

الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي والبحث العلمي



Sétif 1 Université - Ferhat ABBAS
Faculté des Sciences
Département de mathématiques



Mémoire de Master

DOMAINE : Mathématiques et Informatique
FILIÈRE : Mathématique
SPÉCIALITÉ : Mathématiques Appliquées

Thème

Modèles géostatistiques hybrides et Machine-learning
pour la prédiction des propriétés
pétrophysiques des réservoirs sous R

Présenté par :
ROUAI Hichem

encadré par :
Pr. GRICHE Sonia

Mémoire Soutenu le 21 juin 2026, devant le jury composé de :

Présidente : Pr. AMRANI Naima
Encadrante : Pr. GRICHE Sonia
Examineur : Pr. KHABER Alarbi

Université Sétif 1 Ferhat Abbas
Université Sétif 1 Ferhat Abbas
Université Sétif 1 Ferhat Abbas

Promotion : 2025/2026

Chapitre 1

Introduction générale et cadre méthodologique

1.1 Contexte général et cadre numérique

1.2 Problématique

1.3 Objectifs de l'étude

1.4 Méthodologie générale

1.4.1 Rôle principal de l'environnement R dans l'étude

1.4.2 Données utilisées

1.4.3 Organisation du mémoire

1.4.4 Conclusion du chapitre

Chapitre 2

Analyse exploratoire des données (approche sous R)

2.1 Introduction

2.2 Description des données

2.3 Préparation des données

2.4 Analyse statistique descriptive

2.5 Analyse des distributions

2.6 Transformation des variables

2.7 Analyse des corrélations

2.8 Analyse spatiale des données

2.9 Détection des valeurs aberrantes

2.10 Première approche de la continuité spatiale

2.11 Synthèse générale du chapitre

2.12 Conclusion

Chapitre 3

Modélisation de la structure spatiale par le variogramme sous R

3.1 Introduction

3.2 Variable régionalisée et dépendance spatiale

3.3 Hypothèse de stationnarité

3.4 Définition du variogramme

3.5 Préparation spatiale des données sous R

3.6 Calcul du variogramme expérimental

3.7 Paramètres du variogramme

3.7.1 Effet pépite

3.7.2 Palier

3.7.3 Portée

3.8 Modélisation théorique du variogramme

3.8.1 Modèle sphérique

3.8.2 Modèle exponentiel

3.8.3 Modèle gaussien

3.9 Ajustement du modèle sous R

3.10 Choix du modèle le plus adapté

3.11 Analyse de l'anisotropie₄

3.12 Discussion des résultats

Chapitre 4

Interpolation spatiale par krigeage

4.1 Introduction

4.2 Définition rigoureuse du krigeage

4.3 Condition de non-biais

4.4 Minimisation de la variance de l'erreur

4.5 Système du krigeage ordinaire

4.6 Les différents types de krigeage

4.6.1 Krigeage simple

4.6.2 Krigeage ordinaire

4.6.3 Krigeage universel

4.6.4 Krigeage indicateur

4.6.5 Co-krigeage

4.7 Mise en œuvre avancée sous R

4.7.1 Krigeage ordinaire : prédiction et incertitude

4.7.2 Validation croisée (Leave-One-Out)

4.7.3 Analyse des erreurs

4.7.4 Krigeage universel (avec dérive)

4.7.5 Anisotropie

4.7.6 Krigeage avec covariables

4.8 Discussion scientifique

4.9 Conclusion enrichie

Chapitre 5

Modélisation des propriétés pétrophysiques par apprentissage automatique et approche hybride

5.1 Introduction

5.2 Définitions générales du Machine Learning

5.3 Intérêt du Machine Learning dans l'étude des réservoirs

5.4 Données et formulation du problème

5.5 Principaux modèles de Machine Learning utilisés

5.5.1 Régression linéaire

5.5.2 Forêt aléatoire (Random Forest)

5.5.3 XGBoost

5.6 Implémentation sous R

5.6.1 Préparation des données sous R

5.6.2 Séparation apprentissage / test

5.6.3 Régression linéaire sous R

5.6.4 Random Forest sous R

5.6.5 Évaluation des performances sous R

5.6.6 Importance des variables sous R

5.7 Interprétation des résultats des modèles

5.8 Limites des approches purement Machine Learning

5.9 Approche hybride : principe général

Chapitre 6

Analyse de l'incertitude et simulation géostatistique

6.1 Introduction

6.2 Notion d'incertitude en géostatistique

6.3 Limites du krigeage pour la variabilité

6.4 Simulation gaussienne séquentielle (SGS)

6.5 Implémentation sous R

6.6 Validation croisée des modèles

6.7 Cartes d'incertitude

6.8 Analyse des risques

6.9 Recommandations

6.10 Discussion générale

6.11 Conclusion

Chapitre 7

Discussion générale, limites et perspectives

7.1 Introduction

7.2 Synthèse des résultats obtenus

7.3 Apports de la géostatistique

7.4 Apports du Machine Learning

7.5 Intérêt de l'approche hybride

7.6 Apport de la simulation géostatistique

7.7 Limites du travail

7.8 Perspectives de recherche

7.9 Conclusion du chapitre

7.10 Importance des environnements R et Python dans la modélisation géostatistique et l'apprentissage automatique

7.10.1 Importance de R en géostatistique

7.10.2 Importance de Python dans le Machine Learning

7.10.3 Complémentarité entre R et Python

7.10.4 Apport scientifique de l'utilisation conjointe

7.10.5 Conclusion sur R et Python

Conclusion générale

Annexe A

Préparation et traitement des données

A.1 Chargement des données sous R

A.2 Nettoyage des données sous R

A.3 Transformation logarithmique de la perméabilité

Annexe B

Calcul et ajustement du variogramme sous R

B.1 Variogramme expérimental

B.2 Ajustement du modèle théorique

B.3 Analyse de l'anisotropie

Annexe C

Krigeage

C.1 Krigeage sous R

C.2 Cartes de prédiction

C.3 Cartes de l'incertitude

C.4 Validation croisée

Annexe D

Machine Learning sous R

D.1 Régression linéaire

D.2 XGBoost sous R

D.3 Random Forest sous R

Annexe E

Simulation géostatistique sous R

E.1 Simulation gaussienne géostatistique sous R

E.2 Cartes de l'incertitude

E.3 Cartes des scénarios

Annexe F

Validation et métriques sous R

F.1 RMSE

F.2 Coefficient de R

Annexe G

Organisation des fichiers

Glossaire

Bibliographie

[1]

Résumé

Ce mémoire vise à développer une méthodologie intégrée pour améliorer la modélisation des propriétés pétrophysiques des réservoirs pétroliers, notamment la perméabilité et la porosité, en combinant les approches géostatistiques (telles que le krigeage) et les techniques de Machine Learning (telles que la forêt aléatoire et XGBoost), en utilisant principalement l'environnement R et Python de manière complémentaire.

La méthodologie adoptée comprend plusieurs étapes progressives : analyse exploratoire des données, modélisation de la structure spatiale par variogramme, interpolation par krigeage, développement de modèles de Machine Learning, construction d'une approche hybride combinant les avantages des deux méthodes, et simulation géostatistique pour l'analyse de l'incertitude et l'évaluation des risques.

Les résultats montrent que les variables, en particulier la perméabilité, présentent une distribution asymétrique et une forte variabilité spatiale, nécessitant une transformation logarithmique. L'analyse du variogramme a confirmé l'existence d'une structure spatiale significative. Les modèles de Machine Learning, notamment la forêt aléatoire, surpassent la régression linéaire en précision prédictive.

L'approche hybride a fourni les meilleurs résultats, combinant la puissance prédictive du Machine Learning avec la cohérence spatiale du krigeage, produisant des cartes plus précises et réalistes. La simulation géostatistique a permis d'identifier les zones à forte incertitude et d'évaluer les risques via le calcul des percentiles P10 et P90.

Cette méthodologie peut être appliquée à d'autres domaines tels que l'hydrogéologie, la gestion des ressources naturelles et l'environnement. Elle ouvre également des perspectives pour le développement de modèles plus avancés comme les réseaux de neurones profonds et les jumeaux numériques des réservoirs pétroliers.

Mots clés : Géostatistique, Machine Learning, modèle hybride, krigeage, forêt aléatoire, simulation géostatistique, perméabilité, porosité, R.

Abstract

This thesis aims to develop an integrated methodology to improve the modeling of petrophysical properties of oil reservoirs, particularly permeability and porosity, by combining geostatistical approaches (such as kriging) and Machine Learning techniques (such as Ran-

dom Forest and XGBoost), primarily using the R environment with Python as a complementary tool.

The adopted methodology includes several progressive steps : exploratory data analysis, spatial structure modeling using variogram, kriging interpolation, development of Machine Learning models, construction of a hybrid approach combining the advantages of both methods, and geostatistical simulation for uncertainty analysis and risk assessment.

Results show that the variables, especially permeability, exhibit an asymmetric distribution and high spatial variability, requiring a logarithmic transformation. Variogram analysis confirmed the existence of a significant spatial structure. Machine Learning models, particularly Random Forest, outperform linear regression in predictive accuracy.

The hybrid approach provided the best results, combining the predictive power of Machine Learning with the spatial consistency of kriging, producing more accurate and realistic maps. Geostatistical simulation helped identify high-uncertainty zones and assess risks through the calculation of P10 and P90 percentiles.

This methodology can be applied to other fields such as hydrogeology, natural resource management, and environmental studies. It also opens perspectives for developing more advanced models such as deep neural networks and digital twins of oil reservoirs.

Keywords : Geostatistics, Machine Learning, hybrid model, kriging, Random Forest, geostatistical simulation, permeability, porosity, R.

الملخص

يهدف هذا البحث إلى تطوير منهجية متكاملة لتحسين نمذجة الخصائص البتروفيزيائية للمكامن النفطية، وبشكل خاص النفاذية والمسامية، وذلك من خلال الجمع بين الأساليب الجيواستاتيسية (مثل الكريجياج) وتقنيات التعلم الآلي (مثل الغابة العشوائية و XGBoost) بشكل تكميلي Python بشكل أساسي و R باستخدام بيئة البرمجة.

تعتمد المنهجية المتبعة على عدة مراحل متدرجة: بدءاً بالتحليل الاستكشافي للبيانات، مروراً بنمذجة البنية المكانية عبر ثم التنبؤ بالكريجياج، وتطوير نماذج التعلم الآلي، وأخيراً بناء نموذج هجين يجمع بين مزايا الطريقتين، إلى جانب Variogram، المحاكاة الجيواستاتيسية لتحليل عدم اليقين وتقييم المخاطر.

أظهرت النتائج أن المتغيرات، وخاصة النفاذية، تتسم بتوزيع غير طبيعي وتباين مكاني كبير، مما استلزم تحويلها لوغاريتمياً. كما وجود بنية مكانية معنوية. وقد تفوقت نماذج التعلم الآلي، خصوصاً الغابة العشوائية، على الانحدار Variogram أكد تحليل الخطي من حيث الدقة التنبؤية.

أما المقاربة الهجينة فقد قدمت أفضل النتائج، حيث جمعت بين القوة التنبؤية للتعلم الآلي والاتساق المكاني للكريجياج، مما أفضى إلى خرائط أكثر دقة وواقعية. وساهمت المحاكاة الجيواستاتيسية في تحديد مناطق عدم اليقين العالي وتقدير المخاطر عبر حساب P10 و P90 المئينات.

يمكن تطبيق هذه المنهجية في مجالات أخرى مثل الهيدروجيولوجيا وإدارة الموارد الطبيعية والبيئة، كما تفتح آفاقاً لتطوير نماذج أكثر تقدماً كالشبكات العصبية العميقة والتوأمة الرقمية للمكامن النفطية.

، كلمات مفتاحية: جيواستاتيسيك، تعلم الآلة، نموذج هجين، كريجياج، غابة عشوائية، محاكاة جيواستاتيسية، نفاذية، مسامية R.