

N° Ordre/Faculé/UMBB/2021.

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

UNIVERSITE M'HAMED BOUGARA DE BOUMERDES



Faculté des Hydrocarbures et de la Chimie

Département Gisements Miniers et Pétroliers

Mémoire de fin d'études

En vue de l'obtention du diplôme :

MASTER

Présenté par :

Abderrahmane TAIBI

Filière : Hydrocarbures

Option : Génie pétrolier : Production des hydrocarbures.

Thème

Prévision de la production des puits dans les réservoirs de schiste par l'apprentissage automatique et l'analyse des courbe de déclin.

Devant le jury :

BENYOUNES Khaled	Professeur	UMBB	Président
GARECH Mourad	Professeur	UMBB	Examineur
SAIFI Redah	MAA	UMBB	Examineur

ZERAIBI Noureddine	Professeur	UMBB	Encadreur
--------------------	------------	------	-----------

Année Universitaire : 2020/2021

Remerciements

Tout au long de la rédaction de cet mémoire, j'ai reçu beaucoup de soutien et d'aide.

En tout premier lieu, je remercie le bon Dieu, tout puissant, de m'avoir donné la force pour survivre, ainsi que l'audace pour dépasser toutes les difficultés,

Je tiens tout d'abord à remercier mon encadreur, le Pr. ZERAIBI Nouredine, dont l'expertise m'a été précieuse pour formuler les questions de recherche et la méthodologie.

Vos commentaires perspicaces m'ont poussée à affiner ma réflexion et ont porté mon travail à un niveau supérieur.

J'aimerais particulièrement distinguer mon superviseur à SMARTTEST, Eng. AOUANOUK Lotfi. Lotfi, je tiens à vous remercier pour votre soutien patient et pour toutes les occasions qui m'ont été données de faire avancer mes recherches.

Je tiens également à remercier mes professeurs et enseignants, pour leurs précieux conseils tout au long de mes études. Vous m'avez fourni les outils dont j'avais besoin pour choisir la bonne direction et mener à bien ce travail. Et aussi les remercier pour avoir pris le temps d'examiner mon travail, et lui porter leurs jugements.

Je remercie profondément mes parents, pour leur confiance inconditionnelle, leurs encouragements opportuns et leur patience infinie. C'est leur amour qui m'a relevé lorsque j'étais fatigué.

Enfin, je n'aurais pas pu terminer ce travail sans le soutien de mes amis, Ils n'ont jamais lésiné en quoique ce soit à mon égards. Ils m'ont offert des discussions stimulantes ainsi que d'heureuses distractions pour me reposer l'esprit en dehors de mes recherches.

Abderrahmane TAIBI

*Pour ceux qui ont été là depuis le début et qui n'ont
pas pu assister à l'aboutissement de tous cela.*

A.T.

Résumé

Au cours de la dernière décennie, la production d'hydrocarbures provenant des zones de schiste a augmenté de façon spectaculaire, ce qui a eu un impact considérable sur l'industrie pétrolière. Si cette production accrue a profité à l'industrie, ces ressources non conventionnelles ont en même temps posé de nombreux défis aux évaluateurs de réserves de pétrole et de gaz. L'un de ces défis consiste à prévoir la performance et la durée de vie à long terme de la production de schiste, en particulier de manière rapide et fiable. Dans ce travail nous avons relevé le défi, nous avons étudié ce problème de trois points de vue différents. En premier lieu nous avons mis en évidence, l'incapacité des modèles standards de Arps à donner des résultats fiables, ceci même après l'amputation des valeurs aberrantes. Ensuite, nous avons implémenté un autre modèle empirique, le modèle LGM, qui donna des résultats plus fiables. Ensuite nous avons implémenté le modèle stochastique basé sur la méthode de Box-Jenkins, Le modèle SARIMA. Et nous avons fini le travail avec une méthode d'apprentissage profond, l'architecture Bi-LSTM.

Abstract

In the last decade, the hydrocarbon production from shale plays has increased dramatically, which has greatly impacted the petroleum industry. While this increased production has benefited the industry, at the same time these unconventional resources have presented many challenges to oil and gas reserves evaluators. One of these challenges is predicting long-term shale production performance and life, especially in a timely and reliable manner. In this work we have taken up the challenge, we have studied this problem from three different points of view. First we have highlighted the inability of the standard Arps models to give reliable results, even after the outliers have been amputated. Then we implemented another empirical model, the LGM model, which gave more reliable results. Then we implemented the stochastic model based on the Box-Jenkins method, the SARIMA model. And we finished the work with a deep learning method, the Bi-LSTM architecture.

Contents

Remerciement	ii
Dédicace	iii
Introduction	1
1 État de l'art	3
1.1 Étude de Déclin de Production	3
1.2 Courbes de Déclin	4
1.3 Méthodes Modernes	6
1.4 La problématique des formations à faible Perméabilité	7
1.5 Conclusion	7
2 Étude de Déclin et Problématique	8
2.1 Introduction	8
2.2 Les équations de Arps	8
2.2.1 Déclin Exponentiel	9
2.2.2 Déclin Hyperbolique	10
2.2.3 Déclin Harmonique	13
2.2.4 Comparaison entre les trois Déclins	15
2.3 Modèle de croissance logistique - Logistic Growth Model	16
2.4 Mais encore	17
2.4.1 Conclusion	18
3 Modélisation des Séries Temporelles	19
3.1 Définition	19
3.2 Composent d'une série temporelle	19
3.3 L'Analyse	21
3.4 Séries temporelles et processus stochastiques	21
3.5 Concept de stationnarité	22
3.6 Les modèles stochastiques	23
3.6.1 Qu'est-ce que le modèle Box-Jenkins ?	24
3.7 Synthèse	24
4 Outils Mathématiques	25
4.1 Ajustement de courbe	25
4.1.1 Régression Linéaire Simple	25
Méthode des moindres Carrés	25
4.1.2 Régression Linéaire Multiple	27
Gradient Descent	27
4.1.3 Régression non-linéaire	28
Méthodes d'estimation des modèles non-linéaire	28
4.2 Détection de valeurs aberrantes et Réduction des irrégularités	30
4.2.1 Isolation Forest	30
4.2.2 k-Nearest Neighbor	31
4.2.3 LOESS - Régression Locale	32

4.3	Autoregressive Integrated Moving Average	34
4.3.1	Autorégressive Modèle	35
4.3.2	Moving Average	35
4.3.3	ARIMA	35
4.3.4	SARIMA	37
	Autocorrelation et Partial-autocorrelation	37
4.4	Réseau de Neurones Artificiels et Réseaux de Neurones Récurents	39
4.4.1	Introduction aux Réseaux de Neurones	39
	La structure de réseaux de neurones artificiels	39
	Feedforward	40
	Fonction d'activation	41
	Entraînement du réseau	42
4.4.2	Réseaux de Neurones Récurents	43
	Long Short Term Memory	44
	Gated Reccurent Units	45
	La différence entre les RNN standard (Vanille) , LSTM et GRU	46
4.5	Estimateur (statistique)	47
4.6	Implementation	47
5	Méthodologie, Implémentation, Analyse et Résultats	49
5.1	Présentations des Données et la raison de ce Choix	49
5.2	Méthodes Empiriques	50
5.2.1	Formule de Arps Avant Traitements	50
5.2.2	Formule de Arps Après Traitements	59
5.2.3	LGM Avant Traitements	66
5.2.4	LGM Après Traitements	69
5.2.5	Conclusion Partielle	72
5.3	Modèle SARIMA	72
5.3.1	Identification	73
	Rolling Statistics: Moyenne mobile et Écart type mobile	74
	Autocorrelation et Partial-Autocorrelation	75
5.3.2	Le modèle SARIMA	78
5.3.3	Vérifications	78
5.3.4	Conclusion Partielle	80
5.4	Long Short Term Memory (La Variante Bi-LSTM)	80
5.4.1	Architecture du Modèle	81
	Hyper-paramètres	81
	Entraînement et Train-Test Loss	81
5.4.2	Prédictions	83
5.4.3	Conclusion Partielle	83
5.5	Discussion	84
6	Conclusion Générale et Recommandation	86
	References	88
	Annexe	90

List of Figures

2.1	Courbe Typiques de declin exponentiel. (a) Courbe debit-temps ; (b) Courbe debit-production cumulee	10
2.2	Courbes Typiques de declin hyperbolique. (a) Courbe debit-temps ; (b) Courbe debit-production cumulee ; (c) Courbe D-temp(Example).	11
2.3	Courbes Typiques de declin harmonique. (a) Courbe taux-temps ; (b) Courbe taux-production cumulee	14
2.4	Formes de courbe de declin d'Arps pour un trace cartesien debit-temp	15
2.5	Formes des courbes de declin d'Arps pour un trace cartesien du Cumul-temps	16
3.1	Exemple d'une Série Temporelle et de sa décomposition	20
4.1	Régression Linéaire Simple	26
4.2	Exemple d'une Régression Linéaire à deux variables exogènes; La droite se transforme en Plan	28
4.3	Pseudo Code de L'algorithme de L'Isolation Forest	31
4.4	Pseudo Code de l'algorithme du k-Nearest Neighbours	32
4.5	Graphique de la fonction de poids tri-cubique	33
4.6	Procédure d'un modèle ARIMA	34
4.7	AutoCorrelation et Partial Autocorrelation	38
4.8	Diagramme des couches d'un réseau neuronal	40
4.9	Diagramme du pipeline du réseau neuronal	40
4.10	Les Fonctions d'Activation les plus connus	42
4.11	Forward (gauche) et Backward (droite) passes [20]	43
4.12	La Structure d'un Réseau de Neurones Récurrents[20]	44
4.13	La Structure d'un Réseau Long-Short-Term-Memory	45
4.14	Les 3 types de RNN côte à côte	46
5.1	Puits HS01	51
5.2	Puits HS02	51
5.3	Puits HS04	51
5.4	Puits HS10	52
5.5	Les formules de Arps ajustées aux données du puit HS01	55
5.6	Les formules de Arps ajustées aux données du puit HS02	56
5.7	Les formules de Arps ajustées aux données du puit HS04	57
5.8	Les formules de Arps ajustées aux données du puit HS10	58
5.9	L'algorithme de la Régression Local	59
5.10	Les formules de Arps ajustées aux données du puits HS01 Après Traitements	62
5.11	Les formules de Arps ajustées aux données du puits HS02 Après Traitements	63
5.12	Les formules de Arps ajustées aux données du puits HS04 Après Traitements	64
5.13	Les formules de Arps ajustées aux données du puits HS10 Après Traitements	65
5.14	Modèle de croissance logistique ajusté du puits HS01 Avant Traitements	67
5.15	Modèle de croissance logistique ajusté aux données du puits HS02 Avant Traitements	67
5.16	Modèle de croissance logistique ajusté du puits HS04 Avant Traitements	68
5.17	Modèle de croissance logistique ajusté du puits HS10 Avant Traitements	68
5.18	Modèle de croissance logistique ajusté du puits HS01 Après Traitements	70
5.19	Modèle de croissance logistique ajusté du puits HS02 Après Traitements	70
5.20	Modèle de croissance logistique ajusté du puits HS04 Après Traitements	71
5.21	Modèle de croissance logistique ajusté du puits HS10 Après Traitements	71

5.22	Moyenne mobile et Écart type mobile avec une fenêtre de 30 jours	74
5.23	Autocorrelation et Partial-Autocorrelation des 4 puits	76
5.24	Autocorrelation et Partial-Autocorrelation des 4 puits Après 1 ^{ère} Dérivation	77
5.25	Actuel vs. Prédiction SARIMA	79
5.26	Processus d'entraînement	82
5.27	Actuel vs. Prédiction Bi-LSTM	83

List of Tables

2.1	Les Équations de Arps	15
5.1	Les valeurs de la Racine de l'écart quadratique moyen pour chaque modèle- RMSE	52
5.2	Les valeurs du Taux de déclin et de la Caractéristique de base b en Déclin Hyperbolique	52
5.3	Les valeurs du Taux de déclin et de la Caractéristique de base b en Déclin Exponentiel	53
5.4	Les valeurs du Taux de déclin et de la Caractéristique de base b en Déclin Harmonique	53
5.5	Les valeurs de la Racine de l'écart quadratique moyen pour chaque modèle- RMSE Après Traitement avec LOESS	59
5.6	Les valeurs du Taux de déclin et de la Caractéristique de base b en Déclin Hyperbolique	60
5.7	Les valeurs du Taux de déclin et de la Caractéristique de base b en Déclin Exponentiel	60
5.8	Les valeurs du Taux de déclin et de la Caractéristique de base b en Déclin Harmonique	60
5.9	Les valeurs de la Racine de l'écart quadratique moyen pour le modèle LGM- RMSE Avant Traitement avec LOESS	66
5.10	Les valeurs du K - Capacité de charge, n - Exposant hyperbolique, a - Constante Avant Traitement	66
5.11	Les valeurs de la Racine de l'écart quadratique moyen pour le modèle LGM- RMSE Après Traitement avec LOESS	69
5.12	Les valeurs du K - Capacité de charge, n - Exposant hyperbolique, a - Constante Après Traitement	69
5.13	Les valeurs de la Racine de l'écart quadratique moyen pour le modèle SARIMA	78
5.14	Architecture du réseau Bi-LSTM	81
5.15	Les valeurs de la Racine de l'écart quadratique moyen pour le modèle Bi-LSTM	83

"All models are wrong, but some are useful."

George EP Box

Intorduction

La production des hydrocarbures et l'ingénierie de réservoir sont sans aucun doute le coeur de l'industrie pétrolière. Ces disciplines est la partie de l'ingénierie pétrolière qui tente de maximiser la production de pétrole et de gaz de manière rentable. Pour atteindre cet objectif, les ingénieurs doivent avoir une compréhension approfondie du réservoir d'hydrocarbures, des systèmes de production, et tous ce qui les régit avec lesquels ils travaillent.

L'analyse du déclin de la production (PDA), ou analyse de la production ou analyse du débit transitoire, est devenue un point chaud de l'ingénierie des réservoirs ces dernières années. Basée sur la théorie de la filtration, cette technique peut fournir de nouvelles courbes typiques grâce à l'ingénierie des réservoirs et à l'analyse des essais de puits modernes. Elle est également utilisée pour analyser les données de production quotidiennes et quantifier les paramètres du réservoir, les caractéristiques de percolation et l'OOIP (OGIP) avec la méthode de correspondance des courbes types.

La PDA fait appel à quatre types de méthodes, dont (1) la méthode empirique, par exemple Arps, (2) la méthode d'analyse classique, par exemple Fetkovich, (3) l'analyse de correspondance des courbes de type log-log, par exemple Blasingame, et (4) la méthode d'ingénierie des réservoirs, par exemple *Flowing Materila Balance*. Cette technique, ainsi que l'analyse moderne du cycle de vie des puits, est devenue l'une des principales méthodes de description dynamique des réservoirs, et les logiciels d'analyse basés sur les PDA ont été largement utilisés dans les champs pétroliers et gaziers.

Pour le travail qui va suivre, on s'attardera sur:

- Une étude sur les données de production, un traitement de données adéquats par le biais de méthodes qui permettes la détection des valeurs aberrantes (Anomaly detection) et l'impact que cela aura sur la qualité des résultats
- L'utilisation de méthodes conventionnelles - empirique- afin d'étudier le déclin, telle que: Les courbes de de déclin développées par Arps, et le modèle de croissance logistique (Logistic Growth Model).
- Modélisations des séries temporelles par un modèle stochastique et l'utiliser afin de

faire de la prédiction.

- L'utilisation des réseaux de neurones récurrents, plus précisément les architectures LSTM (Long-short-Term-Memory) et GRU (Gated recurrent Networks) afin de remédier au problème de déclin cité avant.

1 État de l'art

1.1 Étude de Déclin de Production

Pour simuler le pétrole, le gaz et/ou les produits non conventionnels, des mécanismes de transport complexes et la présence de fractures naturelles et/ou hydrauliques doivent être pris en compte, ce qui fait de la simulation des réservoirs une tâche difficile. L'existence de différents mécanismes coexistant au cours du processus de production, tels que la diffusion du gaz, le glissement du gaz, et la désorption du gaz...etc. La complexité de ces mécanismes rend les équations classiques de diffusivité inadéquates pour modéliser le transport. En raison de la complexité de ces mécanismes, les équations classiques de diffusivité ne permettent pas de modéliser le transport. Les modèles de simulation existants n'ont pas entièrement rendu compte de ce mécanisme complexe, et de nouveaux modèles complets et fiables doivent encore être construits et étudiés.

Avec des informations pétrophysiques méticuleusement mesurées, les solutions analytiques, semi-analytiques ou numériques basées sur un modèle de réservoir précis pourraient conduire à des prévisions de production raisonnables. Cependant, il est coûteux et long de réaliser des expériences pétrophysiques afin de construire un modèle de réservoir aussi sophistiqué, car de nombreuses données géologiques et pétrophysiques doivent être recueillies et analysées avant d'être intégrées au modèle de réservoir. Par exemple, des puits d'essai doivent être forés et de multiples processus de diagrapie des puits doivent être effectués afin d'obtenir les données nécessaires (par exemple, l'emplacement de la formation, la profondeur, la résistivité, la perméabilité, les propriétés de fracture, les propriétés du gaz) pour le modèle de réservoir. Ces processus sont coûteux et prennent du temps, et certaines techniques de diagrapie ne sont pas assez sensibles pour mesurer les propriétés multi-échelles du schiste. En revanche, la DCA ne nécessite que les données de production, et peu d'informations sur le réservoir sont nécessaires. La DCA est efficace avec une précision adéquate pour répondre aux besoins industriels. Au cours des dernières décennies, plusieurs modèles DCA ont été proposés et comparés à des simulateurs de réservoirs commerciaux.

Les techniques d'analyse des données de production (PDA) se sont considérablement

améliorées au cours des dernières années. Ces techniques sont utilisées pour fournir des informations sur la perméabilité du réservoir, la longueur des fractures, la conductivité des fractures, la zone de drainage du puits, l'OGIP (original-gasin-place), la récupération ultime estimée (EUR) et le skin. Bien qu'il existe de nombreuses méthodes disponibles identifiées afin de caractériser le réservoir, il n'y a pas une seule méthode claire qui donne toujours la réponse la plus fiable.

1.2 Courbes de Déclin

Arps - Équations de Déclin: L'analyse de la courbe de déclin (DCA) est une méthode qui consiste à ajuster les taux de production observés de puits individuels, de groupes de puits ou de réservoirs par une fonction mathématique afin de prédire les performances de la production future en extrapolant la fonction de déclin ajustée. Arps [1] a introduit la méthode d'analyse de la courbe de déclin dans les années 1940 en utilisant des équations mathématiques. La méthode est une équation mathématique sans base physique autre que l'équation montre une tendance au déclin. La fonction introduite par Arps est caractérisée par trois paramètres : le débit initial (q_i), le taux de déclin initial (D_i) et l'exposant de déclin (b). Lorsque $b = 0$, le déclin est exponentiel. Lorsque $b = 1$, la décroissance est harmonique. Lorsque $0 < b < 1$, la décroissance est dite hyperbolique. La méthode d'Arps est toujours utilisée en raison de sa simplicité et, comme il s'agit d'une méthode empirique, elle n'a pas besoin de paramètres de réservoir ou de puits.

Fetkvoich - Courbes Types: Au début des années 1980, Fetkvoich [4] a introduit l'analyse des courbes de déclin par les courbes de type. L'appariement des courbes types est essentiellement une technique graphique permettant de faire correspondre visuellement les données de production à l'aide de courbes pré tracées sur un papier log-log. Fetkovich a utilisé les courbes de déclin d'Arps ainsi que les courbes types pour l'écoulement transitoire radial symétrique de liquides à faible compressibilité à des pressions de fond de puits constantes. Fetkovich a relié les paramètres de déclin d'Arps à certains paramètres d'ingénierie des réservoirs pour la production à des pressions de fond constantes. Fetkovich a recommandé l'utilisation de courbes de déclin harmoniques pour les puits de gaz.

Carter: En 1985, Carter[5] a présenté les résultats du débit de production de gaz sous forme de courbe type pour un système d'écoulement radial et linéaire fini produisant

à une pression de fond de puits constante. Carter a utilisé une variable λ identifiant l'ampleur du rabattement de pression dans les puits de gaz. Une courbe avec une valeur λ de 1 correspond à $b = 0$ dans les courbes de déclin liquide de Fetkovich et représente une courbe de système liquide avec un déclin exponentiel. Les courbes avec $\lambda = 0,5$ et $0,75$ correspondent à des puits de gaz avec une magnitude croissante de la baisse de pression.

Fraim et Wattenbarger - Linéarisation: En 1987, Fraim Wattenbarger [7] ont introduit une fonction de temps normalisé qui linéarise le déclin du taux en fonction du temps normalisé pour les réservoirs de gaz produisant à des pressions de fond constantes pendant un écoulement dominé par les limites. Le calcul des temps normalisés implique un processus itératif.

Palacio - Blasingame: En 1993, Palacio Blasingame [8] ont abordé la question des pressions de fond variables et non constantes dans les puits de gaz. Ils ont introduit de nouvelles méthodes, qui utilisent une fonction temporelle modifiée pour analyser la performance des puits de gaz. Ils ont également présenté un nouvel algorithme avec la fonction de temps modifiée pour calculer le gaz en place, qui sont capables de modéliser le comportement des données de production pour des conditions de taux variable et/ou de chute de pression variable. Comme le temps normalisé, le calcul du temps pseudo-équivalent est un processus itératif.

Agarwal: En 1999, Agarwal et al.[10] ont introduit de nouvelles courbes types, qui représentent un progrès par rapport aux courbes types de Palacio Blasingame car une distinction plus claire peut être faite entre les périodes d'écoulement transitoire et celles dominées par les frontières.

Modèle de Duong:[13] est introduit sur la base d'une règle empirique, à savoir que le tracé log-log de q/G_p (G_p est la production cumulée de gaz) en fonction de t forme une ligne droite. Les tests ont montré que le modèle de Duong est plus précis pour les écoulements linéaires et les écoulements bilinéaires, ils ont également montré qu'il varie le moins en ce qui concerne la longueur de l'historique de production pour tous les puits parmi toutes les méthodes DCA. En outre, le modèle de Duong tend à fournir les résultats les plus conservateurs.

Modèle de croissance logistique: Le modèle de croissance logistique a été utilisé une

fois par Spencer et Coulombe pour modéliser la repousse des foies. Dans leur expérience, les foies de rats étaient réduits à un tiers de leur taille initiale et semblaient se régénérer de manière hyperbolique. Clark et al.[12] ont adopté ce modèle pour les réservoirs de gaz de schiste à très faible perméabilité et ont développé le modèle de croissance logistique (LGM) comme méthode empirique pour prévoir la production de gaz. Dans ce modèle, la production cumulée va jusqu'à une capacité de charge maximale K et il n'y a plus de croissance après avoir atteint cette valeur maximale.

En raison de leurs expressions mathématiques simples, de leur vitesse d'ajustements rapide et de leur faible dépendance aux données géophysiques des réservoirs, les modèles DCA ont été largement appliqués pour prévoir la production des hydrocarbures à l'aide de données. En tant que technique la plus simple pour la prévision de la production, les modèles DCA n'ont pas besoin de données spécifiques sur les réservoirs ou les fractures. Ils ne nécessitent pas la construction d'un modèle de réservoir compliqué. En utilisant uniquement les données de production, les modèles DCA peuvent renvoyer un historique précis correspondant aux données de production par différents types de courbes et prévoir l'EUR -Estimated Ultimate Recovery- en utilisant des courbes interpolées.

1.3 Méthodes Modernes

Néanmoins vu la sensibilité des données que nous traitons, et leurs importances économique, la recherche de méthodes plus précises, plus flexibles et applicables à un domaine plus large de types de réservoirs, sans perte de précision de calcul est toujours un projet en courts.

Entre autre, En 2018, Sun et al. [19], ont conduit une comparaison entre les méthodes conventionnelles - empirique- et les Réseaux de Neurones Récurents avec leurs structure de bases. Et parallèlement, en 2014, Olominu et Sulaimon [14] ont appliqué une analyse sur séries temporelles afin de faire de la d'améliorer la prédiction sur ces dernières.

1.4 La problématique des formations à faible Perméabilité

Le schiste, en tant que roche sédimentaire, est à grain fin et ne se disjoint pas dans l'eau. Certains universitaires affirment que les schistes sont très imperméables, de sorte que l'on peut supposer que le marbre est une éponge par rapport à eux. Le gaz naturel est généralement placé entre des couches épaisses de schiste. Les méthodes de production les plus reconnues pour le gaz de schiste sont la fracturation hydraulique et le forage horizontal. La fracturation hydraulique est un processus qui ouvre des fractures dans les roches en injectant de l'eau à haute pression, puis en les maintenant ouvertes à l'aide de petites particules/grains, notamment de la silice, du sable et du verre. Le forage horizontal décrit une méthode de forage qui commence par un forage droit dans le sol, puis latéral, et enfin parallèle à la direction de la surface.

C'est pour toutes ces raisons citées en dessous, entre la nature des formations aux méthodes d'exploitation, nous avons vu une nécessité à étudier plus en profondeur l'utilisabilité des méthodes plus anciennes et si il était possible de proposer d'autres.

1.5 Conclusion

La consultation de ces articles et la combinaison des connaissances cumulées permettent de conclure que l'étude de déclin de production est un problème qui a fait couler de l'encre pendant des Décennies, bien avant *Arps*, et qui toujours d'actualité surtout avec la classifications de plus en plus spécifiques des types de réservoirs, qu'ils soient conventionnels ou non-conventionnels

La compréhension mathématiques et physique de la nature du problèmes est un outil essentiel à la résolution de dit-problème.

2 Étude de Déclin et Problématique

2.1 Introduction

L'analyse de la courbe de déclin de la production est l'une des plus anciennes méthodes de prévision des réserves de pétrole ou de gaz. L'analyse des courbes de déclin (DCA) est un moyen de prédire la production future de pétrole ou de gaz d'un puits ou d'une série de puits, sur la base d'une extrapolation de l'historique de production. Elles jouent un rôle essentiel dans la détermination de la valeur des puits de pétrole ou de gaz. Si les conditions affectant le taux de production ne sont pas modifiées, la courbe fournira des informations utiles sur la production future du puits. Avec cette connaissance, la valeur d'une propriété peut être jugée et une analyse économique appropriée peut être faite. La DCA est l'une des méthodes les plus courantes pour prévoir la production de pétrole et de gaz. Le principal avantage de l'analyse de la courbe de déclin est qu'elle utilise des données historiques qui sont généralement très faciles à obtenir. Les résultats de la courbe de déclin sont des graphiques simples et faciles à visualiser, à analyser et à comprendre. À ce jour, plusieurs méthodes ont été développées pour l'analyse des courbes de déclin, dont les plus courantes sont les méthodes Arps, Duong et de déclin de la production exponentielle étirée.

2.2 Les équations de Arps

En ce qui concerne les puits ayant un long historique de production et produisant à une pression d'écoulement constante au fond du puits, Arps a dérivé trois types de déclin de la production par la relation débit-temps, y compris le déclin exponentiel, le déclin hyperbolique et le déclin harmonique. Premièrement, les réserves récupérables ultimes doivent être estimées en supposant que les conditions de production resteront inchangées à l'avenir.

2.2.1 Déclin Exponentiel

Le concept de taux de déclin sera d'abord introduit. Il s'agit de la variation du débit par unité de temps, souvent désignée par D , comme le montre l'équation suivante:

$$D = -\frac{\Delta q/q}{\Delta t} \quad (2.1)$$

En ce qui concerne la décroissance exponentielle, la relation taux-temps est la suivante

$$q = q_i e^{-Dt} \quad (2.2)$$

La relation débit-production cumulée est:

$$Q = \int_0^t q dt = \int_0^t q_i e^{-Dt} dt = \frac{q_i - q}{D} \quad (2.3)$$

Sur le tracé semi-logarithmique, $\log q - t$ est une ligne droite, c'est-à-dire que D est une constante. Sur le graphique cartésien, le débit et la production cumulée sont également représentés par une ligne droite, comme le montre la figure ci dessous. En raison de la relation linéaire simple, il est facile d'identifier et d'utiliser la loi de décroissance exponentielle.

Le débit de déclin est utilisé pour déterminer le fluide produit ultime dans le cas d'un déclin exponentiel. Le débit initial q_f au moment de t_f est déterminé par

$$q_f = q_i e^{-Dt_f} \quad (2.4)$$

Soit q_{ab} le débit au moment de l'abandon t_{ab} . Sur la base de l'équation, la production cumulée entre les instants t_f et t_{ab} peut être obtenue par la formule suivante

$$Q_f = \frac{q_f - q_{ab}}{D} \quad (2.5)$$

Par conséquent, le volume récupérable ultime est

$$EUR = N_p + Q_f = N_p + \frac{q_f - q_{ab}}{D} \quad (2.6)$$

Sur la base de l'équation ci-dessus, le temps de production jusqu'à l'abandon peut également être estimé comme suit

$$\Delta t = t_{ab} - t_f = \frac{1}{D} \ln \frac{q_f}{q_{ab}} \quad (2.7)$$

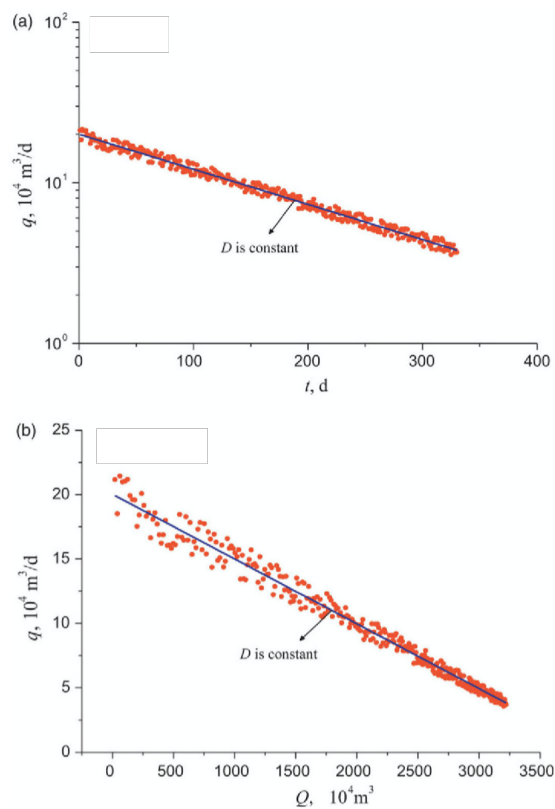


Figure 2.1: Courbe Typiques de declin exponentiel. (a) Courbe debit-temps ; (b) Courbe debit-production cumulee

2.2.2 Déclin Hyperbolique

En ce qui concerne le déclin hyperbolique, $\log q - t$ n'est plus une ligne droite sur le graphique semi-logarithmique ; c'est-à-dire que D est une variable, comme le montre la figure suivante. L'exposant de déclin b est introduit pour décrire une telle variation comme dans l'équation suivante : $b = 0$ correspond à un déclin exponentiel, et $b = 1$

correspond à un déclin harmonique ; plus b est petit, plus le déclin est rapide.

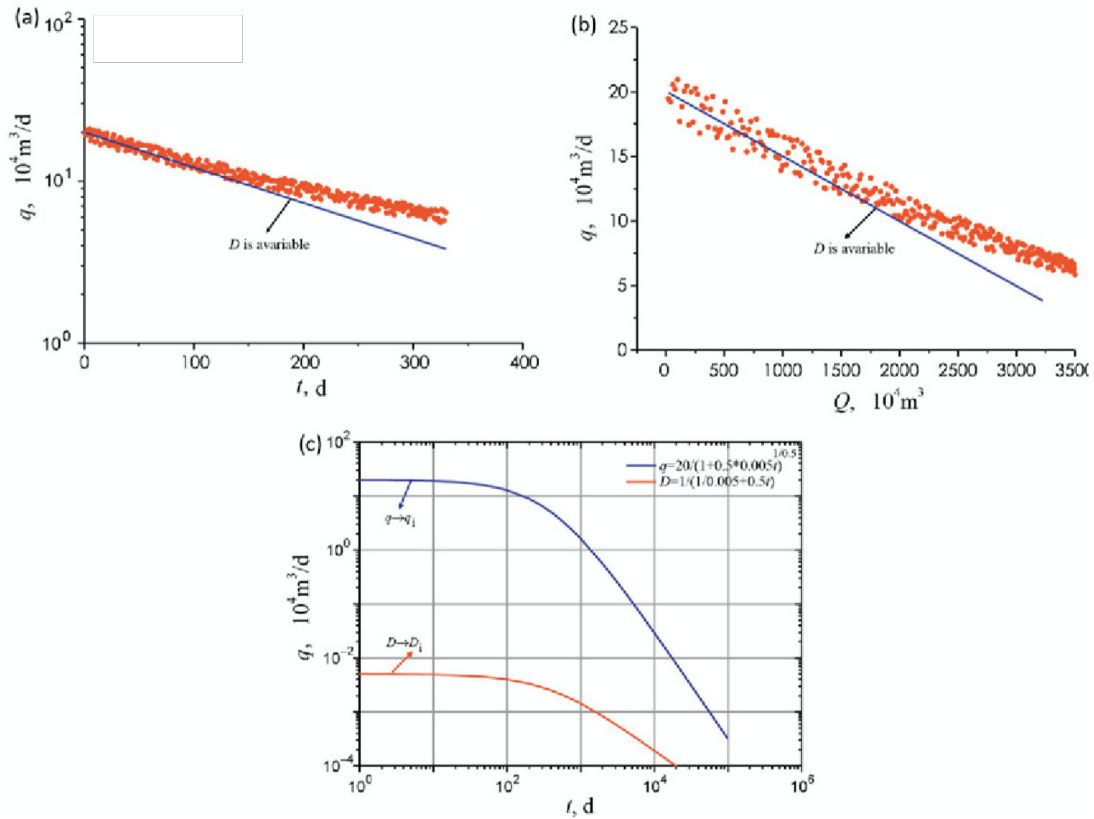


Figure 2.2: Courbes Typiques de declin hyperbolique. (a) Courbe debit-temps ; (b) Courbe debit-production cumulee ; (c) Courbe D-temp(Exemple).

$$q = \frac{q_i}{(1 + bD_i t)^{1/b}} \quad (b \text{ est constante, } 0 < b < 1) \quad (2.8)$$

En intégrant, La relation cumulée production-temps au cours d'une période de temps peut être dérivée de la manière suivante:

$$Q = \int_0^t q dt = \int_0^t \frac{q_i}{(1 + bD_i t)^{1/b}} dt \quad (2.9)$$

De ce fait, en intégrant:

$$Q = \frac{q_i}{D_i(b-1)} (1 + bD_i t)^{1-1/b} \quad (2.10)$$

En substituant les limites de l'intégration, t_0 et t , when $t_0 = 0$:

$$\int_0^t \frac{q_i}{(1 + bD_i t)^{1/b}} dt = \frac{q_i}{(b-1)D_i} \left[\frac{1}{(1 + bD_i t)^{1/b-1}} - 1 \right] = \frac{q_i}{(1-b)D_i} \left[1 - \frac{1}{(1 + bD_i t)^{1/b-1}} \right] \quad (2.11)$$

Nous allons réarranger l'équation du déclin hyperbolique, la substitué dans le dernier résultat, nous obtenons:

$$\int_0^t \frac{q_i}{(1 + bD_i t)^{1/b}} dt = \frac{q_i}{(1-b)D_i} \left[1 - \left(\frac{q}{q_i} \right)^{1-b} \right] \quad (2.12)$$

Le taux de déclin est utilisé pour calculer le fluide produit ultime dans le cas d'un déclin hyperbolique. Le débit initial et le débit au moment de l'abandon sont prédits comme suit:

$$q_f = \frac{q_i}{(1 + bD_i t_f)^{1/b}} \quad (2.13)$$

$$q_{ab} = \frac{q_i}{(1 + bD_i t_{ab})^{1/b}} \quad (2.14)$$

Lorsque $t_0 = t_f$ et $t = t_{ab}$, la production cumulée Q_f de la période de prévision peut être dérivée de la formule suivante:

$$Q_f = \int_{t_f}^{t_{ab}} \frac{q_i}{(1 + bD_i t)^{1/b}} dt = \frac{q_i}{(b-1)D_i} \left[\frac{1}{(1 + bD_i t_{ab})^{1/(b-1)}} - \frac{1}{(1 + bD_i t_f)^{1/(b-1)}} \right] \quad (2.15)$$

Nous réarrangeons,

$$\int_{t_f}^{t_{ab}} \frac{q_i}{(1 + bD_i t)^{1/b}} dt = \frac{q_i^b}{(1-b)D_i} (q_f^{1-b} - q_{ab}^{1-b}) \quad (2.16)$$

Le volume récupérable ultime est de

$$EUR = N_p + Q_f = N_p + \frac{q_i^b}{(1-b)D_i} (q_f^{1-b} - q_{ab}^{1-b}) \quad (2.17)$$

Et encore, le temps de production quand nous atteignons l'abondant:

$$\Delta t = t_{ab} - t_f = \frac{\left(\frac{q_i}{q_{ab}}\right)^b - \left(\frac{q_i}{q_f}\right)^b}{bD_i} \quad (2.18)$$

2.2.3 Déclin Harmonique

La décroissance harmonique est la forme particulière de la décroissance hyperbolique.

Lorsque, $b = 1$ correspond à la décroissance harmonique:

$$q = \frac{q_i}{(1 + D_i t)} \quad (2.19)$$

De la même manière en l'intégrant:

$$Q = \int_{t_0}^t \frac{q_i}{(1 + D_i t)} dt \quad (2.20)$$

Les limites d'intégration t_0 et t sont substituées ce que nous avons trouvé:

$$\int_{t_0}^t \frac{q_i}{(1 + D_i t)} dt = \frac{q_i}{D_i} \ln\left(\frac{1 + D_i t}{1 + D_i t_0}\right) \quad (2.21)$$

Lorsque $t_0 = 0$, la relation cumule-temps peut être dérivée comme suit:

$$Q = \frac{q_i}{D_i} \ln \frac{q_i}{q} \quad (2.22)$$

On sait d'après les équations que l'inverse du débit du déclin harmonique $1/qt$ est linéaire, et que la production cumulée $Q \ln q$ est également linéaire, comme le montre la troisième figure.

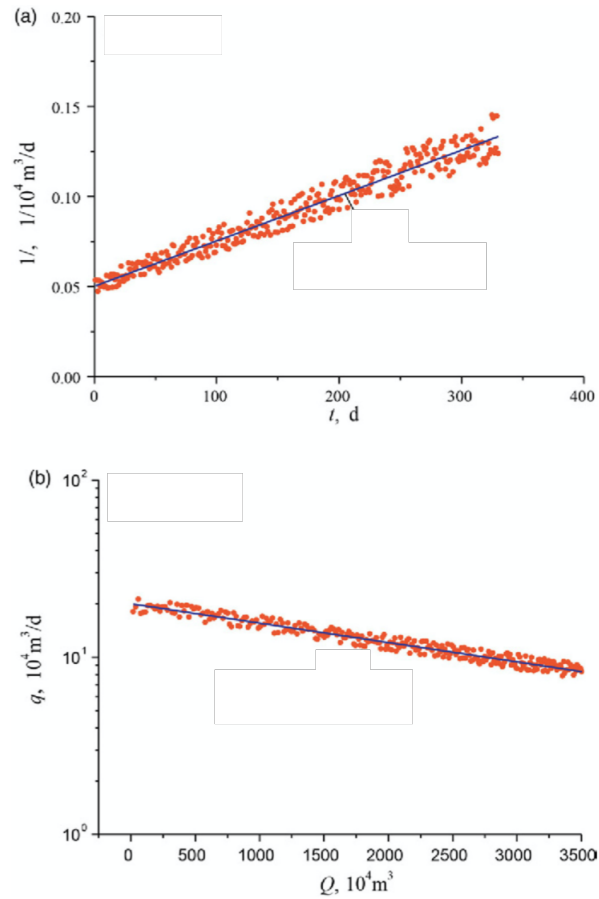


Figure 2.3: Courbes Typiques de declin harmonique. (a) Courbe taux-temps ; (b) Courbe taux-production cumulee

Le débit de déclin est utilisé pour déterminer le fluide produit ultime dans le cas d'un déclin harmonique. Le débit initial et le débit au moment de l'abandon sont prédits comme suit:

$$q_f = \frac{q_i}{1 + Di t_f} \quad (2.23)$$

$$q_{ab} = \frac{q_i}{1 + Di t_{ab}} \quad (2.24)$$

Lorsque $t_0 = t_f$ et $t = t_{ab}$, la production cumulée Q_f est la suivante:

$$Q_f = \frac{q_i}{D_i} \ln \frac{q_f}{q_{ab}} \quad (2.25)$$

Le volume récupérable ultime est de:

$$EUR = N_p + Q_f = N_p + \frac{q_i}{D_i} \ln \frac{q_f}{q_{ab}} \quad (2.26)$$

Le temps de production jusqu'à l'abandon peut être estimé par:

$$\Delta t = t_{ab} - t_f = \frac{q_i}{D_i} \left(\frac{1}{q_{ab}} - \frac{1}{q_f} \right) \quad (2.27)$$

2.2.4 Comparaison entre les trois Déclins

Les relations entre le taux, la production cumulée et le temps de trois courbes de déclin sont indiquées dans le tableau. Avec les mêmes paramètres, les tendances des trois courbes de déclin sont présentées dans la figure. On observe sur la figure que le déclin exponentiel est le plus rapide, mais que la production cumulée prévue est la plus faible ; le déclin harmonique est le plus lent, mais la production cumulée prévue est la plus élevée ; tandis que le déclin hyperbolique se situe entre les deux.

Caractéristiques de base	Type de Declin	Débit-temps	Cummul de production	$D_i t$
$b = 0$	Exponential	$q = q_i e^{-D_i t}$	$Q = (q_i - q) / D_i$	$D_i t = \ln (q_i / q)$
$b = 1$	Harmonique	$q = q_i (1 + b D_i t)^{-1/b}$	$Q = \frac{q_i}{(1-b)D_i} \left[1 - \left(\frac{q}{q_i} \right)^{1-b} \right]$	$D_i t = \frac{(q_i/q)^b - 1}{b}$
$0 < b < 1$	Hyperbolique	$q = q_i (1 + D_i t)^{-1}$	$Q = q_i \ln (q_i / q) / D_i$	$D_i t = (q_i / q) - 1$

Table 2.1: Les Équations de Arps

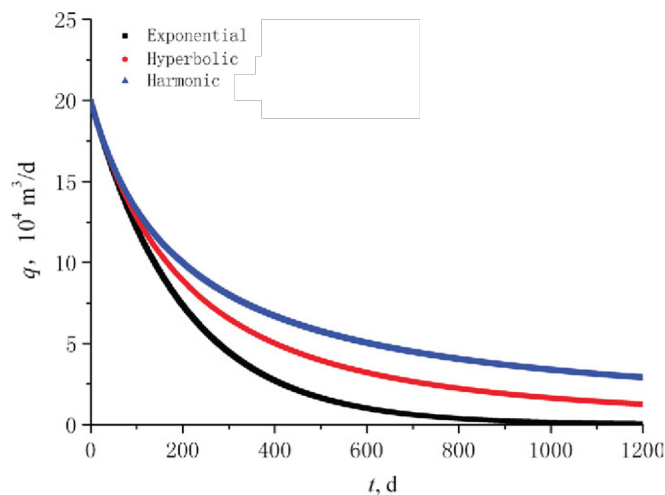


Figure 2.4: Formes de courbe de déclin d'Arps pour un trace cartésien débit-temps

Si les données ne sont pas obtenus pour une période assez longue (dans la figure: 150 jours) et sont utilisées pour l'appariement, les trois méthodes permettent d'obtenir des courbes d'ajustement lentes, mais un écart prodigieux se produira dans la prédiction de la tendance de la production cumulée. Par conséquent, lorsque la méthode Arps est utilisée pour effectuer l'analyse, le temps de production doit être suffisamment long.

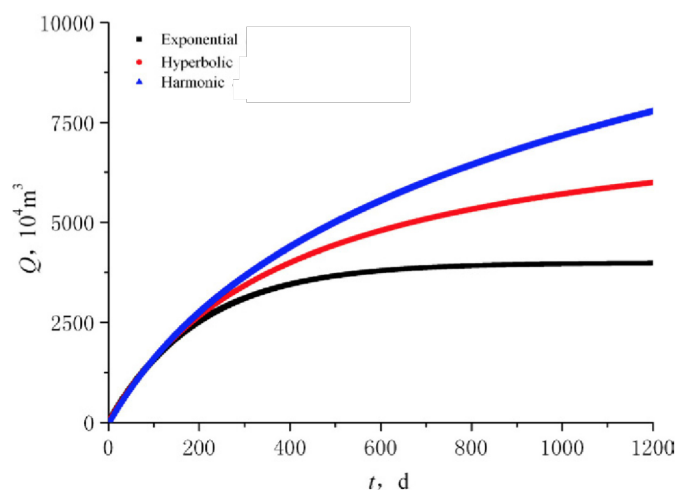


Figure 2.5: Formes des courbes de declin d'Arps pour un trace cartésien du Cumul-temps

2.3 Modèle de croissance logistique - Logistic Growth Model

Une variété de courbes de croissance a été développée pour modéliser à la fois la dynamique des populations intraspécifiques non prédites et la croissance biologique plus générale. Les modèles prédictifs les plus efficaces sont basés sur des formes étendues de l'équation de croissance logistique classique de Verhulst¹.

$$f(t) = \frac{K}{1 + ae^{-rt}} \text{ où } K \text{ et } r \text{ sont des réels positifs et } a \text{ un réel quelconque.}$$

Ce sont les solutions en temps continu du **modèle de Verhulst**.

Le modèle de croissance logistique a été développé par Clark et al [12]. Le modèle est basé sur les courbes de croissance logistiques -cité en dessus- qui sont utilisées pour

¹En dynamique des populations, le modèle de Verhulst est un modèle de croissance proposé par Pierre François Verhulst vers 1840. Verhulst a proposé ce modèle en réponse au modèle de Malthus qui proposait un taux d'accroissement constant sans frein conduisant à une croissance exponentielle de la population.

prévoir la croissance (par exemple la production cumulée de pétrole ou de gaz). Le modèle logistique est basé sur 2 ou 3 paramètres. Le troisième paramètre dépend de la connaissance de la récupération ultime estimée (EUR), qui peut être obtenue indépendamment d'une estimation volumétrique (quantité de production récupérable sans contraintes économiques). Le modèle est borné par l'EUR. Cependant, lorsque l'EUR n'est pas connu, la solution n'est pas unique, car plusieurs bonnes correspondances peuvent être ajustées.

Le LGM peut être représenté comme suit:

$$q(t) = \frac{dQ}{dt} = \frac{K n b t^{n-1}}{(a + t^n)^2} \quad (2.28)$$

où K est la capacité de charge.

Le principal avantage du LGM est que l'estimation de la réserve est inhibée par le paramètre K ainsi que par le taux de production, qui se termine à un moment infini. LE LGM se révèle être très efficace pour l'étude de déclin dans les réservoirs à très faibles perméabilité e.g: gaz de schiste i.e: LGM est très flexible et confident à la modélisation des performances transitoires longues dominées par les frontières des réservoirs non conventionnels (Long transient boundary-dominated).

2.4 Mais encore

En plus des modèles développés plus en détails dans les sous-sections précédentes - Les mêmes que nous allons utilisés dans cette étude, l'introduction d'approches basées sur la pratique, telles que la relation **hyperbolique dite modifiée** et une multitude de relations de déclin modernes: Modèle de Duong[13], loi de puissance[11], Extended Exponential Model[18], ont toutes été développées pour atténuer certains des problèmes associés à la prévision des données de production en fonction du temps dans les systèmes non conventionnels. La plupart de ces modèles ont été développés pour traiter des régimes d'écoulement ou des conditions opérationnelles spécifiques et, à ce stade, tous doivent être considérés comme des relations **empiriques**.

Il convient d'être particulièrement prudent lors de l'application du modèle de déclin le

plus couramment utilisé, à savoir les relations de déclin d'Arps. Cette affirmation se justifie en observant attentivement la liste des hypothèses où l'on peut affirmer que chaque point est violé dans les systèmes non conventionnels. Le manque d'analogues historiques, l'incertitude concernant les changements de régime d'écoulement et le moment où ils se produisent, et une compréhension limitée de la zone de drainage contactée font qu'il est difficile d'établir la confiance dans le fait qu'un modèle particulier représente de manière adéquate le comportement futur de la production. En outre, les stratégies de gestion des étranglements, la re-stimulation des puits, le infill drilling et les d'artificial lift, entre autres pratiques, entraînent une évolution constante des champs, ce qui rend difficile la comparaison des prévisions d'une année sur l'autre.

2.4.1 Conclusion

La prévision de la production présente de nombreux défis pour l'ingénieur praticien. En plus d'avoir une compréhension limitée de la physique des flux qui régissent certains systèmes complexes, de nombreuses techniques de prévision traditionnellement utilisées dans les systèmes conventionnels perdent leur validité lorsqu'elles sont confrontées à des systèmes non conventionnels. Cela est particulièrement vrai lorsque l'on ne prend pas garde à l'application des relations empiriques de déclin qui, au moment de la rédaction de ce document, constituent la norme de l'industrie en matière de prévision de la production pour les évaluations des réserves.

3 Modélisation des Séries Temporelles

La modélisation des séries temporelles est un domaine de recherche dynamique qui a attiré l'attention de la communauté des chercheurs au cours des dernières décennies. Ce modèle est ensuite utilisé pour générer des valeurs futures pour la série, c'est-à-dire pour faire des prévisions. L'étude des séries chronologiques est indispensable pour divers domaines, de l'économie à la science, et l'ingénierie des réservoirs ne fait pas exception.

3.1 Définition

Une série temporelle est un ensemble séquentiel de points de données, mesurés généralement sur des temps successifs. Elle est définie mathématiquement comme un ensemble de vecteurs

$$x(t), t = 0, 1, 2, \dots$$

où t représente le temps écoulé. Une série temporelle contenant des enregistrements d'une seule variable est qualifiée d'univariée. Mais si l'on considère les enregistrements de plus d'une variable, on parle de série multivariée. Par exemple, les relevés de température, le débit d'une rivière, la concentration d'un processus chimique, etc. peuvent être enregistrés sous forme de série temporelle continue. D'autre part, la population d'une ville particulière, la production d'une entreprise, les taux de change entre deux monnaies différentes peuvent représenter des séries temporelles discrètes. Habituellement, dans une série temporelle discrète, les observations consécutives sont enregistrées à des intervalles de temps réguliers tels que des séparations horaires, quotidiennes, hebdomadaires, mensuelles ou annuelles.

3.2 Composent d'une série temporelle

Une série temporelle en général est censée être affectée par quatre composantes principales, qui peuvent être séparées des données observées. Ces composantes sont : **Les composantes tendancielle, cyclique, saisonnière et irrégulière.**

La **tendance** générale d'une série chronologique à augmenter, diminuer ou stagner sur une longue période de temps est appelée tendance séculaire ou simplement tendance. Ainsi, on peut dire que la tendance est un mouvement à long terme dans une série chronologique.

Les **variations saisonnières** dans une série temporelle sont des fluctuations au cours d'une année pendant la saison. La variation saisonnière est un facteur important pour les études économique

La **variation cyclique** dans une série temporelle décrit les changements à moyen terme dans la série, causés par des circonstances qui se répètent en cycles. La plupart des séries chronologiques économiques et financières présentent une certaine forme de variation cyclique.

Les **variations irrégulières** ou aléatoires d'une série chronologique sont causées par des influences imprévisibles, qui ne sont pas régulières et ne se répètent pas non plus selon un schéma particulier.

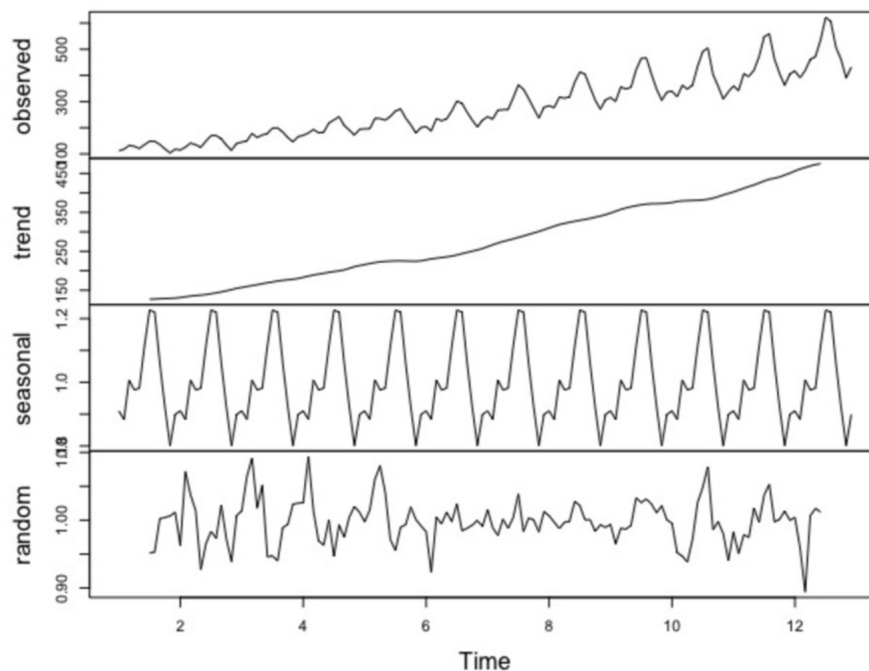


Figure 3.1: Exemple d'une Série Temporelle et de sa décomposition

Compte tenu des effets de ces quatre composantes, deux types de modèles différents sont généralement utilisés pour une série chronologique, à savoir les modèles multiplicatifs et additifs.

$$\text{Modèle Multiplicatif: } Y(t) = T(t) \times S(t) \times C(t) \times I(t)$$

$$\text{Modèle additif: } Y(t) = T(t) + S(t) + C(t) + I(t)$$

Ici $Y(t)$ est l'observation et $T(t)$, $S(t)$, $C(t)$ et $I(t)$ sont respectivement la tendance, la variation saisonnière, cyclique et irrégulière au temps t .

Le modèle multiplicatif est basé sur l'hypothèse que les quatre composantes d'une série chronologique ne sont pas nécessairement indépendantes et qu'elles peuvent s'influencer mutuellement, alors que dans le modèle additif, on suppose que les quatre composantes sont indépendantes les unes des autres.

3.3 L'Analyse

Dans la pratique, un modèle approprié est ajusté à une série chronologique donnée et les paramètres correspondants sont estimés à l'aide des valeurs de données connues. La procédure d'ajustement d'une série temporelle à un modèle approprié est appelée *analyse des séries temporelles*[9]. Elle comprend des méthodes qui tentent de comprendre la nature de la série et est souvent utile pour les prévisions et les simulations futures.

Dans la prévision des séries temporelles, les observations passées sont collectées et analysées pour développer un modèle mathématique approprié qui capture le processus sous-jacent de génération de données pour la série. Les événements futurs sont ensuite prédits à l'aide du modèle. Cette approche est particulièrement utile lorsque l'on connaît mal le modèle statistique suivi par les observations successives ou lorsqu'il n'existe pas de modèle explicatif satisfaisant.

3.4 Séries temporelles et processus stochastiques

L'expression mathématique décrivant la structure de probabilité d'une série temporelle est appelée processus stochastique. Ainsi, la séquence d'observations de la série est en fait un échantillon de réalisation du processus stochastique qui l'a produit. Une hypothèse habituelle est que les variables de la série temporelle x_t sont **indépendantes et identiquement distribuées** suivant la **distribution normale**. Cependant, un point intéressant est que les séries temporelles ne sont en fait pas exactement indépendantes et identiquement distribuées ; elles suivent plus ou moins un **modèle régulier à long terme**. C'est la raison pour laquelle la prévision des séries temporelles, en utilisant une technique appropriée, donne des résultats proches de la valeur réelle.

3.5 Concept de stationnarité

Le concept de stationnarité d'un processus stochastique peut être visualisé comme une forme d'équilibre statistique. Les propriétés statistiques telles que la moyenne et la variance d'un processus stationnaire ne dépendent pas du temps. Il s'agit d'une condition nécessaire à la construction d'un modèle de série chronologique utile pour les prévisions futures. De plus, la complexité mathématique du modèle ajusté diminue avec cette hypothèse.

Un processus $\{x(t), t = 0, 1, 2, \dots\}$ est fortement stationnaire ou strictement stationnaire si la fonction de distribution de probabilité conjointe de $\{x_{t-s}, x_{t-s+1}, \dots, x_t, \dots, x_{t+s-1}, x_{t+s}\}$ est indépendante de t pour tout s . Ainsi, pour un processus fortement stationnaire, la distribution conjointe de tout ensemble possible de variables aléatoires du processus est indépendante du temps.

Cependant, pour les applications pratiques, l'hypothèse de forte stationnarité n'est pas toujours nécessaire et une forme un peu plus faible est donc considérée. Un processus stochastique est dit faiblement stationnaire d'ordre k si les moments statistiques du processus jusqu'à cet ordre ne dépendent que des différences de temps et non du moment d'occurrence des données utilisées pour estimer les moments. Par exemple, un processus stochastique $\{x(t), t = 0, 1, 2, \dots\}$ est stationnaire du second ordre s'il a une moyenne et une variance indépendantes du temps et que les valeurs de covariance $Cov(x_t, x_{t-s})$ ne dépendent que de s .

Il est important de noter que ni la stationnarité forte ni la stationnarité faible n'impliquent l'autre. Cependant, un processus faiblement stationnaire suivant une distribution normale est également fortement stationnaire. Certains tests mathématiques comme celui de Dickey et Fuller sont généralement utilisés pour détecter la stationnarité dans les données d'une série chronologique.

Le concept de stationnarité est une idée mathématique construite pour simplifier le développement théorique et pratique des processus stochastiques. Pour concevoir un modèle approprié, adapté aux prévisions futures, la série chronologique sous-jacente doit être stationnaire. Malheureusement, ce n'est pas toujours le cas. Comme l'ont indiqué Hipel et McLeod[9], plus la durée des observations historiques est longue, plus il y a de chances que la série chronologique présente des caractéristiques non stationnaires.

Cependant, pour des périodes relativement courtes, on peut raisonnablement modéliser la série en utilisant un processus stochastique stationnaire. Habituellement, les séries temporelles qui montrent des tendances ou des modèles saisonniers sont non stationnaires par nature. Dans de tels cas, la différenciation et les transformations de puissance sont souvent utilisées pour supprimer la tendance et rendre la série stationnaire.

3.6 Les modèles stochastiques

Dans les sections précédentes, nous avons abordé les principes fondamentaux de la modélisation et de la prévision des séries chronologiques. La sélection d'un modèle approprié est extrêmement importante car elle reflète la structure sous-jacente de la série et ce modèle ajusté est à son tour utilisé pour les prévisions futures. Un modèle de série chronologique est dit linéaire ou non linéaire selon que la valeur actuelle de la série est une fonction linéaire ou non linéaire des observations passées.

En général, les modèles de séries chronologiques peuvent prendre de nombreuses formes et représenter différents processus stochastiques. Il existe deux modèles de séries chronologiques linéaires largement utilisés dans la littérature, à savoir le modèle autorégressif (AR) [9] et le modèle de moyenne mobile (MA) . En combinant ces deux modèles, les modèles Autoregressive Moving Average (ARMA) et Autoregressive Integrated Moving Average (ARIMA) ont été proposés dans la littérature. Le modèle ARIMA et ses différentes variations sont basés sur le célèbre principe de Box-Jenkins [9] et sont donc également connus sous le nom de modèles de Box-Jenkins.

- **AR (Autoregression)** : Modèle qui montre une variable changeante qui régresse sur ses propres valeurs décalées/précédentes.
- **I (Integrated)** : Différenciation des observations brutes pour permettre aux séries chronologiques de devenir stationnaires.
- **MA (Moving average)** : Dépendance entre une observation et une erreur résiduelle d'un modèle de moyenne mobile

3.6.1 Qu'est-ce que le modèle Box-Jenkins ?

Les modèles de Box-Jenkins sont utilisés pour prévoir une variété de points de données ou de plages de données anticipées, y compris les données commerciales et les prix futurs des titres.

Le modèle Box-Jenkins a été créé par deux mathématiciens : George Box et Gwilym Jenkins. Les deux mathématiciens ont discuté des concepts qui composent ce modèle dans une publication de 1970 intitulée "Time Series Analysis : Forecasting and Control" [15].

L'estimation des paramètres du modèle de Box-Jenkins peut être très compliquée. C'est pourquoi, comme pour les autres modèles de régression des séries temporelles, les meilleurs résultats seront généralement obtenus par l'utilisation d'un logiciel programmable. Le modèle de Box-Jenkins est aussi généralement mieux adapté aux prévisions à court terme, de 18 mois ou moins.²

3.7 Synthèse

L'analyse des séries temporelles est un domaine de recherche en pleine expansion qui offre de nombreuses possibilités de travaux futurs. Dans ce qui a précédé, nous avons donné un aperçu des séries temporelles, de leurs composantes, de leur analyse, nous avons approfondi le concept de stationnarité, nous avons donné une introduction aux modèles stochastiques ainsi que la méthode de Box-Jenkins. Il s'agit d'une méthode utile, que nous utiliserons à bon escient.

²Nous nous intéresserons à la réalisation et la méthodologie dans le chapitre suivant: *Outils Mathématiques*

4 Outils Mathématiques

4.1 Ajustement de courbe

L'ajustement de courbe est le processus de construction d'une courbe, ou fonction mathématique, qui s'adapte le mieux à une série de points de données, éventuellement soumis à des contraintes. L'ajustement de courbe peut impliquer soit une interpolation, lorsqu'un ajustement exact aux données est requis, soit un lissage, dans lequel une fonction "lisse" est construite qui s'ajuste approximativement aux données.

4.1.1 Régression Linéaire Simple

Les modèles de régression décrivent la relation entre les variables en ajustant une ligne aux données observées. Les modèles de régression linéaire utilisent une ligne droite. La régression linéaire simple est utilisée pour estimer la relation entre deux variables quantitatives.

Nous pouvons donc écrire:

$$Y \approx \alpha x + \beta \tag{4.1}$$

Avec α et β , pente et interception respectivement; des constants à déterminer. Pour cela nous devons utiliser des données pour ajuster les deux paramètres.

Méthode des moindres Carrés

Afin de déterminer les constantes citées en dessous, et ainsi pouvoir utiliser l'équation de la droite i.e. La régression linéaire, Il existe plusieurs façons de faire l'approximation. Cependant, l'approche la plus courante consiste à minimiser l'erreur des moindres carrés. En ayant les couples de points suivant:

$$(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)$$

Soit $(\hat{y}_i = \beta + \alpha x_i)$ la prédiction pour Y basée sur la i -ème valeur de X . Ensuite, $(e_i = y_i - \hat{y}_i)$ représente le i ème résidu (distance-différence- entre la droite et un point i)

La somme rsiduelle étant:

$$RSS = e_1^2 + e_2^2 + \dots + e_n^2 \quad (4.2)$$

ou: $e_i = (y_i - \beta - \alpha x_i)^2$

À l'aide d'une démonstration, nous trouvons:

$$\alpha = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (4.3)$$

$$\beta = \bar{y} - \alpha \bar{x} \quad (4.4)$$

ou $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ et $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$

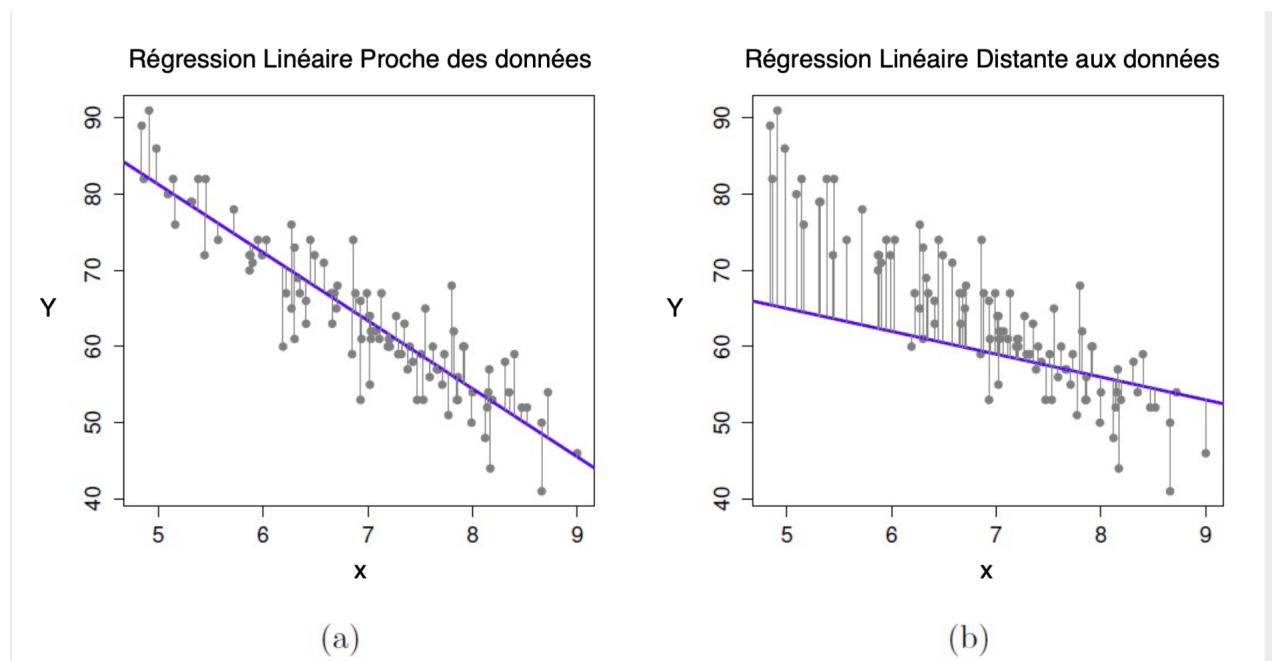


Figure 4.1: Régression Linéaire Simple

Le RSS fournit une mesure absolue de l'écart du modèle par rapport aux données. Mais comme elle est mesurée en unités Y, il n'est pas évident de juger l'accordance du modèle par rapport aux données. La statistique **R-carrée** fournit une autre mesure de l'adéquation. Elle prend la forme d'une proportion. Elle prend donc toujours une valeur comprise entre 0 et 1, et est indépendante de l'échelle de Y. Pour calculer R^2 , nous utilisons la formule:

$$R^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2 - (y_i - \hat{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4.5)$$

4.1.2 Régression Linéaire Multiple

La régression linéaire multiple (RLM), également appelée simplement régression multiple, est une technique statistique qui utilise plusieurs variables explicatives pour prédire le résultat d'une variable de réponse. La formulation mathématiques étant:

$$Y = \beta + \sum_{i=1}^n \alpha_i x_i + \epsilon \quad (4.6)$$

Gradeint Descent

À l'instar de de la régression linéaire simple, les coefficient nous sont inconnues, et nous aurons besoins de les déterminer, et nous utiliserons la même approche des moindres carrés :

$$RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \beta + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \dots + \alpha_p x_{ip})^2 \quad (4.7)$$

La formule permettant l'estimation des coefficients qui *minimisent* la formule en dessus, et plus complexe que celle permettant la régression simple . Qui dit **Minimisation** dit un problème d'optimisation, et pour ce problème là l'algorithme du **Gradient Descent** est tout trouvé.

Gradient Descent: La descente de gradient est un algorithme d'optimisation itératif du premier ordre permettant de trouver un minimum local d'une fonction différentiable. L'idée est d'effectuer des étapes répétées dans la direction opposée au gradient (ou gradient approximatif) de la fonction au point actuel, car c'est la direction de la descente la plus raide. Sa formulation mathématique (a étant un hyper-paramètre³):

$$\alpha_n = \alpha_n - a \frac{\partial RSS}{\partial \alpha_n} \quad (4.8)$$

³En apprentissage automatique, un hyperparamètre est un paramètre dont la valeur est utilisée pour contrôler le processus d'apprentissage. En revanche, les valeurs des autres paramètres (généralement les poids des nœuds) sont dérivées par l'entraînement. En d'autre termes, il est laissé aux choix de l'ingénieur

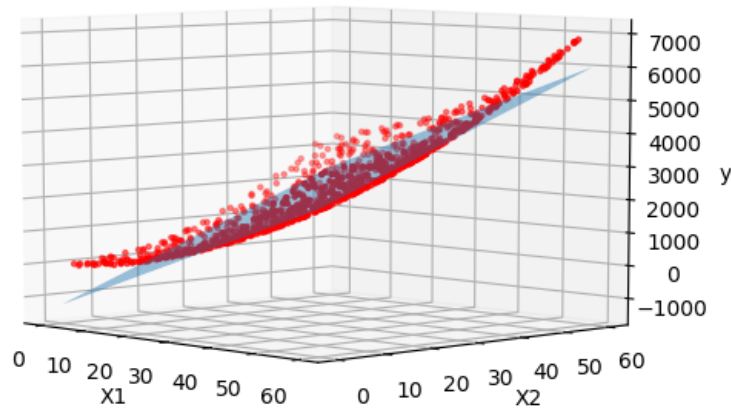


Figure 4.2: Exemple d'une Régression Linéaire à deux variables exogènes; La droite se transforme en Plan

4.1.3 Régression non-linéaire

Un modèle est linéaire si ses paramètres et ses variables sont linéaires, ainsi sa forme fonctionnelle sera linéaire. Un paramètre est linéaire s'il n'est pas multiplié avec un autre paramètre ou s'il n'est pas élevé à une puissance supérieure à 1. De même pour une variable, elle sera linéaire si la sensibilité d'une autre variable à la suite de sa variation est constante, c'est-à-dire indépendante de la variable à la base du changement ou d'une autre variable quelconque dans le modèle. ⁴

Entre les modèles non linéaire bien connu, le modèle **Verhulst** dont nous avons parler dans le chapitre: *Étude de Déclin et Problématique*

Méthodes d'estimation des modèles non-linéaire

Notons qu'il y a deux approches qui permettent d'estimer des modèles non linéaires, à savoir :

- *Approches intuitives* Elles consistent à émettre des hypothèses sur les valeurs du

⁴Nous nous intéresserons pas aux modèles linéarisables, du fait qu'ils sont très peu rencontrés dans notre domaines

niveau de saturation: **L'analogie géographique, La méthode de balayage**

- *Approches ou méthodes de résolution numérique* tel que l'algorithme de Gauss-Newton

Algorithme de Gauss-Newton

Soit le modèle non-linéaire:

$$Y_t = f(X_t, a) + e_t \quad (4.9)$$

Avec : Y_t = variable dépendante ; X_t = vecteur de variables explicatives (matrice des observations des variables explicatives de dimension « $n, k + 1$ ») ; a = vecteur de paramètres à estimer (dimension « $k + 1$ »), et e_t = terme aléatoire.

L'estimateur des Moindres Carrés pour le modèle non linéaire est « a » qui minimise la RSS ($S(a) = e'e = [Y_t - f(X_t, a)]'[Y_t - f(X_t, a)]$) comme suit (on a $k + 1$ conditions du 1er ordre : $\frac{\partial S}{\partial a} = 0$):

$$\frac{\partial S}{\partial a} = -2 \frac{\partial f(X_t, a)}{\partial a} [Y_t - f(X_t, a)] = 0 \quad (4.10)$$

$$\frac{\partial f(X, a)}{\partial a} = Z(a) = \begin{bmatrix} \frac{\partial f(x_1, a)}{\partial a_0} & \cdots & \frac{\partial f(x_1, a)}{\partial a_k} \\ \vdots & \ddots & \\ \frac{\partial f(x_n, a)}{\partial a_0} & \cdots & \frac{\partial f(x_n, a)}{\partial a_k} \end{bmatrix} \quad (4.11)$$

Notons $Z(a^*)$ = matrice calculée pour les valeurs particulières de a ($a = a^*$). Grâce au développement limité de Taylor au voisinage de « a^* », nous obtenons l'approximation suivante à la « t -ième observation »

$$f(x_t, a) \simeq f(x_t, a^*) + \left\{ \frac{\partial f(x_t, a)}{\partial a_0} \Big|_{a=a^*} \cdots \frac{\partial f(x_t, a)}{\partial a_k} \right\} (a - a^*) \quad (4.12)$$

Autrement, en notation matricielle :

$$f(X, a) \approx f(X, a^*) + Z(a^*)(a - a^*) + e \quad (4.13)$$

En posant: $\bar{y}(a^*) = y - f(X, a^*) + Z(a^*)a^*$, l'équation se réduit à:

$$\hat{y}(a^*) = Z(a^*)a + e \quad (4.14)$$

L'estimateur de ce modèle est de:

$$a^2 = [Z(a^*)' Z(a^*)]^{-1} Z(a^*)' \bar{y}(a^*) = a^* + [Z(a^*)' Z(a^*)]^{-1} Z(a^*)' [y - f(X, a)] \quad (4.15)$$

Nous aurons, $k + 1$ nouvelles valeurs pour le vecteur ($a = a^2$), la convergence atteinte lorsque ($\hat{a} = a^p \approx a^{p-1}$) (stabilité des coefficients à la p -ième itération).

Notons que « l'algorithme de Newton-Raphson » raffine l'algorithme de Newton en utilisant un développement de Taylor du second ordre.

4.2 Détection de valeurs aberrantes et Réduction des irrégularités

Une anomalie : un ou des points s'écartent significativement des autres dans une base de données. Ils sont épars et localisés dans une zone peu dense des données (s'ils forment un groupe compact, on ne peut pas vraiment parler d'anomalies)

Le problème n'est pas simple parce que nous ne connaissons pas nécessairement au préalable ce qui caractérise une anomalie. C'est alors à l'algorithme d'apprendre une métrique appropriée pour les détecter au sein des données.

Un algorithme de détection d'anomalies: a pour objectif principal de repérer des données atypiques et non conformes aux autres données.

4.2.1 Isolation Forest

L'isolation forest (moins connu sous son nom français « forêt d'isolement ») calcule un score d'anomalie pour chaque observation du dataset. Ce score donne une mesure de la normalité de chaque observation en fonction de l'ensemble des données. Pour calculer ce score, l'algorithme isole la donnée en question de manière récursive : il choisit une variable

au hasard et fixe un seuil de coupure au hasard, puis il évalue si cela permet d'isoler une observation en particulier. Ceci est son pseudo-code

Algorithm 1 : $iForest(X, t, \psi)$

Inputs: X - input data, t - number of trees, ψ - sub-sampling size
Output: a set of t $iTrees$

- 1: **Initialize** $Forest$
- 2: set height limit $l = ceiling(\log_2 \psi)$
- 3: **for** $i = 1$ to t **do**
- 4: $X' \leftarrow sample(X, \psi)$
- 5: $Forest \leftarrow Forest \cup iTree(X', 0, l)$
- 6: **end for**
- 7: **return** $Forest$

Figure 4.3: Pseudo Code de L'algorithme de L'Isolation Forest

Ce type d'algorithme est à privilégier dans le cas de problèmes mathématiques de grandes dimensions (grand nombre d'observations et de variables).

4.2.2 k-Nearest Neighbor

k-Nearest Neighbours (k-NN voire KNN ou méthode des k plus proches voisins) : k-NN est un algorithme standard de classification qui repose exclusivement sur le choix de la métrique de classification. Il est "non paramétrique" (seul k doit être fixé) et se base uniquement sur les données d'entraînement.

L'idée est la suivante : à partir d'une base de données étiquetées, on peut estimer la classe d'une nouvelle donnée en regardant quelle est la classe majoritaire des k données voisines les plus proches (d'où le nom de l'algorithme). Le seul paramètre à fixer est k, le nombre de voisins à considérer .

Les métriques les plus souvent choisies sont la distance usuelle dite **euclidienne** et la distance de **Mahalanobis** (qui tient compte de la variance – du point de vue statistique – et de la corrélation entre les données). Bien que l'algorithme puisse fonctionner avec ces métriques par défaut, il est généralement bien meilleur quand il est utilisé avec une métrique adaptée aux données, métrique qui peut être calculée à partir d'heuristiques connues liées au problème (par exemple la distance euclidienne pondérée).

```

k-Nearest Neighbor
Classify (X, Y,  $x$ ) // X: training data, Y: class labels of X,  $x$ : unknown sample
for  $i = 1$  to  $m$  do
    Compute distance  $d(\mathbf{X}_i, x)$ 
end for
Compute set  $I$  containing indices for the  $k$  smallest distances  $d(\mathbf{X}_i, x)$ .
return majority label for  $\{\mathbf{Y}_i \text{ where } i \in I\}$ 

```

Figure 4.4: Pseudo Code de l’algorithm du k-Nearest Neighbours

Les points faibles de cet algorithme sont : Son coût en puissance de calcul (pour prédire l’image d’un nouveau point, on doit calculer sa distance à tous les autres).

4.2.3 LOESS - Régression Locale

Cet algorithme estime la fonction latente de manière ponctuelle. Pour chaque valeur de x , nous estimons la valeur de $f(x)$ en utilisant ses valeurs échantillonnées voisines (connues). Cet algorithme est assez similaire à l’algorithme KNN, où k , la taille de la fenêtre, est un paramètre réglable et, dans ce cas particulier, déterminera la douceur de l’estimation résultante. En un sens, k est le bouton de réglage du biais par rapport à la variance. De grandes valeurs de k entraîneront un biais plus élevé et des valeurs plus faibles induiront une variance plus élevée.

La première étape consiste à collecter la valeur de x pour laquelle nous voulons estimer y . Appelons-les x' et y' . En alimentant l’algorithme LOESS avec x' , et en utilisant les valeurs x et y échantillonnées, nous obtiendrons une estimation y' . En ce sens, LOESS est un algorithme non paramétrique⁵ qui doit utiliser tout l’ensemble de données pour l’estimation. Maintenant que nous avons x' , nous devons trouver ses k plus proches voisins en utilisant une simple distance euclidienne. Appelons l’ensemble ordonné résultant D . L’étape suivante consiste à convertir l’ensemble D des k distances en un ensemble ordonné W contenant les poids qui seront utilisés ultérieurement dans le processus de régression linéaire. Ces poids sont calculés à l’aide d’une fonction de poids spécialisée qui attribue

⁵La statistique non paramétrique est un domaine de la statistique qui ne repose pas sur des familles de loi de probabilité paramétriques. Bien que le nom de non paramétriques soit donné à ces méthodes, elles reposent en vérité sur l’estimation de paramètres. Dans les méthodes non paramétriques, le nombre de paramètres estimés croît avec le nombre de données disponibles ; dans les méthodes classiques, ce nombre est décidé à l’avance.

une importance à chacun des k voisins de x en fonction de sa distance à x' .

Les poids de distance sont calculés en utilisant la fonction tri-cubique:

$$w(x) = \begin{cases} (1 - |x|^3)^3 & |x| < 1 \\ 0 & |x| \geq 1 \end{cases}$$

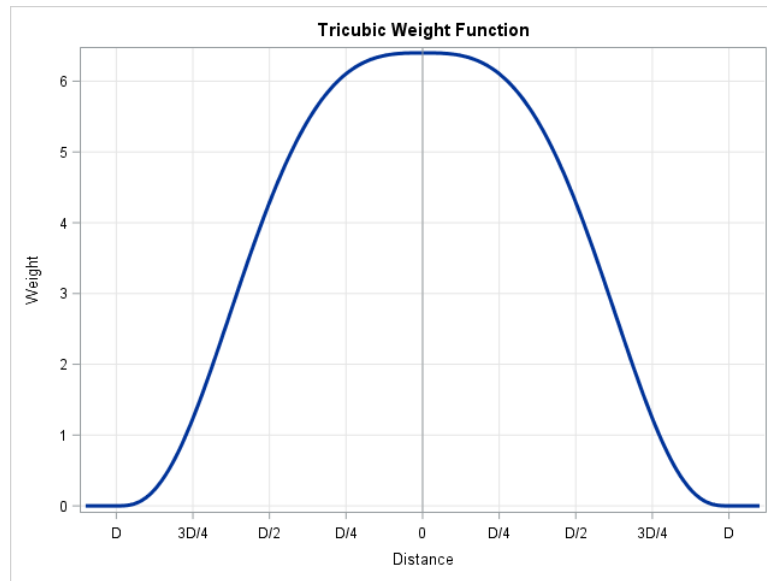


Figure 4.5: Graphique de la fonction de poids tri-cubique

Les valeurs de cette fonction de sont entre -1 et 1 . Comme cette fonction n'a de résultats positifs que pour $-1 \leq x \leq 1$, nous devons normaliser la distance en la divisant par la valeur maximale observée dans D . Plus concrètement,

$$w(x) = \begin{cases} \left(1 - \left|\frac{d(x,x')}{\max_i d(x_i,x')}\right|^3\right)^3 & |x| < 1, x_i \in D \\ 0 & |x| \geq 1 \end{cases}$$

Nous désignons ici par $d(x, x')$ la distance entre x , l'un des k plus proches voisins, et x' . L'effet de la normalisation est que les distances plus grandes seront associées à des poids plus faibles. À l'extrême, le point correspondant à la distance maximale aura un poids de zéro, et le point à distance nulle aura le poids le plus élevé possible - un. C'est ainsi que l'effet de "localité" est obtenu, en attribuant une plus grande importance aux données d'apprentissage les plus proches de l'endroit où nous voulons que la prédiction

soit calculée.

Nous sommes maintenant prêts à calculer l'estimation en utilisant une **régression linéaire pondérée** simple qui est entraînée avec les valeurs x de D , et les valeurs y correspondantes.

4.3 Autoregressive Integrated Moving Average

Le processus de prévision commence par l'identification du modèle de données en utilisant la moyenne mobile intégrée autorégressive. Ensuite, cette recherche doit développer l'estimation du paramètre pour le modèle ARIMA choisi. Le résidu est la différence entre la valeur observée et la valeur estimée de la quantité d'intérêt.

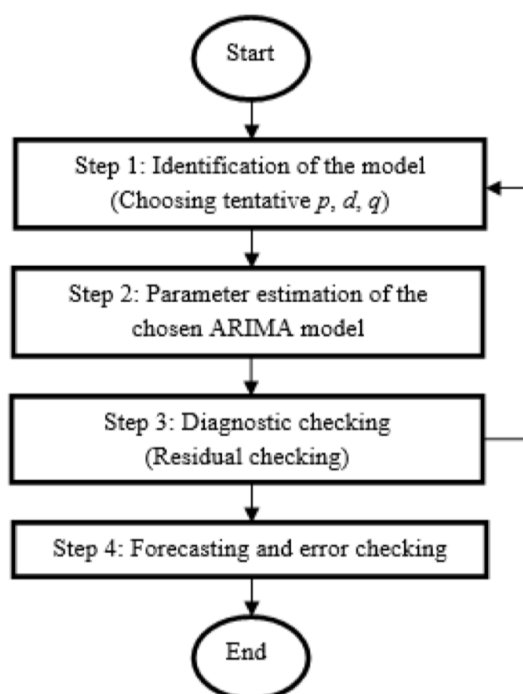


Figure 4.6: Procédure d'un modèle ARIMA

Cette section décrit la dérivation mathématique de la moyenne mobile intégrée autorégressive. Cette méthode combine la méthode autorégressive et la méthode des moyennes mobiles. Ensuite, cette méthode hybride est intégrée aux données du processus de différence. Tout d'abord, cette étude décrit la dérivation de la méthode autorégressive.

4.3.1 Autorégressif Modèle

Un **modèle autorégressif** est une représentation d'un type de processus aléatoire. La notation AR (p) indique un modèle autorégressif d'ordre p. Le modèle AR (p) est défini comme suit :

$$X_t = c + \sum_{i=1}^p \phi_i X_{t-i} + \epsilon_t \quad (4.16)$$

où ϕ_1, \dots, ϕ_p les paramètres du modèle, c la constante, et ϵ_t le bruit blanc.

4.3.2 Moving Average

Ensuite, nous dérivons l'équation de la **moyenne mobile- Moving Average-** . Le modèle de moyenne mobile spécifie que la variable de sortie dépend linéairement de la valeur actuelle et de diverses valeurs passées d'un terme stochastique.

$$X_t = \mu + \epsilon_t + \sum_{i=1}^q \theta_i \epsilon_{t-i} \quad (4.17)$$

où μ est la moyenne de la série, $\theta_1, \dots, \theta_q$ sont les paramètres du modèle, et $\epsilon_t, \dots, \epsilon_{t-q}$ sont des termes d'erreur de bruit blanc. La valeur de q est appelée l'ordre du modèle MA.

Après cela, nous faisons l'addition des deux modèles.

4.3.3 ARIMA

Ensuite, nous avons procédé à la dérivation de la moyenne mobile intégrée autorégressive (ARIMA). Étant donné une série chronologique de données X_t où t est un indice entier et les X_t sont des nombres réels. Un ARMA (p', q), après réarrangement :

$$X_t - \alpha_1 X_{t-1} - \dots - \alpha_{p'} X_{t-p'} = \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}$$

$$\left(1 - \sum_{i=1}^{p'} \alpha_i L^i\right) X_t = \left(1 + \sum_{i=1}^q \theta_i L^i\right) \epsilon_t \quad (4.18)$$

où L est l'opérateur de décalage, α_i sont les paramètres de la partie autorégressive du modèle, θ_i sont les paramètres de la partie moyenne mobile et ϵ_t sont les termes d'erreur. Les termes d'erreur sont généralement supposés être des variables indépendantes, identiquement distribuées, échantillonnées à partir d'une distribution normale avec une moyenne nulle. Dans l'analyse des séries temporelles, l'opérateur de retard, L ou opérateur de décalage arrière opère sur un élément d'une série temporelle pour produire l'élément précédent.

Dans l'analyse des séries temporelles, l'opérateur de décalage, L ou backshift, opère sur un élément d'une série temporelle pour produire l'élément précédent. Par exemple, étant donné une série temporelle :

$$X = \{X_1, X_2, \dots\}$$

Alors, $LX_t = X_{t-1}$ pour tous les $t > 1$. où L est l'opérateur de décalage. Notez que l'opérateur de décalage peut être élevé à des puissances entières arbitraires de sorte que :

$$L^k X_t = X_{t-k}$$

Maintenant, nous supposons que le polynôme $\left(1 - \sum_{i=1}^{p'} \alpha_i L^i\right)$ a une racine unitaire (un facteur $(1 - L)$ de multiplicité d). Elle peut alors être réécrite :

$$\left(1 - \sum_{i=1}^{p'} \alpha_i L^i\right) = \left(1 - \sum_{i=1}^{p'-d} \alpha_i L^i\right) (1 - L)^d \quad (4.19)$$

Un processus ARIMA(p,d,q) exprime cette propriété de factorisation polynomiale avec $p = p' - d$, et est donné par :

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right) (1 - L)^d X_t = \left(1 + \sum_{i=1}^q \theta_i L^i\right) \epsilon_t$$

L'équation peut être généralisée de la manière suivante, (4.20)

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right) (1 - L)^d X_t = \delta + \left(1 + \sum_{i=1}^q \theta_i L^i\right) \epsilon_t$$

Ceci définit l'équation d'un processus ARIMA (p,d,q) avec dérive $\frac{\sigma}{(1-\sum\phi_i)}$.

4.3.4 SARIMA

Le modèle SARIMA reprends les prémisses du modèle ARIMA, en y rajoutant 4 (Quatre) Hyper-paramètres, qui sont là afin de modeler la seasonalité du processus stochastique:

- P : Ordre autorégressif saisonnier.
- D : Ordre de différence saisonnière.
- Q : Ordre de moyenne mobile saisonnière.
- m : Le nombre de pas de temps pour une seule période saisonnière.

Les éléments de tendance peuvent être choisis par une analyse attentive des graphiques ACF et PACF en examinant les corrélations des pas de temps récents (par exemple 1, 2, 3).

De même, les diagrammes ACF et PACF peuvent être analysés pour spécifier les valeurs du modèle saisonnier en examinant la corrélation des pas de temps du décalage saisonnier.

Autocorrelation et Partial-autocorrelation

Autocorrelation Function La corrélation statistique résume la force de la relation entre deux variables.

Nous pouvons supposer que la distribution de chaque variable correspond à une distribution gaussienne (courbe en cloche). Si c'est le cas, nous pouvons utiliser le coefficient de corrélation de Pearson pour résumer la corrélation entre les variables.

Le coefficient de corrélation de Pearson est un nombre compris entre -1 et 1 qui décrit une corrélation négative ou positive respectivement. Une valeur de zéro indique l'absence de corrélation.

Nous pouvons calculer la corrélation des observations de la série temporelle avec les observations des étapes temporelles précédentes, appelées décalages (lags). Comme la corrélation des observations de la série temporelle est calculée avec les valeurs de la même série à des moments antérieurs, on parle de corrélation sérielle, ou d'auto-corrélation.

Le tracé de l'autocorrélation d'une série temporelle par décalage est appelé fonction d'autocorrélation, ou ACF. Ce graphique est parfois appelé corrélogramme ou graphique

d'autocorrélation.

Partial Autocorrelation Function: Une autocorrélation partielle est un résumé de la relation entre une observation dans une série temporelle avec des observations à des étapes temporelles antérieures, les relations des observations intermédiaires étant supprimées.

L'autocorrélation pour une observation et une observation à un pas de temps antérieur est composée à la fois de la corrélation directe et des corrélations indirectes. Ces corrélations indirectes sont une fonction linéaire de la corrélation de l'observation, avec les observations des pas de temps intermédiaires.

Ce sont ces corrélations indirectes que la fonction d'autocorrélation partielle cherche à éliminer. Sans entrer dans les mathématiques, voici l'intuition de l'autocorrélation partielle.

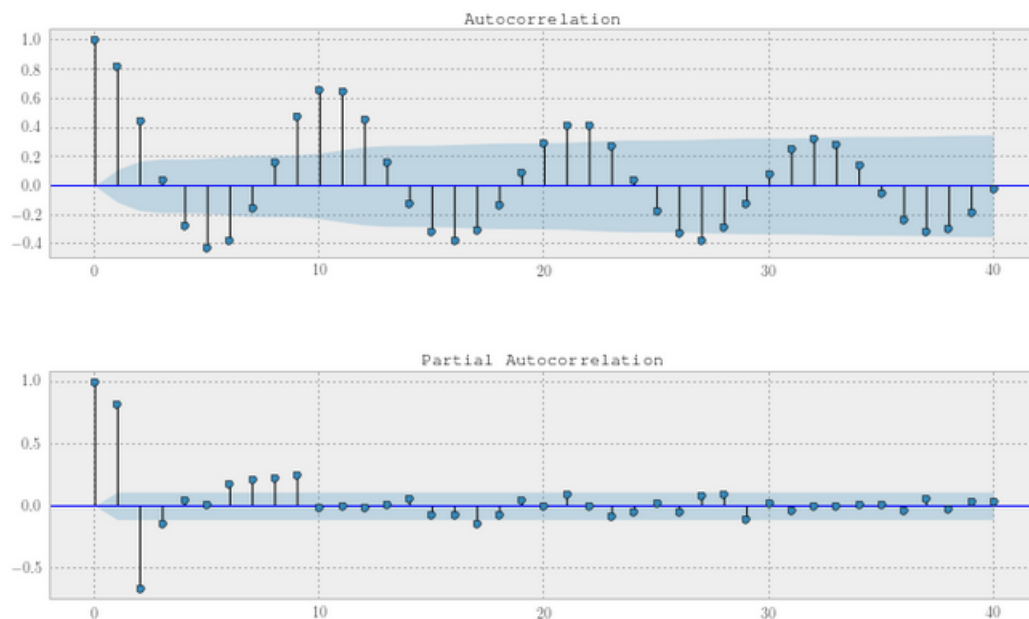


Figure 4.7: AutoCorrelation et Partial Autocorrelation

Quand confronté à ceux-là, voici certains indices auxquels il faut être attentifs:

- **Décroissance exponentielle vers zéro** : Modèle autorégressif (utilisez le graphique d'autocorrélation partielle pour identifier l'ordre p)
- **Oscillations amorties décroissant (exponentiellement) vers zéro** : Modèle autorégressif

- **Un ou plusieurs pics, le reste étant essentiellement nul** : Modèle de moyenne mobile (ordre q identifié par l'endroit où la courbe d'autocorrélation devient nulle)
- **Décroissance exponentielle commençant après quelques retards** : Modèle mixte autorégressif et moyenne mobile
- **Pas d'autocorrélations significatives (zéro ou proche de zéro)** : bruit blanc
- **Valeurs élevées à intervalles fixes** : Inclure des termes autorégressifs saisonniers saisonniers
- **Pas de décroissance vers zéro ou décroissance très lente** : Non-stationnarité ou effets de mémoire longue

4.4 Réseau de Neurones Artificiels et Réseaux de Neurones Récurrents

4.4.1 Introduction aux Réseaux de Neurones

Un réseau neuronal est un assemblage interconnecté d'éléments de traitement simples, d'*unités* ou de *nœuds*, dont la fonctionnalité est vaguement basée sur le neurone animal. La capacité de traitement du réseau est stockée dans les forces de connexion entre les unités, ou *poids*, obtenues par un processus d'adaptation ou d'apprentissage à partir d'un ensemble de modèles de d'entraînement.

La structure de réseaux de neurones artificiels

Les réseaux de neurones artificiels sont des systèmes de computation utilisés pour trouver la correspondance entre une **-input-** et une sortie **-output-**, ils sont constitués de plusieurs couches **-Layers-** et chaque couche contient un certain nombre de neurones **-unities-** où chaque neurone de chaque couche est connecté à tous les neurones de la couche précédente et de la couche suivante . chaque neurone de chaque couche reçoit une entrée des neurones de la couche précédente, multiplie le vecteur par quelques poids **-weights-** et somme le résultat puis applique une **fonction d'activation** linéaire. Les deux diagrammes de pipeline et de couches sont assez représentatif.

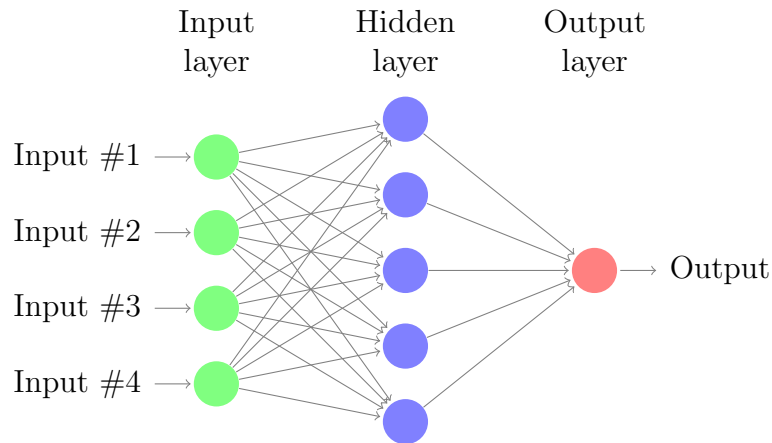


Figure 4.8: Diagramme des couches d'un réseau neuronal

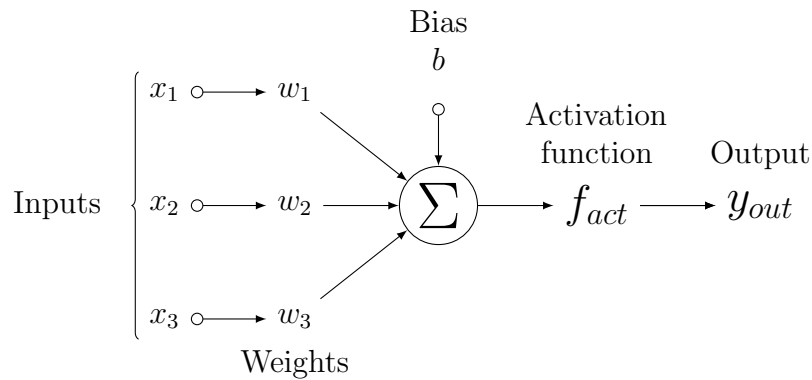


Figure 4.9: Diagramme du pipeline du réseau neuronal

Feedforward

Dans le calcul feed-forward, les couches successives s'alimentent les unes les autres dans le sens de l'avancement de l'entrée vers la sortie. L'architecture par défaut des réseaux à action directe suppose que tous les nœuds d'une couche sont connectés à ceux de la couche suivante. Suivant la formule (l'opération se répète pour chaque couche jusqu'à la couche de sortie.) :

$$a^{[l]} = g^{[l]} (W^{[l]} a^{[l-1]} + b^{[l]}) \quad (4.21)$$

Avec :

$g^{[l]}$: La fonction d'activation de la couche l .

$W^{[l]}$: La matrice des poids.

$a^{[l]}$: Les activations de la couche l .

b : la valeur du biais (est utile lorsque la somme devient 0).

Fonction d'activation

Une fonction d'activation dans un réseau neuronal définit comment la somme pondérée de l'entrée est transformée en une sortie à partir d'un ou plusieurs nœuds dans une couche du réseau. La fonction d'activation est parfois appelée "fonction de transfert". Si la plage de sortie de la fonction d'activation est limitée, elle peut être appelée "fonction d'écrasement". De nombreuses fonctions d'activation sont non linéaires et peuvent être appelées "non-linéarité" dans la couche ou la conception du réseau.

Un réseau peut comporter trois types de couches : les couches d'entrée qui prennent les données brutes du domaine, les couches cachées qui prennent les données d'une autre couche et transmettent les données de sortie à une autre couche, et les couches de sortie qui font une prédiction.

Toutes les couches cachées utilisent généralement la même fonction d'activation. Cela est nécessaire étant donné que les réseaux neuronaux sont généralement formés à l'aide de l'algorithme de rétro-propagation (Backpropagation of Error), qui requiert la dérivée de l'erreur de prédiction afin de mettre à jour les poids du modèle.










Name	Plot	Equation	Derivative
Identity		$f(x) = x$	$f'(x) = 1$
Binary step		$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{for } x \neq 0 \\ ? & \text{for } x = 0 \end{cases}$
Logistic (a.k.a Soft step)		$f(x) = \frac{1}{1 + e^{-x}}$	$f'(x) = f(x)(1 - f(x))$
Tanh		$f(x) = \tanh(x) = \frac{2}{1 + e^{-2x}} - 1$	$f'(x) = 1 - f(x)^2$
ArcTan		$f(x) = \tan^{-1}(x)$	$f'(x) = \frac{1}{x^2 + 1}$
Rectified Linear Unit (ReLU)		$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$
Parameteric Rectified Linear Unit (PReLU) ^[2]		$f(x) = \begin{cases} \alpha x & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} \alpha & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$
Exponential Linear Unit (ELU) ^[3]		$f(x) = \begin{cases} \alpha(e^x - 1) & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} f(x) + \alpha & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$
SoftPlus		$f(x) = \log_e(1 + e^x)$	$f'(x) = \frac{1}{1 + e^{-x}}$

Figure 4.10: Les Fonctions d'Activation les plus connus

Entraînement du réseau

La dérivée d'une fonction donne la direction dans laquelle la fonction augmente, et sa négative, la direction dans laquelle la fonction diminue.

L'apprentissage d'un modèle consiste simplement à minimiser la fonction de perte, et pour la minimiser, il faut se déplacer dans la direction négative de la dérivée. La rétropropagation est le processus de calcul des dérivées et la descente de gradient est le processus de descente par le gradient, c'est-à-dire l'ajustement des paramètres du modèle pour descendre dans la fonction de perte.

La rétropropagation^[6] est appelée ainsi car pour calculer la dérivée, vous utilisez la règle de la chaîne de la dernière couche (qui est celle directement connectée à la fonction de perte, car c'est elle qui fournit la prédiction) à la première couche, qui est celle qui prend les données d'entrée. Vous vous "déplacez de l'arrière vers l'avant".

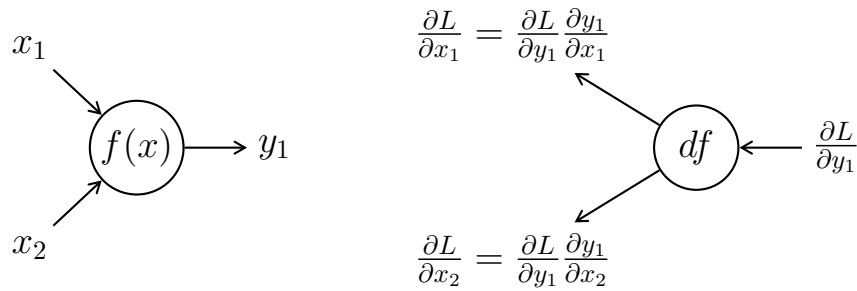


Figure 4.11: Forward (gauche) et Backward (droite) passes [20]

Dans la descente de gradient, on essaie d'atteindre le minimum de la fonction de perte par rapport aux paramètres en utilisant les dérivées calculées dans la micropropagation arrière. La méthode la plus simple consiste à ajuster les paramètres en soustrayant la dérivée correspondante multipliée par un taux d'apprentissage, qui régule le degré de progression dans la direction du gradient. Mais il existe des algorithmes plus avancés comme par exemple ADAM ou le NADAM, car le premier présente des inconvénients tels que :

- Peut dévier dans la mauvaise direction en raison de mises à jour fréquentes
- perdre les avantages de la vectorisation puisque nous traitons une observation à la fois.

4.4.2 Réseaux de Neurones Récurrents

Les réseaux récurrents sont des réseaux de neurones dans lesquels l'information peut se propager dans les deux sens, y compris des couches profondes aux premières couches. Pour cette raison, les RNNs sont particulièrement adaptés aux applications faisant intervenir le contexte, et plus particulièrement au traitement des séquences temporelles comme l'apprentissage et la génération de signaux, c'est à dire quand les données forment une suite et ne sont pas indépendantes les unes des autres. Néanmoins, pour les applications faisant intervenir de longs écarts temporels, cette « mémoire à court-terme » n'est pas suffisante. En effet, les RNNs « classiques » ne sont capables de mémoriser que le passé dit proche, et commencent à « oublier » au bout d'une cinquantaine d'itérations environ.

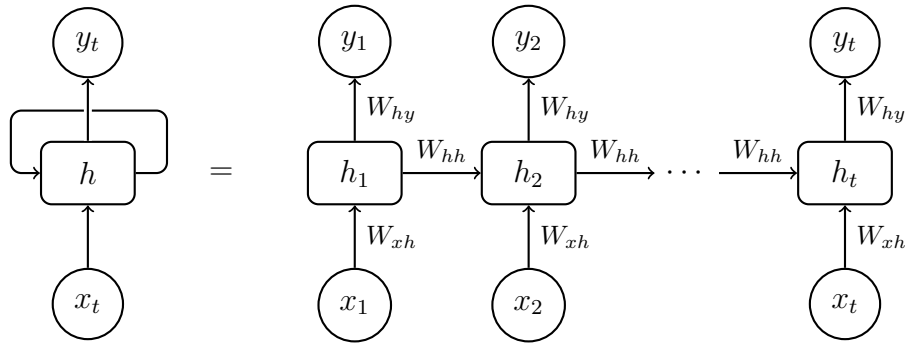


Figure 4.12: La Structure d'un Réseau de Neurones Récurrents[20]

Long Short Term Memory

LSTM outre-passe les problèmes liés à les RNN simples en utilisant un état cellulaire qui permet aux réseaux LSTM d'apprendre les dépendances à long terme. Les cellules LSTM possèdent également trois types de portes (gates) qui, ensemble, contrôlent le flux d'informations à l'intérieur de la cellule :

- **Porte d'oubli** : Contrôle quelles informations sont conservées ou rejetées au moment t
- **Porte d'entrée** : Contrôle les informations à stocker dans l'état cellulaire au moment t
- **Porte de sortie** : Contrôle la sortie finale de la cellule au moment t

Les valeurs des portes d'entrée, d'oubli et de sortie sont calculées à l'aide des équations(Figure 4.13 pour la notation):

$$\begin{aligned}
 i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \\
 f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \\
 o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o)
 \end{aligned}
 \tag{4.22}$$

Les portes d'entrée, d'oubli et de sortie sont calculées au temps t en utilisant des ensembles de poids et de biais $(W_{xi}, W_{hi}, W_{ci}, b_i)$, $(W_{xf}, W_{hf}, W_{cf}, b_f)$ et $((W_{xo}, W_{ho}, W_{co}, b_o)$ qui contrôle comment chacun de x_t , h_{t1} et c_{t1} affecte la valeur du porte respectivement. Une fonction sigmoïde est utilisée pour convertir les valeurs dans l'intervall de 0 à 1.

À chaque étape t , un état de cellule candidate \tilde{c}_t est calculé en utilisant des poids et un

terme de biais (W_{xc}, W_{hc}, b_c) qui fait correspondre les valeurs de l'entrée x_t et de l'état caché précédent h_{t-1} à l'état de la cellule candidate. Comme le nom l'indique, \tilde{c}_t sert de candidat pour remplacer l'état actuel de la cellule c_t au moment t . L'état de la cellule au moment t est obtenu en utilisant la porte d'oubli pour contrôler quelles informations sont conservées de l'état de la cellule précédente et la porte d'entrée pour contrôler quelles informations sont conservées de l'état de la cellule candidate. La sortie finale de la cellule est calculée en utilisant l'état actuel de la cellule c_t et la porte de sortie selon l'équation :

$$\begin{aligned}\tilde{c}_t &= \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \\ c_t &= f_t c_{t-1} + i_t \tilde{c}_t \\ h_t &= o_t \tanh(c_t)\end{aligned}\tag{4.23}$$

Les portes contrôlent le flux d'informations à l'intérieur de la cellule LSTM, leurs valeurs vont de 0 à 1 (fonction sigmoïde) et contrôlent les informations à conserver et celles à écarter lors de la mise à jour de l'état de la cellule (portes d'oubli et d'entrée) et lors du calcul de la sortie de la cellule LSTM (porte de sortie).

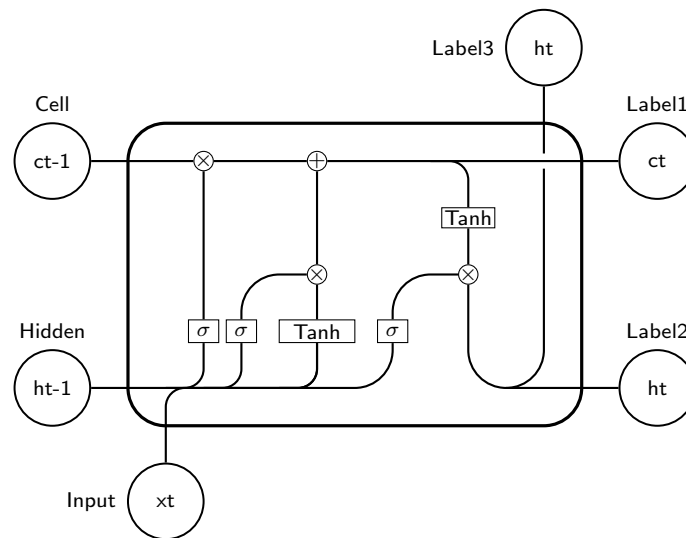


Figure 4.13: La Structure d'un Réseau Long-Short-Term-Memory

Gated Recurrent Units

Les unités récurrentes à porte (GRU) sont un mécanisme de porte dans les réseaux neuronaux récurrents, introduit en 2014 par Kyunghyun Cho et al. Le GRU est comme une mémoire à long court terme (LSTM) avec une porte d'oubli mais a moins de paramètres

que la LSTM, car il manque une porte de sortie. Les performances du GRU sur certaines tâches de modélisation de la musique polyphonique, de modélisation du signal de la parole et de traitement du langage naturel se sont avérées similaires à celles du LSTM .

La différence entre les RNN standard (Vanille) , LSTM et GRU

Tous les RNN ont des boucles de rétroaction dans la couche récurrente. Cela leur permet de conserver les informations en ”**mémoire**” au fil du temps.

Mais il peut être difficile d’entraîner des **RNN standard** à résoudre des problèmes qui nécessitent l’apprentissage de dépendances temporelles à *long terme*. En effet, le gradient de la fonction de perte décroît de manière exponentielle avec le temps (c’est ce qu’on appelle le problème du gradient évanescent).

Les réseaux LSTM sont un type de RNN qui utilise des unités spéciales en plus des unités standard. Les unités LSTM comprennent une ”**cellule mémoire**” qui peut conserver des informations en mémoire pendant de longues périodes. Un ensemble de portes est utilisé pour contrôler quand l’information entre dans la mémoire, quand elle est sortie et quand elle est oubliée. Cette architecture leur permet d’apprendre des dépendances à plus long terme.

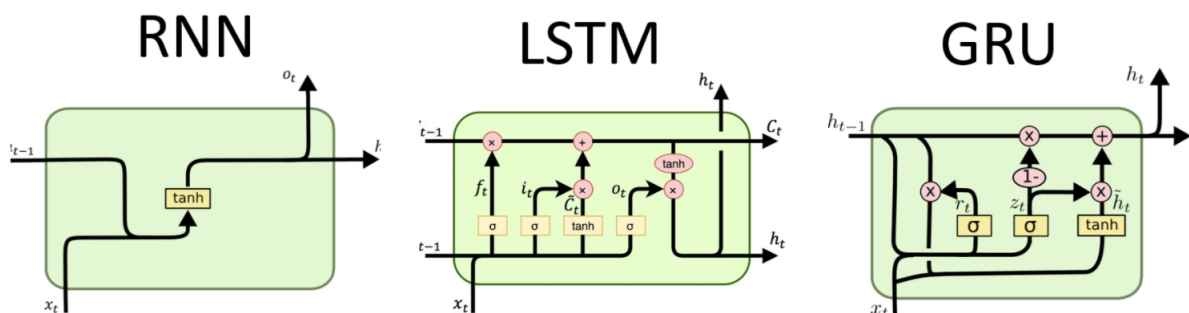


Figure 4.14: Les 3 types de RNN côte à côte

Les GRUs sont similaires aux LSTMs, mais utilisent une structure simplifiée. Ils utilisent également un ensemble de portes pour contrôler le flux d’informations, mais ils n’utilisent pas de cellules de mémoire séparées et ils utilisent moins de portes.

4.5 Estimateur (statistique)

Les métriques d'évaluation sont utilisées pour mesurer la qualité du modèle statistique ou d'apprentissage automatique. Il existe de nombreux types de métriques d'évaluation disponibles pour tester un modèle. Il existe d'autres métriques d'évaluation qui peuvent être utilisées et qui n'ont pas été répertoriées. L'évaluation des métriques consiste à utiliser une combinaison de ces métriques d'évaluation individuelles pour tester un modèle ou un algorithme.

Dans notre cas, nous serons plus intéressés par : Les métriques d'évaluation des **modèles de régression** Les métriques MSE, MAE, RMSE et R au carré sont principalement utilisées pour évaluer les taux d'erreur de prédiction et les performances du modèle dans l'analyse de régression.

- **MAE (Mean absolute error)** représente la différence entre les valeurs originales et prédites extraites en calculant la moyenne de la différence absolue sur l'ensemble des données.
- **MSE (Mean Squared Error)** représente la différence entre les valeurs originales et prédites extraites par le carré de la différence moyenne sur l'ensemble des données.
- **RMSE (Root Mean Squared Error)** est le taux d'erreur par la racine carrée de MSE.
- **R-squared (Coefficient de détermination)** représente le coefficient de l'adéquation des valeurs par rapport aux valeurs originales. La valeur de 0 à 1 est interprétée comme des pourcentages. Plus la valeur est élevée, plus le modèle est bon.

4.6 Implementation

Afin de pouvoir optimiser notre travail, tous les algorithmes et outil mathématique cités dans ce chapitre seront implémentés à l'aide du langage de programmation **Python**: qui est un langage de programmation interprété, multi-paradigme et multiplateformes. Il favorise la programmation impérative structurée, fonctionnelle et orientée objet. Il est doté d'un typage dynamique fort, d'une gestion automatique de la mémoire par ramasse-miettes

et d'un système de gestion d'exceptions. Il est s'avère utile pour le calcul scientifique et l'apprentissage automatique.

5 Méthodologie, Implémentation, Analyse et Résultats

5.1 Présentations des Données et la raison de ce Choix

Pour les besoins de ce travail nous avons utilisé les données de production proposées par la **Society of Petroleum Engineers - SPE** à des fins de recherches scientifiques. Le Dépôt d'archives de données est constitués de 53 puits en fichiers **.csv** contenant les débits de production de l'huile, du gaz et de l'eau, ainsi que divers données de pression (e.g. casing pressure, annular pressure ...etc) en plus des cumuls de productions.

Tous les puits sont situés aux U.S.A, dans différents états(Texas, Californie...etc) et produisent de différents types de formations rocheuses qui vont du **EAGLE FORD** au **HAYNESVILLE SHALE**.

Après une analyse des recherches passées, et des données présent nous avons constatés que:

- les formations à très faible perméabilité *e.g.* les schistes causaient encore des problèmes quand nous voulons faire des prévisions
- les méthodes empiriques conventionnelles *i.e.* les formules de Arps, présentent des résultats peu fiables quand confrontées à des formations non-conventionnelles.
- En présence d'irrégularités dans les données ou de valeurs aberrantes, l'ajustement des courbes (**Curve fitting**) donnera des résultats qui divergent des données réelles.

Sur la base de ces constatations nous allons diriger notre travail afin de remédier à ces problématiques.

Nous allons:

- Étudier les données de production à partir des formations de **HAYNESVILLE SHALE**, étant des Shales, Elles représentent une excellente bases d'études.
- Nous allons faire une étude de détection des valeurs aberrantes et l'effet de leurs

amputations sur la qualité du fitting et des prévisions en utilisant les méthodes empiriques anciennes et récentes.

- L'utilisation d'un modèle stochastique afin de faire des prédictions plus précises.
- L'utilisation des modèles de l'apprentissage profonds afin de modéliser des séries temporelles.
- Études Comparatives entre les trois approches: Empiriques, Stochastiques et L'apprentissage Profonds.

Dans la formation **HAYNESVILLE SHALE**, Le dépôt de données contient 14 puits producteurs de Gaz de schiste, néanmoins nous nous concentrerons que sur 4 d'entre eux que nous avons renommée à des fin des fin de praticité (HS01, HS02, HS04, HS10) car:

- Ces 4 là représentent un excellent échantillon qui permettra l'étude de la présence de valeurs aberrantes sur la qualité des résultats à l'utilisation des méthodes empiriques et le modèle stochastique.
- Des tailles diverses de jeu de données allant d'un très grand jeu de données (HS01), deux jeux de données avec une taille moyenne (HS04 et HS10) et un autre petit (HS02). Ce qui permettra d'avoir une étude en différentes circonstances réelles.

5.2 Méthodes Empiriques

5.2.1 Formule de Arps Avant Traitements

Pour cette partie là, Nous avons implémenter les trois variantes de Arps: **Hyperbolique**, **Exponentielle** et **Harmonique** sur script **Python**, ainsi que le modèle **LGM - Logistique Growth Model** qui est connu pour avoir fait ses preuves quand confronté aux données de formation à faibles perméabilité.

Nous trouvons ci-joint les courbes Débit en fonction du temps ($q(t)$) des quatre puits:

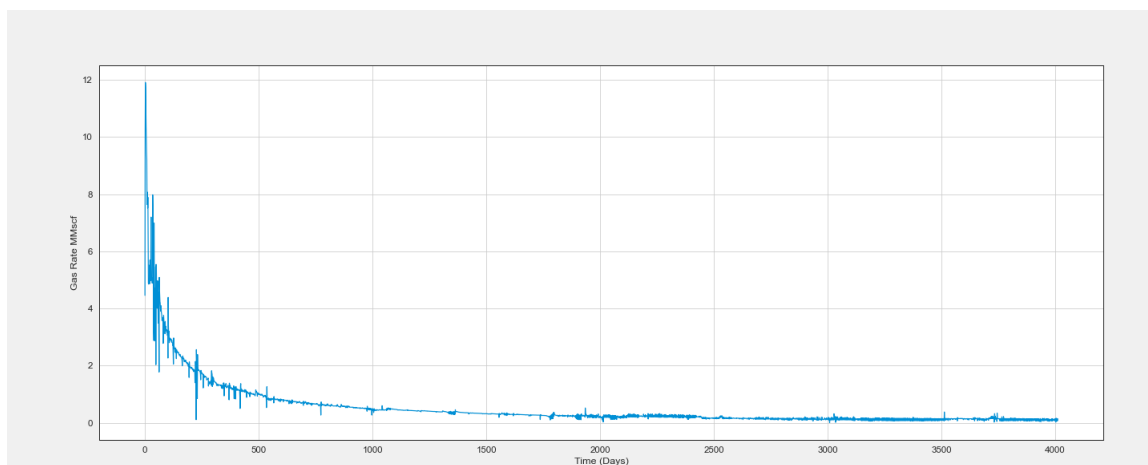


Figure 5.1: Puits HS01

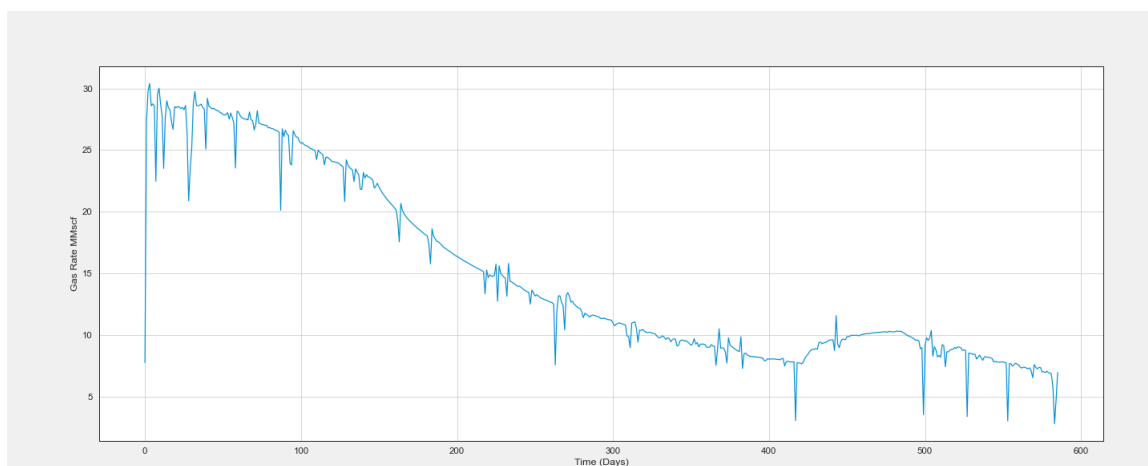


Figure 5.2: Puits HS02

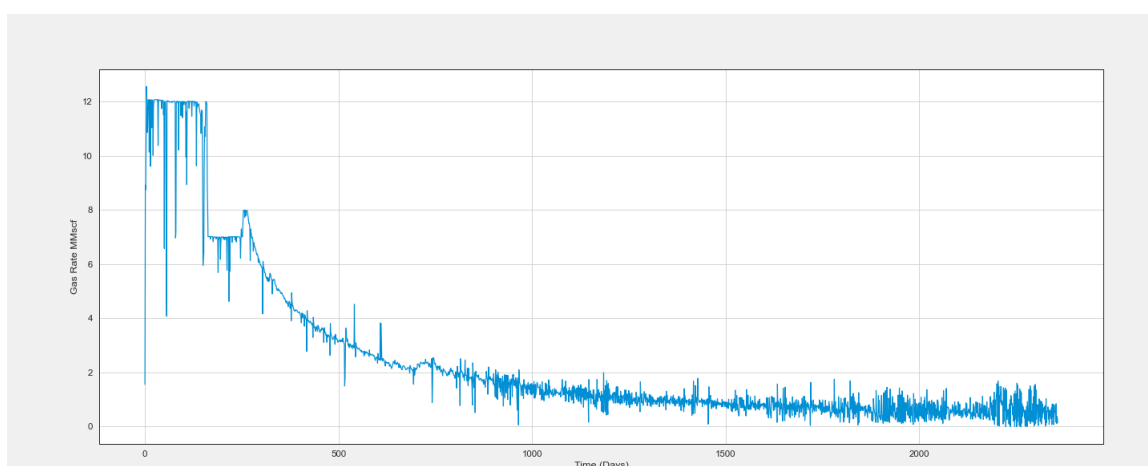


Figure 5.3: Puits HS04

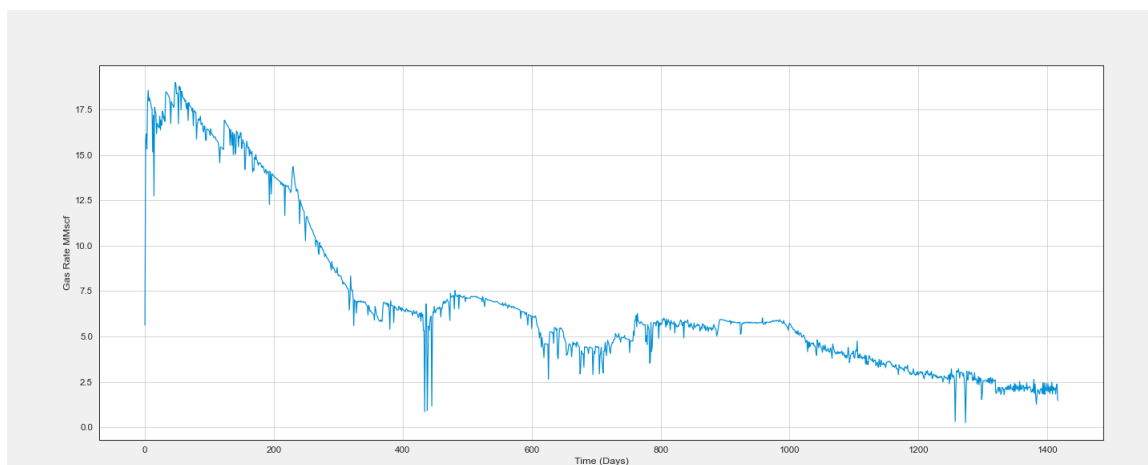


Figure 5.4: Puits HS10

Nous constatons des courbes assez diverses dans leurs structures allant de la plus lisse à celle présentant le plus d'irrégularités, ainsi que la plus courtes en termes de temporalité à des courbes qui représentent des +10 ans de productions.

Nous avons ajusté les paramètres de chaque modèle afin de caler au mieux aux données réelles sans aucun traitements au préalable. Nous obtenons les **Racine de l'écart quadratique moyen - RMSE** ce qui suit:

<i>Modèles</i>	<i>Puits</i>			
	HS01	HS02	H04	HS10
Hyperbolique	0.060876	4.73291	0.55372	2.9366
Exponentiel	0.347794	4.7356	0.61620	3.7758
Harmonique	0.072997	7.3819	1.14155	3.1418

Table 5.1: Les valeurs de la Racine de l'écart quadratique moyen pour chaque modèle-RMSE

<i>Puits</i>	<i>Les paramètres d'ajustements</i>	
	Taux de Déclin	b
HS01	0.0382	1.227
HS02	0.00289	0.0249
HS04	0.0027	0.222
HS10	0.00260	0.616

Table 5.2: Les valeurs du Taux de déclin et de la Caractéristique de base b en Déclin Hyperbolique

<i>Puits</i>	<i>Les paramètres d'ajustements</i>	
	Taux de Déclin	b
HS01	0.0132	0
HS02	0.00285	0
HS04	0.00235	0
HS10	0.00176	0

Table 5.3: Les valeurs du Taux de déclin et de la Caractéristique de base b en Déclin Exponentiel

<i>Puits</i>	<i>Les paramètres d'ajustements</i>	
	Taux de Déclin	b
HS01	0.0288	1
HS02	0.00453	1
HS04	0.00506	1
HS10	0.00328	1

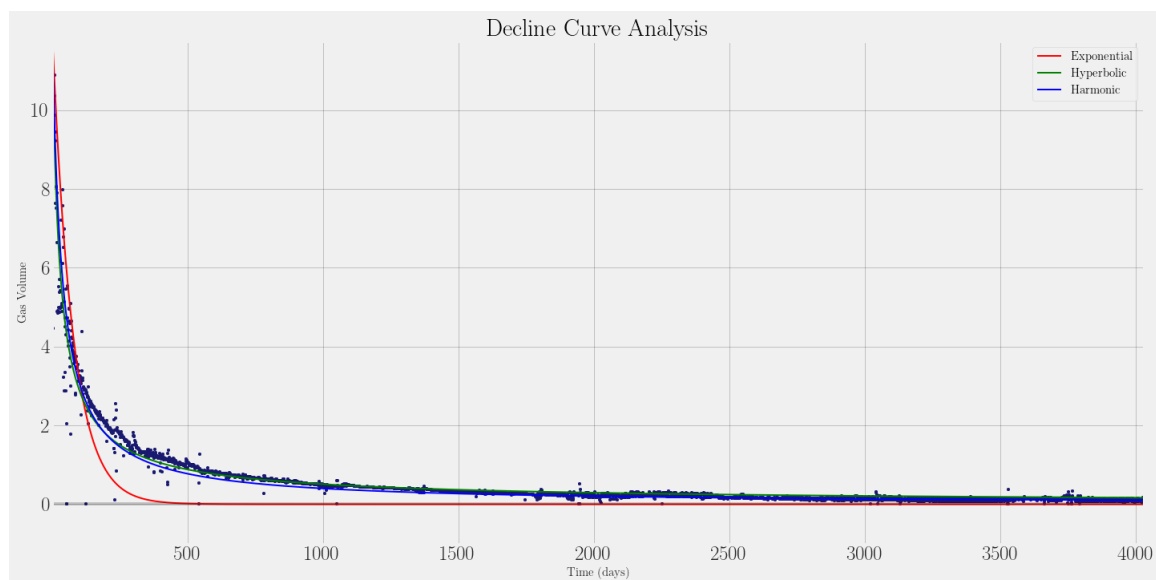
Table 5.4: Les valeurs du Taux de déclin et de la Caractéristique de base b en Déclin Harmonique

Observations

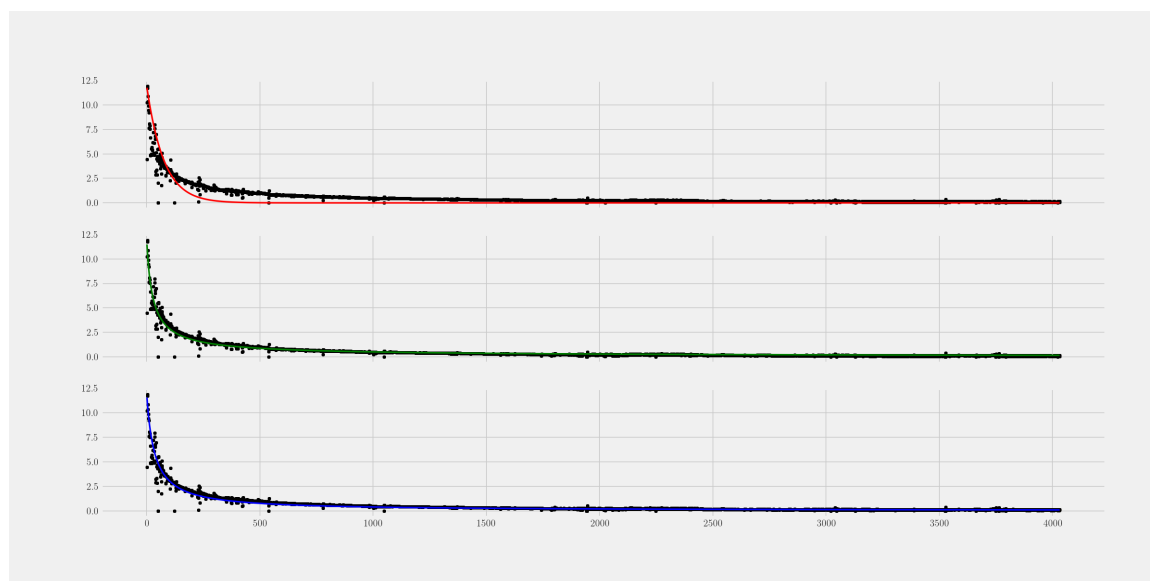
- Nous constatons que les résultats les plus précis après le **fitting** sont pour les données présentant le moins d'irrégularités et le moins de valeurs aberrantes, en outre les données de productions des puits: **HS01, HS04**.
- Nous ne remarquons pas de tendance particulière pour les puits produisant à partir des Formations de schiste (Exponentielles ou Harmonique), les meilleurs fitting sont **Hyperboliques**.
- Un autre constat que nous pouvons faire est que la valeur du b qui devrait $0 < b < 1$ a une valeur supérieur à 1 dans certains cas (e.g. 1.227) dans un ajustement avec une courbe hyperbolique⁶ et Les ajustements de courbe hyperbolique avec une constante de déclin (b) supérieure à 1 impliquent généralement que la production est influencée par un comportement transitoire, que nous retrouvons fréquemment

⁶la valeur de b a un impact important dans le temps tardif, elle conduira à des estimations différentes de EUR. La fiabilité de l'estimation de b augmente avec la maturité des données de production. La valeur de b rend compte d'un grand nombre d'événements et de processus physiques. Un grand nombre de publications sont consacrées à ce sujet.

dans les réservoirs non conventionnels(Ce qui est le cas ici, et cela est apparent dans dans le puits **HS01** car il a produit plus longtemps que tout les autres).

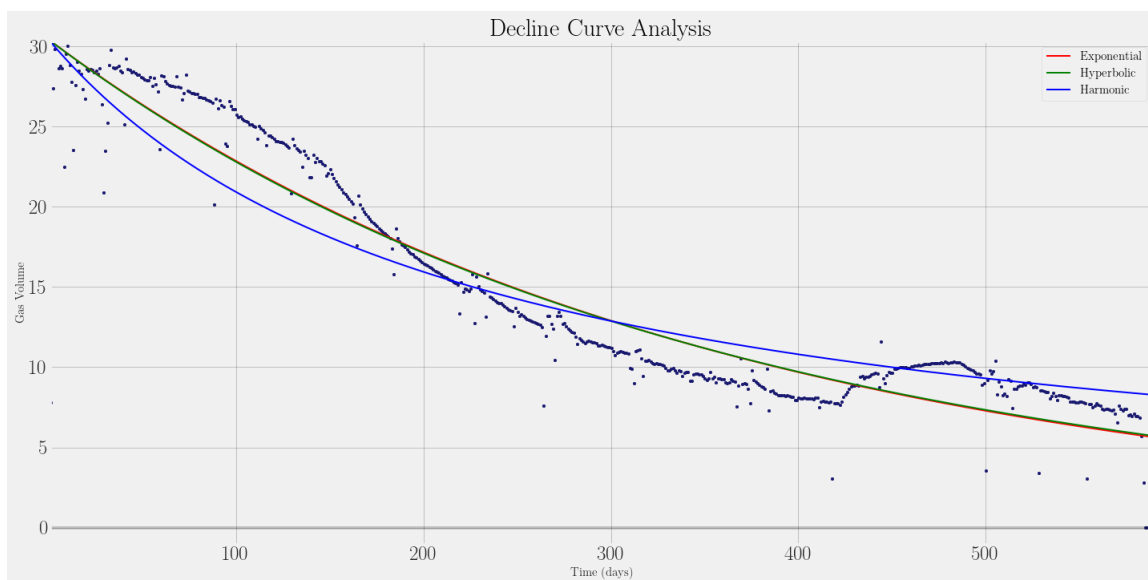


(a) Les Courbes de Arps

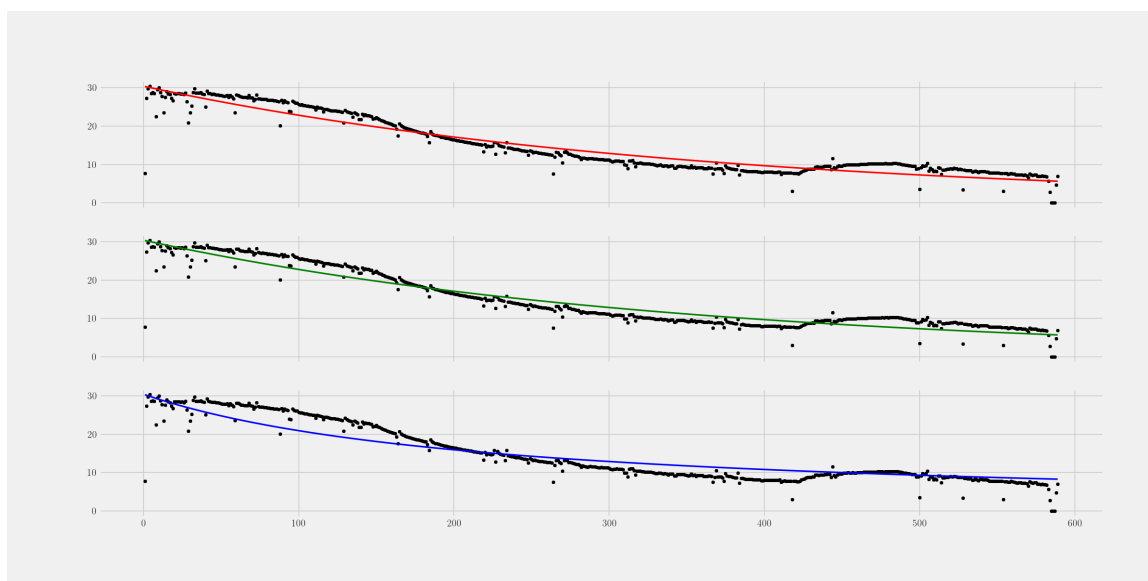


(b) Les courbes tracées individuellement

Figure 5.5: Les formules de Arps ajustées aux données du puit HS01

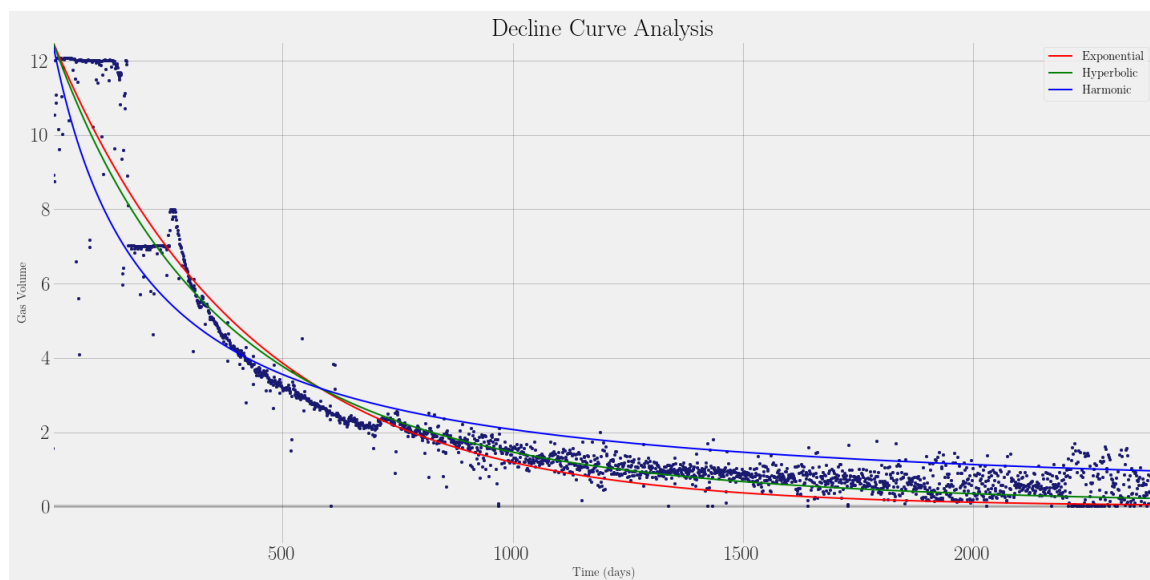


(a) Les Courbes de Arps

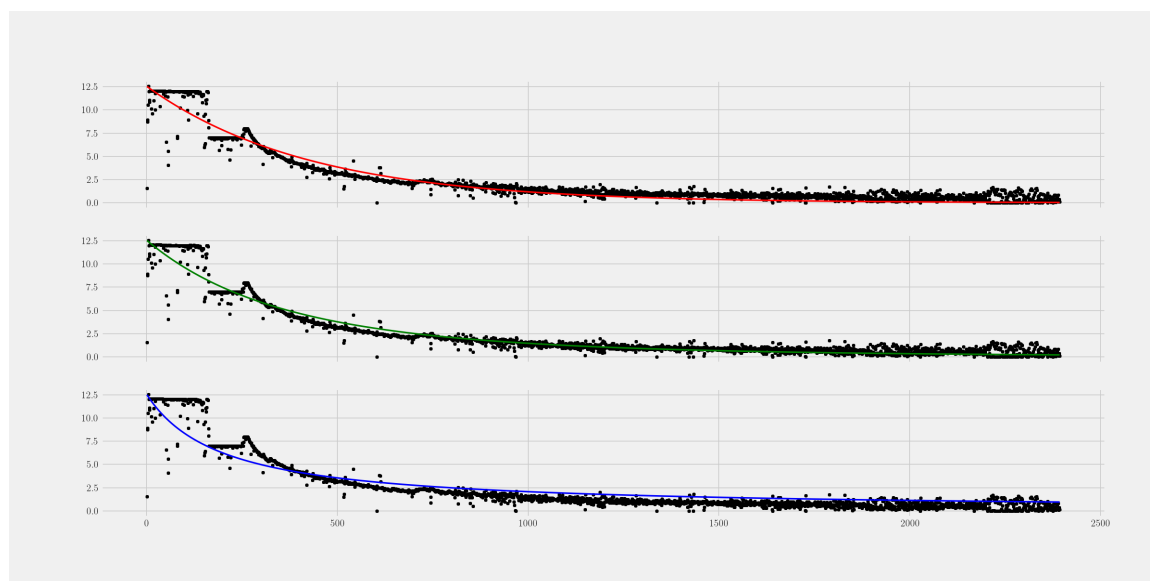


(b) Les courbes tracées individuellement

Figure 5.6: Les formules de Arps ajustées aux données du puit HS02

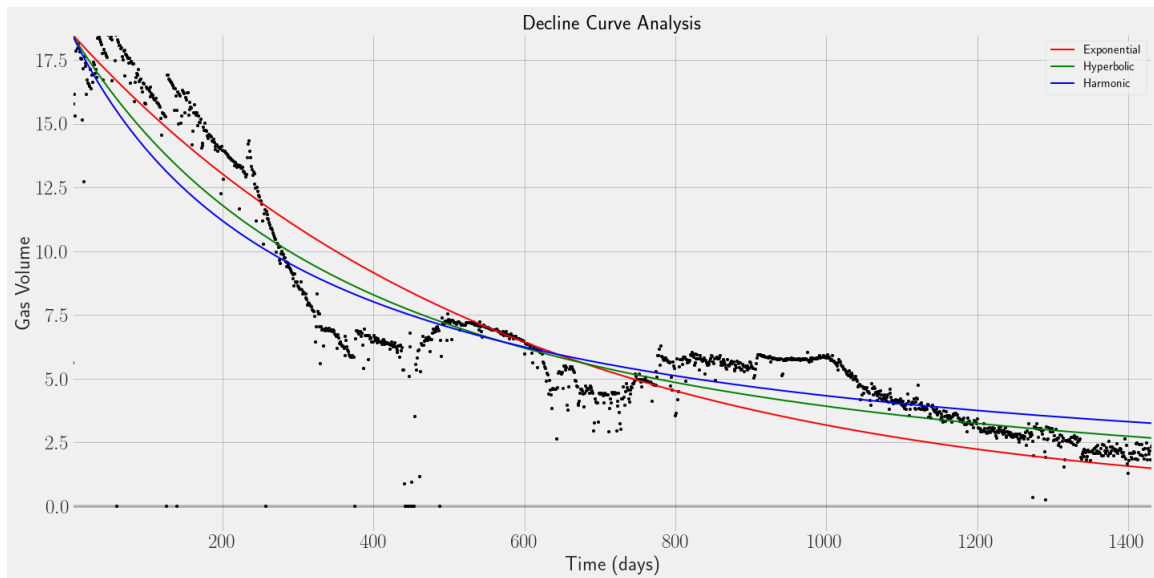


(a) Les Courbes de Arps

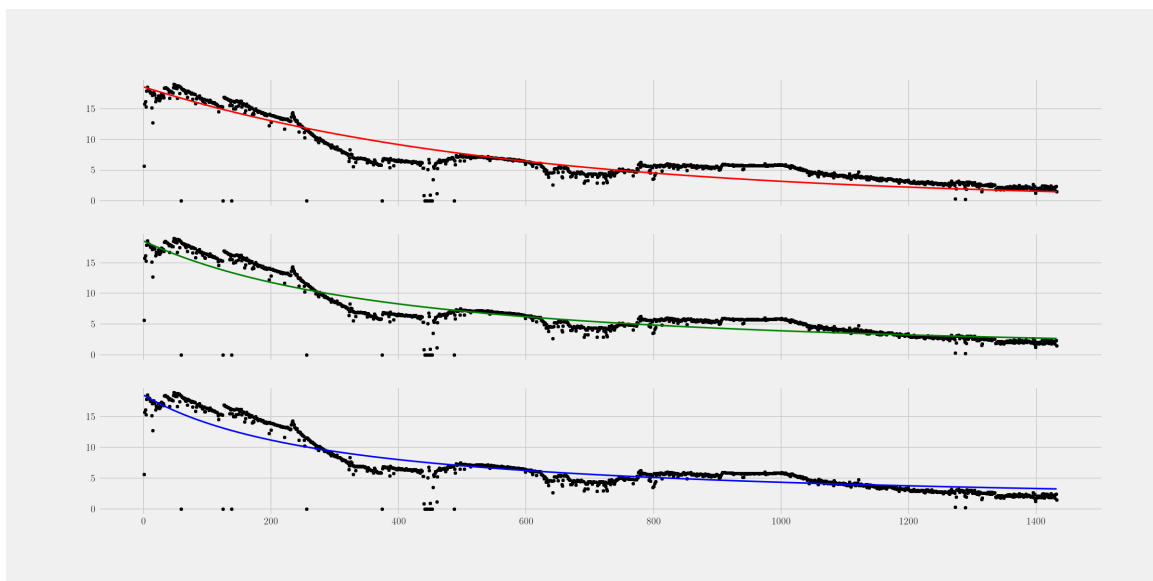


(b) Les courbes tracées individuellement

Figure 5.7: Les formules de Arps ajustées aux données du puit HS04



(a) Les Courbes de Arps



(b) Les courbes tracées individuellement

Figure 5.8: Les formules de Arps ajustées aux données du puit HS10

5.2.2 Formule de Arps Après Traitements

Nous avons adapté l'**Algorithme LOESS** à nos données de production, La régression locale (en anglais LOcal regrESSion : LOESS[3]) permet d'imputer des données manquantes. Pour cela, un polynôme de degré faible est ajusté autour de la donnée manquante par moindres carrés pondérés, en donnant plus de poids aux valeurs proches de la donnée manquante. Soit **une observation** à laquelle il manque **q** valeurs manquantes. On impute ces données manquantes par régression locale en suivant l'algorithme Expliqué ci-pres:

Algorithme LOESS

1. Obtention des k plus proches voisins $Y_{(i_1)}, \dots, Y_{(i_k)}$
2. Création des matrices $A \in \mathbb{R}^{k \times (n-q)}$, $B \in \mathbb{R}^{k \times q}$ et $w \in \mathbb{R}^{(n-q) \times 1}$ de sorte que :
 - Les lignes de A correspondent aux voisins privés des valeurs aux indices des données manquantes de Y_{i^*}
 - Les colonnes de B correspondent aux valeurs des voisins aux indices des données manquantes de Y_{i^*}
 - Le vecteur w correspond aux $(n - q)$ valeurs observées de Y_{i^*} :
 $w_j = (y_{i^*j})_{obs}$
3. Résolution du problème des moindres carrés

$$\min_{x \in \mathbb{R}^k} \| A^\top x - w \|$$

où $\| \cdot \|$ est la norme quadratique de \mathbb{R}^k .

4. Le vecteur u des données manquantes s'exprime alors par

$$u = B^\top x = B^\top (A^\top)^{-1} w$$

avec $(A^\top)^{-1}$ la matrice pseudo-inverse de A^\top .

Figure 5.9: L'algorithme de la Régression Local

Modèles	Puits			
	HS01	HS02	H04	HS10
Hyperbolique	0.08732	3.1360	0.2152	1.4451
Exponentiel	0.4305	3.1467	0.2722	2.1534
Harmonique	0.1167	0.8344	1.14155	1.6875

Table 5.5: Les valeurs de la Racine de l'écart quadratique moyen pour chaque modèle-RMSE Après Traitement avec LOESS

<i>Puits</i>	<i>Les paramètres d'ajustements</i>	
	Taux de Déclin	b
HS01	0.05065	1.394
HS02	0.002881	0.0501
HS04	0.00272	0.2092
HS10	0.00245	0.5771

Table 5.6: Les valeurs du Taux de déclin et de la Caractéristique de base b en Déclin Hyperbolique

<i>Puits</i>	<i>Les paramètres d'ajustements</i>	
	Taux de Déclin	b
HS01	0.01258	0
HS02	0.00281	0
HS04	0.00235	0
HS10	0.00171	0

Table 5.7: Les valeurs du Taux de déclin et de la Caractéristique de base b en Déclin Exponentiel

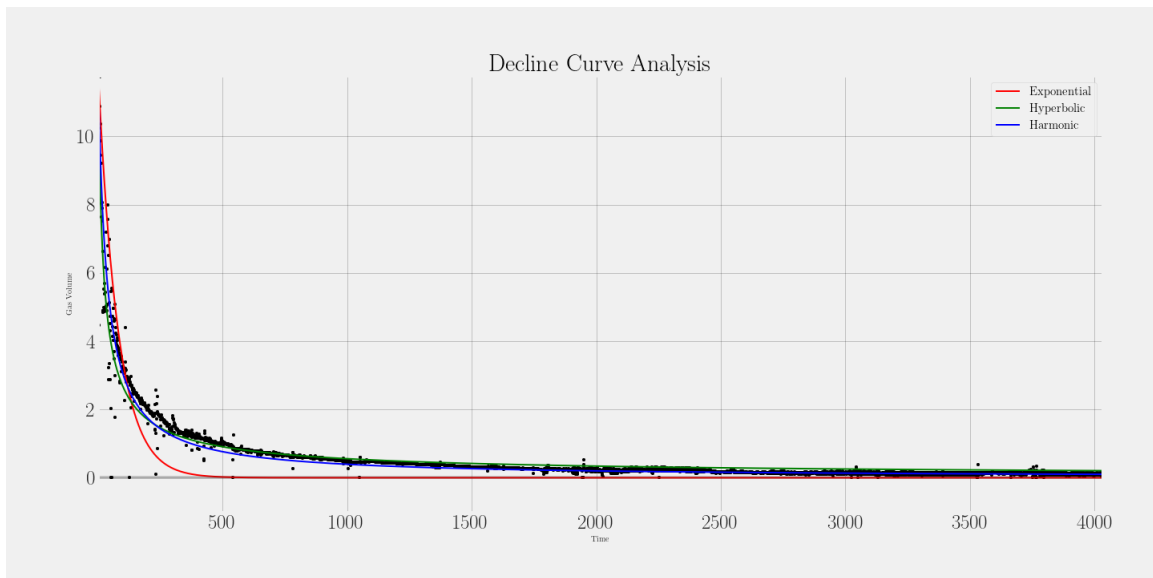
<i>Puits</i>	<i>Les paramètres d'ajustements</i>	
	Taux de Déclin	b
HS01	0.02922	1
HS02	0.00443	1
HS04	0.00506	1
HS10	0.00315	1

Table 5.8: Les valeurs du Taux de déclin et de la Caractéristique de base b en Déclin Harmonique

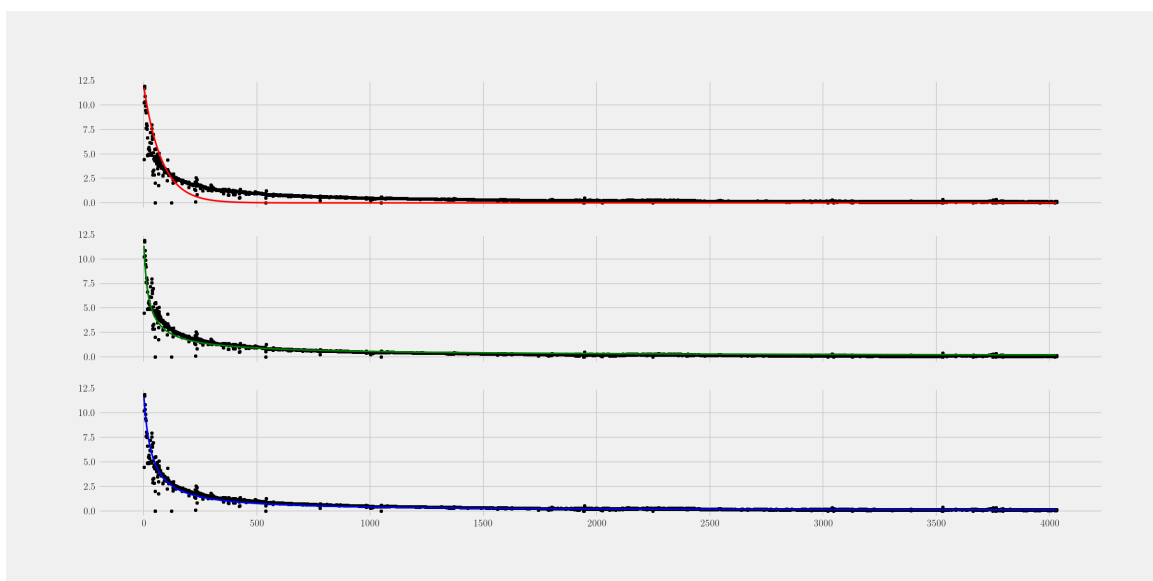
Observations

- Le premier constat qui se fait est que l'algorithme ne donne pas d'amélioration significative quand les données sont peu bruitées, comme le cas est des données du puits **H01** ou nous remarquons même une perte de précision ce qui pourrait s'expliquer par le fait que l'algorithme n'ai pas repéré d'irrégularité particulière.
- Néanmoins nous remarquons un gain en précision de **fitting** après pour les données bruitées tel que: **H02 et HS10**

-
- Néanmoins cela n'affecte pas l'obtention d'une valeur dans certains cas $b > 1$, ce qui pourrait suggérer le besoin d'un autre modèle.

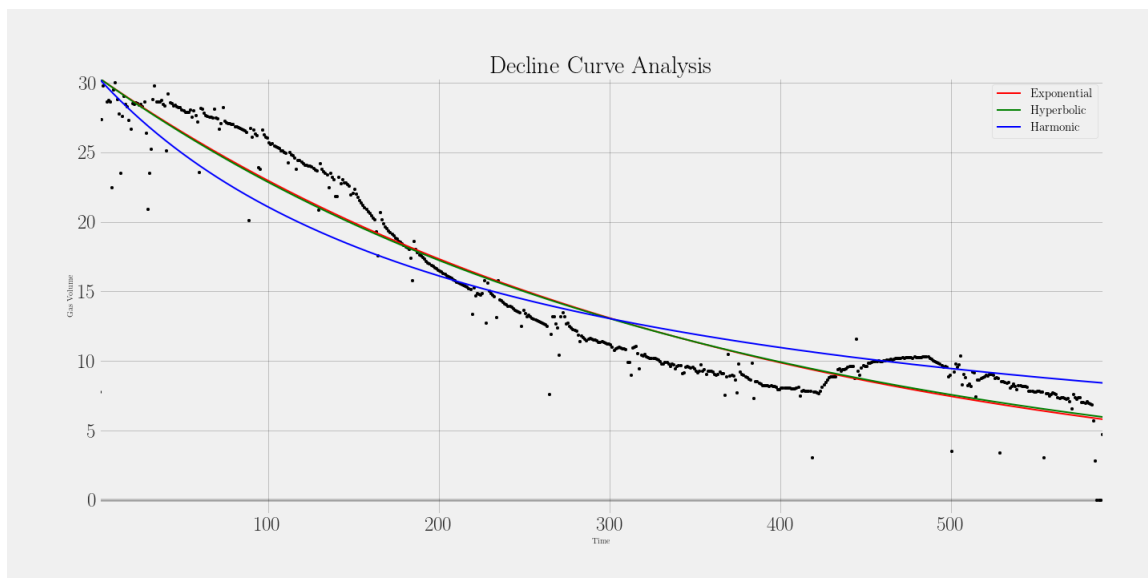


(a) Les Courbes de Arps

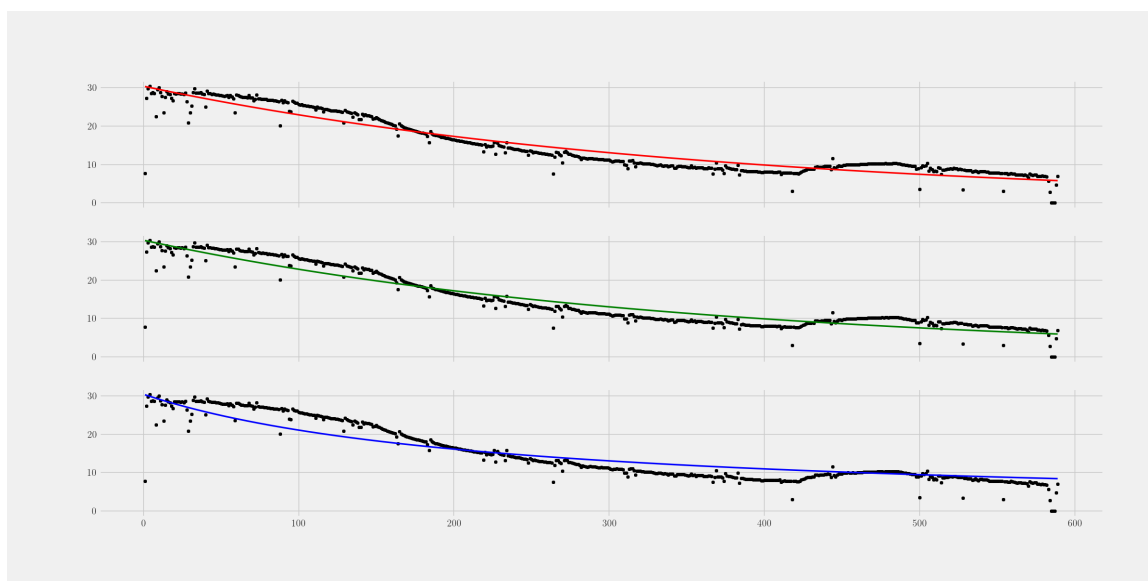


(b) Les courbes tracées individuellement

Figure 5.10: Les formules de Arps ajustées aux données du puits HS01 Après Traitements

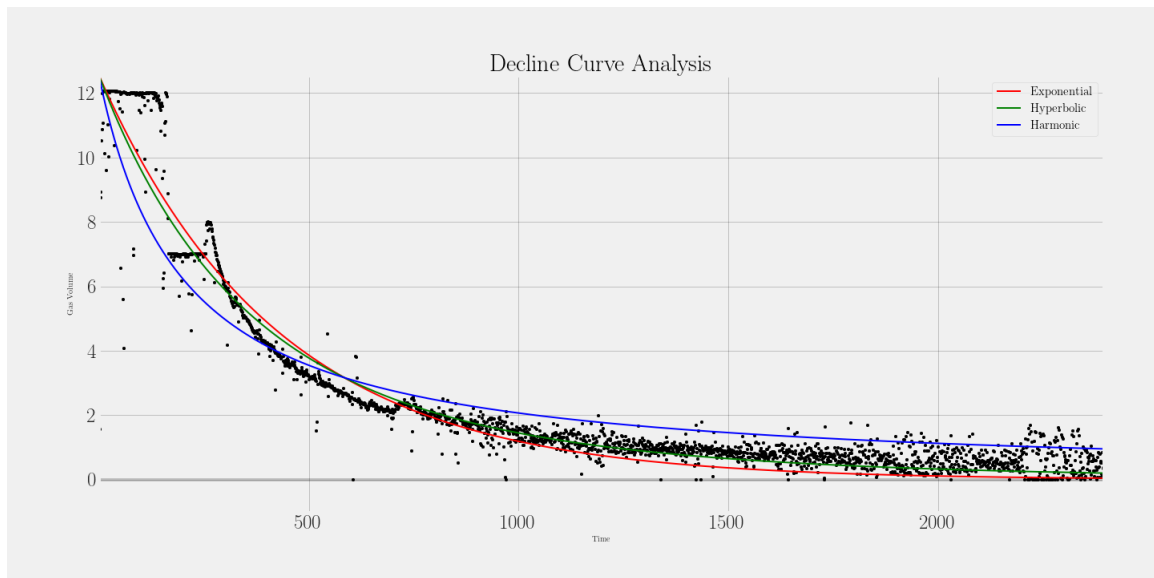


(a) Les Courbes de Arps

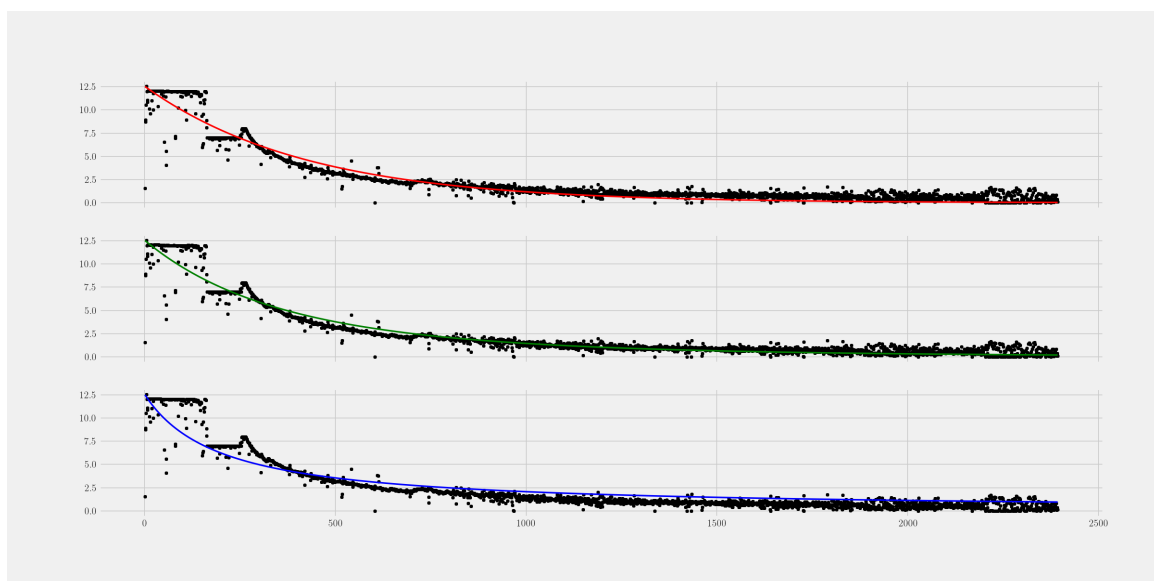


(b) Les courbes tracées individuellement

Figure 5.11: Les formules de Arps ajustées aux données du puits HS02 Après Traitements

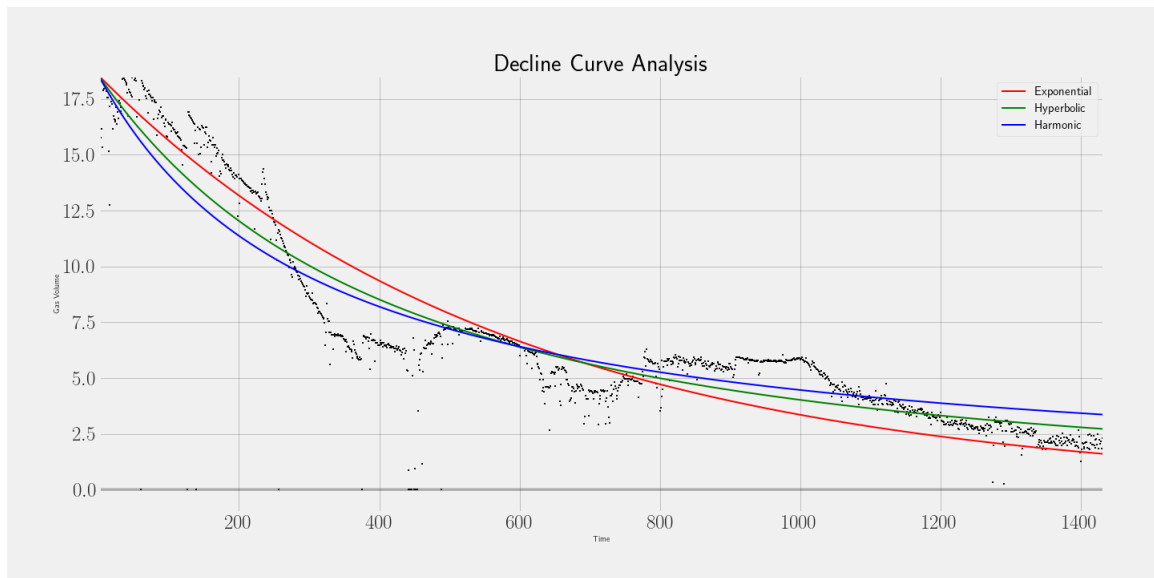


(a) Les Courbes de Arps

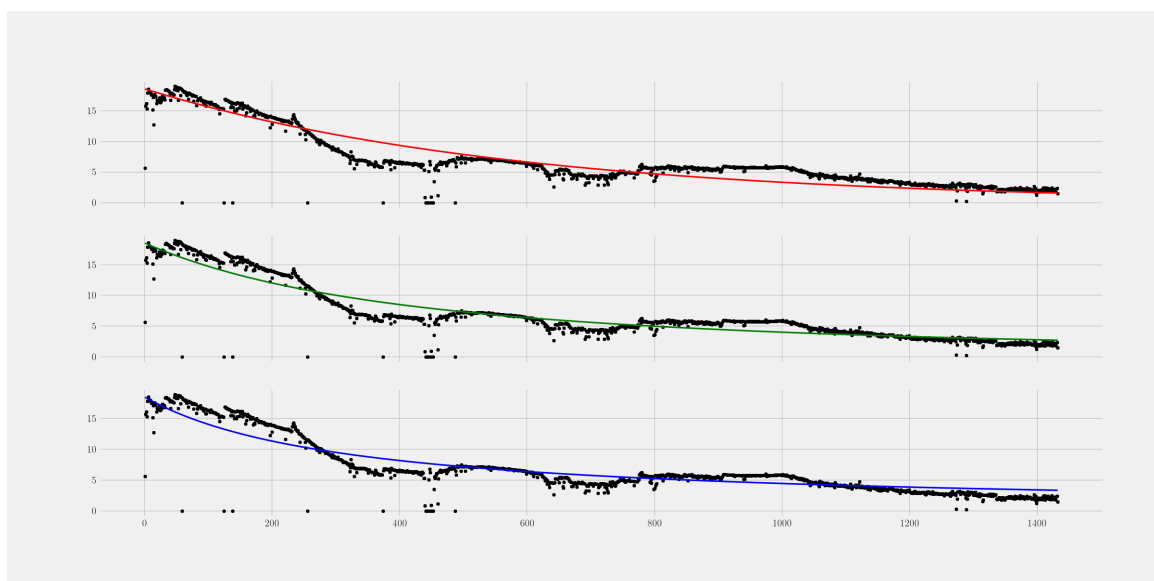


(b) Les courbes tracées individuellement

Figure 5.12: Les formules de Arps ajustées aux données du puits HS04 Après Traitements



(a) Les Courbes de Arps



(b) Les courbes tracées individuellement

Figure 5.13: Les formules de Arps ajustées aux données du puits HS10 Après Traitements

5.2.3 LGM Avant Traitements

<i>Modèles</i>	<i>Puits</i>			
	HS01	HS02	H04	HS10
LGM	0.0603	3.170	0.371	2.911

Table 5.9: Les valeurs de la Racine de l'écart quadratique moyen pour le modèle LGM-RMSE Avant Traitement avec LOESS

<i>Puits</i>	<i>Les paramètres d'ajustements</i>		
	K - Capacité de charge	n - Exposant hyperbolique	a - Constante
HS01	390864.884	0.8101	161.678
HS02	9317338.231	1.1309	706.531
HS04	5629361.023	1.1539	900.132
HS10	12353554.279	1.0265	855.437

Table 5.10: Les valeurs du K - Capacité de charge, n - Exposant hyperbolique, a - Constante Avant Traitement

Observations

Même sans pré-traitement, nous réalisons des résultats beaucoup plus précis et qui cale aux données réelles.

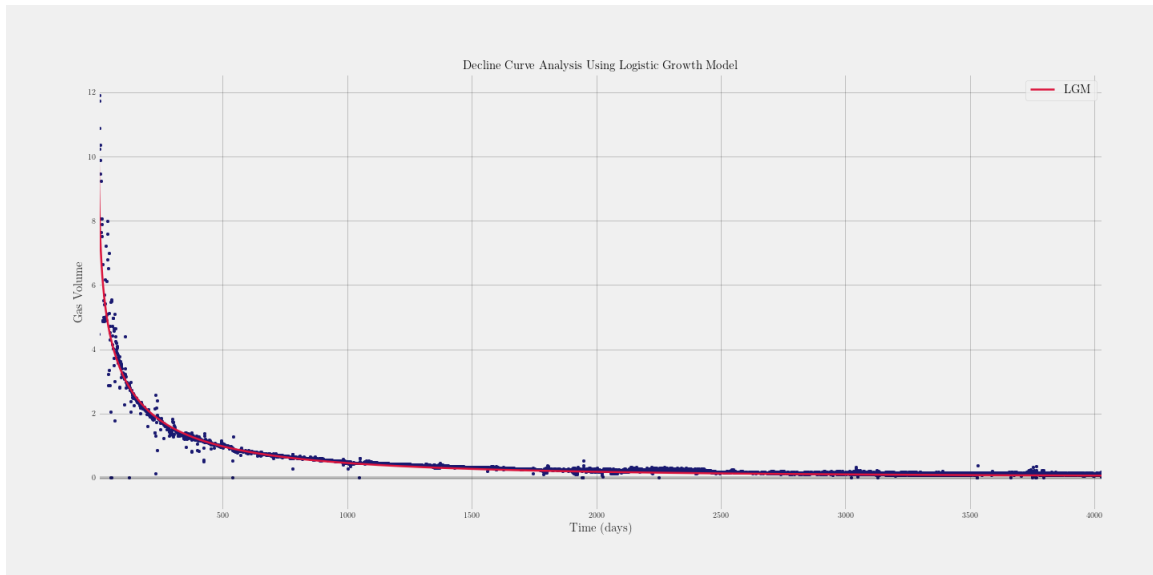


Figure 5.14: Modèle de croissance logistique ajusté du puits HS01 Avant Traitements

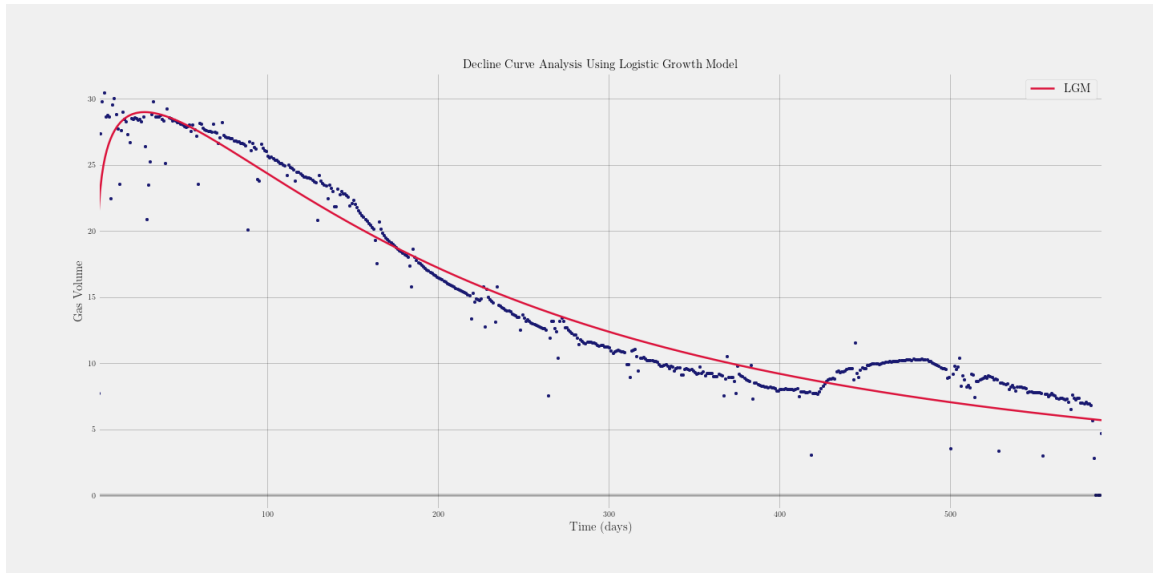


Figure 5.15: Modèle de croissance logistique ajusté aux données du puits HS02 Avant Traitements

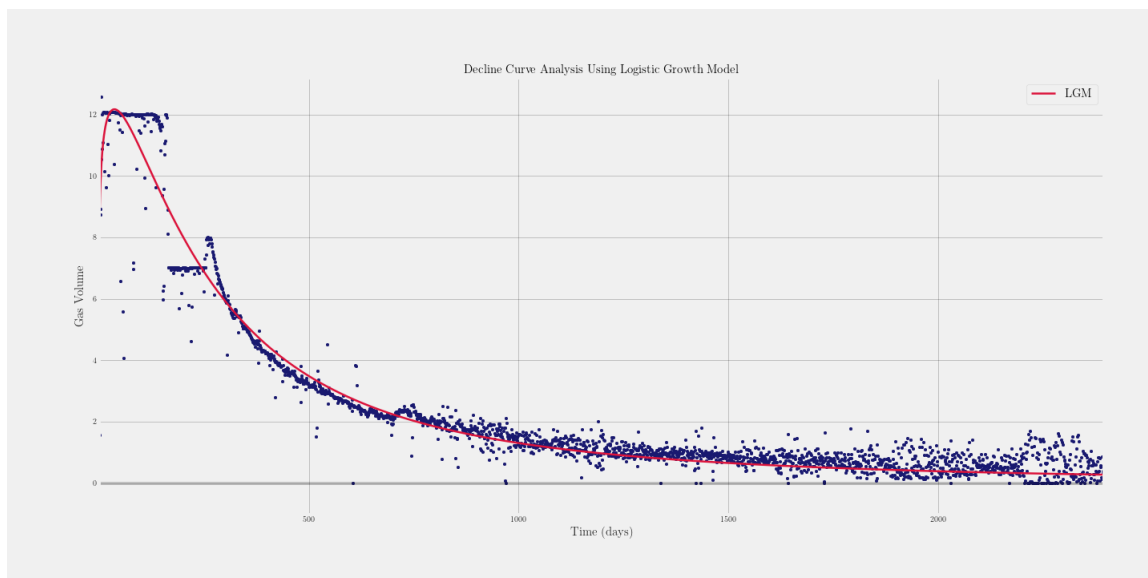


Figure 5.16: Modèle de croissance logistique ajusté du puits HS04 Avant Traitements

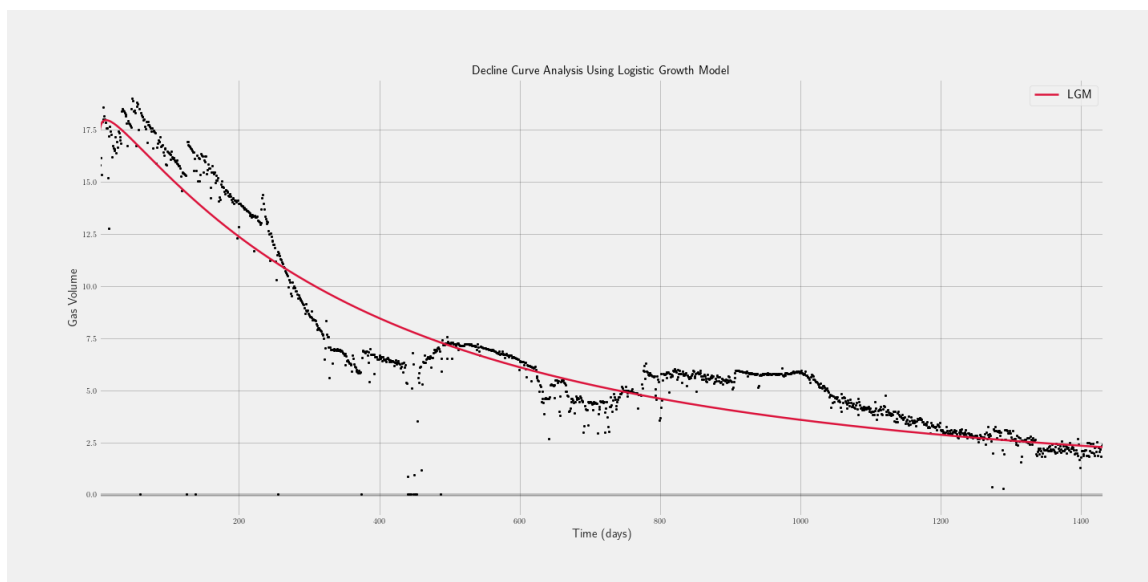


Figure 5.17: Modèle de croissance logistique ajusté du puits HS10 Avant Traitements

5.2.4 LGM Après Traitments

<i>Modèles</i>	<i>Puits</i>			
	HS01	HS02	H04	HS10
LGM	0.008	1.862	0.0689	1.408

Table 5.11: Les valeurs de la Racine de l'écart quadratique moyen pour le modèle LGM-RMSE Après Traitement avec LOESS

<i>Puits</i>	<i>Les paramètres d'ajustemnets</i>		
	K - Capacité de charge	n - Exposant hyperbolique	a - Constante
HS01	714143.361	0.9503	325.226
HS02	8950396.632	1.1066	651.934
HS04	4763458.252	1.1151	742.657
HS10	13274464.395	1.0263	887.678

Table 5.12: Les valeurs du K - Capacité de charge, n - Exposant hyperbolique, a - Constante Après Traitement

Observations

Comme prévu, Les résultats après amputation des valeurs aberrantes, en connu des une nette améliorations aux niveau de la métrique de précision.

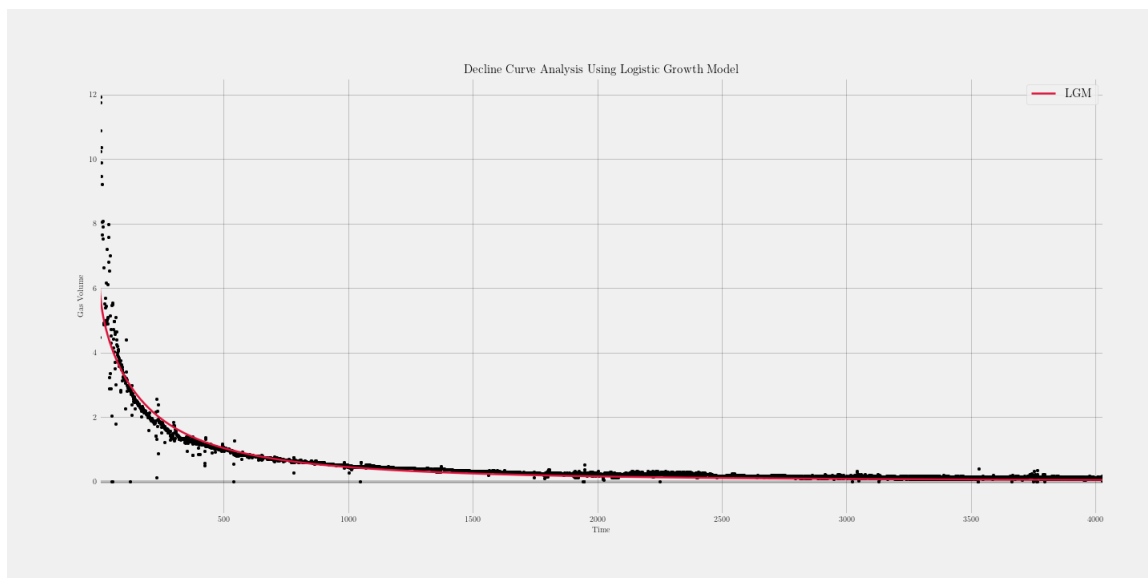


Figure 5.18: Modèle de croissance logistique ajusté du puits HS01 Après Traitements

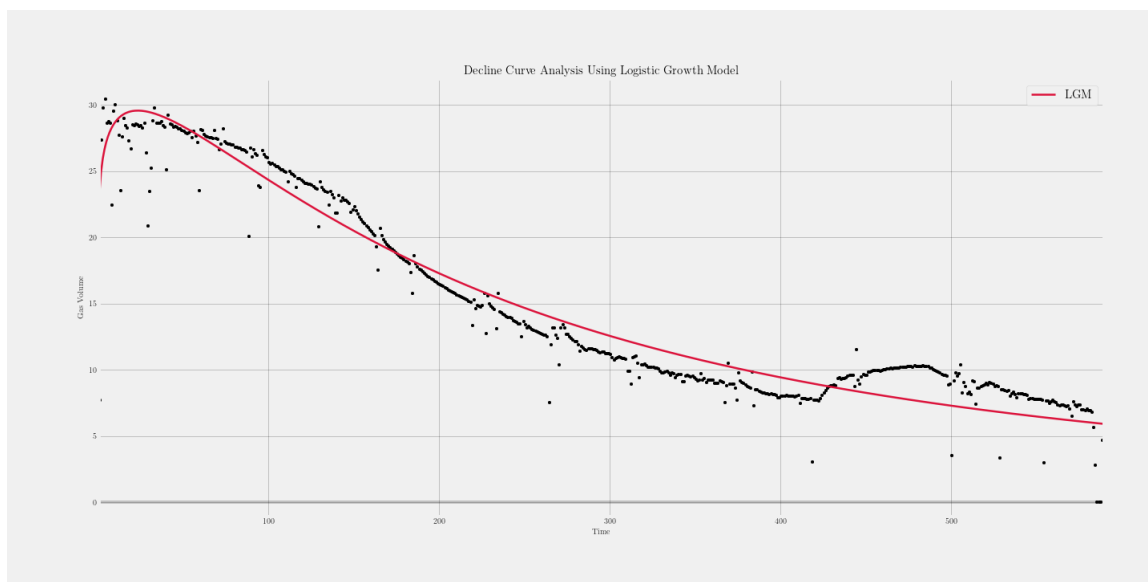


Figure 5.19: Modèle de croissance logistique ajusté du puits HS02 Après Traitements

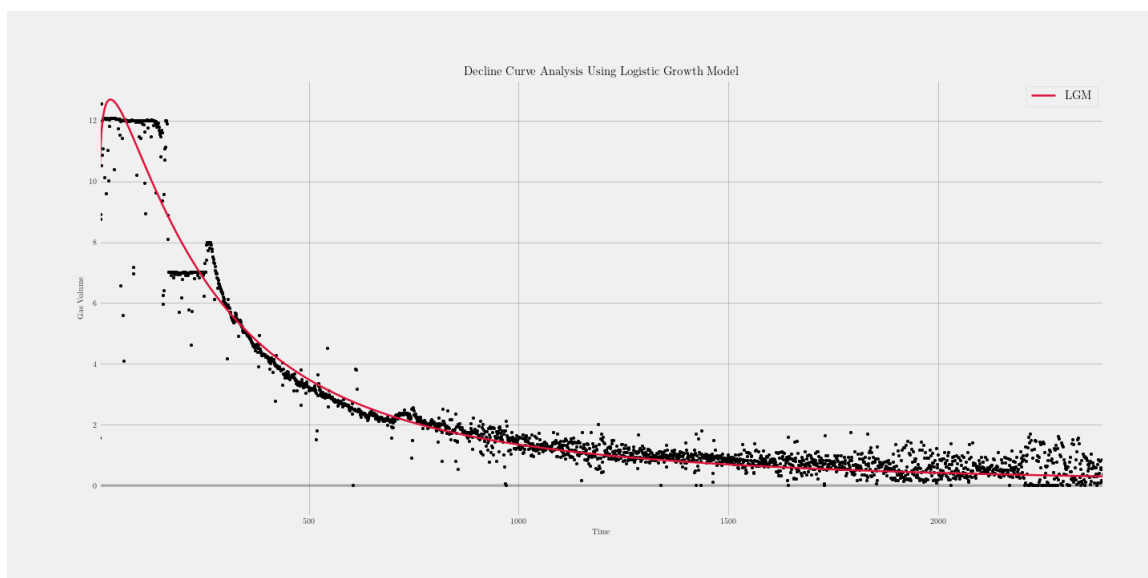


Figure 5.20: Modèle de croissance logistique ajusté du puits HS04 Après Traitements

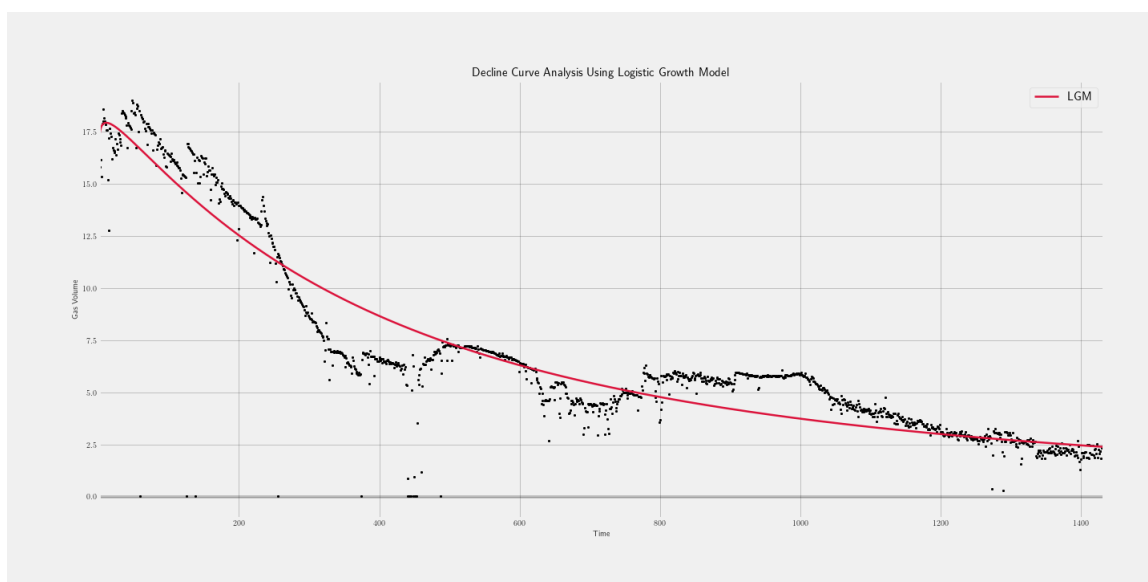


Figure 5.21: Modèle de croissance logistique ajusté du puits HS10 Après Traitements

5.2.5 Conclusion Partielle

Le modèle logistique permet d'établir une tendance très précise du comportement de déclin de ces puits. Il utilise le concept de la capacité de charge pour limiter la valeur totale de la production cumulée prédite. Le taux de production extrapolé finira toujours par atteindre 0. Et le modèle LGM couplé au LOESS améliore les résultats jusqu'à + de 20% plus précis.

5.3 Modèle SARIMA

Afin d'utiliser les modèles stochastiques, une certaine méthodologie est de rigueur. Pour cette fin là, Nous suivrons la méthode mise en place par Box et Jenkins[15] qui ont proposé une technique de prévision pour une série univariée laquelle est fondée sur la notion du processus ARIMA (SARIMA étant une Extension dans le cas de saisonnalité). Cette technique possède trois étapes: **identification**, **estimation** et **vérification**.

La première étape consiste à identifier le modèle $ARIMA(p, d, q)$ qui pourrait engendrer la série . Elle consiste, d'abord en transformer la série afin de la rendre stationnaire (le nombre de différenciations détermine l'ordre d'intégration: d), et ensuite d'identifier le modèle $ARMA(p, q)$ de la série transformée avec l'aide du corrélogramme (ACF) et du corrélogramme partiel (PACF). Le graphique des coefficients d'autocorrélation (corrélogramme) et d'autocorrélation partielle (corrélogramme partiel) donnent information sur l'ordre du modèle ARMA. Ainsi, si on observe que les deux premiers coefficients d'autocorrélation sont significatifs on identifiera le modèle suivant: MA(2). Et par la suite donner une estimation des paramètres de saisonnalité

La deuxième étape consiste à estimer le modèle ARIMA en utilisant une méthode non linéaire (moindres carrés non-linéaires ou maximum de vraisemblance). Ces méthodes sont appliquées en utilisant les degrés p , d et q trouvés dans l'étape d'identification.

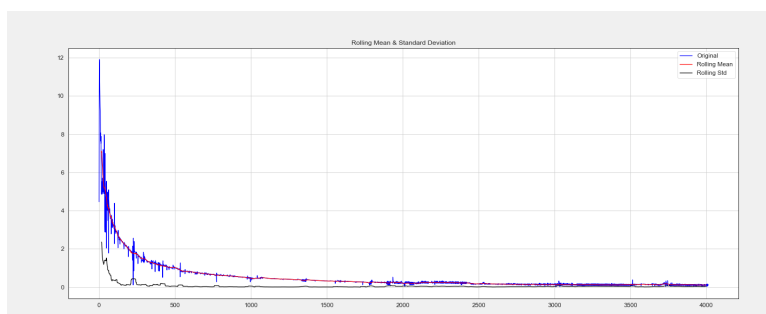
La troisième étape consiste à vérifier si le modèle estimé reproduit le modèle qui a engendré les données.

5.3.1 Identification

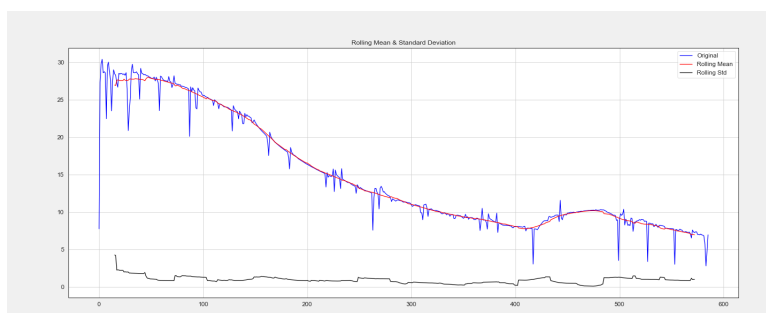
Afin de procéder à l'identification des paramètres cités en dessous, certaines plots représentent une aide précieuse, parmi cela, on trouve: ACF (Autocorrelation), PACF (Partial Autocorrelation), Moyenne mobile et Écart type mobile (Rolling Statistics).

Note: En pratique, l'identification de ces paramètres l'aide de l'ACF et du PACF implique une approche par tâtonnement... avec plus ou moins de subjectivité dans l'interprétation de ces fonctions : les "données réelles" présentent rarement des schémas simples. Et le plus souvent, Les outils graphiques peuvent avoir plusieurs interprétation du à certains schémas sous-jascents.

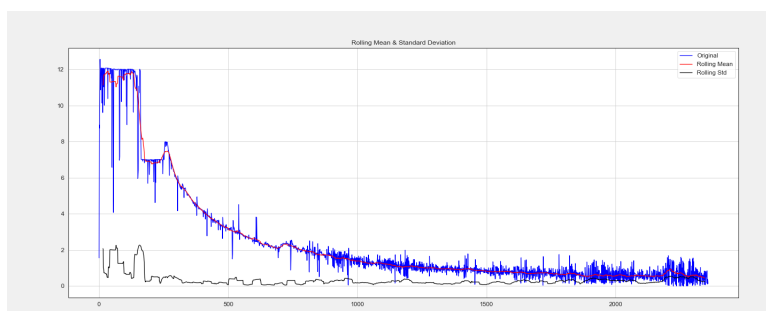
Rolling Statistics: Moyenne mobile et Écart type mobile



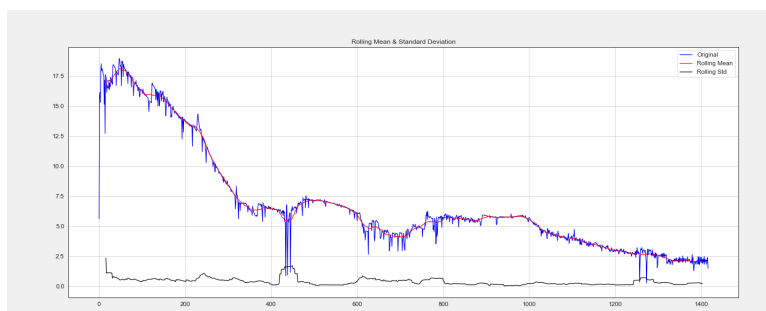
(a) HS01 Moyenne mobile et Écart type mobile



(b) HS02 Moyenne mobile et Écart type mobile



(c) HS04 Moyenne mobile et Écart type mobile



(d) HS10 Moyenne mobile et Écart type mobile

Figure 5.22: Moyenne mobile et Écart type mobile avec une fenêtre de 30 jours

Observations

Nous observons que la moyenne mobile et l'écart-type ne sont pas constants par rapport au temps (tendance décroissante).

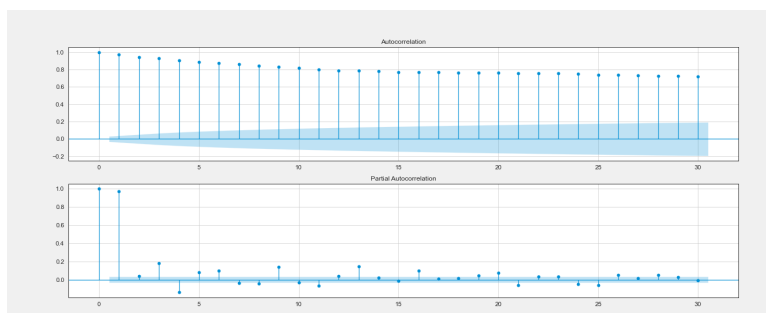
Cela peut signifier une non-stationnarité de ces séries temporelles (Une non-stationnarité que nous pouvons confirmer dans la section suivante.)

Autocorrelation et Partial-Autocorrelation

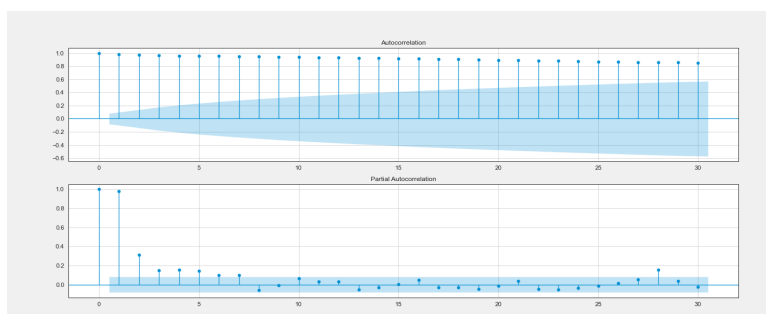
En étudiant les ACF et PACF, nous pouvons constater⁷:

- Les premiers groupes de plot représentent les couples ACF et PACF de chaque puits étudiés sans un traitement aux préalables. Nous remarquons que toutes les courbes **ACF** présentent une: *lente décroissance de la corrélation résiduelle qui est un signe que la différenciation peut être nécessaire car non-stationnaire, ce qui confirme nos soupçons rencontrés avec les moyennes et écart types mobiles.*
- Nous pouvons résoudre le problème de non-stationnarité des séries temporelles en dérivant une fois ce qui permet de donner les résultats obtenues dans le deuxième lot de ACF et PACF qui montre que la série est maintenant **stationnaire**, et donc $d = 1$
- Et le plot PACF se "**coupe - cuts off -**" après le 1er décalage -Lag- ce qui signifie: $AR(1)$
- Un ACF avec une autocorrélation significative uniquement au premier Lag est un indicateur d'un possible modèle $MA(1)$.
- Même avec le premier ordre de différenciation, nous observons qu'il y a toujours une lente décroissance résiduelle dans les graphiques ACF à une période de décalage saisonnière. Ceci suggère donc l'application d'une différence saisonnière.
- Nous avons remarqué qu'après une modélisation du modèle $SARIMA(P, D, Q)_m$, une période saisonnière de $m = 2$ donnez le meilleurs fitting, et les autres P, D, Q Sont analogues à ceux du $ARIMA(p, d, q)$

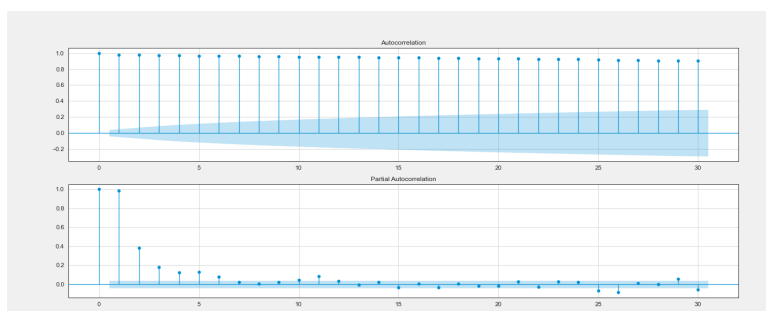
⁷La détermination des AR se fait mieux avec le PACF, et la détermination du MA se fait mieux avec le ACF



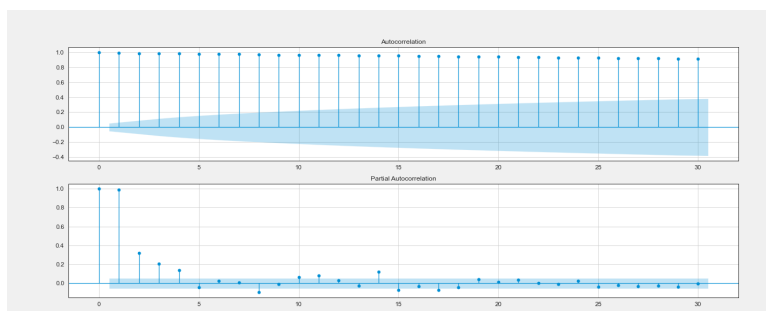
(a) Autocorrelation et Partial-Autocorrelation HS01



(b) Autocorrelation et Partial-Autocorrelation HS02

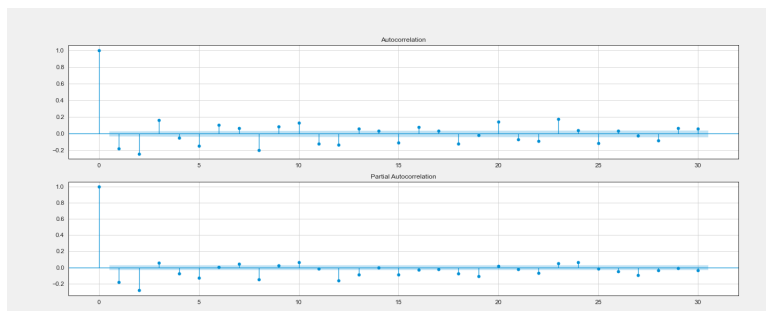


(c) Autocorrelation et Partial-Autocorrelation HS04

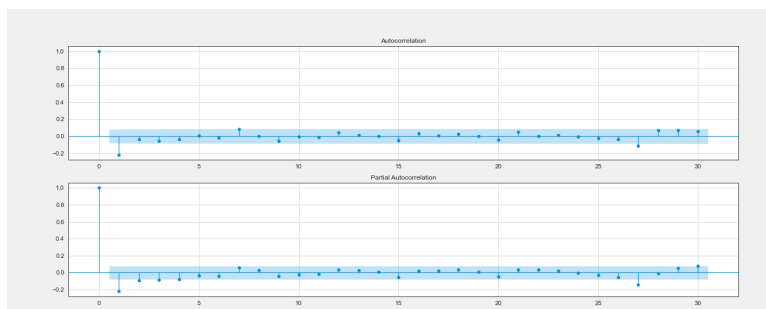


(d) Autocorrelation et Partial-Autocorrelation HS10

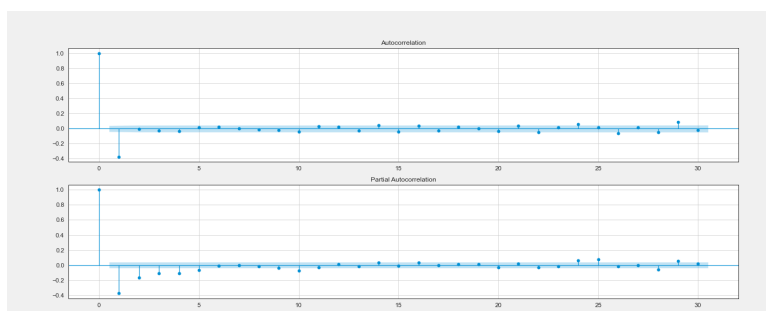
Figure 5.23: Autocorrelation et Partial-Autocorrelation des 4 puits



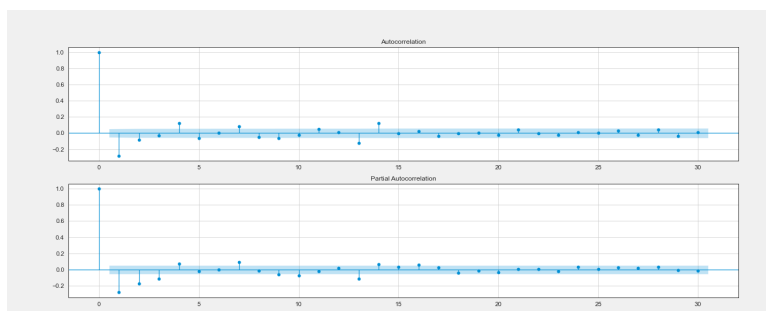
(a) Autocorrelation et Partial-Autocorrelation HS01 Après 1ère Dérivation



(b) Autocorrelation et Partial-Autocorrelation HS02 Après 1ère Dérivation



(c) Autocorrelation et Partial-Autocorrelation HS04 Après 1ère Dérivation



(d) Autocorrelation et Partial-Autocorrelation HS10 Après 1ère Dérivation

Figure 5.24: Autocorrelation et Partial-Autocorrelation des 4 puits Après 1^{ère} Dérivation

5.3.2 Le modèle SARIMA

Afin de l'ajuster, nous utilisons un algorithme de **moindre carrés non-linéaire**.

Nous déterminons un modèle $ARIMA(1, 1, 1)$ qui est notre modèle candidats, Nous remarquons l'impossibilité de l'ajuster correctement, et même la prédiction de valeurs négatives dans certains cas - Ce qui physiquement impossible vu que nous travaillons avec des débit - .

Néanmoins ce problème est résolu après avoir ajouter un terme de saisonnalité, période $m = 2$, et donc $SARIMA(1, 1, 1)(1, 1, 1)_2$ saisonnalité ce qui donnera des résultats précis et qui se cale correctement au donner de test.

Ce qui nous donne deux constats:

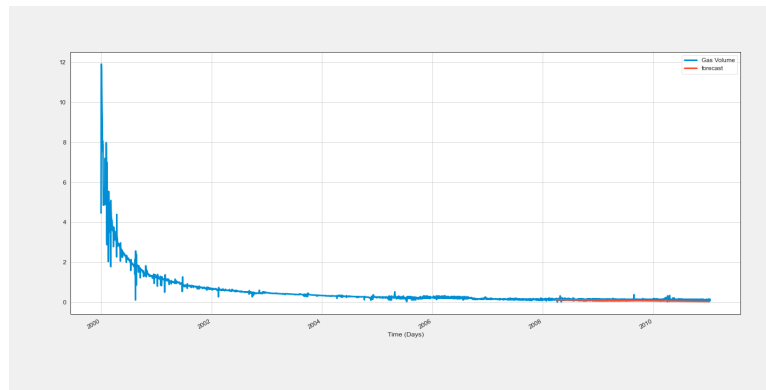
- Les données présentent mathématiquement une saisonnalité.
- L'étude faite au préalables doit être compléter par un tâtonnage qui est laisser au grès de l'expérimentateur et/ou de l'ingénieur, et cela est du à la nature stochastiques de ces modèles.

5.3.3 Vérifications

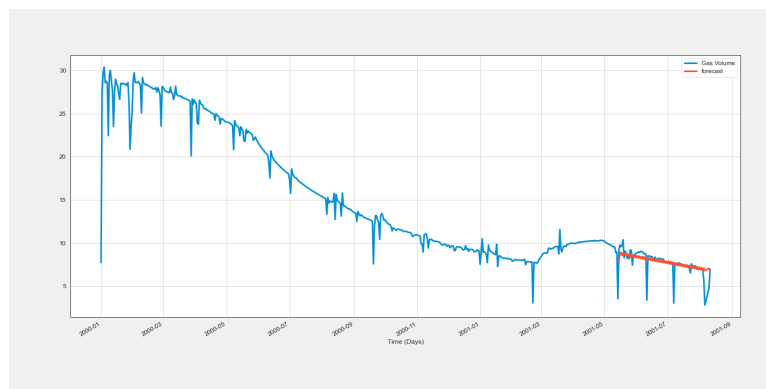
Nous avons préserver une partie des donner afin de pratiquer le test dessus, ce qui donnera les résultats suivants:

<i>Modèles</i>	<i>Puits</i>			
	HS01	HS02	H04	HS10
SARIMA(1,1,1)(1,1,1) ₂	0.0635	0.8458	0.4530	0.3963

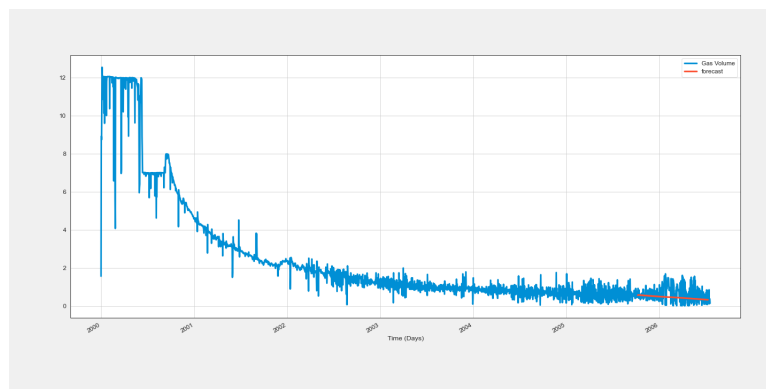
Table 5.13: Les valeurs de la Racine de l'écart quadratique moyen pour le modèle SARIMA



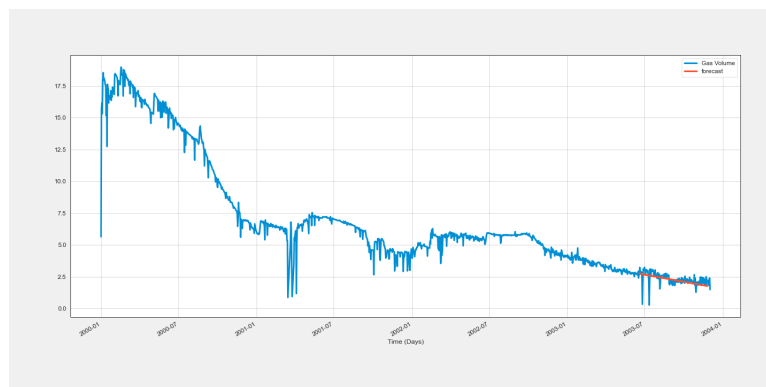
(a) Actuel vs. Prédiction puits HS01



(b) Actuel vs. Prédiction puits HS02



(c) Actuel vs. Prédiction puits HS04



(d) Actuel vs. Prédiction puits HS10

Figure 5.25: Actuel vs. Prédiction SARIMA

5.3.4 Conclusion Partielle

- L'analyse des séries temporelles peut être utilisée pour modéliser les performances de la production pétrolière et permettre des prédictions. Les prédictions peuvent servir de données d'entrée pour le processus de gestion des réservoirs.
- la stimulation des puits ou l'utilisation de données de production précoce peuvent entraîner des coefficients hors limites et des prédictions peu fiables, car ces méthodes sont très sensibles aux lags les plus récents.
- Les modèles doivent être adaptés à l'historique dès que les données de production sont disponibles, ce qui augmente la précision du modèle ajusté.

5.4 Long Short Term Memory (La Variante Bi-LSTM)

Les réseaux de neurones entièrement connectés constituent un outil puissant pour la modélisation d'un large éventail de problèmes, mais l'utilisation de cette architecture pour la prédiction de séries chronologiques, peut donner un résultat très bruyant car le réseau est *acyclique* et chaque instant dans le temps est évalué séparément et ne tient pas compte des prédictions précédentes. Long Short-Term Memory est un outil puissant pour modéliser de tels problèmes et pour faire des prédictions plus fiables et moins bruyantes. Dans cette section, Nous userons de l'architecture décrite dans la sous-section suivantes.

En utilisant les LSTM bidirectionnels, nous alimentons l'algorithme d'apprentissage avec les données originales une fois du début à la fin et une fois de la fin au début. Il y a des débats ici, mais cette approche apprend généralement plus rapidement que l'approche unidirectionnelle, bien que cela dépende de la tâche.

5.4.1 Architecture du Modèle

Couche (type)	Forme de la sortie	Param
LSTM1 (Bidirectional LSTM)	(None, 1, 200)	160800
Dropout (Dropout)	(None, 1, 200)	0
LSTM2 (Bidirectional LSTM)	(None, 1, 200)	240800
Dropout (Dropout)	(None, 1, 200)	0
Dense1 (Dense)	(None, 1)	201
Total params : 401,801		
Trainable params : 401,801		
Non-trainable params : 0		

Table 5.14: Architecture du réseau Bi-LSTM

Hyper-paramètres

Les hyper-paramètres comme définies dans le chapitre "*Outils mathématiques*" sont des paramètres définis par l'expérimentateur qui permettent de maîtriser l'entraînement du modèle. Les plus prominant dans le cas des réseaux de neurones récurrents:

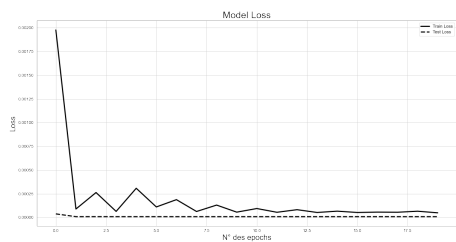
- **Look-Back** = 100: Une "période de recul" définit le nombre de pas de temps précédents utilisés pour prédire le pas de temps suivant.
- **Nombre de couches - et Nombre d'unités par couches** respectivement 2 *couches* et 100 *unités* par couches.
- **Dropout** = 0.2: Le dropout est une technique de régularisation qui consiste à ignorer des neurones sélectionnés au hasard pendant la formation. Cela signifie que leur contribution à l'activation des neurones en aval est temporairement supprimée lors du passage en avant et que les mises à jour de poids ne sont pas appliquées au neurone lors du passage en arrière.
- **Batch Size** = 128: un terme utilisé dans l'apprentissage automatique et fait référence au nombre d'exemples d'apprentissage utilisés dans une itération.

Entraînement et Train-Test Loss

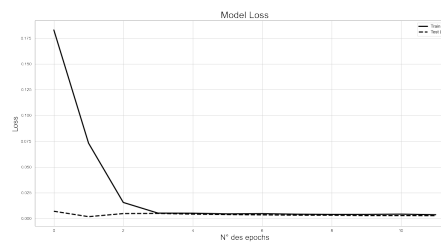
En apprentissage automatique, l'arrêt précoce (Early stopping) est une forme de régularisation utilisée pour éviter le surajustement (overfitting) lors de la phase d'entraînement d'un modèle avec une méthode itérative, telle que la descente de gradient (et

ses variantes). Ces méthodes mettent à jour le modèle de manière à ce qu'il s'adapte mieux aux données d'apprentissage à chaque itération. Jusqu'à un certain point, cela améliore les performances de l'apprenant sur les données en dehors de l'ensemble d'apprentissage. Passé ce point, cependant, l'amélioration de l'adaptation du modèle aux données d'apprentissage se fait au prix d'une augmentation de l'erreur de généralisation. Les règles d'arrêt précoce fournissent des indications sur le nombre d'itérations à effectuer avant que l'apprenant ne commence à trop s'adapter. Les règles d'arrêt précoce ont été utilisées dans de nombreuses méthodes d'apprentissage automatique différentes, avec des fondements théoriques variables.

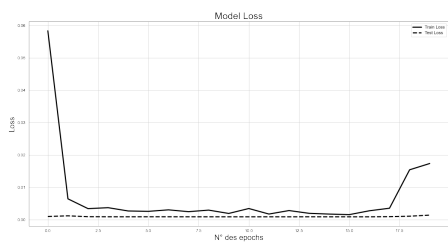
Dans notre cas nous avons conditionné qu'il fallait: **Arrêtez l'entraînement lorsque notre mesure surveillée -Validation Loss- a cessé de s'améliorer après 10 epochs.**



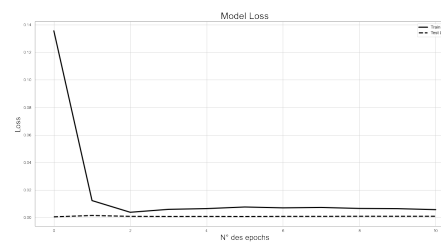
(a) Processus d'entraînement des données du puits HS01



(b) Processus d'entraînement des données du puits HS02



(c) Processus d'entraînement des données du puits HS04



(d) Processus d'entraînement des données du puits HS10

Figure 5.26: Processus d'entraînement

<i>Jeu de données</i>	<i>Puits</i>			
	HS01	HS02	H04	HS10
jeu d'entraînement	0.0703	1.2830	1.1756	0.8890
jeu de test	0.0353	1.3951	0.4635	0.5486

Table 5.15: Les valeurs de la Racine de l'écart quadratique moyen pour le modèle Bi-LSTM

5.4.2 Prédictions

Nous traçons la partie test, Actuelle vs. Prédictions pour nos quatre puits.

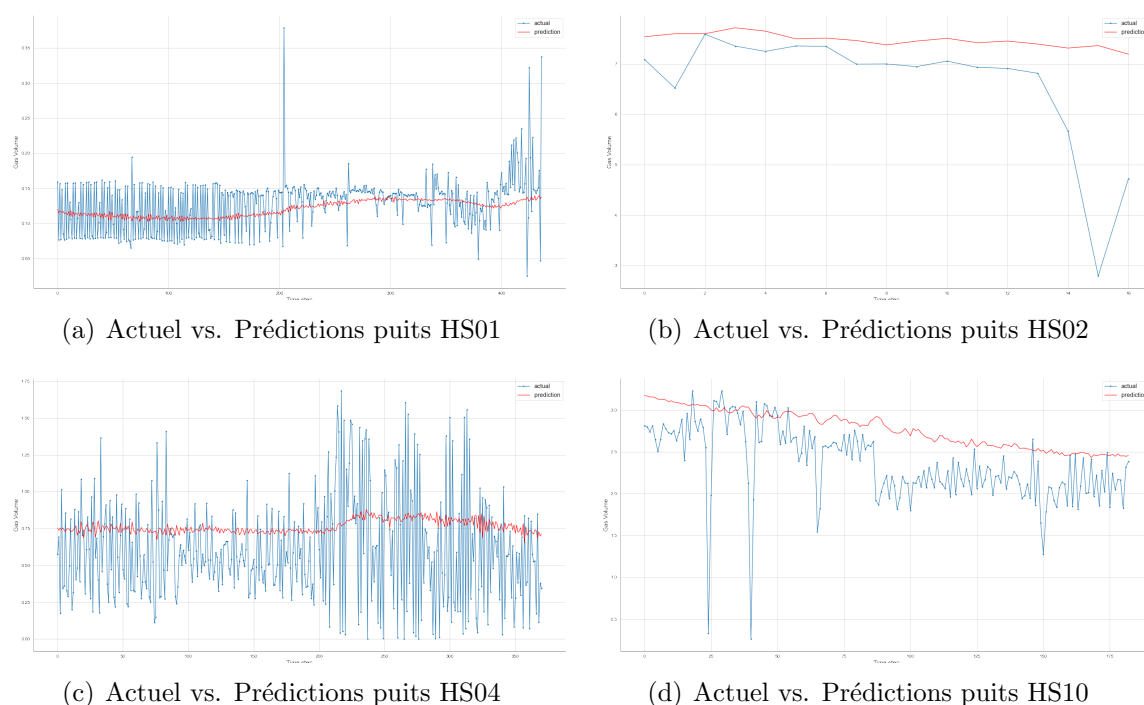


Figure 5.27: Actuel vs. Prédictions Bi-LSTM

5.4.3 Conclusion Partielle

nous avons mis en œuvre une approche Data-Driven pour évaluer les débits de production des puits pour des modèles de séries temporelles de puits produisant à partir de formation de shale, basés sur les algorithmes Bi-LSTM. Cette approche est validée et comparée à l'historique des données réelles et aux méthodes DCA traditionnelles, et montre une bien meilleure prévision de la tendance et moins d'erreurs calculées.

Cette approche ne dépend pas fortement des calculs de principes d'ingénierie et donne des

solutions plus rapidement. Bien que cette approche basée sur les données soit dépourvue de physique, elle constitue une bonne alternative pour fournir des résultats rapides et robustes en plus des simulations empiriques et numériques.

Par contre cette approche dépend fortement de la taille de jeu de données (Plus les données sont disponibles, mieux est la précision des prédictions).

5.5 Discussion

- Les modèles empiriques de Arps présentent des déficiences dans les prédictions en présentant des valeurs de b trop positives dans quand nous produisant à partir des formations à faibles perméabilité qui induit le modèles en erreurs dans l'estimation du EUR, en donnant un résultat qui en s'approche jamais de 0
- Le modèle basé sur la croissance logistique, résout le problème des formules de Arps en introduisant le terme de capacité de charge qui limite la capacité de production maximale et ainsi donner de meilleurs résultats de prédictions.
- L'algorithme **LOESS** s'avère d'une utilité surprenante quand à la détection des valeurs aberrantes et la réduction de leurs impacts sur les résultats finaux.
- Les approches se basant sur la méthode Box-Jenkins, ARIMA (SARIMA), pour les prévisions à court terme avec des données à haute fréquence, les résultats peuvent être difficiles à battre. Ils ont également l'avantage d'être moins sensibles aux hypothèses sous-jacentes de la nature des fluctuations des données que de nombreux autres systèmes.
- Contrairement aux modèles naïfs simples ou aux modèles de lissage, il n'y a pas de fonction de mise à jour automatique lorsque de nouvelles données sont disponibles ; toute la procédure de modélisation doit être répétée. cela s'explique par le fait que ces modèles là sont très sensibles aux observations -lags- plus récentes -Ce qui dans certains cas peut représenter un avantage qui est la détection des petites variations -.
- Les estimations par le biais des modèles ARIMA et leurs variantes est faites pour des prévisions à court termes seulement, à long termes les déviations de la réalité sont de plus en plus accentuées.

- La variantes Bi-LSTM, quoique gourmandes en ressources de calculs, elle s'adapte au mieux, même en présence d'irrégularités, et à la conditions que le jeu de données soit assez grand, les résultats finaux présentés par les Bi-LSTM est le plus précis⁸.

⁸La variante Bi-directionnelle est plus robustes car elle alimentent l'architecture des réseaux de neurones dans les deux sens -*Passé au futur et Futur au Passé*- ce qui lui permet de comprendre le *contexte*

6 Conclusion Générale et Recommandation

L'étude de déclin est un processus qui a suscité et qui suscite jusqu'à aujourd'hui, l'intérêt chez les ingénieurs réservoirs et pétroliers. en raison de la portée économique à laquelle il touche.

L'analyse du déclin de la production (PDA), est devenue un point chaud du génie pétrolier ces dernières années. Avec Arps, qui en 1945 a mis en place les bases de cette branches avec ses trois formules de déclin empiriques: *Exponentielle*, *Harmonique* et *Hyperbolique* -qui ont fait un excellent parcours et qui ont représenté jusqu'à très récemment le standard dans l'industrie. Néanmoins, comme chaque formules empiriques elle va par la suite présenter des déficiences quand confronté: *soit à*, des valeurs aberrantes qui peuvent induire ces modèles en erreurs, *ou bien à* des bases physiques sur lesquelles elles n'étaient pas fondées telles que les réservoirs à très faibles perméabilité i.e. schistes.

Pour le premier cas de figures, nous avons usé d'une variante du local polynomial Regression, **LOcally Estimated Scatterplot Smoothing**, communément appelée **LOESS- Filtre de Savitzky-Golay-**, qui reprends les bases de l'algorithme KNN dans la détection des valeurs aberrantes et elle adapte des modèles simples à des sous-ensembles localisés de données afin de construire une fonction qui décrit la partie déterministe de la variation des données, point par point, et ainsi réduire l'impacte des outliers.

Pour le second problème, nous avons opté pour trois types de modèles:

- Un modèle empirique mis en place sur les bases du *modèle de Verhulst sur la croissance logistique* qui s'avère plus adéquat que les modèles de Arps quand confronté aux formations à faible perméabilité.
- Un second modèle basé sur les travaux de George Box and Gwilym Jenkins, d'où *La méthode de Box-Jenkins* sur les modèles stochastiques, La variante **SARIMA** donnent des résultats plus que satisfaisants et adaptatifs souvent plus précis que tous les modèles implémentés en haut, néanmoins présentant une certaines sensibilités aux observations -lags - plus récentes.
- L'outil Bi-LSTM a prouvé son efficacité pour la modélisation des séries temporelles et produire des prédictions plus proches de la réalité avec une robustesse contre les

bruits, même si la taille de jeu de données restent un frein majeur à cette méthode.

Nous recommandons d'approfondir ce travail par ce qui va suivre:

- L'utilisation de l'apprentissage ensembliste qui use de plusieurs algorithmes d'apprentissage pour obtenir de meilleures prédictions entre les modèles statistique et/ou d'apprentissage.
- Implémenter et tester d'autre modèles empiriques tel que le modèle de Duong, Extended Exponential Decline Curve Analysis...etc.
- Mettre en place une méthodologie qui permettra de tirer avantage des méthodes: empiriques, probabiliste et d'apprentissage automatique.

References

- [1] Arps, J. J. Analysis of decline curves. *Transactions of the AIME*, 160(01):228–247, 1945.
- [2] Blumberg, A. Logistic growth rate functions. *Journal of Theoretical Biology*, 21(1):42–44, 1968.
- [3] Cleveland, W. S. Robust locally weighted regression and smoothing scatterplots. *Journal of the American statistical association*, 74(368):829–836, 1979.
- [4] Fetkovich, M. J. Decline curve analysis using type curves. *Journal of Petroleum Technology*, 32(06):1065–1077, 1980.
- [5] Carter, R. D. Type curves for finite radial and linear gas-flow systems: constant-terminal-pressure case. *Society of Petroleum Engineers Journal*, 25(05):719–728, 1985.
- [6] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.
- [7] Fraim, M. Gas reservoir decline-curve analysis using type curves with real gas pseudopressure and normalized time. *SPE Formation Evaluation*, 2(04):671–682, 1987.
- [8] Doublet, L., Pande, P., McCollum, T., and Blasingame, T. Decline curve analysis using type curves—analysis of oil well production data using material balance time: Application to field cases. In *International petroleum conference and exhibition of Mexico*. OnePetro, 1994.
- [9] Hipel, K. W. and McLeod, A. I. *Time series modelling of water resources and environmental systems*. Elsevier, 1994.
- [10] Agarwal, R. G., Gardner, D. C., Kleinsteiber, S. W., and Fussell, D. D. Analyzing well production data using combined-type-curve and decline-curve analysis concepts. *SPE Reservoir Evaluation & Engineering*, 2(05):478–486, 1999.
- [11] Ilk, D., Rushing, J. A., Perego, A. D., and Blasingame, T. A. Exponential vs. hyperbolic decline in tight gas sands: understanding the origin and implications for reserve estimates using arps’ decline curves. In *SPE annual technical conference and exhibition*. OnePetro, 2008.
- [12] Clark, A. J., Lake, L. W., and Patzek, T. W. Production forecasting with logistic growth models. In *SPE annual technical conference and exhibition*. OnePetro, 2011.
- [13] Duong, A. N. Rate-decline analysis for fracture-dominated shale reservoirs. *SPE Reservoir Evaluation & Engineering*, 14(03):377–387, 2011.
- [14] Olominu, O. and Sulaimon, A. A. Application of time series analysis to predict reservoir production performance. In *SPE Nigeria Annual International Conference and Exhibition*. OnePetro, 2014.
- [15] Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

-
- [16] Sun, H. *Advanced production decline analysis and application*. Gulf professional publishing, 2015.
- [17] Chaudhary, N. L. and Lee, W. J. Detecting and removing outliers in production data to enhance production forecasting. In *SPE/IAEE Hydrocarbon Economics and Evaluation Symposium*. OnePetro, 2016.
- [18] Zhang, H., Nelson, E., Olds, D., Rietz, D., and Lee, W. J. Effective applications of extended exponential decline curve analysis to both conventional and unconventional reservoirs. In *SPE Annual Technical Conference and Exhibition*. OnePetro, 2016.
- [19] Sun, J., Ma, X., and Kazi, M. Comparison of decline curve analysis dca with recursive neural networks rnn for production forecast of multiple wells. In *SPE Western Regional Meeting*. OnePetro, 2018.
- [20] Latrach, A. *Machine Learning for Predictive Maintenance*. Master's thesis, Faculty of Hydrocarbons and Chemistry (EX:INH), Algeria, 2020.
- [21] Xiong, H., Kim, C., and Fu, J. A data-driven approach to forecasting production with applications to multiple shale plays. In *SPE Improved Oil Recovery Conference*. OnePetro, 2020.

Annexe