

EMI : Prédire et suivre l'évolution quotidienne du niveau des nappes grâce à l'intelligence artificielle

Jean LUCHIER
Product Manager et
Hydrogéologue

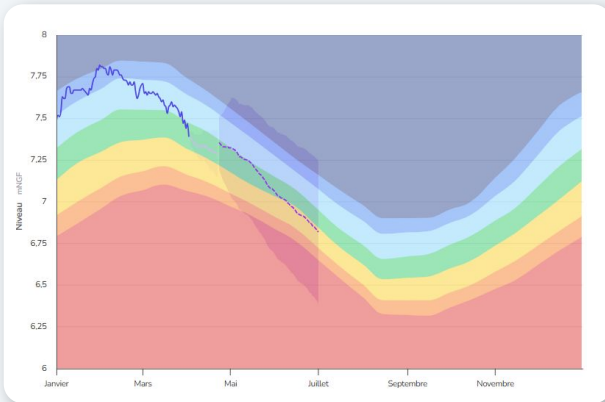


Faustine BOUSQUET
Data Scientist

04/03/2026

EMI : Mieux anticiper pour mieux gérer

Anticipation du
niveau des nappes



IO⁺

Aide à la décision
à l'appui de
données visuelles



Comité
Sécheresse et seuils
arrêtés cadre



Impact sur
l'exploitation des
captages



Opérations
préventives

Quels défis ?

Des problématiques diverses

- Limites des modèles hydrogéologiques à base physique
- Aquifères méconnus / mal contraints / hétérogènes
- Littérature ML fournie mais souvent mono-site / expérimentale

**Vers une prévision
piézométrique opérationnelle
à grande échelle**

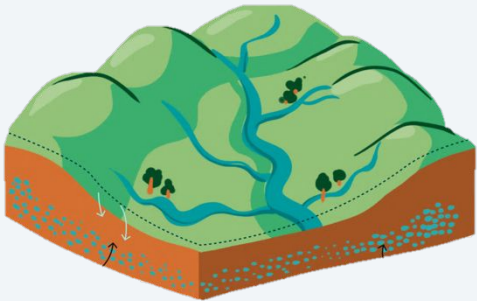
Notre solution

- Approche multi-sites
- Prévision quotidienne
- Déploiement opérationnel sur notre application EMI

Architecture de la prévision piézométrique

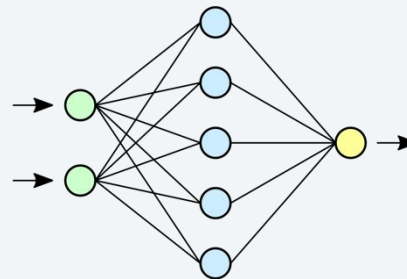
Données d'entrée

Piézométrie
Données météo
Hydrométrie



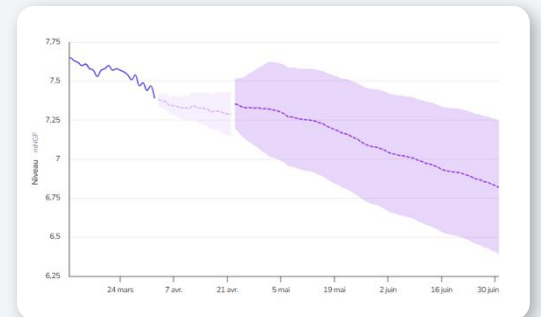
Modélisation du niveau des nappes

Modèles court terme
Modèle moyen terme



Prévisions sur EMI

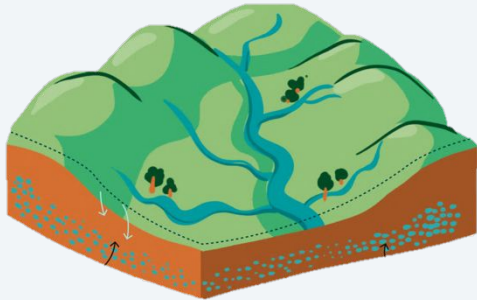
Prévisions quotidiennes
Intervalles d'incertitude



Architecture de la prévision piézométrique

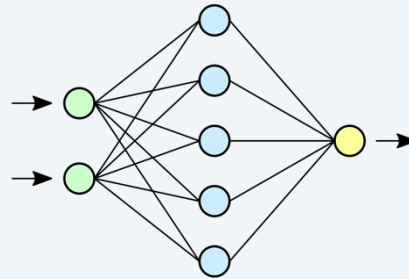
Données d'entrée

Piézométrie
Données météo
Hydrométrie



Modélisation du niveau des nappes

Modèles court terme
Modèle moyen terme



Prévisions sur EMI

Prévisions quotidiennes
Intervalles d'incertitude



Construction des prédicteurs

à partir des données hydrométéorologiques

Données d'entrée

- y_t : niveau piézométrique
- Pluie, T° , ETR
- Débit rivière

Vecteur explicatif au temps t

$$x_t \in \mathbb{R}^p$$

$$x_t = (\text{lags, cumul pluie, moyennes débit, } T^\circ, \dots)$$

t : date de la prédiction

p : nombre de prédicteurs sélectionnés localement

Feature engineering

- Lags dépendant de l'horizon h
- Sommes/moyennes glissantes
- Indicateurs cumulatifs

Objectif : expliciter les dépendances temporelles.

Sélection locale

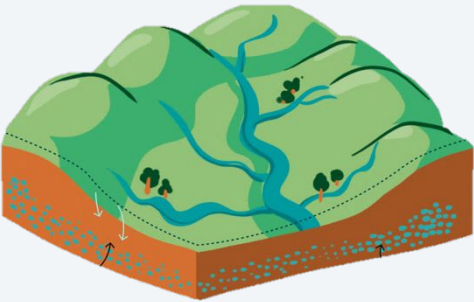
- Espace initial redondant
- Sélection automatique
- Adaptation par site

Objectif : répondre à la forte hétérogénéité inter-stations

Architecture de la prévision piézométrique

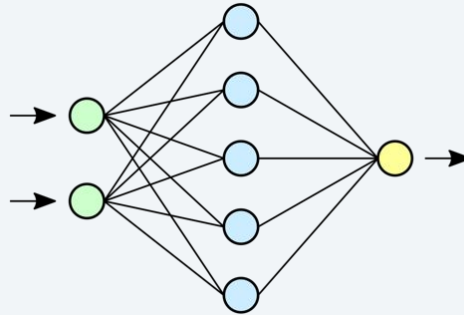
Données d'entrée

Piézométrie
Données météo
Hydrométrie



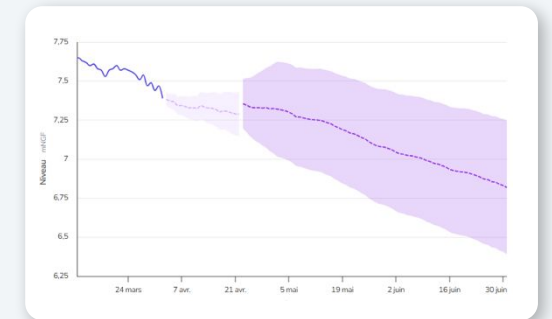
Modélisation du niveau des nappes

Modèles court terme
Modèle moyen terme



Prévisions sur EMI

Prévisions quotidiennes
Intervalles d'incertitude



Architecture de prévision multi-horizons

y_t : niveau piézométrique

$x_t \in \mathbb{R}^P$: vecteur de prédicteurs

Court terme (1-20 jours) : Modèles horizon-spécifiques

$$\hat{y}_{t+h} = f_h(x_t), \quad h = 1, \dots, H_{CT}$$

- Un modèle par horizon
- Sélection par validation croisée temporelle
- MLP / Gradient Boosting

Moyen terme (20-90 jours) : Modèle additif multi-horizons

$$(\hat{y}_{t+H_{CT}+1}, \dots, \hat{y}_{t+H_{LT}}) = f_{LT}(x_t)$$

- Prédiction conjointe des horizons $H_{CT}+1$ à H_{LT}
- Modèle additif (Prophet)
- Tendence + saisonnalité + régresseurs

Intervalle de prédiction : estimé à partir des erreurs historiques

→ Calcul du quantile de l'erreur absolue

$$q(h) = \text{Quantile}_{\alpha} (|y_{t+h} - \hat{y}_{t+h}|)$$

Métriques & protocole d'évaluation

Erreur moyenne absolue

$$MAE = \frac{1}{n} \sum |y_i - \hat{y}_i|$$

- Exprimée en mètres
- Interprétable physiquement



Normalisation

MAE / amplitude saisonnière (%)

- Comparaison inter-sites
- Faible % = bon modèle

Critère de persistance

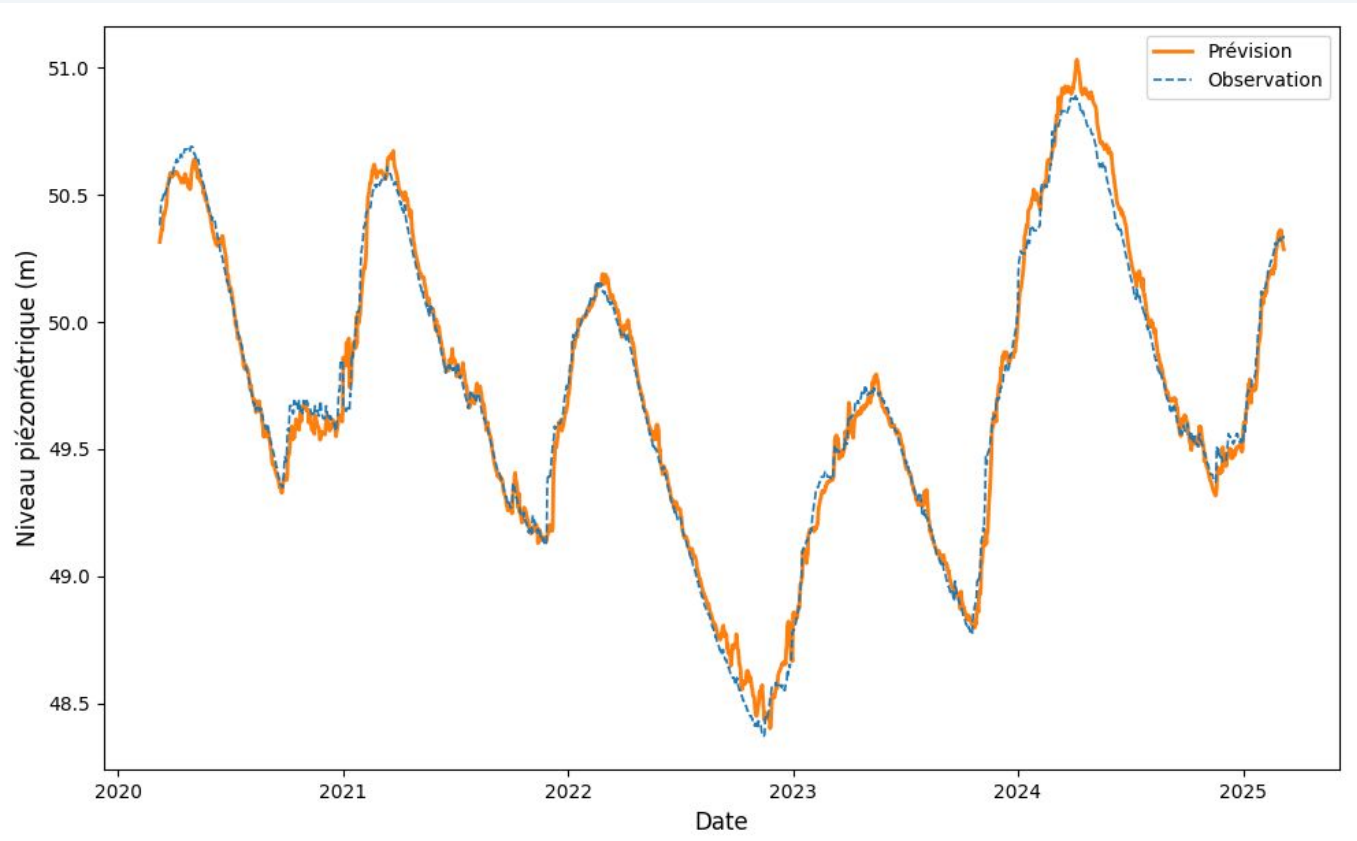
$$Persistence = 100 \times \left(1 - \frac{MAE_{\text{modèle}}}{MAE_{\text{naïf}}} \right)$$

- > 0 : modèle apporte de la valeur
- = 0 : équivalent naïf
- < 0 : moins bon que modèle naïf

Modèle naïf

- **Prévision court terme** : valeur observée décalée de h jours où h correspond à l'horizon de prédiction
- **Prévision moyen terme** : moyenne historique observée à la même période de l'année

Performance sur 60 piézomètres



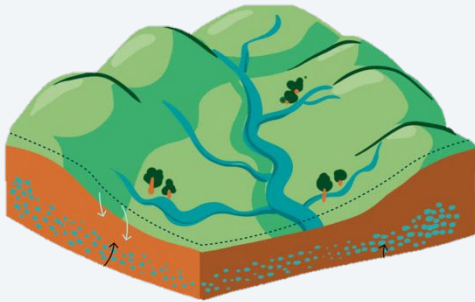
Prévision à l'horizon 10 jours pour un piézomètre : observations (pointillé bleu) et prévisions (trait orange)

Court terme (1–20 jours)	Moyen terme (21–90 jours)
MAE / Variation saisonnière 4.5 %	MAE / Variation saisonnière 18 %
98 % des stations > modèle naïf	78 % des stations > modèle naïf
Persistance moyenne 40%	Persistance moyenne 27%
Amélioration quasi systématique par rapport au modèle naïf	Dégradation attendue mais valeur prédictive maintenue

Architecture de la prévision piézométrique

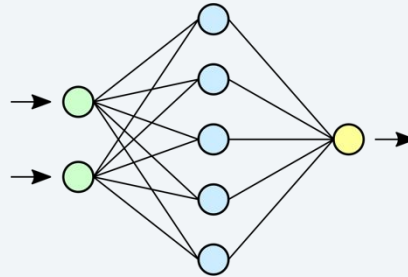
Données d'entrée

Piézométrie
Données météo
Hydrométrie



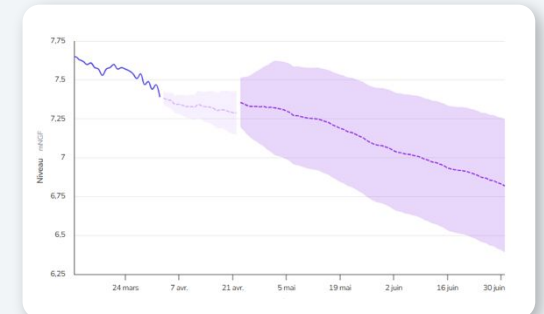
Modélisation du niveau des nappes

Modèles court terme
Modèle moyen terme



Prévisions sur EMI

Prévisions quotidiennes
Intervalles d'incertitude



Du modèle à l'inférence opérationnelle



Pipeline Python
reproductible

Entraînement automatisé des modèles

- Preprocessing automatique par site
- Lancement programmé des jobs d'entraînement
- Stockage et versioning des artefacts modèles

➡ Passage d'un entraînement manuel à une orchestration cloud.

Inférence quotidienne orchestrée

- Déclenchement automatique journalier
- Reconstruction dynamique du vecteur de prédicteurs
- Génération de prévisions multi-horizons (1-90 jours)
- Inférence distribuée par piézomètre

➡ Chaîne d'inférence entièrement automatisée, sans intervention humaine.

Infrastructure cloud (AWS)

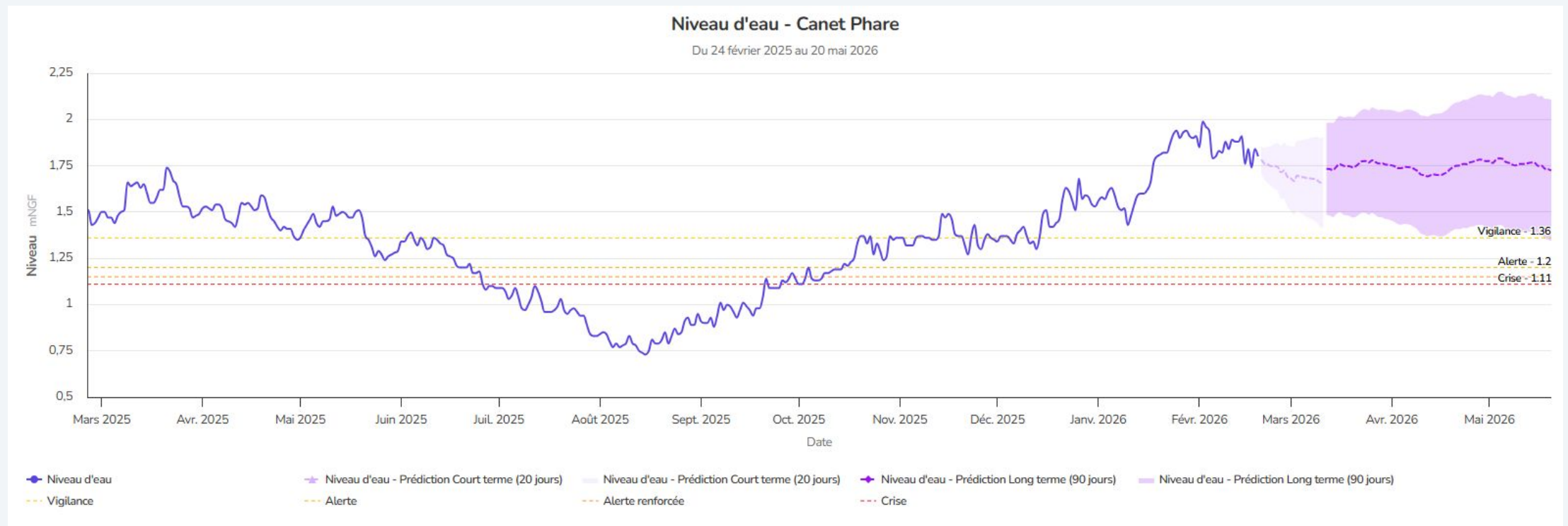
- Orchestration des workflows
- Conteneurisation des traitements Python
- Stockage centralisé des données et modèles
- Scalabilité adaptée au nombre de sites



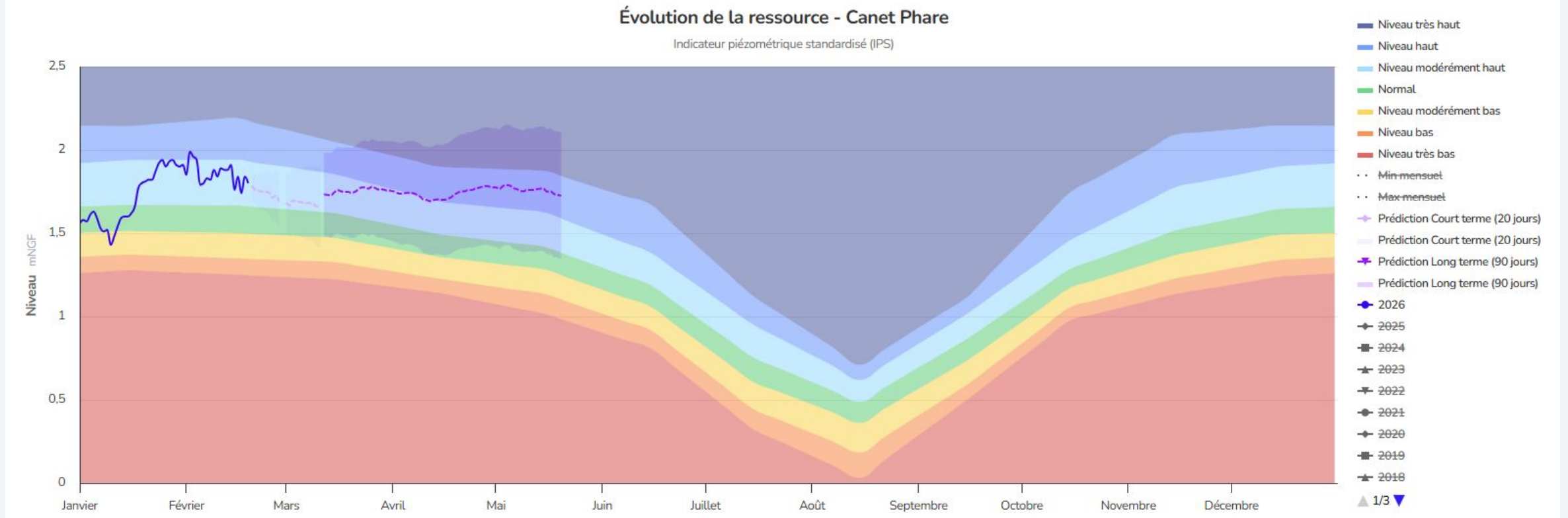
✓ **60** piézomètres avec prédiction, actualisés **quotidiennement** dans notre application **EMI**

✓ Utilisé par des Conseils départementaux, EPTB, Syndicats et Régies

Prévision sur EMI – Seuils ACS



Prévision sur EMI – Seuils IPS



Conclusion et perspectives

Une prévision piézométrique opérationnelle

- Prévisions multi-horizons (1–90 jours) produites quotidiennement
- Intégration directe dans les cadres de gestion (ACS, IPS)
- Inférence quotidienne sur plus de 60 piézomètres
- Déploiement en environnement réel, utilisé par les collectivités

➔ **La prévision devient un outil d'aide à la décisions**

Les enjeux techniques de l'industrialisation

- Automatisation complète de la chaîne entraînement → inférence
- Orchestration cloud robuste et reproductible
- Architecture scalable permettant le déploiement sur un nombre croissant de piézomètres
- Supervision et maintien des performances dans le temps

Et la suite ?

- Amélioration des prévisions à moyen et long terme via une **thèse CIFRE** encadrée par l'IMT Mines Alès (Doctorante : Houa Bedjil - Directrice de thèse : Anne Johannet)
- Extension de ces travaux vers les forages exploités



Des questions ?



#missionwater