

Modèle de prédiction de niveau piézométrique basé sur Transformers

Mohamed Louay Rabah*, Nedra Mellouli**, Imed Riadh Farah*

*Laboratoire RIADI, Université Manouba, Manouba 2010, Tunisie
louayrabah@gmail.com

imedriadh.farah@isamm.uma.tn

**Laboratoire LIASD, Université Paris 8, 93200 Paris, France
n.mellouli@iut.univ-paris8.fr

Résumé. Au cours des dernières années, les réseaux de neurones ont été utilisés comme solution alternative aux approches classiques afin de faciliter des tâches telles que la classification ou la prédiction. Le but de notre étude est de prédire les niveaux piézométriques de 18 stations dispersées en France et d'explorer l'applicabilité de différentes approches de prédiction. Une corrélation entre les données piézométriques, les différences de valeurs, les précipitations, l'évaporation et le débit des cours d'eau a été fournie aux modèles comparés en tant qu'entrée pendant la phase d'apprentissage. Les modèles ont montré des résultats encourageants mentionnant l'application d'ARIMA, LSTM, GRU et les Transformers qui sont nouvellement utilisés pour la prévision des séries temporelles (TS). Les résultats de cette étude peuvent être utilisés pour orienter le futur traitement des données sur les ES. De plus, ce travail peut être une référence utile pour d'autres problèmes de TS dans le domaine géotechnique.

1 Introduction

L'eau phréatique est le type d'eau trouvée sous la surface de la terre qui est présent dans les espaces poreux du sol ainsi que dans les roches. Lorsque les cavités et les vides des roches sont totalement saturés d'eau, ils forment ce qu'on appelle une nappe phréatique. Les eaux souterraines (ES) ont toujours été l'une des ressources en eau les plus utiles connues de l'humanité en raison de sa disponibilité et de sa non-vulnérabilité à la pollution, contrairement aux eaux de surface, de ce fait, elles sont fréquemment utilisées. Afin d'être bien exploitées, les ES ont toujours fait l'objet de différentes études afin de connaître les techniques et approches les plus adaptées à leur gestion. Dans cet article, nous aborderons la prévision des niveaux d'ES pour 18 stations hydrauliques en France, ce qui est considéré comme une tâche importante pour les autorités afin qu'elles puissent élaborer une planification plus précoce pour l'utilisation et la distribution des ES. Plusieurs méthodes de solutions numériques et d'études analytiques ont été réalisées pour analyser le comportement des ES dans différentes conditions dans le but de créer des modèles de prédiction basés sur des lois physiques. Malgré cela, les résultats des études numériques n'ont pas pu correspondre complètement aux observations en raison de la non-linéarité des données et des conditions (Zhan et Ng (2004)). Diverses

techniques d'apprentissage automatique (ML) ont été utilisées afin de résoudre des problèmes géotechniques tels que la prédiction de niveau piézométrique introduites dans plusieurs articles de recherche.

(Rankovic et al. (2013)) a proposé un modèle de prédiction utilisant des réseaux de neurones FNN et uniquement des mesures piézométriques du barrage Iron Gate 2 situé entre la Serbie et la Roumanie. (Derbela et Nouir (2020)) ont prédit le niveau d'eau dans l'aquifère de Nebhana en Tunisie en proposant un modèle ANN utilisant les précipitations mensuelles, l'évaporation et le niveau piézométrique du mois précédent. Dans (Wei et al. (2020)), les auteurs ont comparé 3 architectures de réseaux neuronaux récurrents, le RNN standard, la mémoire à long court terme LSTM et l'unité récurrente à portes GRU, en utilisant des observations de précipitations, évaporation et de pression d'eau à partir d'une pente de terrain naturel dans la région de Hong Kong. . (Solgi et al. (2021)) ont prédit le niveau des ES dans l'aquifère d'Edwards au Texas en utilisant un réseau de neurones LSTM en l'appliquant sur des mesures piézométriques depuis 1932 en insistant sur le fait que les observations piezométriques sont suffisantes. Les auteurs de (Pandey et al. (2020)) ont introduit un modèle hybride ANN optimisé avec un algorithme génétique pour prédire la profondeur où l'aquifère avait différents types de roches et de sols en prenant les données pluviométriques quotidiennes de 21 ans et la profondeur de 38 capteurs. (Kombo et al. (2020)) ont utilisé un modèle hybride K-Nearest Neighbor-Random Forest à l'aide de données ES et de données climatiques collectées dans l'est du Rwanda. Les auteurs de (Hussein et al. (2020)) ont comparé différents outils ML en utilisant des images satellites mensuelles des ES collectées entre 2002 et 2019. Les modèles comparés sont MLR, MLP, XGB, RF et SVM. (Gutierrez Pacheco et al. (2021)) ont utilisé des mesures bimensuelles dans un site de tourbière situé à Saint-Modeste, dans l'Est du Canada, en les appliquant sur les architectures de GLM, KNN, SVM, arbre de décision, RF et ADABOOST. L'objectif de cet article est d'appliquer l'architecture Transformers à la prédiction des ES et de la comparer à d'autres modèles ML. À la connaissance des auteurs, il s'agit de la première étude sur l'application de Transformers pour la prédiction d'ES.

2 Méthodologies

2.1 ARIMA et VAR

En statistique, et en particulier en analyse de TS, un modèle de moyenne mobile intégré autorégressif (ARIMA) est une généralisation d'un modèle de moyenne mobile autorégressive (ARMA). Ces deux modèles sont ajustés aux données de TS soit pour mieux comprendre les données, soit pour prédire les points futurs de la série. Un modèle (VAR) est une généralisation du modèle autorégressif univarié pour la prévision d'une série multivariée. Dans ce cadre, toutes les variables sont traitées de manière symétrique. Ils sont tous modélisés comme s'ils s'influencent tous de manière égale. Dans une terminologie plus formelle, toutes les variables sont désormais traitées comme « endogènes ».

2.2 Long Short Term Memory et Gated Recurrent Unit

Les réseaux de neurones récurrents (RNN) sont des réseaux de neurones dynamiques adaptés au traitement des problèmes de TS. Différent de ANN, les neurones cachés sont connectés

dans l'architecture RNN. Les données d'entrée sont transmises séquentiellement à travers des unités cachées via des liens. RNN est à la fois simple et efficace, cependant, en réalité, entraîner un modèle qui a des problèmes avec les longues séquences est un peu difficile. Les entrées initiales décroissent avec le temps pendant la forward propagation et le réseau finit par oublier l'entrée initiale et le problème du vanishing gradient se produit lors de la back propagation en essayant d'optimiser la fonction de perte qui entraîne progressivement la disparition du gradient. Ainsi, le RNN standard ne peut apprendre qu'à court terme.

La mémoire à long court terme (LSTM) est une variante du RNN standard. Le concept d'unités fermées est introduit dans un premier temps par Hochreiter et Schmidhuber (1997) pour surmonter le problème du vanishing gradient dans RNN. L'unité récurrente à portes GRU est la nouvelle génération de variations RNN qui est très similaire à LSTM. Ayant trois portes, la conception LSTM est complexe dans une certaine mesure, ce qui peut affecter son efficacité. GRU n'a que deux portes, une porte de réinitialisation (Reset gate RG) et une porte de mise à jour (Update gate UG) et a utilisé l'état caché (hidden state) pour transférer les informations. La RG contrôle l'association des données d'entrée avec les informations historiques tandis que l'UG spécifie la quantité qui sera conservée à partir des données historiques.

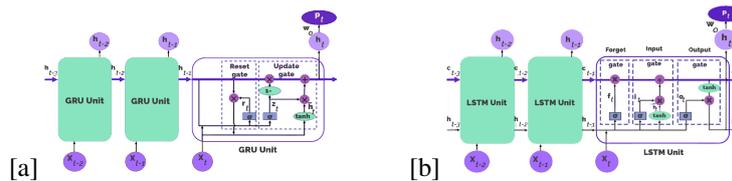


FIG. 1 – Structure de [a] GRU et [b] LSTM.

2.3 Transformers

Les auteurs de (Vaswani et al. (2017)) ont présenté les Transformers et ce qu'on appelle une architecture séquence à séquence. Séquence à séquence est un réseau de neurones qui transforme un groupe d'éléments, comme une séquence de mots en une nouvelle séquence. Le Transformer est un modèle qui utilise l'attention pour augmenter la vitesse à laquelle les modèles peuvent être entraînés et sa conception est considérée comme l'évolution de l'architecture encodeur-décodeur. L'attention est un mécanisme qui a amélioré les performances des réseaux de neurones nouvelle application. Selon leur papier, non seulement les transformeurs se sont entraînés plus rapidement, mais ils ont également obtenu de meilleurs résultats, comme indiqué par les auteurs. Dans la plupart des cas, les modèles récurrents organisent le calcul selon les séquences d'entrée et de sortie, et en plaçant les positions à des pas de temps dans le calcul, ils produisent une séquence d'états cachés, ayant pour paramètre l'état caché précédent et l'entrée pour la position t . Et à cause des restrictions de mémoire que cela limite dans le lot d'échantillons, et à cause de son comportement séquentiel, la parallélisation est naturellement interdite surtout pour les séquences plus longues. Alors que le modèle Transformers a un encodeur qui prend une séquence en entrée, la transforme et le décodeur génère la sortie.

Modèle de prédiction de niveau piézométrique basé sur Transformers

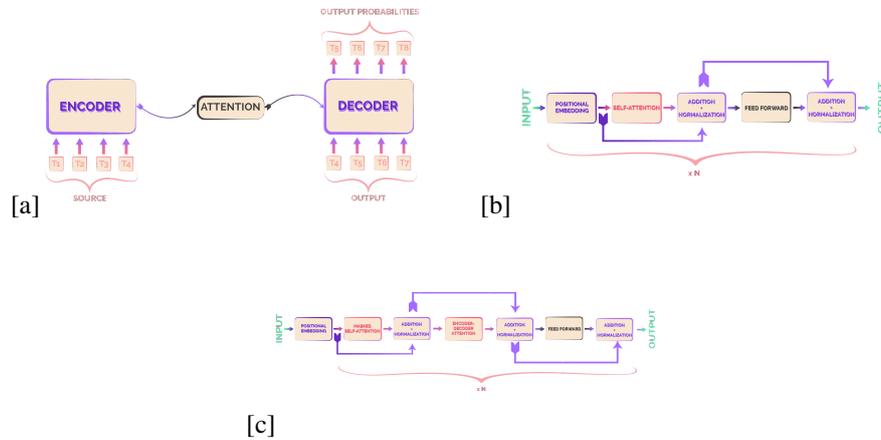


FIG. 2 – [a] Structure du modèle Transformers adaptée aux TS. [b] Structure de l'encodeur. [c] Structure du décodeur.

3 Experimentation

3.1 Description et analyse des données

La première étape consistait à effectuer des tests statistiques classiques sur chaque piézomètre pour extraire le maximum, le minimum, la moyenne et l'écart-type, puis en parcourant les mesures, nous avons observé non seulement l'instabilité du pattern mais aussi le manque de données qui nous incitent à faire plus de tests et à passer à la deuxième phase de data engineering et features extraction. Pour le rééchantillonnage et l'interpolation des données, nous avons essayé d'expérimenter avec différents pas de temps (figure 3). Les TS peuvent être non

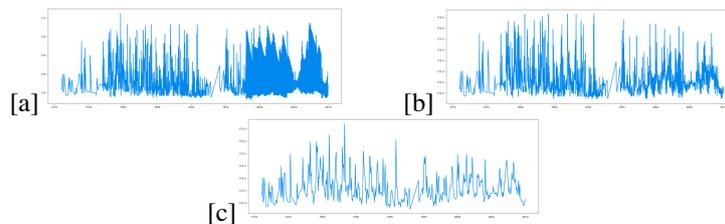


FIG. 3 – La figure montre les données rééchantillonnées et interpolées pour le pas de temps [a] quotidiens [b] hebdomadaires et [c] mensuels pour le piézomètre 02648X0020/S1

stationnaires et nous pouvons les identifier visuellement, mais il existe des méthodes plus formelles pour effectuer ce travail. Le test de Dicky-Fuller augmenté (ADF) est la méthode la plus souvent utilisée. Comme pour notre cas, nous avons effectué ADF résultant un tuple contenant *Test statistic* (plus il pourrait être négatif, plus il est probable que les données soient stationnaires), *p-value* (affecté par la "Test statistic" et doit être inférieur à 0,05 afin que nous rejetions l'hypothèse nulle et supposons que notre série soit stationnaire) qui sont les 2 variables qui

nous aident à décider si la série est stationnaire. Après l'exécution du test sur les 18 stations on observe la p-value est inférieure à 0,05 dans la majorité dont on suppose qu'ils sont stationnaires. En ce qui concerne les TS non stationnaires, nous pouvons effectuer les transformations de différenciation ou logarithmiques pour les rendre stationnaires. L'étude de la saisonnalité

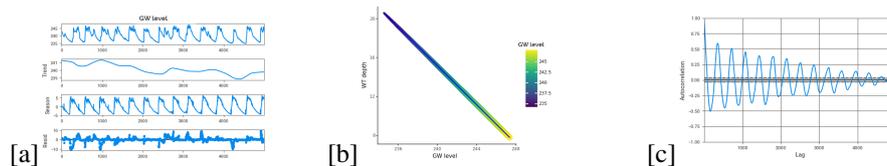


FIG. 4 – La figure [a] montre décomposition saisonnière [b] montre la corrélation entre le niveau des ES et la profondeur de la nappe phréatique et [c] montre l'autocorrélation du niveau d'eau pour le piézomètre 03124X0088/F avec une version décalée de l'observation. La figure montre que la plupart des pics sortent des limites de confiance de 95% et 99%, ce qui prouve qu'il existe des autocorrélations significatives.

est également une tâche importante lors de la gestion des TS, pour ce faire, nous devons appliquer la décomposition qui est une technique qui divise la série en plusieurs composantes et chaque composante représente une catégorie, par exemple la tendance, définit si la TS est en déclin, stable ou croissante avec le temps, la saisonnalité, se réfère au signal périodique dans le TS, et résiduel, ou bruit qui définit ce qui reste après la séparation des données temporelles. Nous avons utilisé la méthode de Loess qui permet de contrôler à la fois le changement de composante saisonnière au fil du temps et la régularité de la tendance-cycle en modifiant la périodicité, comme le montre la figure 4. Nous avons mené plusieurs autres opérations pour extraire des caractéristiques comme l'entropie, l'asymétrie, le kurtosis, l'autocorrélation et la corrélation entre le niveau des ES et la profondeur de la nappe phréatique.

3.2 Expérimentations

Pour cet article, nous avons eu comme données les TS collectées à partir de 18 capteurs piézométriques que nous avons obtenus d'une API provenant du portail ADES (portail national d'accès aux données sur les ES pour la France). Dans cette étude, l'apprentissage s'est fait principalement en utilisant les données des piézomètres mais nous avons comme features les données climatiques et hydrogéologiques ainsi extraites de différents portails français et européens. Nous avons entraîné les modèles sur Google Colaboratory à l'aide d'un GPU NVIDIA Tesla K80 avec 14 Go de RAM. Chaque époque a pris environ 2 secondes et nous avons entraîné les modèles pour un total de 2000 époques par modèle qui ont duré environ 2 heures. Dans la phase du training, nous avons utilisé Adam (Kingma et Ba (2015)) en tant qu'optimiseur avec un taux d'apprentissage de $1e-5$ en utilisant le framework d'IA de Facebook pytorch.

3.3 Comparaison

La fluctuation des ES s'est avérée être la conséquence d'une combinaison d'un certain nombre de facteurs. En mettant tous les problèmes de côté, nous avons commencé le pipeline de prétraitement dont nous avons discuté précédemment, puis nous sommes passés à la phase

Modèle de prédiction de niveau piézométrique basé sur Transformers

training et test que nous avons commencée en évaluant les modèles à différentes époques et optimiseurs jusqu'à ce que nous avons décidé d'utiliser Adam comme optimiseur, le batch size 32 et le dropout rate 0,5 et 2000 époques. Contrairement aux autres méthodes utilisées, le modèle VAR avait une approche différente pour le training, tandis qu'en apprentissage profond nous essayons différents hyper paramètres afin d'ajuster un modèle, une approche statistique comme le VAR est principalement basée sur des équations et chaque variable est modélisé en fonction des valeurs passées et comme l'ensemble de données est multivarié, il est modélisé comme un système d'équations avec une équation par variable. En plus de tester et de transformer les données, pour un modèle d'autorégression, nous devons trouver l'ordre optimal et nous avons donc testé le fitting avec plusieurs lags. Figure 5 montre la comparaison entre les observations et la prédiction avec VAR. Une étude comparative a été réalisée sur le modèle GRU entre différentes fonctions d'activation car nous avons testé les fonctions tanh et ReLu avec 2 optimiseurs différents afin d'étudier les effets en termes de fonctions d'activation comme le montre la figure 6. Le tableau montre les valeurs comparées pour l'erreur absolue moyenne (MAE) en utilisant à la fois les optimiseurs Adam et RMSprop pour la formation. En plus du GRU, nous exécutons également nos ensembles de données sur le modèle LSTM afin de le comparer aux autres modèles utilisés dans l'étude en comparant les résultats d'un modèle LSTM classique des deux optimiseurs. GRU et LSTM étaient supérieurs dans cette étude par rapport au modèle VAR et Transformers qui nous a certainement donné à la fois de bonnes et de mauvaises prédictions, mais il a montré des résultats respectueux et pourrait être amélioré en l'adaptant pleinement à la prédiction des TS.

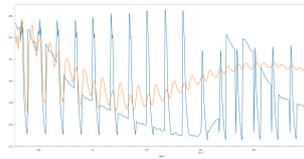


FIG. 5 – Prédiction des TS à l'aide du modèle VAR pour le piézomètre 03124X0088/F.

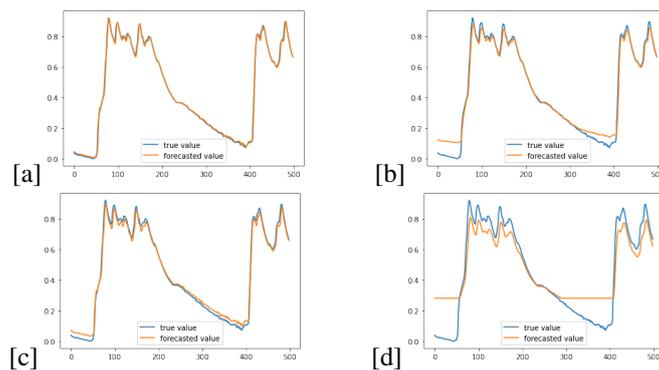


FIG. 6 – Comparaison des modèles [a] GRU-tanh-Adam et [b] GRU-ReLu-Adam [c] GRU-tanh-RMSprop et [d] GRU-ReLu-RMSprop pour le piézomètre 03124X0088/F.

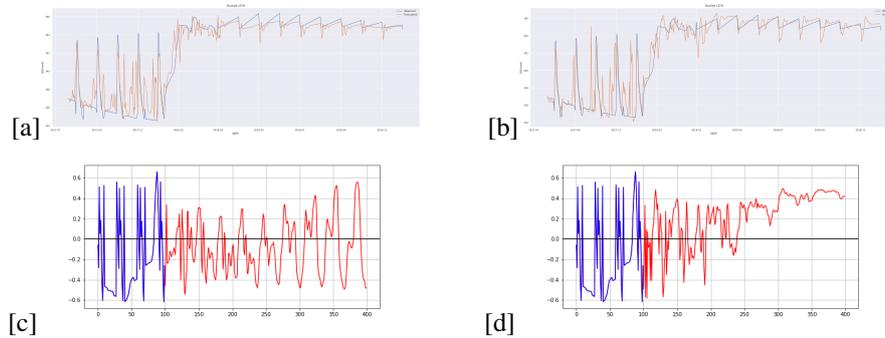


FIG. 7 – Comparaison des modèles [a] LSTM-Adam et [b] LSTM-RMSprop et la prédiction des séries à l'aide du modèle Transformers avec [a] le résultat obtenu avec Adam et [b] avec RMSprop pour le piézomètre 03124X0088/F.

Modèle Deep learning	MAE (Adam)	MAE (RMSprop)
GRU-tanh	0.33116474	0.33815362
GRU-ReLu	0.29645575	0.31061599
LSTM	0.33787793	0.38402377
Transformers	0.31713313	0.48000625
Modèle classique		MAE
VAR	2.71584665	

TAB. 1 – Comparaison des différents modèles pour le piézomètre 03124X0088/F.

3.4 Discussion

Les résultats du modèle Transformers ne sont pas aussi bons que nous l'espérons, surtout compte tenu des meilleures prédictions du GRU, mais ils sont toujours encourageants. Dans une certaine mesure, dans l'ensemble de notre étude, les résultats de l'approche des Transformers n'étaient pas efficaces en surestimant légèrement, mais sur certains des bons exemples avec lesquels nous avons travaillé, les prévisions correspondent aux observations originales et éliminent rigoureusement les fluctuations bruyantes. Réobserver les résultats, changer les hyper-paramètres pourrait peut-être nous aider à obtenir de meilleurs résultats.

4 Conclusion

Les transformers sont maintenant très populaires et extrêmement recommandés par la communauté des passionnés d'apprentissage automatique (Machine Learning ML) puisqu'il s'agit d'un sujet de pointe largement discuté dans la littérature et très populaire pour une multitude d'applications ML, il est donc tout à fait raisonnable que ils seront employés pour les tâches TS. Les transformers pour TS peuvent être un modèle assez gourmand en données mais ils sont indispensables dans votre arsenal d'apprentissage automatique en raison de leur adaptabilité et leur scalabilité et également leur utilisation pour différentes applications telles que la NLP, la vision par ordinateur et maintenant les TS. On ne s'attendait pas à ce que les résultats

soient exempts d'erreurs, mais des ajustements supplémentaires pourraient l'améliorer, mais cela pourrait constituer la base pour une utilisation future.

Références

- Derbela, M. et I. Nour (2020). Intelligent approach to predict future groundwater level based on artificial neural networks (ann). *Springer Nature*.
- Gutierrez Pacheco, S., R. Lagacé, S. Hugron, S. Godbout, et L. Rochefort (2021). Estimation of daily water table level with bimonthly measurements in restored ombrotrophic peatland. *Sustainability*.
- Hussein, E., C. Thron, M. Ghaziasgar, A. Bagula, et M. Vaccari (2020). Groundwater prediction using machine-learning tools. *Algorithms*.
- Kingma, D. et J. Ba (2015). Adam: A method for stochastic optimization.
- Kombo, O., S. Kumaran, Y. Sheikh, A. Bovim, et A. Jayavel (2020). Long-term groundwater level prediction model based on hybrid knn-rf technique. *Hydrology*.
- Pandey, K., S. Kumar, A. Malik, et A. Kuriqi (2020). Artificial neural network optimized with a genetic algorithm for seasonal groundwater table depth prediction in uttar pradesh, india. *Sustainability*.
- Rankovic, V., A. Novakovic, N. Grujovic, D. Divac, et N. Milivojevic (2013). Predicting piezometric water level in dams via artificial neural networks. *Springer Verlag*.
- Solgi, R., H. Loáiciga, et M. Kram (2021). Long short-term memory neural network (lstm-nn) for aquifer level time series forecasting using in-situ piezometric observations. *Journal of Hydrology*.
- Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, L. Kaiser, et I. Polosukhin (2017). Attention is all you need.
- Wei, X., L. Zhang, H. Yang, L. Zhang, et Y. Yao (2020). Machine learning for pore-water pressure time-series prediction: Application of recurrent neural networks. *Geoscience Frontiers*.
- Zhan, T. et C. Ng (2004). Analytical analysis of rainfall infiltration mechanism in unsaturated soils. *International Journal of Geomechanics*.

Summary

In this work we present Transformers as an innovative model for dealing with groundwater time series and compare it to other more common prediction adapted models. Herein lies our goal to explore the applicability of the models on multivariate data and predict farther steps for 18 stations scattered in France. In order to fulfil our purpose we followed a classic pipeline for the preprocessing of the data for the feature engineering also experimenting with different optimizers helped in partially avoiding the vanishing gradient problem and there is plenty of room to experiment with the model's other settings.