} ; determines how each candidate allocation distributes to producers.
Forecast water cut	From the machine-learning model; one value per producer; converted to saturation and hence to the oil fraction $1 - f_w$.
Current field rates	Latest producer oil and water totals; enter the voidage-replacement constraint; STB/d.
Physical limits	Budget 12,000 STB/d, pump cap 11,000 STB/d, fracture ceiling 5564 psia, voidage band [0.25, 1.50].
Injectivity indices	$J = 5.0$ (N_A7X), 4.0 (N_A12X) STB/d/psi; map rate to bottomhole pressure.

Table IV.13 Detailed output specification of the injection optimizer

Output	Description and use
I_7, I_{12}	Recommended water-injection rates for the next month; STB/d. The actionable decision.
Incremental oil Φ	Objective value at the optimum; STB/d.
Efficiency E	Incremental oil per barrel injected; dimensionless headline metric.
Bottomhole pressures	P_{wf} for both injectors; verifies the fracture constraint with margin.
Voidage ratio	VRR_w at the optimum; verifies the pressure-maintenance band.

Channeling penalty	Value of P_{chan} at the optimum; quantifies residual channeling exposure.
Converged flag	Boolean audit that the solver reached a feasible optimum.

IV.2.4.6 Optimization result

1) The optimal allocation

For the current field state the optimizer recommends $N_{A7X} = 3819$ STB/d and $N_{A12X} = 3055$ STB/d (**Figure IV.13**), a combined 6874 STB/d. The allocation favours N_{A7X} for two reinforcing reasons: it has the higher field-wide connectivity (notably the strong N_{P18X} – N_{A7X} pair) and the higher injectivity index, so it can accept more water before reaching the fracture-pressure ceiling. At this allocation the predicted incremental oil volume in a year is be **222285 STB**.

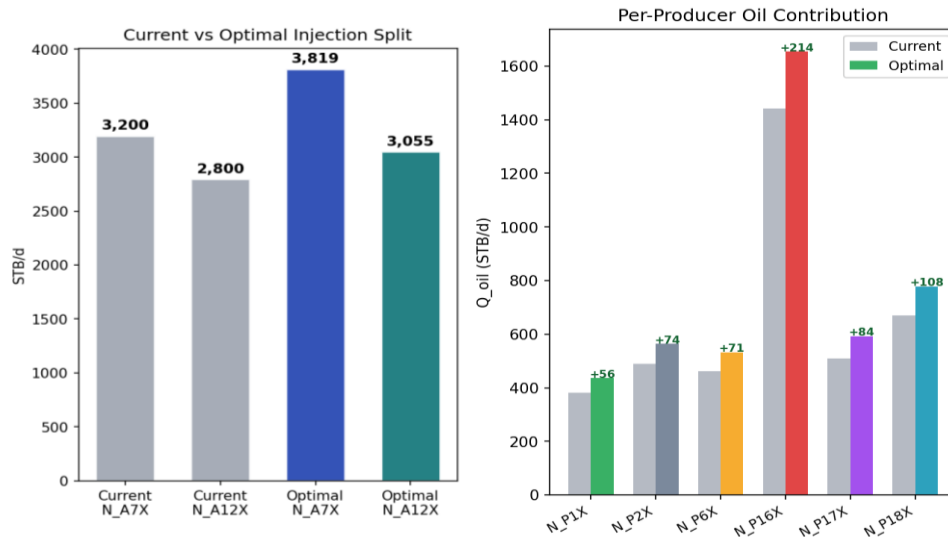


Figure IV.13 The optimum I njection-allocation and oil rate gain

Conclusion

This chapter presented a comprehensive, hybrid Machine Learning framework for waterflooding optimization in the EMN sub-field of the El Merk field. By integrating a physics-informed Capacitance-Resistance Model (CRM) for injector-producer connectivity, an LSTM-based fractional flow forecaster, a field-trained Physics-Informed Neural Network (PINN) for solving the heterogeneous Buckley-Leverett frontal advance equation, and a constrained non-linear optimizer,

the developed system addresses the core challenges of inefficient injection allocation, premature water breakthrough, and suboptimal sweep efficiency.

The modular architecture successfully combines data-driven pattern recognition with reduced-physics models and hard operational constraints. Key outcomes include:

- Robust estimation of inter-well connectivity and response times via the hybrid CRM, revealing distinct drainage behaviors across the six producers (notably fast-channeling indicators in N_P16X).
- High-accuracy monthly water-cut forecasts from the LSTM model ($R^2 = 0.944$ on the test set).
- Detailed saturation profiles, flood-front tracking, and sweep efficiency diagnostics from the PINN models.
- An actionable monthly optimization workflow using SLSQP that respects injection budget, pump limits, voidage replacement, and most critically the formation fracture pressure constraint.

For the current field state, the optimizer recommends an optimal injection allocation of 3,819 STB/d for N_A7X and 3,055 STB/d for N_A12X (total 6,874 STB/d). This allocation is expected to deliver approximately 222,285 STB of incremental oil over the next year while maintaining safe operating margins and mitigating channeling risks.

The proposed system represents a practical step toward real-time, data-driven reservoir management. It moves beyond traditional decline-curve or simulation-heavy approaches by delivering transparent, auditable, and frequently updated injection recommendations that directly maximize injection efficiency (incremental oil per barrel injected).

General Conclusion

This dissertation developed a hybrid physics-informed machine learning system for optimizing waterflooding in the EMN sub-field of El Merk field. The framework integrates a Capacitance-Resistance Model (CRM) for injector-producer connectivity, an LSTM model for fractional flow forecasting ($R^2 = 0.944$), Physics-Informed Neural Networks (PINN) for solving the heterogeneous Buckley-Leverett equation ($R^2 = 0.978$), and a constrained SLSQP optimizer.

The modular system operates on a monthly receding-horizon basis and successfully addresses key challenges such as inefficient injection allocation, premature water breakthrough, and suboptimal sweep efficiency. For current field conditions, the optimizer recommends 3,819 STB/d for N_A7X and 3,055 STB/d for N_A12X, predicting an incremental oil gain of 222,285 STB over one year while respecting fracture pressure and other operational constraints.

This work demonstrates that combining reduced-physics models with data-driven machine learning provides a practical, fast, and accurate alternative to traditional reservoir simulation for real-time waterflood management. The system enhances decision-making, improves injection efficiency, and supports higher oil recovery in mature Algerian fields.

Future work will focus on field implementation, continuous recalibration, and extended waterflooding optimization scope and features.

This hybrid approach represents a valuable contribution to digital transformation in reservoir engineering and offers strong potential for increasing recovery factors while reducing operational costs.

References

Ahmed, T. (2019). *Reservoir engineering handbook (5th ed.)*. Gulf Professional Publishing.

Buckley, S. E., & Leverett, M. C. (1942). *Mechanism of fluid displacement in sands*. *Transactions of the AIME*, 146(1), 107–116.

Corey, A. T. (1954). *The interrelation between gas and oil relative permeabilities*. *Producers Monthly*, 19(1), 38–41.

Craig, F. F. (1971). *The reservoir engineering aspects of waterflooding*. Society of Petroleum Engineers.

Green, D. W., & Willhite, G. P. (2018). *Enhanced oil recovery (2nd ed.)*. Society of Petroleum Engineers.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning with applications in R*. Springer.

Lake, L. W., Johns, R. T., Rossen, W. R., & Pope, G. A. (2014). *Fundamentals of enhanced oil recovery*. Society of Petroleum Engineers.

Muskat, M. (1949). *Physical principles of oil production*. McGraw-Hill.

Nocedal, J., & Wright, S. J. (2006). *Numerical optimization (2nd ed.)*. Springer.

Occidental Petroleum Corporation. (2025). *Algeria fast facts*.

https://www.oxy.com/siteassets/documents/publications/fast-facts/oxy_fast_facts_algeria.pdf

Pandey, Y. N., Rastogi, A., Kainkaryam, S., Bhattacharya, S., & Saputelli, L. (2020). *Machine learning in the oil and gas industry: Including geosciences, reservoir engineering, and production engineering with Python*. Apress.

Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). *Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear PDEs*. *Journal of Computational Physics*, 378, 686–707.

Saoud, H. (2013). *Rapport de stage*. Kasdi Merbah University.

Satter, A., & Thakur, G.C. (1998). *Integrated Waterflood Asset Management*. PennWell Books.

Scikit-Learn Developers. (n.d.). *Preprocessing data documentation*. Scikit-Learn. <https://scikit-learn.org>

Seright, R. S. (2017). *How much polymer should be injected during a polymer flood? Review of previous and current practices*. SPE Journal, 22(1), 1–18.

Sydansk, R. D., & Romero-Zerón, L. (2011). *Reservoir conformance improvement*. Society of Petroleum Engineers.

TensorFlow Developers. (n.d.). *TensorFlow documentation*. <https://www.tensorflow.org>

Terwilliger, P. L., Wilsey, L. E., Hall, H. N., Bridges, P. M., & Morse, R. A. (1951). *An experimental and theoretical investigation of gravity segregation performance*. Transactions of the AIME, 192, 285–298.

Thomas, S., Mahoney, J., & Winter, G. (1989). *Enhanced oil recovery: An assessment of the potential for enhanced oil recovery from known fields in Alberta*. Alberta Energy Resources Conservation Board.

Welge, H. J. (1952). *A simplified method for computing oil recovery by gas or water drive*. Transactions of the AIME, 195(1), 91–98.

Willhite, G. P. (1986). *Waterflooding*. Society of Petroleum Engineers.

You, J., Ampomah, W., Sun, Q., & Balch, R. (2021). *Capacitance resistance models in reservoir engineering: A review*. Journal of Petroleum Science and Engineering, 204, 108744.

APPENDIX

A1																			
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	ELM_Well	Tests																	
2	WELLS	TIME	QOP_W	QWP_W	QGP_W	QGLG_W	GOR_W	WCUT_W	WHP_W	WHT_W									
3	WELLS	DATE	STB / DAY	STB / DAY	MSCF / DAMSCF	DA MSCF / STB	FRACTION	psia	F										
4	N_P1X	#####	7666	0	0	0	0	1	1638.65										
5	N_P1X	#####	7525	0	0	0	0	1	1639.65										
6	N_P1X	#####	7525	0	5215	0	693	0	1639.65	161									
7	N_P1X	#####	6734	0	6512	0	967	0	1654.65	179									
8	N_P1X	#####	3156	0	2970	0	941	0	1714.65	165									
9	N_P1X	#####	3213	0	2933	0	913	0	1711.65	165									
10	N_P1X	#####	6085	0	3949	0	649	0	1651.65	177									
11	N_P1X	#####	3909	0	2513	0	643	0	1687.65	170									
12	N_P1X	#####	3991	0	2522	0	632	0	1687.65	170									
13	N_P1X	#####	6815	0	5234	0	768	0	1608.65	180									
14	N_P1X	#####	8930	0	5751	0	644	0	1542.65	184									
15	N_P1X	#####	1402	0	789	0	563	0	1689.65	143									
16	N_P1X	#####	9849	0	6825	0	693	0	1479.65	167									
17	N_P1X	#####	4333	0	2348	0	542	0	1664.65	171									
18	N_P1X	#####	5934	0	2955	0	498	0	1645.65	176									
19	N_P1X	#####	9984	0	3614	0	362	0	1516.65	174									

EMN well test data

A1																			
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	LGR																		
2	ROOT	31	1	1	1337434	10994300	-8843.02		144.3988										
3	ROOT	23	2	1	1334181	10990153	-8813.48		143.7406										
4	ROOT	26	2	1	1335577	10991543	-8823.63		155.4346										
5	ROOT	27	2	1	1336042	10992008	-8822.02		109.4846										
6	ROOT	28	2	1	1336506	10992470	-8822.78		183.6839										
7	ROOT	29	2	1	1336970	10992930	-8822.66		0.0473										
8	ROOT	30	2	1	1337434	10993393	-8821.19		18.3655										
9	ROOT	31	2	1	1337897	10993855	-8822.28		17.0919										
10	ROOT	32	2	1	1338359	10994315	-8826.06		178.6334										
11	ROOT	34	2	1	1339284	10995238	-8834.77		0.0976										
12	ROOT	35	2	1	1339746	10995700	-8843.7		0.0473										
13	ROOT	36	2	1	1340207	10996160	-8851.7		0.01										
14	ROOT	21	3	1	1333709	10988778	-8779.76		25.3987										
15	ROOT	22	3	1	1334175	10989243	-8791.04		221.8593										
16	ROOT	23	3	1	1334641	10989708	-8801.39		203.6322										
17	ROOT	24	3	1	1335107	10990173	-8808.27		21.3024										
18	ROOT	27	3	1	1336504	10991563	-8807.55		37.7264										
19	ROOT	28	3	1	1336969	10992025	-8806.32		93.4097										

EMN porosity data

A1																			
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	LGR																		
2	ROOT	31	1	1	1337434	10994300	-8843.02		0.0572										
3	ROOT	23	2	1	1334181	10990153	-8813.48		0.0518										
4	ROOT	26	2	1	1335577	10991543	-8823.63		0.0864										
5	ROOT	27	2	1	1336042	10992008	-8822.02		0.1125										
6	ROOT	28	2	1	1336506	10992470	-8822.78		0.0568										
7	ROOT	29	2	1	1336970	10992930	-8822.66		0.0881										
8	ROOT	30	2	1	1337434	10993393	-8821.19		0.0936										
9	ROOT	31	2	1	1337897	10993855	-8822.28		0.064										
10	ROOT	32	2	1	1338359	10994315	-8826.06		0.064										
11	ROOT	34	2	1	1339284	10995238	-8834.77		0.0369										
12	ROOT	35	2	1	1339746	10995700	-8843.7		0.0363										
13	ROOT	36	2	1	1340207	10996160	-8851.7		0.0093										
14	ROOT	21	3	1	1333709	10988778	-8779.76		0.04										
15	ROOT	22	3	1	1334175	10989243	-8791.04		0.0393										
16	ROOT	23	3	1	1334641	10989708	-8801.39		0.0447										
17	ROOT	24	3	1	1335107	10990173	-8808.27		0.0354										
18	ROOT	27	3	1	1336504	10991563	-8807.55		0.1164										
19	ROOT	28	3	1	1336969	10992025	-8806.32		0.1185										

EMN permeability data

```
hydrocarbon_PVT.txt - Notepad
File Edit Format View Help
Hydrocarbon PVT, no free gas initially:
BOF ~1.483 rb/stb
Oil density ~41.3 lb/ft3
RS ~835 scf/stb
```

EMN PVT data

```
C:\Users\abdel\Downloads...
<?xml version="1.0" encoding="UTF-8"?>
<b208_SIM_WELLS>
  <VERSION Number="1.00"/>
  <WELLLIST>
    <WELL KB="237.3" UTMX="3350714.72" UTMX="412242.87" Station="F-EMNT-02" Res="EMN" REGION="T-EMN-N" CurrentStatus="OIL" SimName="N_P1X" Name="EMN-001"/>
    <WELL KB="248.6" UTMX="3347613.83" UTMX="409227.87" Station="F-EMNT-01" Res="EMN" REGION="T-EMN-S" CurrentStatus="OIL" SimName="N_P2X" Name="EMN-002"/>
    <WELL KB="245.2" UTMX="3345152.19" UTMX="407288.13" Station="F-EMNT-01" Res="EMN" REGION="T-EMN-S" CurrentStatus="OIL" SimName="N_P6X" Name="EMN-006"/>
    <WELL KB="238.5" UTMX="3349618.41" UTMX="409388.64" Station="F-EMNT01" Res="EMN" REGION="T-EMN-N" CurrentStatus="WI" SimName="N_A7X" Name="EMN-007"/>
    <WELL KB="241.9" UTMX="3348280.24" UTMX="407569.92" Station="F-EMNT01" Res="EMN" REGION="T-EMN-S" CurrentStatus="OBS" SimName="N_A8X" Name="EMN-008"/>
    <WELL KB="255.77" UTMX="3350990.08" UTMX="411245.15" Station="F-EMNT01" Res="EMN" REGION="T-EMN-N" CurrentStatus="WI" SimName="N_A12X" Name="EMN-012"/>
    <WELL KB="279.5" UTMX="3346345.08" UTMX="408222.06" Station="G-EMN" Res="EMN" REGION="T-EMN-S" CurrentStatus="GI" SimName="N_A10X" Name="EMN-010"/>
    <WELL KB="286.4" UTMX="3349494.14" UTMX="411259.785" Station="F-EMNT-02" Res="EMN" REGION="T-EMN-N" CurrentStatus="OIL" SimName="N_P16X" Name="EMN-016"/>
    <WELL KB="280.88" UTMX="3348339.46" UTMX="410130.22" Station="F-EMNT-01" Res="EMN" REGION="T-EMN-N" CurrentStatus="OIL" SimName="N_P17X" Name="EMN-017"/>
    <WELL KB="283.68" UTMX="3351569.98" UTMX="413128.958" Station="F-EMNT-02" Res="EMN" REGION="T-EMN-N" CurrentStatus="OIL" SimName="N_P18X" Name="EMN-018"/>
    <WELL KB="253.57" UTMX="3352919.77" UTMX="414555.57" Station="G-EMN" Res="EMN" REGION="T-EMN-N" CurrentStatus="GI" SimName="N_A19X" Name="EMN-019"/>
    <WELL KB="276.59" UTMX="3349591.03" UTMX="409985.09" Station="G-EMN" Res="EMN" REGION="T-EMN-N" CurrentStatus="GI" SimName="N_A13X" Name="EMN-013"/>
    <WELL KB="237.3" UTMX="3350714.72" UTMX="412242.87" Station="G-EMN" Res="EMN" REGION="T-EMN-N" CurrentStatus="GI" SimName="N_A1X" Name="EMN-001"/>
  </WELLLIST>
  <OBSLIST/>
</b208_SIM_WELLS>
```

EMN well Coordonates data

```
relpmOilLeg (1).dat - Notepad
File Edit Format View Help
!
WOTABLE
SW KRW KROW PCWO
0.15000 0.0000000000 0.8000000000 240.010000
0.15195 1.1314e-6 0.7910400000 227.600000
0.15390 5.1986e-6 0.7821600000 216.090000
0.15780 2.3886e-5 0.7646200000 195.610000
0.16560 0.0001097500 0.7304800000 162.580000
0.17340 0.0002678000 0.6975300000 136.830000
0.18120 0.0005042900 0.6657500000 116.570000
0.18900 0.0008239200 0.6351100000 100.390000
0.20460 0.0017273000 0.5771200000 77.198000
0.22800 0.0037857000 0.4979400000 54.736000
0.26700 0.0092374000 0.3850100000 34.290000
0.30600 0.0173950000 0.2930900000 22.912000
0.34500 0.0284200000 0.2192100000 16.303000
0.38400 0.0424440000 0.1607100000 12.147000
0.42300 0.0595800000 0.1151300000 9.372900
0.46200 0.0799250000 0.0803100000 7.435300
0.50100 0.1035700000 0.0542900000 6.031300
0.54000 0.1305800000 0.0353550000 4.973700
0.57900 0.1610500000 0.0220060000 4.167200
0.61800 0.1950200000 0.0129530000 3.929100
0.65700 0.2325700000 0.0071023000 3.720300
0.69600 0.2737600000 0.0035492000 3.534100
0.73500 0.3186300000 0.0015625000 3.366600
0.77400 0.3672400000 0.0005724300 3.215200
0.81300 0.4196400000 0.0001568600 3.077600
0.89100 0.5359700000 1.118e-6 2.839000
0.90660 0.5611100000 1.1224e-7 2.797100
0.92220 0.5868800000 8.e-10 2.755200
0.93000 0.6000000000 0.0000000000 2.734200
```

EMN relative permeability data

K10																			
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	WELLS	TIME	QOP_W	COP_W	QWP_W	CWP_W	QGP_W	CGP_W	GOR_W	WCUT_W	QGLG_W	CGLG_W	QDW_W	CDW_W	QWI_W	CWI_W	QGI_W	CGI_W	THP_W
2	WELLS	DATE	STB / DAY	MSTB	STB / DAY	MSTB	MSCF / DA	MMSCF	SCF / STB	FRACTION	MSCF / DA	MMSCF	STB / DAY	MSTB	STB / DAY	MSTB	MSCF / DA	MMSCF	psia
3	N_P1X	#####	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	N_P1X	#####	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	N_P1X	#####	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	N_P1X	#####	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	N_P1X	#####	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	N_P1X	#####	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	N_P1X	#####	506.49	0.5065	0	0	351	0.351	693	0	0	0	0	0	0	0	0	0	1653
10	N_P1X	#####	0	0.5065	0	0	0	0.351	0	0	0	0	0	0	0	0	0	0	0
11	N_P1X	#####	0	0.5065	0	0	0	0.351	0	0	0	0	0	0	0	0	0	0	0
12	N_P1X	#####	0	0.5065	0	0	0	0.351	0	0	0	0	0	0	0	0	0	0	0
13	N_P1X	#####	0	0.5065	0	0	0	0.351	0	0	0	0	0	0	0	0	0	0	0
14	N_P1X	#####	0	0.5065	0	0	0	0.351	0	0	0	0	0	0	0	0	0	0	0
15	N_P1X	#####	0	0.5065	0	0	0	0.351	0	0	0	0	0	0	0	0	0	0	0
16	N_P1X	#####	0	0.5065	0	0	0	0.351	0	0	0	0	0	0	0	0	0	0	1646
17	N_P1X	#####	0	0.5065	0	0	0	0.351	0	0	0	0	0	0	0	0	0	0	0
18	N_P1X	#####	1599.71	2.1062	0	0	1108.6	1.4596	693	0	0	0	0	0	0	0	0	0	1716
19	N_P1X	#####	5726.7	7.8329	0	0	3968.6	5.4282	693	0	0	0	0	0	0	0	0	0	1700

EMN production data

K14																			
	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
1	WCUT_W	QGLG_W	CGLG_W	QDW_W	CDW_W	QWI_W	CWI_W	QGI_W	CGI_W	THP_W	BHP_W								
2	FRACTION	MSCF / DA	MMSCF	STB / DAY	MSTB	STB / DAY	MSTB	MSCF / DA	MMSCF	psia	psia								
3	0	0	0	3592	3.592	0	0	0	0	638									
4	0	0	0	4761	8.353	0	0	0	0	1639									
5	0	0	0	10734	19.087	0	0	0	0	1827									
6	0	0	0	7724	26.811	0	0	0	0	1920									
7	0	0	0	2663	29.474	0	0	0	0	1521									
8	0	0	0	2342	31.816	0	0	0	0	1425									
9	0	0	0	5380	37.196	0	0	0	0	1636									
10	0	0	0	7082	44.278	0	0	0	0	1800									
11	0	0	0	33	44.311	0	0	0	0	1165									
12	0	0	0	3500	47.811	0	0	0	0	1470									
13	0	0	0	10069	57.88	0	0	0	0	1845									
14	0	0	0	7846	65.726	0	0	0	0	1813									
15	0	0	0	4544	70.27	0	0	0	0	1834									
16	0	0	0	3820	74.09	0	0	0	0	1665									
17	0	0	0	10408	84.498	0	0	0	0	1856									
18	0	0	0	3917	88.415	0	0	0	0	1900									
19	0	0	0	2735	91.15	0	0	0	0	1827									

EMN water injection data

A1																			
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	WELLS	TIME	BHP_W																
2	WELLS	DATE	PSIA																
3	N_P1X	#####	4351.28																
4	N_P1X	#####	4256.13																
5	N_P1X	#####	4265.76																
6	N_P1X	#####	4259.41																
7	N_P1X	#####	4274.97																
8	N_P1X	#####	4350.31																
9	N_P1X	#####	4345.44																
10	N_P1X	#####	4280.54																
11	N_P1X	#####	4170.34																
12	N_P1X	#####	4145.94																
13	N_P1X	#####	4136.49																
14	N_P1X	#####	4284.42																
15	N_P1X	#####	4286.43																
16	N_P1X	#####	4246.56																
17	N_P1X	#####	4237.29																
18	N_P1X	#####	4231.18																
19	N_P1X	#####	4228.3																

EMN well producers BHFP