

Reconstruction des données piézométriques manquantes par autoencodeur LSTM

Application au bassin de la Seine (France)

Projet ANR Hydrolink — **Hydrolink-ANR-24-CE01-0140**

Houssam Eddine Rahimi Abderrahim Jardani Matthieu Fournier

M2C (UMR 6143) Université de Rouen Normandie / Université de Caen Normandie / CNRS

Congrès Eau & IA 2026 ■ Grenoble (World Trade Center) ■ 4–6 mars 2026



Pourquoi reconstruire les niveaux piézométriques manquants ?

Les lacunes biaisent l'analyse et limitent la modélisation prédictive.

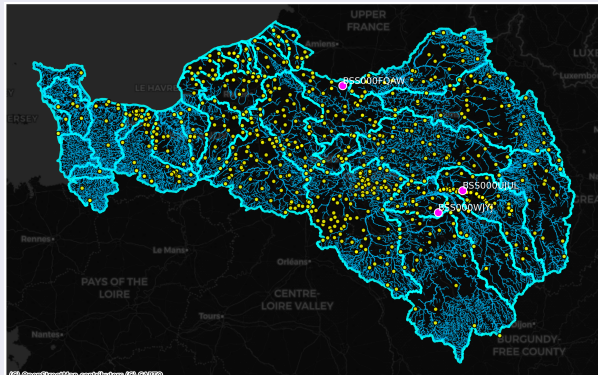
- Les séries temporelles de niveau de nappe (GWL) sont essentielles pour la gestion de la ressource, l'évaluation des risques et le calage des modèles.
- Les archives de suivi présentent souvent des lacunes (pannes de capteurs, maintenance, changements d'instrumentation, filtrage qualité).
- À l'échelle du bassin, les variables de forçage (météo, pompage) ne sont pas toujours disponibles partout à une résolution homogène.
- **Objectif** : montrer qu'il est possible de reconstruire les niveaux de nappe à l'échelle du bassin en s'appuyant uniquement sur les données piézométriques disponibles.

Jeu de données du bassin de la Seine (mensuel)

Source des données & sélection

- Hub'Eau *Piézométrie* (données issues d'ADES)
- janv. 1970 – mars 2025 (mensuel)
- 652 piézomètres avec ≥ 120 observations mensuelles
- **Lacunes typiques** : points manquants intermittents + longues lacunes continues

Répartition spatiale (piézomètres sélectionnés)



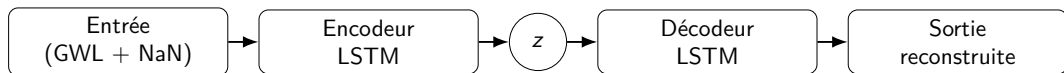
Basemap : ©CARTO — données : ©OpenStreetMap contributors.

Chaîne de traitement proposée (échelle bassin)



- Entraîner un modèle unique à l'échelle du bassin sur la matrice incomplète des séries piézométriques.
- Utiliser une **perte masquée** afin que les entrées manquantes (NaN) ne contribuent pas à l'optimisation.

Modèle : autoencodeur LSTM (LSTM-AE)



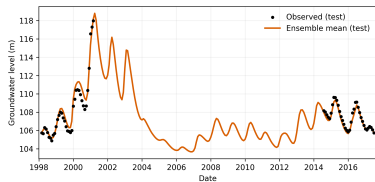
- Architecture encodeur–goulot d'étranglement–décodeur, entraînée de bout en bout sur des séquences.
- Normalisation des séquences avant l'entraînement.
- **MSE masqué** : masque binaire $\Rightarrow$ les NaN ne contribuent pas à la perte.
- Sortie : séries mensuelles **comblées** pour chaque piézomètre.

Protocole d'évaluation représentatif

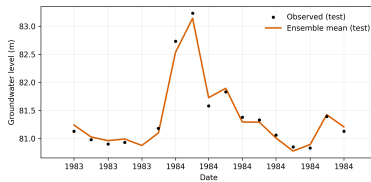
- L'hétérogénéité du bassin implique des dynamiques contrastées et des disponibilités de données très inégales.
- L'évaluation doit refléter des interruptions réalistes (ex. lacunes continues artificielles de durées différentes).
- Les métriques sont calculées **uniquement sur les observations masquées** :

MAE	RMSE	NSE
Erreur absolue moyenne	Racine de l'erreur quadratique moyenne	Efficacité de Nash–Sutcliffe ($\rightarrow 1$ est optimal)

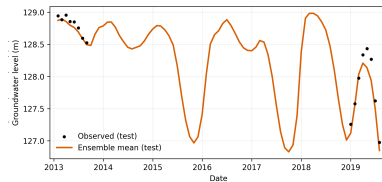
Résultats : exemples de reconstruction (test)



BSS000FQAW



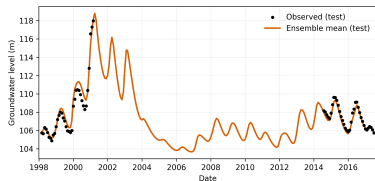
BSS000UJUL



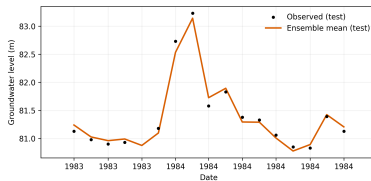
BSS000WJYJ

Points noirs : observations (test). Courbe orange : reconstruction LSTM-AE (test).

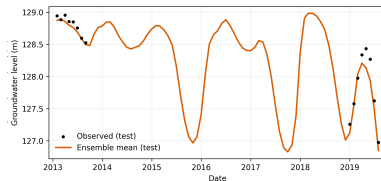
Résultats : mêmes exemples + métriques (test)



BSS000FQAW



BSS000UJUL



BSS000WJYJ

Code BSS	n	MAE (m)	RMSE (m)	NSE
BSS000FQAW	73	0.395	0.523	0.956
BSS000UJUL	16	0.080	0.089	0.982
BSS000WJYJ	16	0.103	0.137	0.951

- Bonne qualité de reconstruction sur ces exemples de validation/test.
- Baisse de performance pour les séries courtes ou les dynamiques sous-représentées à l'entraînement.

Limites & perspectives

Limites (version actuelle)

- Résolution temporelle mensuelle uniquement
- Aucune contrainte physique ou hydrogéologique explicite
- Axé sur la reconstruction, et non sur la prévision

Perspectives (Hydrolink)

- Intégration d'autres données exogènes (météorologie, géologie, etc.)
- Extension aux données journalières sur les eaux souterraines
- Couplage avec des modèles physiques (CAWAQS)
- Intégration d'approches hybrides ou basées sur la physique
- Transition de la reconstruction aux prévisions à court et moyen terme

- Un autoencodeur LSTM *head-only* peut combler les niveaux de nappe mensuels manquants à l'échelle du bassin.
- Points forts : simple (pas d'entrées exogènes), exploite les dépendances entre piézomètres, bon niveau de précision sur des cas tests représentatifs.
- À consolider : séries clairsemées, dynamiques rares, passage au journalier et intégration d'info physique.

Merci !

Questions / discussion

Contact

houssam-eddine.rahimi@univ-rouen.fr