

Application IA

Rapport d'analyse

Prévision du niveau des nappes par réseau de neurones



Nom et Prénom: **GENOUIZ Mustapha**
Filière: **IA**

1 Introduction

Ce rapport représente une analyse complète sur les différentes étapes qui nous ont guidé à résoudre un problème de la vie réelle qui est de prédire les niveaux des nappes phréatiques pour les 6 prochains mois en Australie à partir d'une observation de quatre caractéristiques (P : précipitation en cumul, T : température moyenne mensuelle de la surface de sol, ET : évapotranspiration moyenne mensuelle et NDVI : indice de végétation), pendant ce rapport on va expliquer les différentes analyses de données qu'on a faites à la première partie, puis, dans la deuxième partie, on va expliquer les différentes méthodes qu'on a utilisées pour attaquer à ce problème.

2 Partie 1 : Visualisation et étude de corrélation

2.1 A. Affichage des courbes non normalisées :

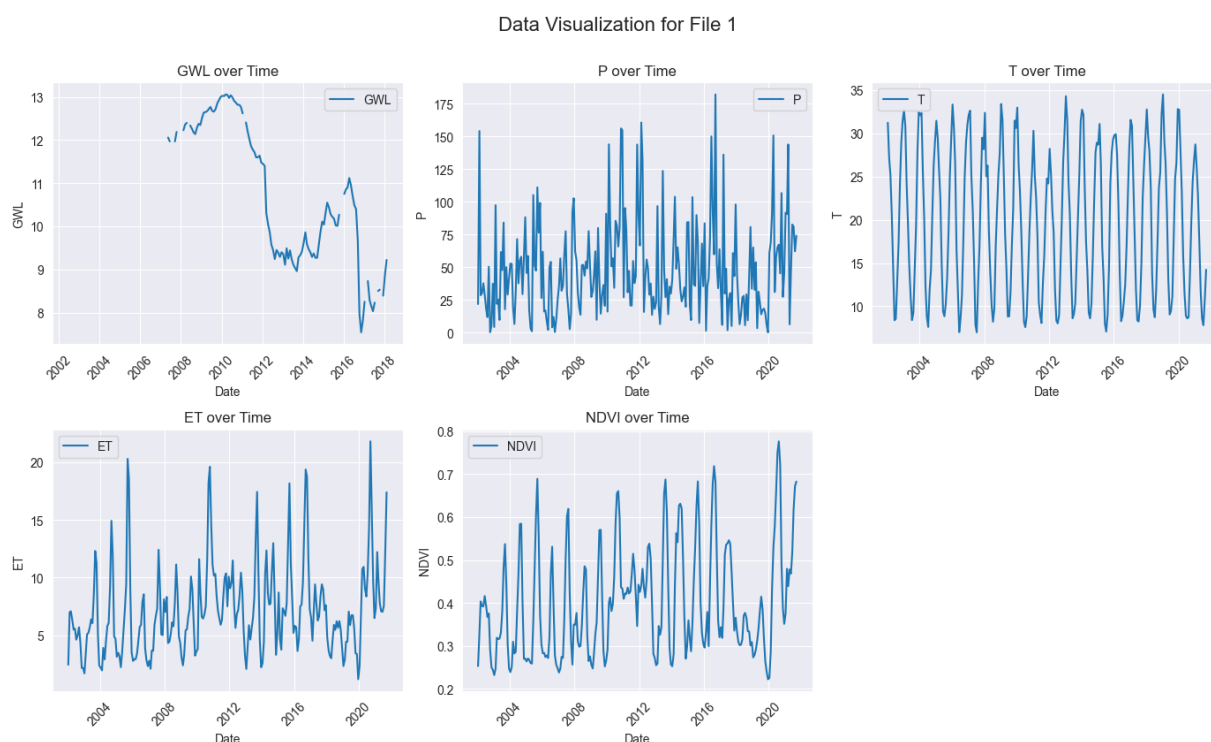


Figure 1: Graphe de courbes non normalisées du fichier 1

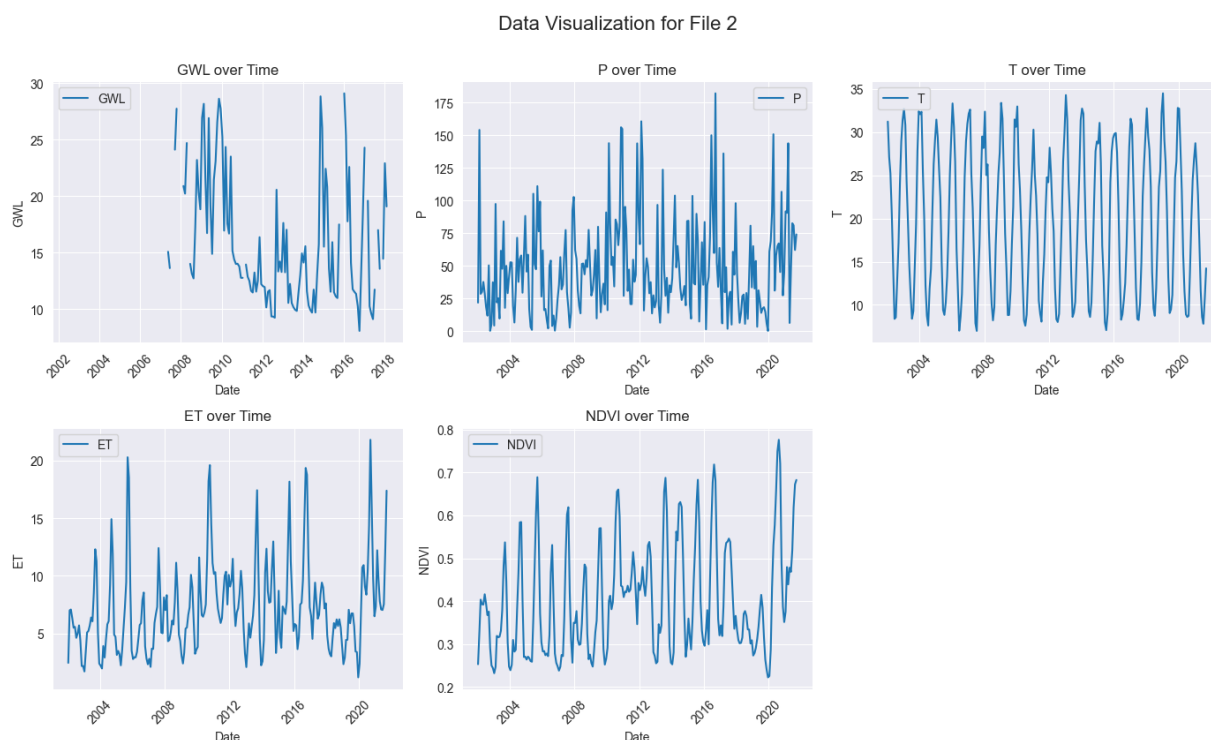


Figure 2: Graphe de courbes non normalisées du fichier 2

On remarque d'après la visualisation de la figures 1, qu'il y'a des similarités entre la courbe d'ET et celle de NDVI, aussi pour ET et P, pour la visualisation du deuxième fichier 2, on remarque des similarités presque identique à celle vu à la visualisation du deuxième fichier, qui nous indique une potentielle relation entre ces deux ensemble : ET et P, ET et NDVI, aussi, on peut remarquer des correspondance au niveau de GWL et T.

2.2 B. Affichage des courbes normalisées :

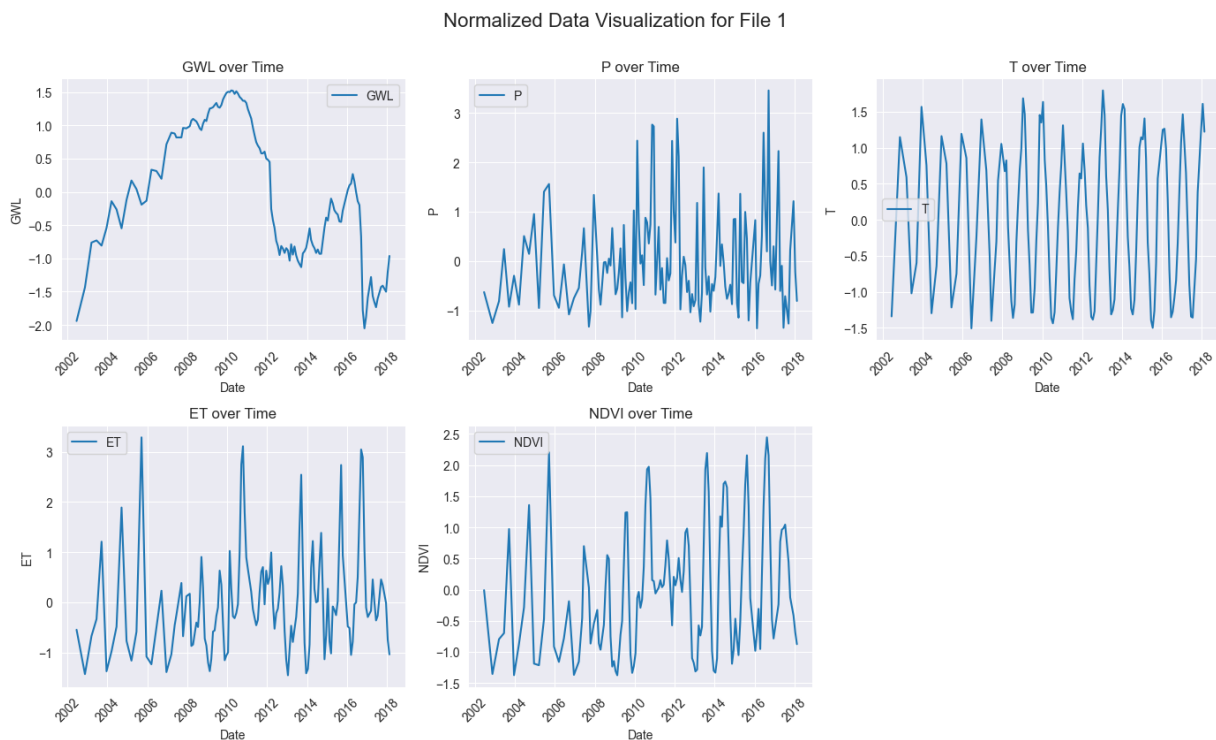


Figure 3: Graphe de courbes normalisées du fichier 1

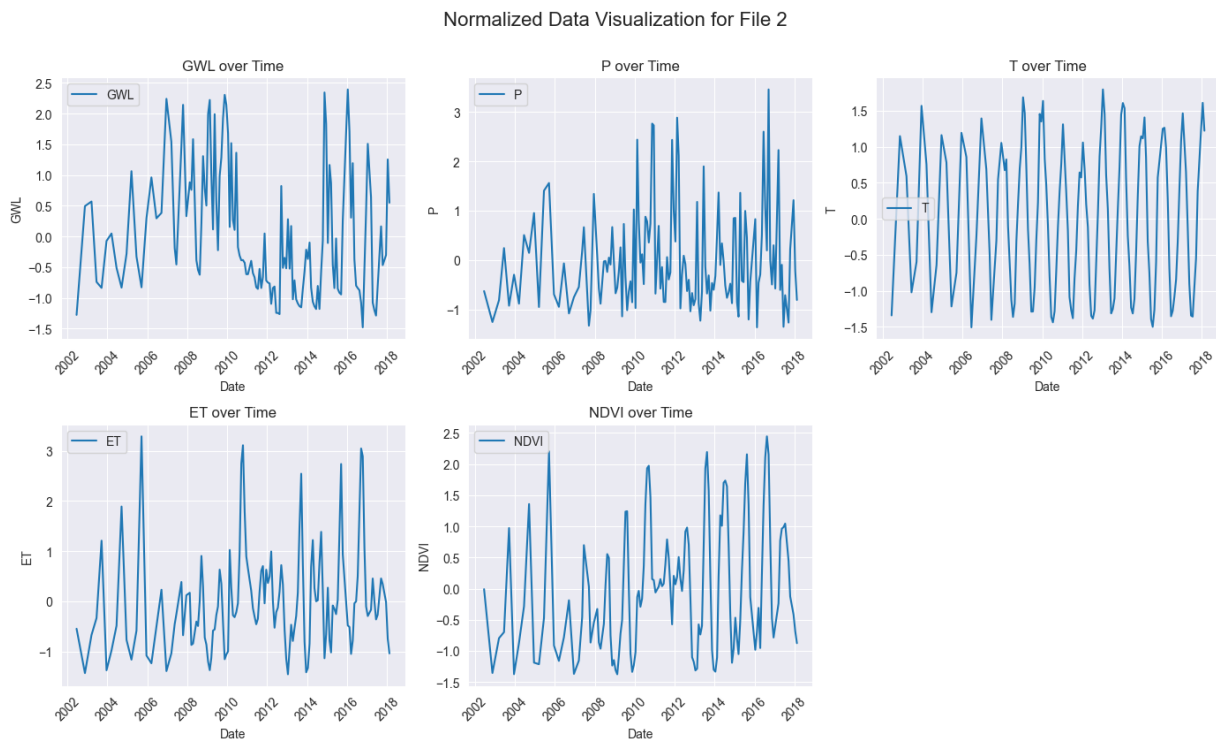


Figure 4: Graphe de courbes normalisées du fichier 2

En visualisant les deux figures 3 et 4, on peut remarquer les mêmes tendances vu au visualisation non normalisées, mais, d'une façon plus claire visuellement.

2.3 C. D. Calcul de CCC pour toutes les paires de variables possible

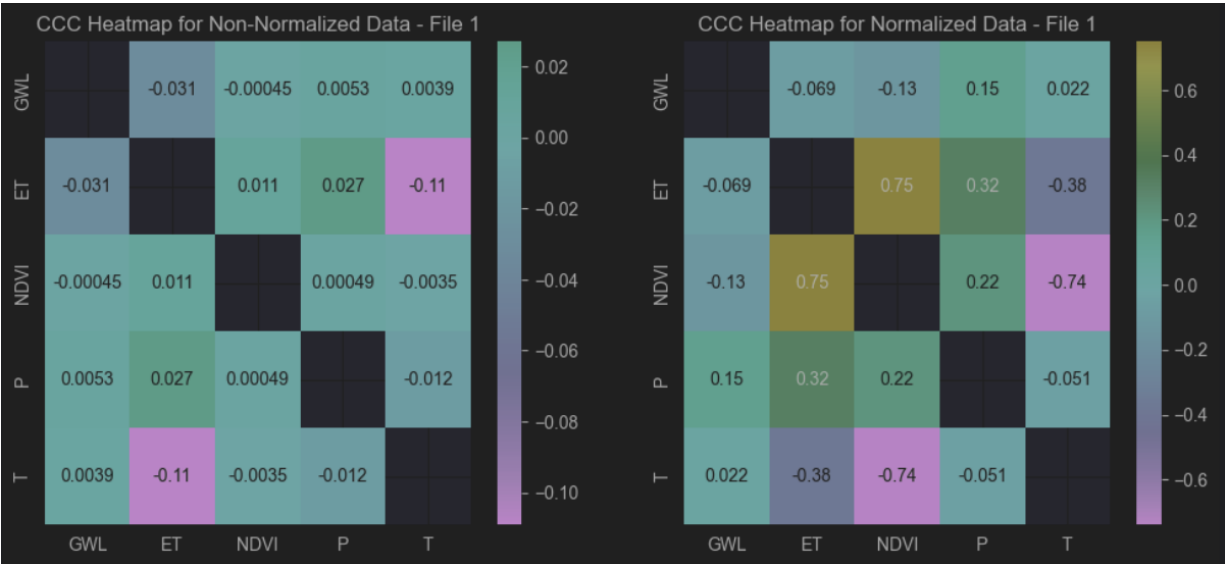


Figure 5: Heatmap des CCC du fichier 1

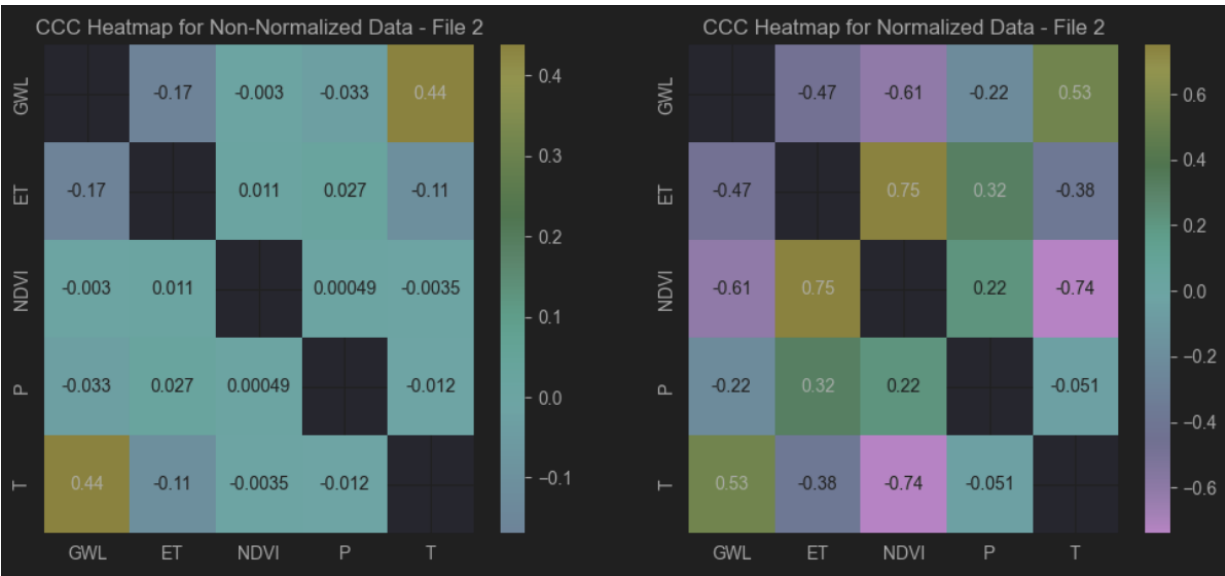


Figure 6: Heatmap des CCC du fichier 2

En visualisant les deux figures de Heatmap de CCC (coefficient de corrélation de concordance) 5 et 6, on confirme que nos interprétation faite d'après la visualisation des figures est correcte avec le CCC entre ET et NDVI égale à 0.75 , ET et P à 0.32 et GWL et T à 0.53.

3 Partie 2. Prédiction du niveau des nappes

Dans cette partie on expliquera brièvement les méthodes qu'on à appliquer pour entraîner nos modèle et prédire les niveau de nappes phréatiques.

3.1 a) construction d'un modèle spécifique à un puit

Dans cette première méthode, on à commencer tout d'abord par la préparation des données, en commençant par le traitement des données manquantes pour la colonne GWL, en remplaçant les valeur qui se trouve à NaN en appliquant la méthode d'interpolation linéaire, avec un traitement spécifique lorsque les valeur manquantes se trouve aux extrimités par la moyenne globale de ce fichier, cette technique m'as permis d'obtenir les meilleures résultats de traitement de GWL manquantes parmi d'autres méthode, comme la moyenne ou bien même l'implémentation d'un modèle de prédiction regressif. Ensuite, on crée la classe et on prépare les séquence qui répondent au problème de prédiction des 6 prochaines mois, en se basant sur la dernière visualisation.

On divise notre ensemble des données en données de train, validation et test, avec les 12 dernier pour le test, les 12 qui les précèdent pour la validation, et le reste pour l'entraînement, ce modèle (Cette opération se répète pour chaque puit séparément sans les concaténés).

Après, on met en place notre modèle LSTM, en s'assurant qu'il retourne les état cachées dans la fonction forward(), puis, on entrîne notre modèle avec application de la technique de Best Epoch, pour retourner l'époque qui à donner les meilleures résultats sur l'ensemble de validation, en prenant en considération les états cachées du batch précédent.

L'input de ce modèle sont les 4 métrique P, T, ET, NDVI et l'output est la prédiction de GWL pour les 6 prochaines mois, en prédictant ces 6 mois en même temps.

On a mit 200 epochs pour chaque puit avec un learning rate de 0.001, et on à obtenu la résultat suivantes sur les 20 premiers fichiés :

- Average Test MAE for all wells: 2.5883

3.2 b) Modèle générale simple

Pour ce modèle la, on à appliquer les mêmes techniques de traitement de données que le modèle spécifique à un puit, sauf qu'on à maintenant pris la totalité de notre corpus de données qui est de 2510 fichiés, la divisé en deux ensemble de données, données de test que représente les 12 dernier mois de chaque fichié, et le reste pour l'entraînement, on à concaténé les données d'entraînement pour avoir un modèle générale.

On à garder les mêmes démarches d'entraînement et de création de modèle que celle du modèle spécifique à un puit, mais, en éliminant la technique de best epoch.

avec un nombre d'epoch égale à 150 et learning rate de 0.001, on à eu les métriques

suivantes :

```
Overall Avg MAE: 8.0519
Overall Std MAE: 5.3884
Overall Min MAE: 1.1794
Overall Max MAE: 37.5441
Model saved to AI Model/model_general.pth
```

Figure 7: Résultats du modèle générale

On constate d'après ces résultats, que la moyenne des erreurs absolue à augmenter qui est logique vu le grand corpus de données, ainsi que la similarités des données entre les régions, on obtenue aussi un MIN de 1.1 qui veut dire qu'il y'as des puits qui on été prédit avec un erreur bien réduit, et d'autre (Max), qui sont moins fiable.

On à fini par sauvegarder ce modèle par ce que on l'aurais besoin dans les méthodes prochaines de fine-tuning.

3.3 c) Modèle général fine-tuned par région

Pour cette méthodes mieux adapté aux spéciftés de chaque région, on applique la technique de fine-tuning sur notre modèle générale pré-entraîné et sauvegarder.

On commence tout d'abord par le chargement et la préparation des données, en associant l'identifiant de chaque well (puits) avec sa région (on à concentré sur les deux régions VIC (victoria) et QLD (Queensland)).

Puis, on à préparer les données pour le fine-tuning, charger le modèle pré-entraîné, et, on à raffiner notre modèle général, pour être un modèle général spécifique à des régions choisis.

On à utiliser 50 epoch et un learning rate de 0.0001 qui est plus petit que les modèle précédent vu qu'on veut pas apporter des modifications importantes des poids du modèle pré entraîné.

Avec cette méthode de fine tune par région, on à obtenu les résultats suivantes :

D'après ces résultat, on observe que le MAE pour la région de victoria à augmenté que celui du modèle générale simple, tandis que, les wells de la région de Queensland à diminier que celle du modèle générale, avec un minimum MAE obtenu dans les deux régions de 0.0005, qui indique qu'il y'as des puits qui ont été prédite d'une manière presque parfaite.

```
Évaluation pour la région Victoria
VIC: Avg MAE: 10.5810, Std MAE: 7.0708, Min MAE: 0.0005, Max MAE: 40.4808
Model fine-tuned for VIC saved to AI Model/model_fine_tuned_vic.pth
Évaluation pour la région Queensland
QLD: Avg MAE: 7.8593, Std MAE: 6.3599, Min MAE: 0.0005, Max MAE: 45.2309
Model fine-tuned for QLD saved to AI Model/model_fine_tuned_qld.pth
```

Figure 8: Résultats du modèle régional

On à finis par l'enregistrement de ces deux modèles fine-tuner, pour l'utilisation ultérieure dans la partie d) de fine-tune par puit.

3.4 d) Modèle général fine-tuned par puit

3.4.1 Modèle général fine-tuned par puit depuis le modèle régional

Pour ce modèle, on va utiliser les deux modèles régionaux qu'on à pre-entraîné pour les adapté aux spécifité de chaque puit de ces deux régions.

On commence tout d'abord par le chargement de ces deux modèles, puis, à charger les données pour chaque puit, puis, appliquer le fine-tune pour chaque puit.

En appliquant 200 epoch par puit et un learning rate de 0.0001, on à obtenu les résultats suivantes :

```
Overall VIC: Avg MAE: 5.4444
Overall QLD: Avg MAE: 3.5800
Overall Avg MAE for all wells: 5.1568
Overall Std MAE for all wells: 2.7897
Overall Min MAE for all wells: 0.9394
Overall Max MAE for all wells: 13.1025
```

Figure 9: Résultats du modèle spécifique à un puit depuis le modèle régional

On remarque les deux métriques MAE on améliorée que celle de puit régional, avec une amélioration imprévisive surtout au niveau de MAX MAE, qui c'est améliorée de 40-45 à 13 via ce modèle, qui indique que la marge d'erreur à significativement diminuer.

3.4.2 Modèle général fine-tuned par puit depuis le modèle général

Pour ce modèle, on va suivre presque les mêmes étapes que le modèle précédent, mais, en se basant sur le modèle général simple pré-entraîné précédemment.

On s'est focalisé sur les Wells (puits) des deux régions vu précédemment, pour pouvoir comparer d'une façon plus précise après, et on a fine-tuné le modèle général pour qu'il soit un modèle général, fine-tuné aux spécificités de chaque puit des deux régions (VIC) et (QLD).

On trouve après l'entraînement les résultats suivants :

```
Overall VIC: Avg MAE: 5.1249
Overall QLD: Avg MAE: 2.8970
Overall Avg MAE for all wells: 4.7812
Overall Std MAE for all wells: 2.9479
Overall Min MAE for all wells: 0.5879
Overall Max MAE for all wells: 15.5704
```

Figure 10: Résultats du modèle spécifique à un puit depuis le modèle général

On remarque d'après les résultats que ce modèle s'est amélioré en termes de MAE par rapport au modèle général fine-tuné par puit depuis le modèle régional, atteignant un MAE proche du MAE qu'on a trouvé à la partie a), ce qui indique des meilleurs résultats pour cette approche que ça précédente.

3.5 la meilleure approche pour prédire le niveau des nappes phréatiques

Pour répondre à cette question en terme de MAE, on trouve que c'est le premier modèle a), qu'on a eu avec les meilleurs résultats, sauf que en pratique, ce modèle est pas vraiment fiable pour exploiter des nouveaux puits non connus précédemment, cette approche est plutôt adaptée à une prévision future des Wells existantes, je pense que la meilleure approche selon moi est la dernière méthode du Modèle général fine-tuné par puit depuis le modèle général, qui a donné les meilleurs résultats en terme de généralisation et de modèle général, et qui est très pratique pour des futures exploitations des puits non connus déjà, en ajoutant la région que ce futur puits est localisé pour des résultats optimisés.

4 Conclusion

Ce projet à été une très bonne expérience pour pratiquer mes connaissance en terme de traitement de données numériques, surtout avec la taille importantes du corpus de données, j'ai à réussi à surmonter beaucoup de challenges au niveau de ce pré-traitement.

Aussi, l'occasion de tester une telle variété de modèle avec des technique différentes m'as appris beaucoup de chose qui vont me servir dans mes futures études, ainsi que, l'application du technique de fine-tune qui était un black box pour moi avant.