

Régionalisation de modèles neuronaux appliquée aux aquifères fracturés et karstiques

Manon Erguy (HSM, IMT Mines Alès), Sébastien Morilhat (CEA), Guillaume Artigue (HSM, IMT Mines Alès – MGVA Ingénierie), Julien Trincal (CEA), Anne Johannet (HSM, IMT Mines Alès), Severin Pistre (HSM, U Montpellier)

Sommaire

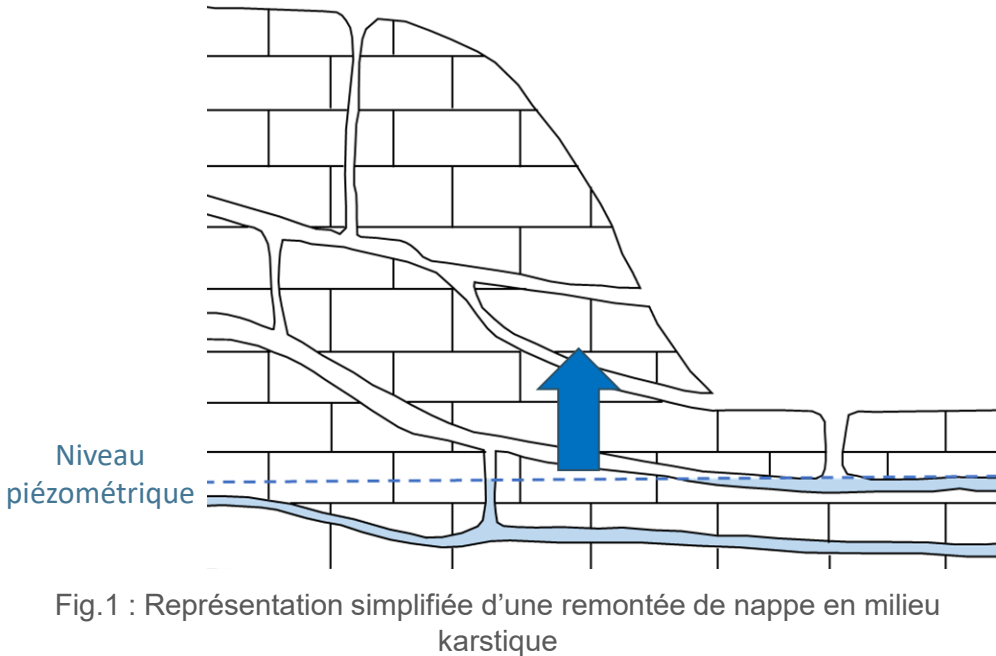
1. Introduction
2. Site d'étude – hydrosystème de Cadarache
3. Conception de modèle à réseaux de neurones
4. Méthodes de régionalisation
5. Résultats de la régionalisation
6. Conclusion

1. Introduction

Contexte

Aquifères karstiques :

- ▶ Hydrosystèmes complexes, anisotropes et hétérogènes liés à des processus non-linéaires
- ▶ Associés à différents enjeux : gestion de la ressource, transfert de polluants, inondations notamment par remontées de nappe



Remontée de nappe sur le site CEA de Cadarache :



Fig.2 : Phénomène de débordement résultant d'une remontée de nappe (photo S. Morilhat)

- ▶ Site industriel en région méditerranéenne
- ▶ Mises en charge locales de plus de 40 m à la suite d'épisodes pluvieux intenses

→ Fort enjeu de sûreté associé à la prévision opérationnelle de ces crues souterraines à l'échelle du site

1. Introduction

Contexte

Aquifères karstiques :

- ▶ Hydrosystèmes complexes, anisotropes et hétérogènes liés à des processus non-linéaires
- ▶ Recharge rapide → mises en charge hétérogènes selon perméabilités des structures

→ remontées de nappe localisées

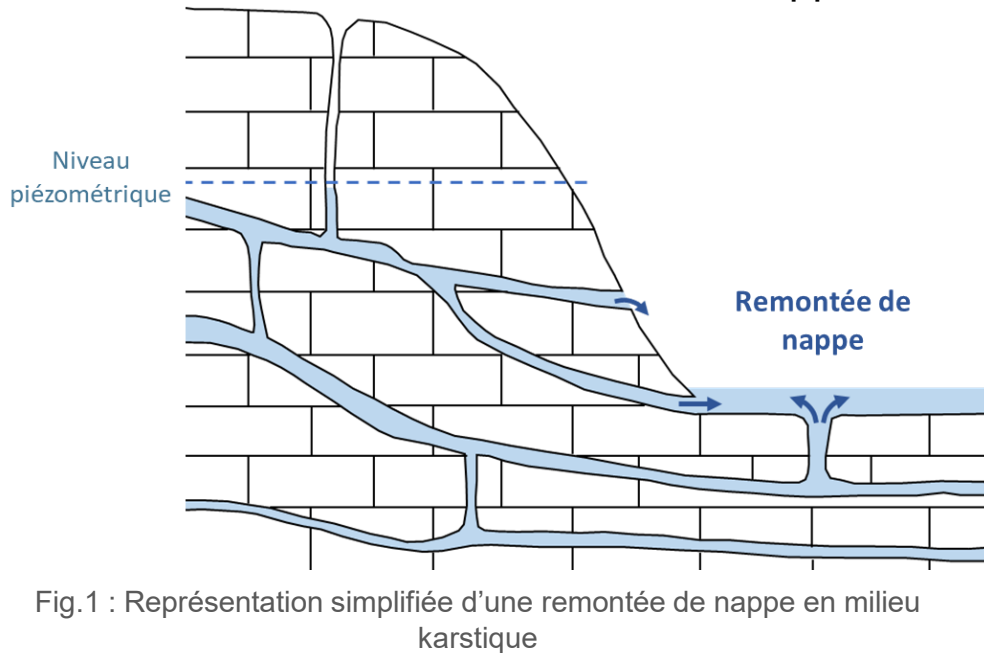


Fig.1 : Représentation simplifiée d'une remontée de nappe en milieu karstique

Remontée de nappe sur le site CEA de Cadarache :



Fig.2 : Phénomène de débordement résultant d'une remontée de nappe (photo S. Morilhat)

- ▶ Site industriel en région méditerranéenne
- ▶ Mises en charge locales de plus de 40 m à la suite d'épisodes pluvieux intenses

→ Fort enjeu de sûreté associé à la prévision opérationnelle de ces crues souterraines à l'échelle du site

1. Introduction

Problématique

Apport des réseaux de neurones artificiels à la prévision du risque de crue karstique sur le site du CEA de Cadarache (Thèse Erguy, 2024)

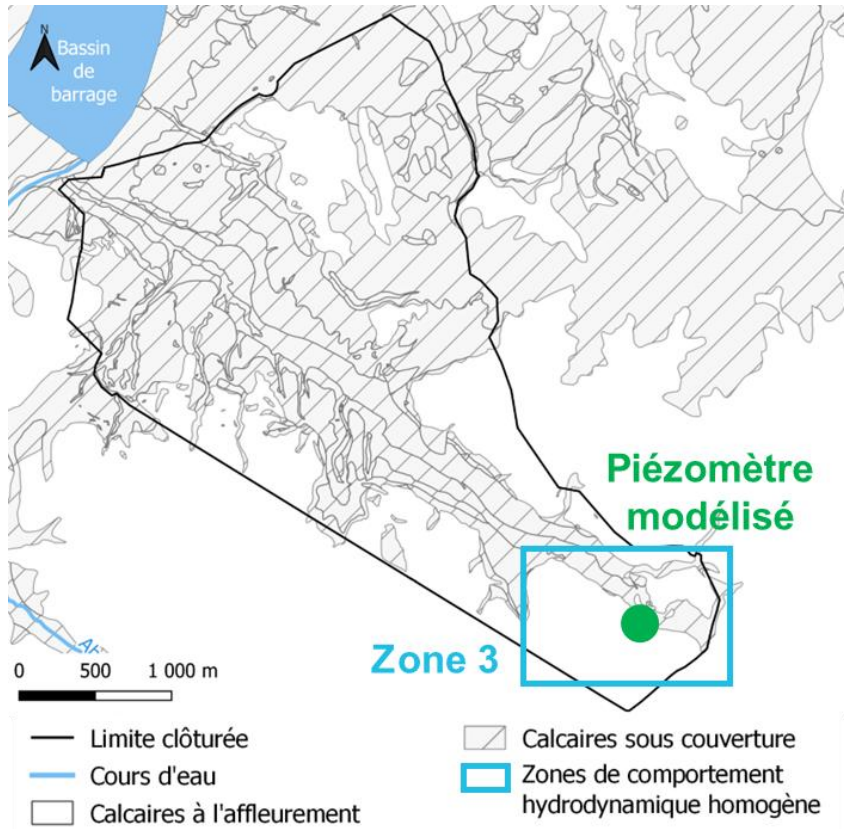


Fig.3 : Localisation du piézomètre modélisé

Travaux préalables :

- Modèle à réseau de neurones artificiels au pas de temps journalier et à horizon de prévision 24h
- Conçu pour un piézomètre « pilote » de la zone sud enregistrant les mises en charge les plus importantes en crue

Contraintes :

- Aquifère karstique hétérogène spatialement et temporellement
- Nécessité de prévoir les remontées de nappe à l'échelle du site

Objectif : régionalisation du modèle à réseau de neurones sur l'ensemble du site

1. Introduction

Travaux préalables - modèle réalisé

- Modèle neuronal au pas de temps journalier et à horizon de prévision 24h ($k+1$) de type perceptron multicouche dirigé (Erguy, 2024)

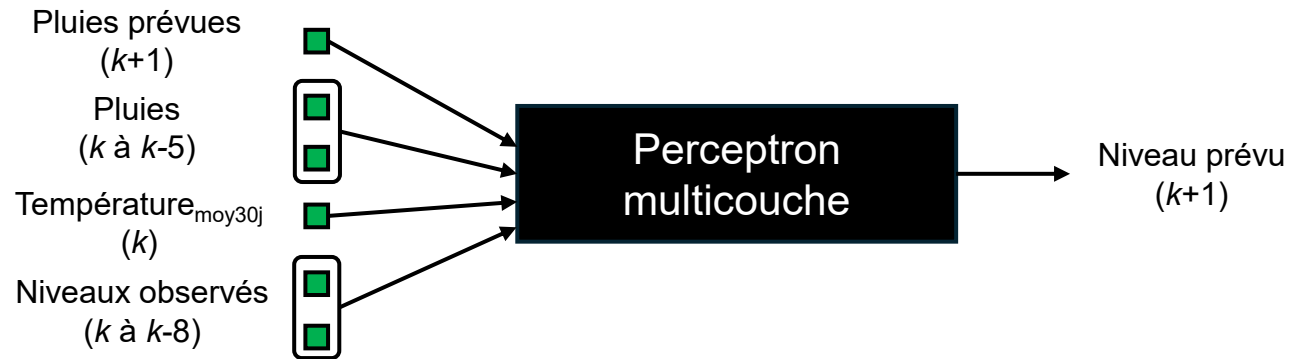


Fig.4 : Schématisation du modèle neuronal conçu sur le piézomètre pilote (Erguy, 2024)

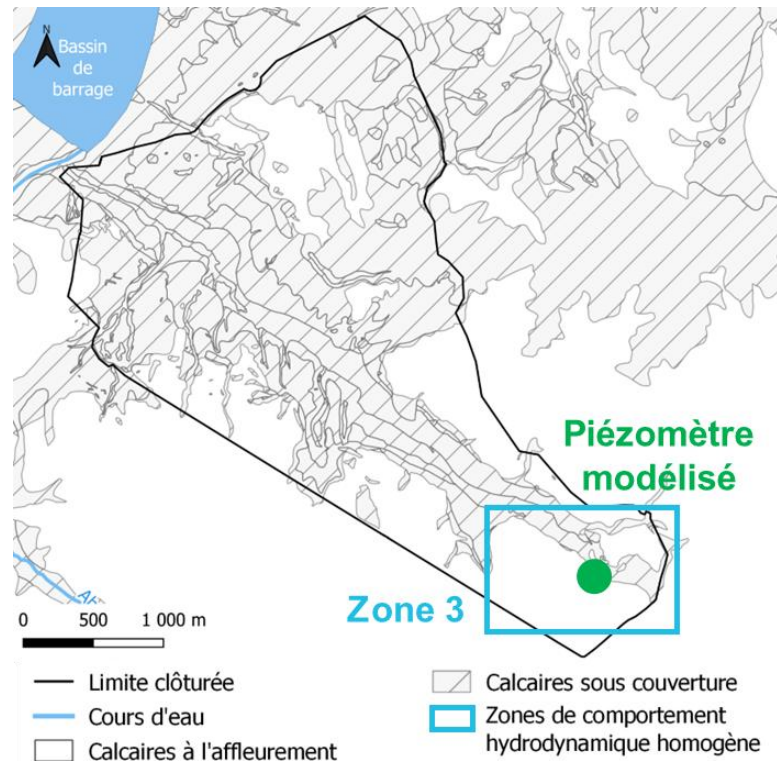


Fig.3 : Localisation du piézomètre modélisé

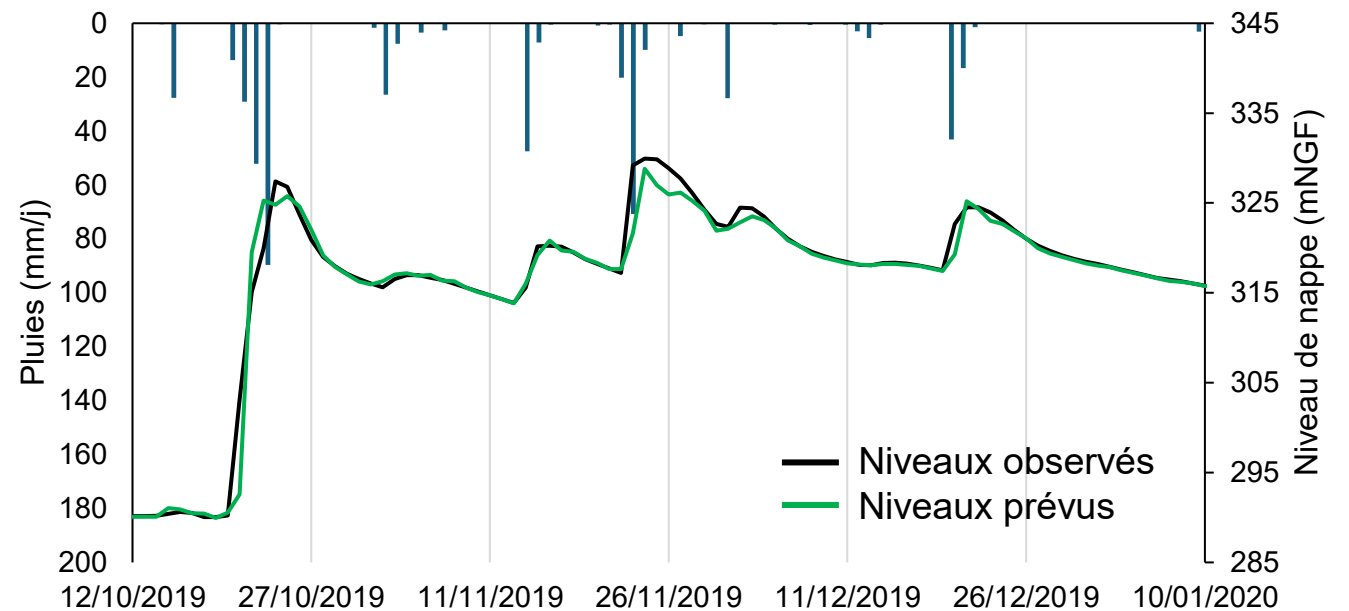


Fig.5 : Prévision de l'épisode test majeur d'octobre-novembre 2019 par le modèle neuronal journalier

1. Introduction

Travaux préalables - modèle réalisé

- Modèle neuronal au pas de temps journalier et à horizon de prévision 24h ($k+1$) de type perceptron multicouche dirigé (Erguy, 2024)

$$NSE = 1 - \frac{\sum_{k=1}^n (H_{obs}^k - H_{sim}^k)^2}{\sum_{k=1}^n (H_{obs}^k - \overline{H_{obs}})^2}$$

$$C_p = 1 - \frac{\sum_{k=1}^n (H_{obs}^{k+h} - H_{sim}^{k+h})^2}{\sum_{k=1}^n (H_{obs}^{k+h} - H_{obs}^k)^2}$$

Avec H : niveau piézométrique, k : temps discret, n : nombre d'échantillons, h : horizon de prévision

Scores de validation croisée		Critères de performance sur le test	
NSE	C_p	NSE	C_p
0,97	0,68	0,98	0,58

Tableau 1 : Performances du modèle

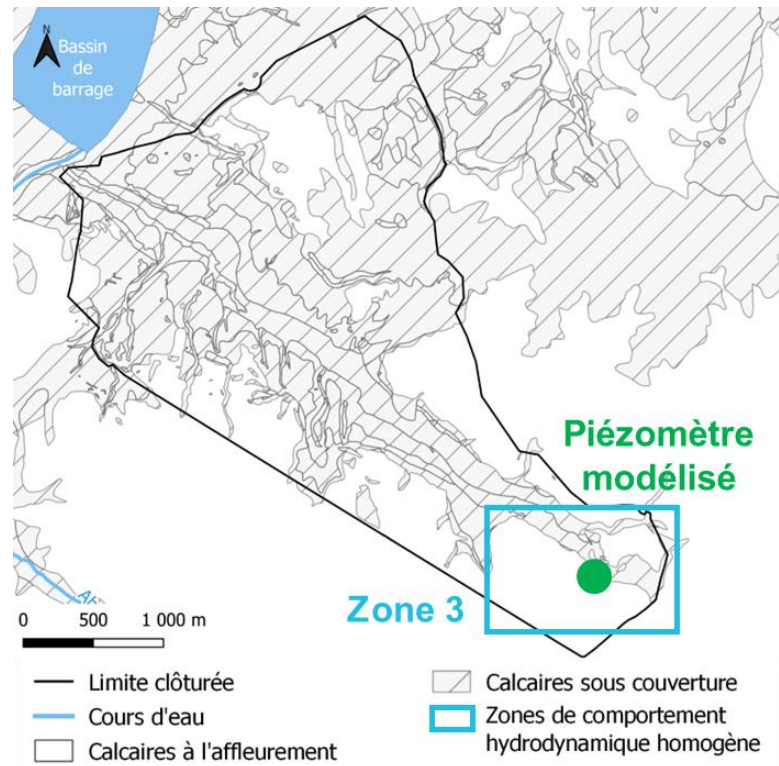


Fig.3 : Localisation du piézomètre modélisé

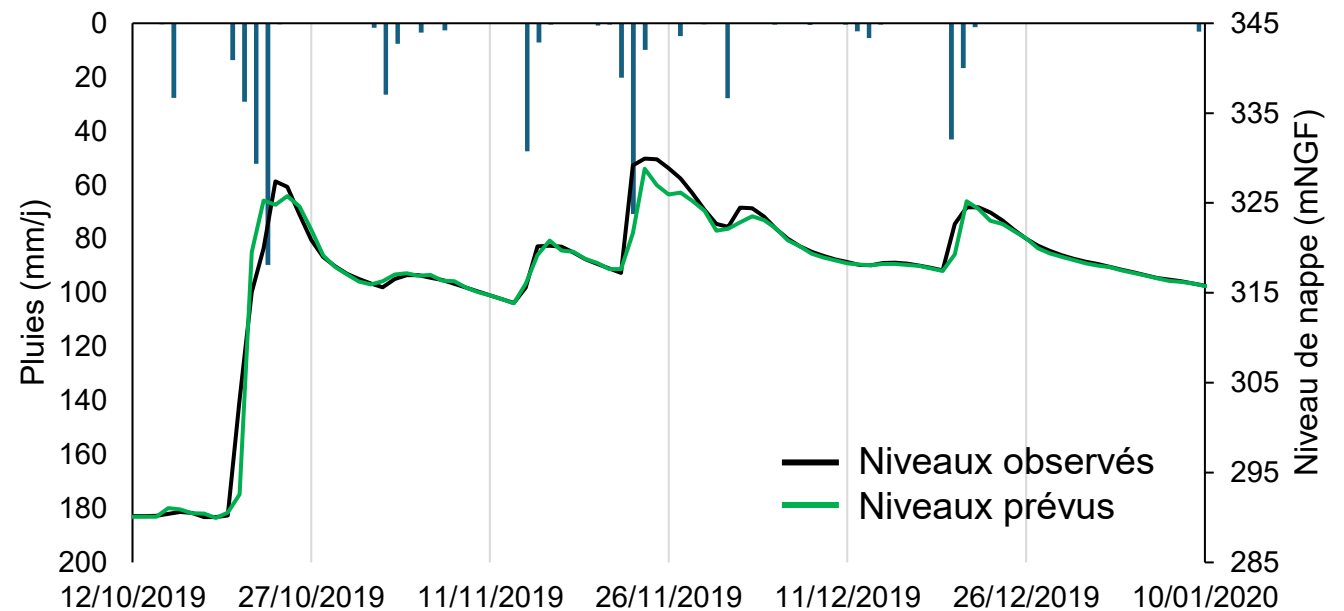
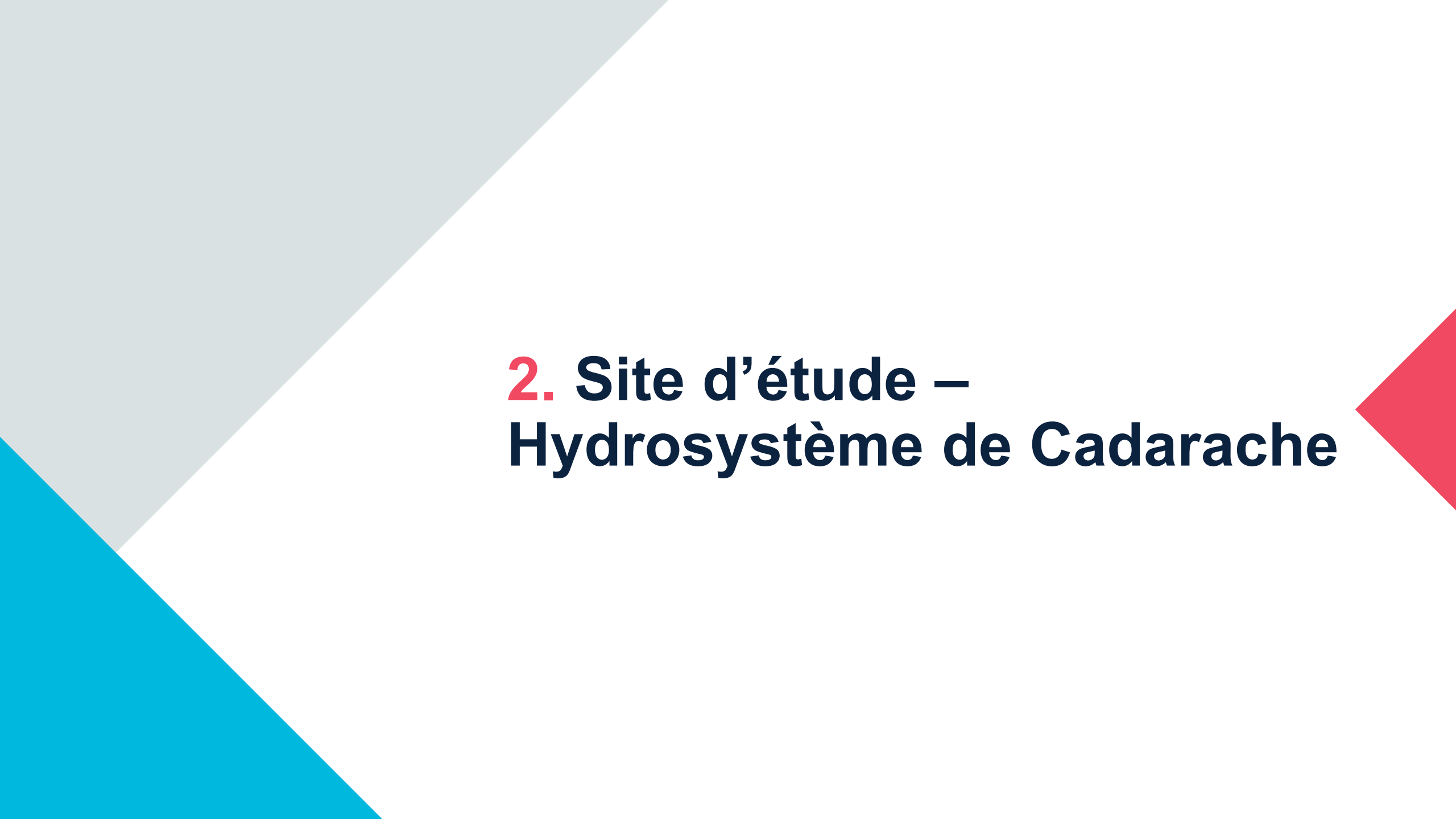


Fig.5 : Prédiction de l'épisode test majeur d'octobre-novembre 2019 par le modèle neuronal journalier



2. Site d'étude – Hydrosystème de Cadarache

2. Site d'étude – hydrosystème de Cadarache

Contexte hydrogéologique et réseau de mesure hydrométéorologique

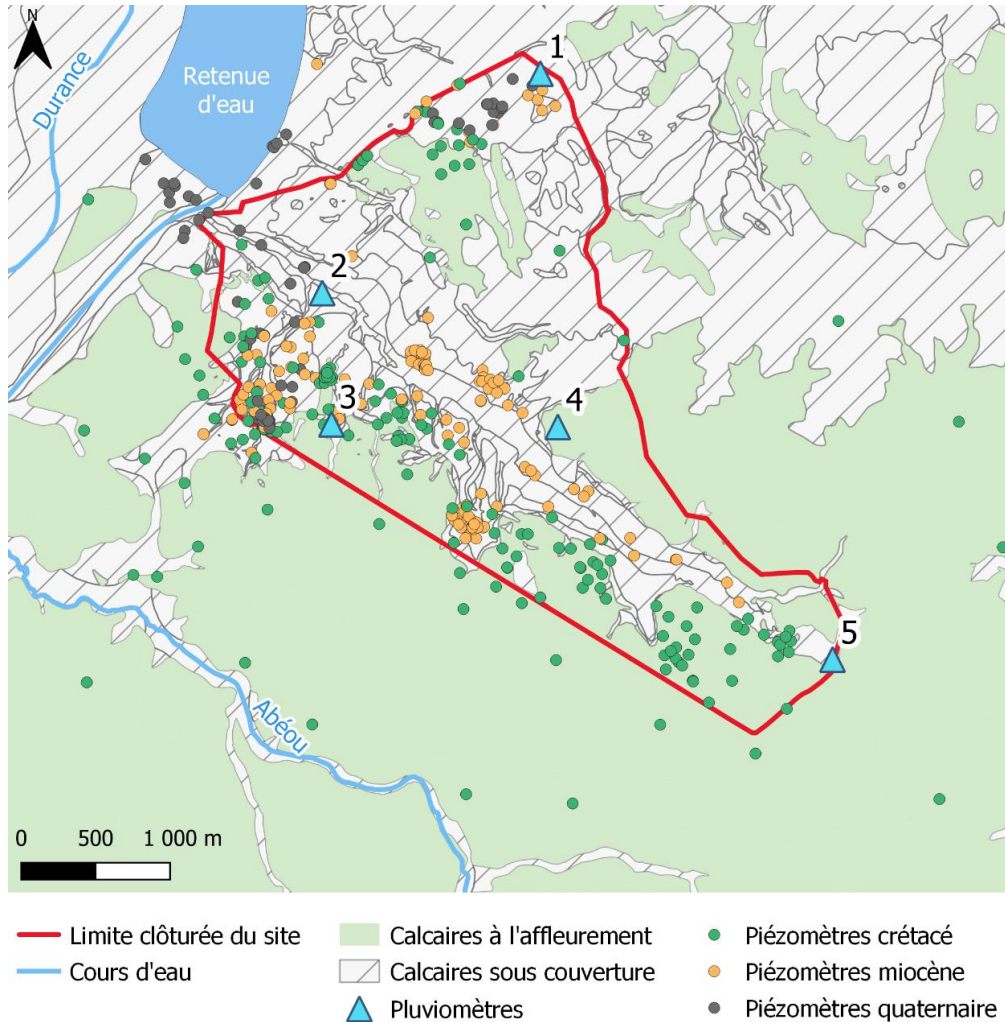


Fig.6 : Carte géologique simplifiée du site de Cadarache et réseau de mesure

Aquifère crétacé :

- ▶ Limite du centre : 9 km²
- ▶ Réseau karstique peu développé et fracturation subverticale
- ▶ Echanges entre aquifères encore mal connus
- ▶ Comportement fortement non linéaire (réponses aux précipitations)

Données piézométriques :

- ▶ Plus de 400 piézomètres :
 - ◆ **280** équipés de capteurs de pression
 - ◆ **115** mesurant la nappe crétacée

→ **Concevoir un modèle pour chaque piézomètre : trop chronophage**

2. Site d'étude – hydrosystème de Cadarache

Hétérogénéité de l'aquifère crétacé

Etude précédente - hauteurs d'eau classées :

- ▶ **Clustering** semi-automatique basé sur les courbes de hauteurs d'eau classées
- ▶ Forte hétérogénéité spatiale
- ▶ Identification des **zones** de comportements hydrodynamiques homogènes

Erguy M, Morilhat S, Artigue G, et al. (2023). A new simple statistical method for the unsupervised clustering of the hydrodynamic behavior at different boreholes: analysis of the obtained clusters in relation to geological knowledge. Environ Earth Sci 82:451.

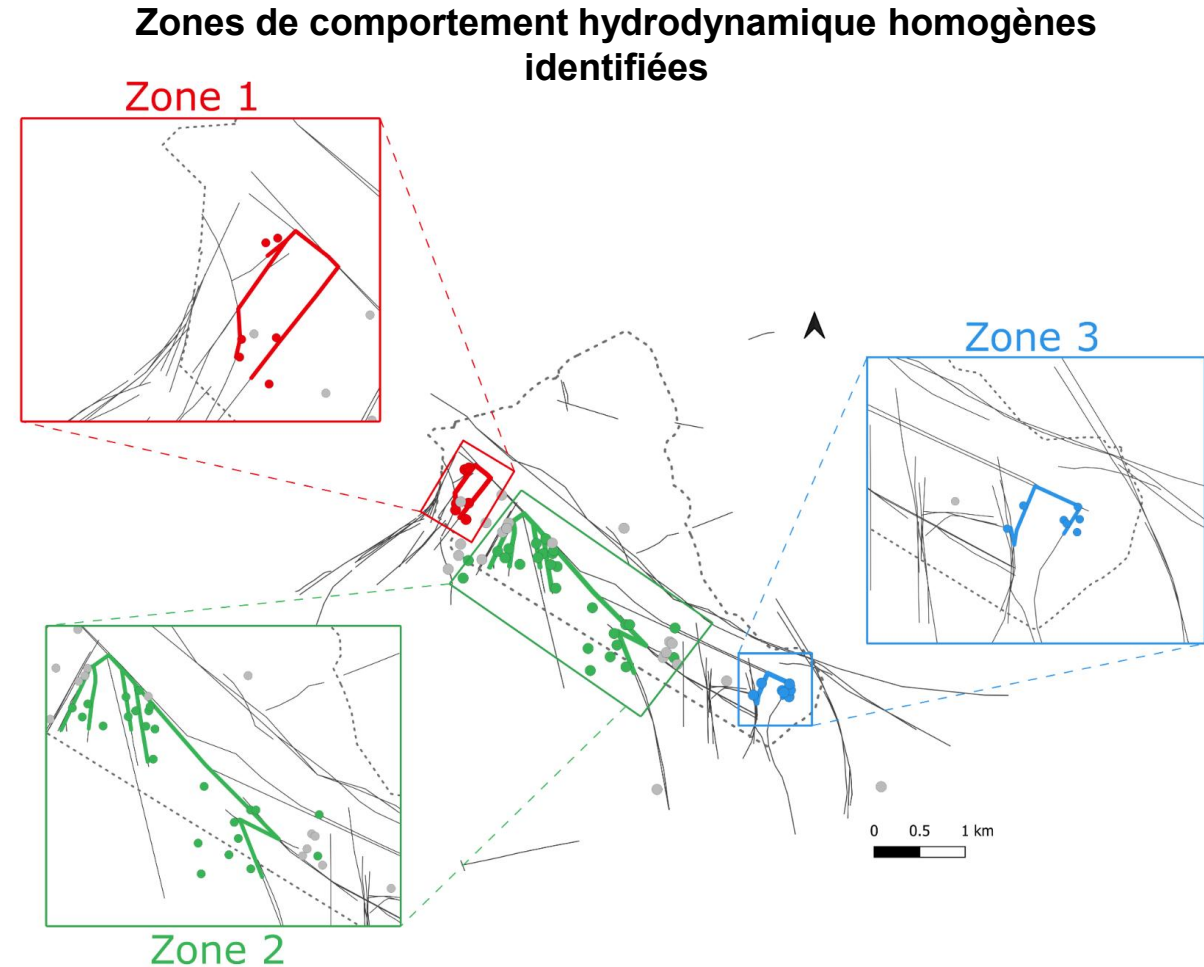
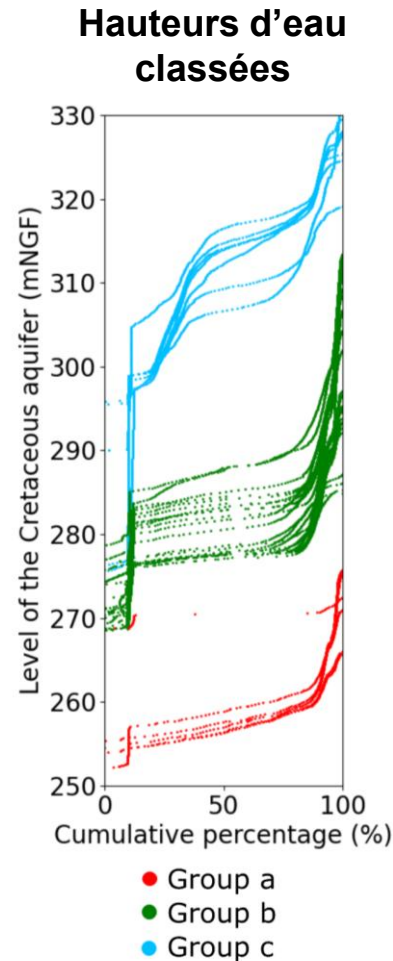


Fig.7 : Représentation spatiale du clustering réalisé sur l'évènement de crue karstique de novembre 2011 (Erguy et al., 2023)

3. Conception de modèle à réseaux de neurones

3. Modèle à réseaux de neurones

Principe des réseaux de neurones artificiels

Neurone formel

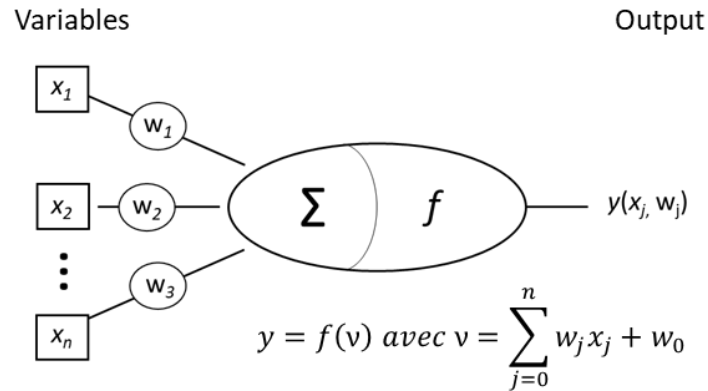


Fig. 8 : Représentation schématique d'un neurone formel avec $\mathbf{x}$: entrée, $\mathbf{w}$: paramètres, $f(v)$: fonction d'activation, y : sortie

Propriétés :

- Approximation universelle (Hornik, 1989)
- Parcimonie (Barron, 1993)

Perceptron multicouche

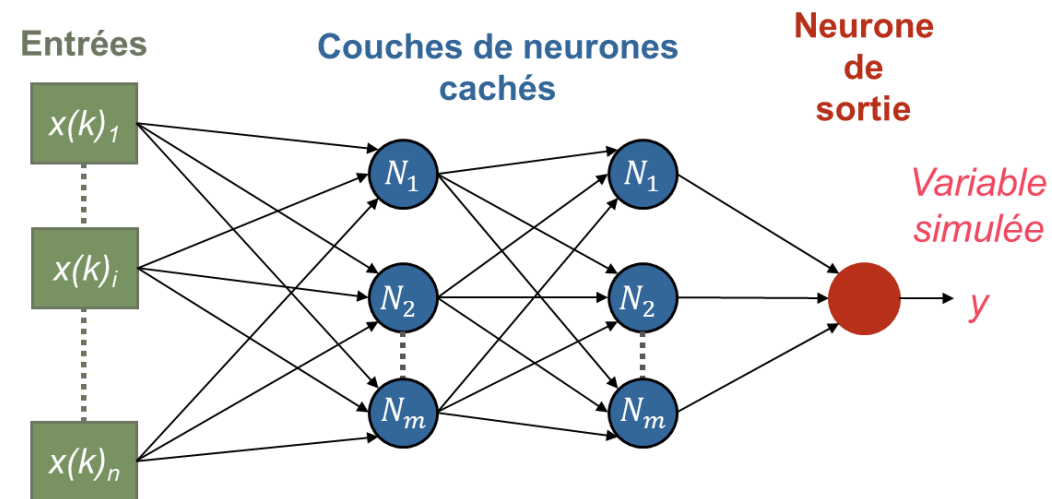


Fig.9 : Schématisation d'un perceptron multicouche

3. Modèle à réseaux de neurones

Principe des réseaux de neurones artificiels

Neurone formel

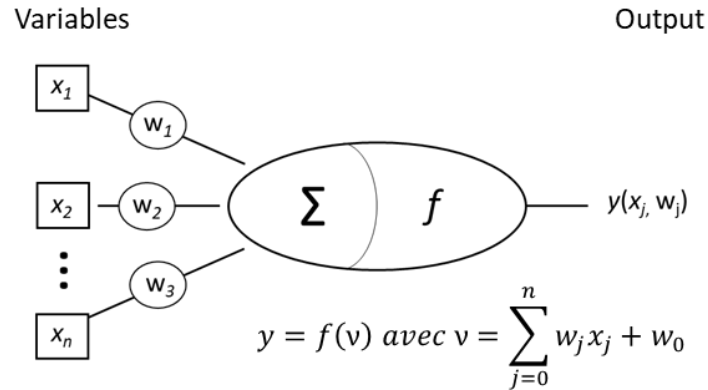


Fig. 8 : Représentation schématique d'un neurone formel avec $\mathbf{x}$: entrée, $\mathbf{w}$: paramètres, $f(v)$: fonction d'activation, y : sortie

Propriétés :

- Approximation universelle (Hornik, 1989)
- Parcimonie (Barron, 1993)

Apprentissage

Apprentissage : minimiser une fonction de coût en modifiant les paramètres $\mathbf{w}$

Perceptron multicouche

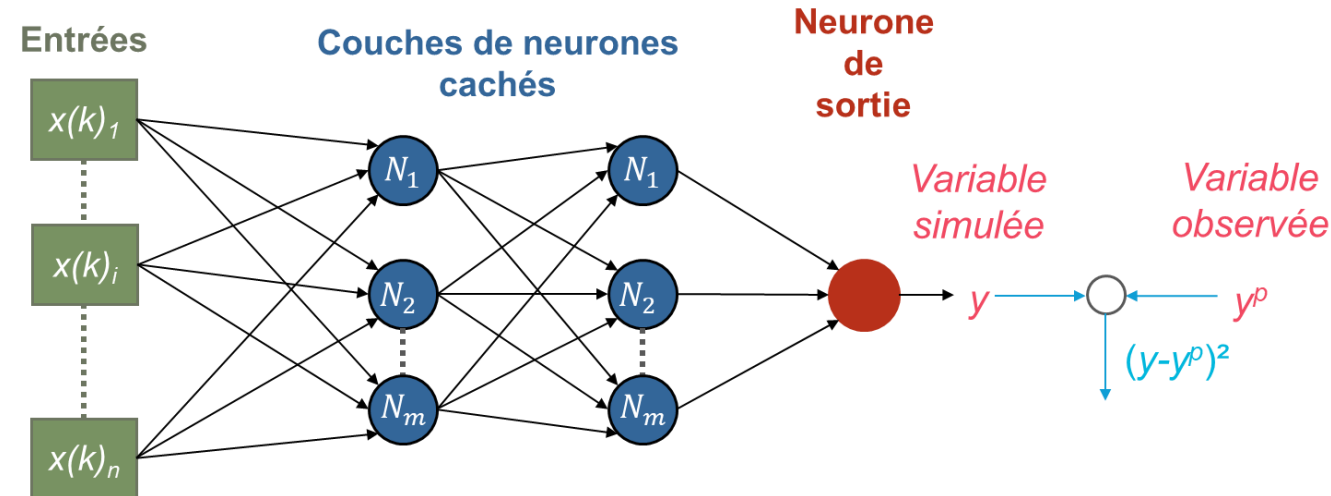


Fig.9 : Schématisation d'un perceptron multicouche et de l'apprentissage

3. Modèle à réseaux de neurones

Principe des réseaux de neurones artificiels

Neurone formel

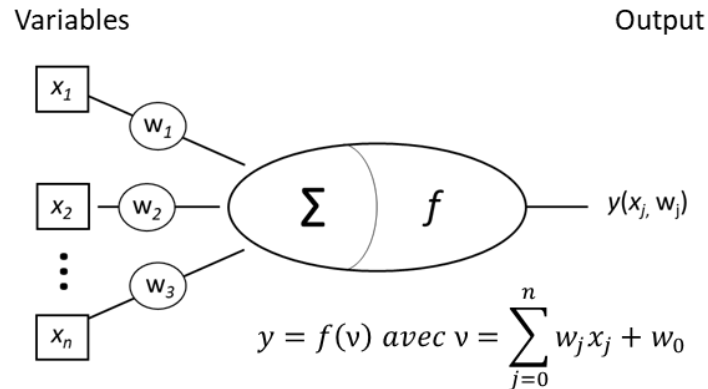


Fig. 8 : Représentation schématique d'un neurone formel avec $\mathbf{x}$: entrée, $\mathbf{w}$: paramètres, $f(v)$: fonction d'activation, y : sortie

Propriétés :

- Approximation universelle (Hornik, 1989)
- Parcimonie (Barron, 1993)

Apprentissage

Apprentissage : minimiser une fonction de coût en modifiant les paramètres $\mathbf{w}$

→ **Méthode itérative** (ici : algorithme de Levenberg-Marquardt)

Perceptron multicouche

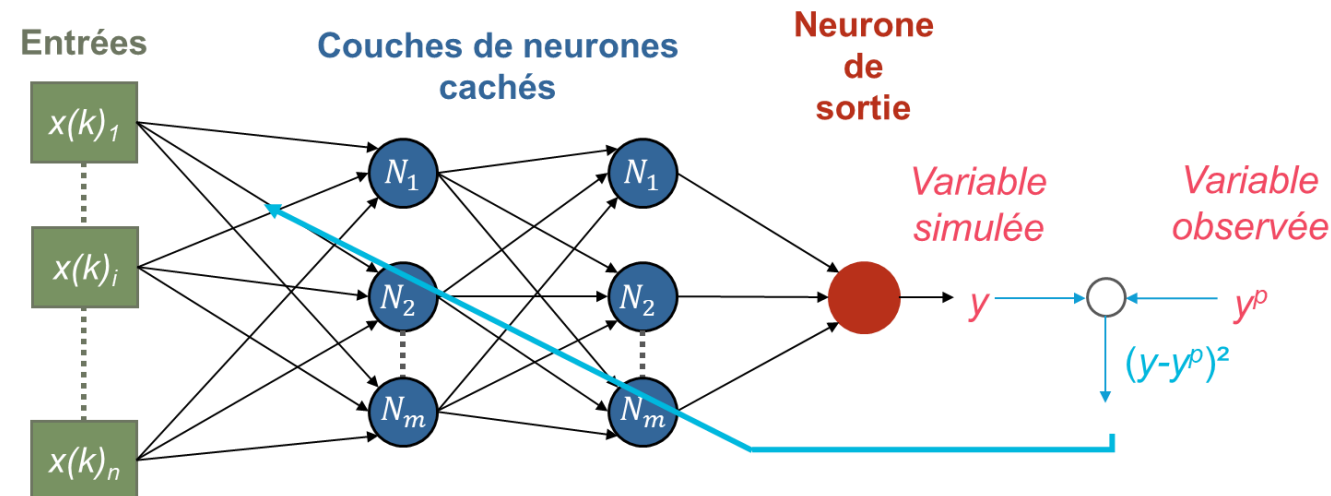


Fig.9 : Schématisation d'un perceptron multicouche et de l'apprentissage

Méthodes de régularisation – Dilemme biais/variance

Définir la complexité optimale pour éviter le surajustement

- Arrêt précoce (Sjöberg et al., 1995)
- Validation croisée (Stone, 1974)
- Modèles d'ensemble (Darras, 2018)

→ **Modèles adaptés pour la prévision des extrêmes**
(hors de la gamme de valeurs de l'ensemble d'apprentissage)

3. Modèle à réseaux de neurones

Conception d'un modèle neuronal

1. Division de la base de données en sous-ensembles (apprentissage, test, arrêt)

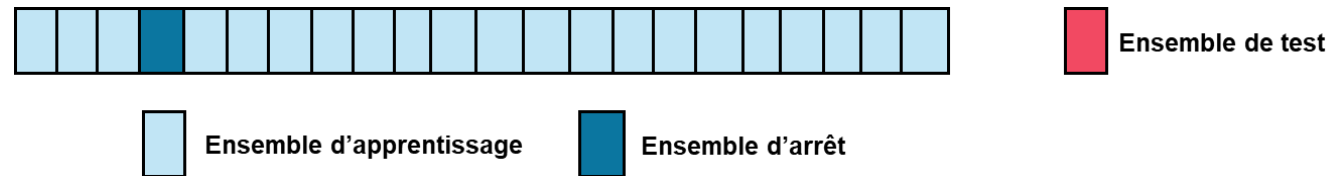


Fig.10 : Représentation schématique de la base de données

2. Choix d'une architecture postulée

3. Sélection des variables d'entrée et des hyperparamètres par la validation croisée

4. Apprentissage et évaluation des performances



4. Méthodes de régionalisation

4. Méthodes de régionalisation

Mise en application

Principe et base de données :

- ▶ **Régionalisation** : application du modèle conçu sur le piézomètre pilote sur des piézomètres « cibles »
- ▶ Base de données : **12 piézomètres** répartis dans les zones identifiées par le clustering : **6 du groupe sud**, **5 du groupe central** et **1 du groupe nord-ouest**

Cas étudiés :

- ▶ 2 cas de régionalisation (Erguy, 2024) :
 - Régionalisation « faible » : même zone que le piézomètre pilote
 - Régionalisation « forte » : autres zones que le piézomètre pilote

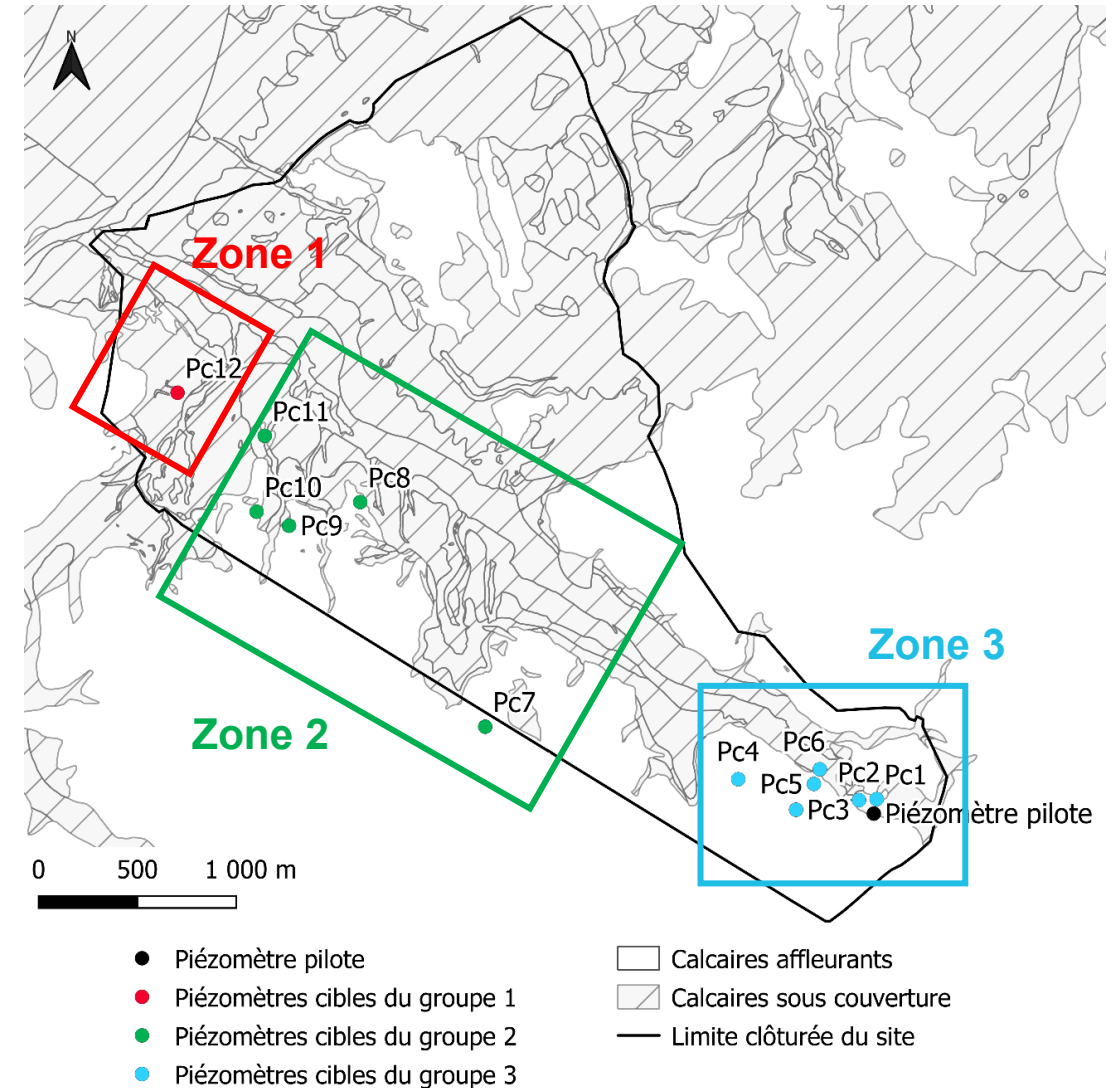


Fig.11 : Localisation des piézomètres cibles et du piézomètre pilote

4. Méthodes de régionalisation

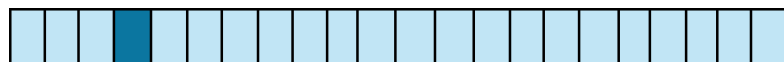
Principes des méthodes

Conception d'un modèle à réseau de neurones sur le **piézomètre pilote**

Méthode 1

Apprentissage sur le **piézomètre pilote**

Régionalisation : généralisation sur le **piézomètre cible**



Ensemble d'apprentissage

P_{pilote}



Ensemble d'arrêt

P_{pilote}



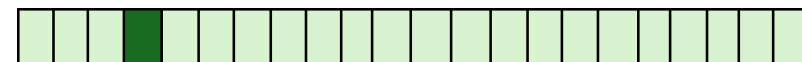
Ensemble de test

P_{cible}

Méthode 2

Apprentissage sur le **piézomètre cible**

Régionalisation : généralisation sur le **piézomètre cible**



Ensemble d'apprentissage

P_{cible}



Ensemble d'arrêt

P_{cible}



Ensemble de test

P_{cible}



5. Résultats de régionalisation

5. Résultats de régionalisation

Régionalisation du modèle à la zone sud - méthode 1

Méthode 1 - zone 3 :

(Généralisation au sein de la zone du piézomètre pilote)

- Prévisions satisfaisantes
- Différence d'amplitude de P_{c2} bien reproduites

Piézomètres	Critères de performance sur le test	
	NSE	C_p
Pilote	0,980	0,58
P_{c1}	0,976	0,50
P_{c2}	0,971	0,47
P_{c3}	0,967	0,57
P_{c4}	0,951	0,15
P_{c5}	0,924	-0,34
P_{c6}	0,94	-2,49
Médiane	0,97	0,47

Tableau 2 : Critères de performance sur le test de 2019

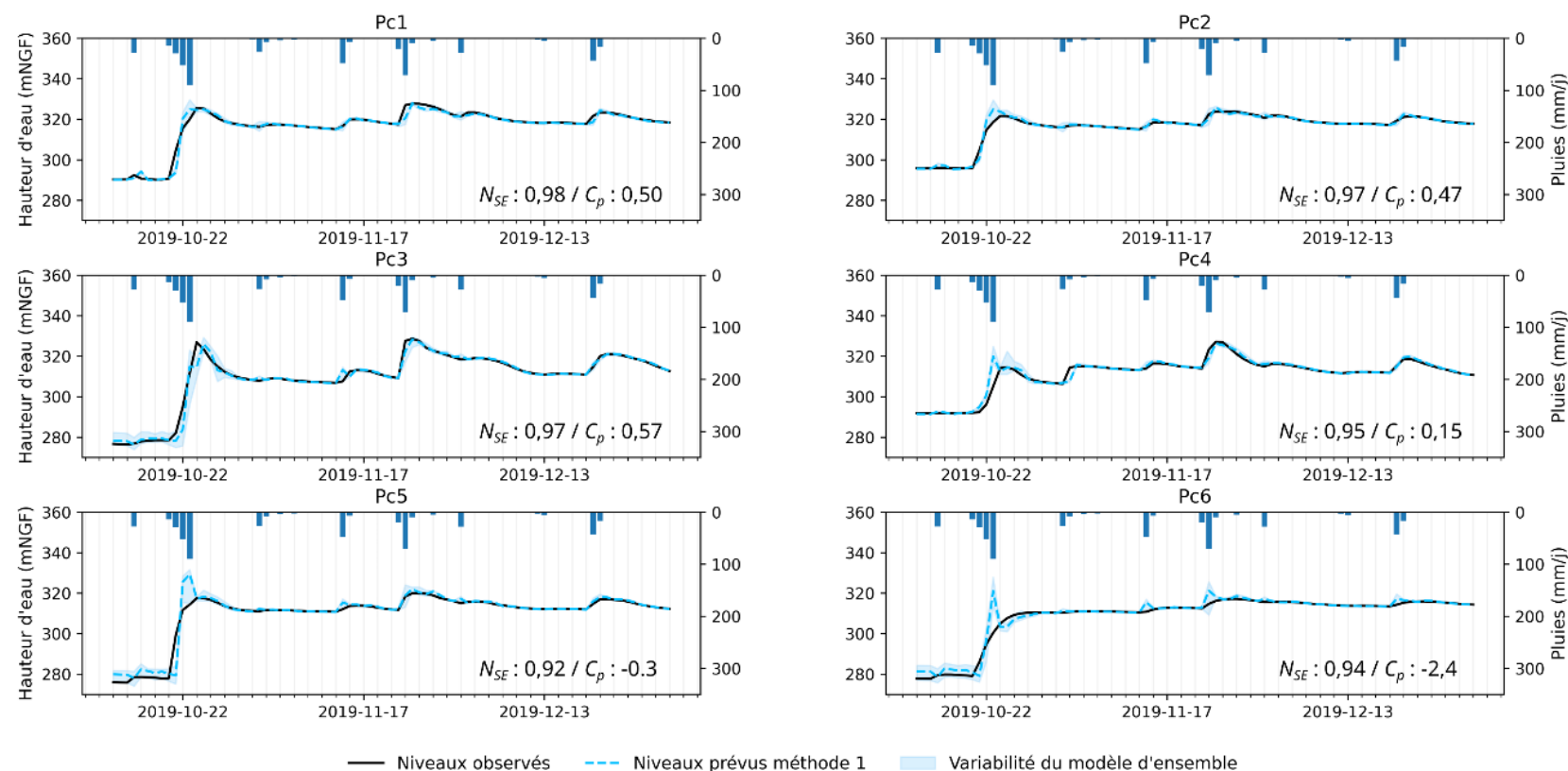
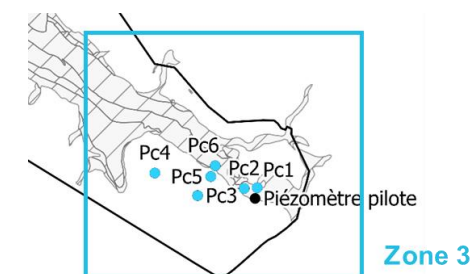
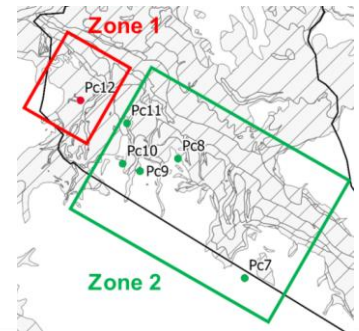


Fig.12 : Niveaux prévus à 24h pour les piézomètres cible de la zone sud (groupe 3) en appliquant la méthode 1

5. Résultats de régionalisation

Régionalisation du modèle aux autres zones - méthode 1 et 2



► Méthode 1 - sans apprentissage sur le forage cible :

- Performances faibles (NSE entre -4 et 0,7 et C_p négatifs sur les forages cibles)
- Surestimation des pics de crue (P_{c9} et P_{c10})

► Méthode 2 - avec apprentissage sur le forage cible :

- Amélioration des performances (NSE entre 0,86 et 0,97 et C_p entre 0,37 et 0,68 sur les forages cibles)
- Pics de crue sont bien restitués (sauf P_{c12} en retard)

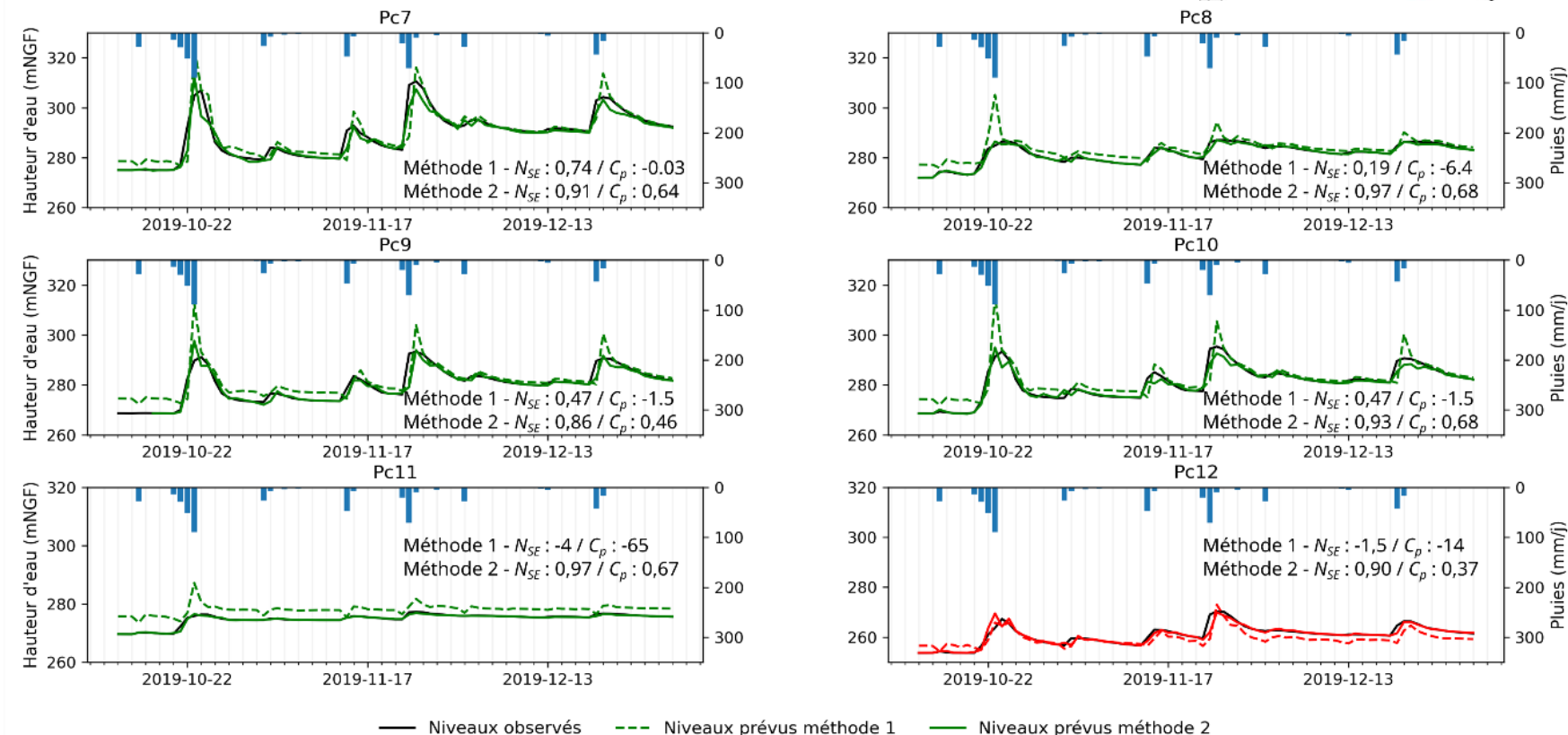


Fig.13 : Niveaux prévus à 24h pour les piézomètres cible des zones nord (groupe 1) et centrale (groupe 2) en appliquant les méthodes 1 et 2



Conclusion

6. Conclusions

Contexte et enjeux :

- ▶ Les aquifères karstiques présentent des comportements complexes pouvant générer des crues souterraines, dont la prévision constitue un enjeu de sûreté majeur sur le site CEA de Cadarache ainsi que pour d'autres installations enterrées (parking, tunnels...)
- ▶ Objectif : régionaliser à l'échelle du site un modèle neuronal pluie-niveau à 24h développé sur un piézomètre pilote afin d'éviter la multiplication de modèles locaux

Résultats clés :

- ▶ Deux méthodes de régionalisation :
 - Généralisation directe efficace pour des réponses hydrodynamiques similaires
 - Réapprentissage ciblé performant pour des comportements différents
- ▶ Régionalisation d'un modèle de prévision des niveaux piézométriques → information locale
- ▶ Ces résultats montrent qu'un modèle neuronal conçu de manière rigoureuse est transférable à d'autres points de mesure

Perspectives :

- ▶ Vers une prévision opérationnelle des niveaux piézométriques des hydrosystèmes karstiques et complexes : robuste, régionalisable, à moindres coûts et ressources de calcul

Références

- Barron, A.R., 1993. Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information Theory* 39, 930–945. <https://doi.org/10.1109/18.256500>
- Darras, T., Johannet, A., Vayssade, B., Kong-A-Siou, L., Pistre, S., 2018. Ensemble model to enhance robustness of flash flood forecasting using an Artificial Neural Network: case-study on the Gardon Basin (south-eastern France). *Boletín geológico y minero* 129, 565–578.
- Erguy, M., 2024. Apport des réseaux de neurones artificiels dans la prévision du risque de crue karstique sur le site du Commissariat à l’Energie Atomique de Cadarache (thesis). [theses.fr. IMT Mines Alès](https://theses.fr/IMT_Mines_Alès).
- Erguy, M., Morilhat, S., Artigue, G., Trincal, J., Johannet, A., Pistre, S., 2023. A new simple statistical method for the unsupervised clustering of the hydrodynamic behavior at different boreholes: analysis of the obtained clusters in relation to geological knowledge. *Environ Earth Sci* 82, 451. <https://doi.org/10.1007/s12665-023-11066-z>
- Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. *Neural Networks* 2, 359–366. [https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8)
- Sjöberg, J., Zhang, Q., Ljung, L., Benveniste, A., Delyon, B., Glorennec, P.-Y., Hjalmarsson, H., Juditsky, A., 1995. Nonlinear black-box modeling in system identification: a unified overview. *Automatica, Trends in System Identification* 31, 1691–1724. [https://doi.org/10.1016/0005-1098\(95\)00120-8](https://doi.org/10.1016/0005-1098(95)00120-8)
- Stone, M., 1974. Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society: Series B (Methodological)* 36, 111–133. <https://doi.org/10.1111/j.2517-6161.1974.tb00994.x>



MERCI

Manon Erguy - manon.erguy@mines-ales.fr

Guillaume Artigue – guillaume.artigue@mgva-ingenierie.fr

Sébastien Morilhat – sebastien.morilhat@cea.fr

Julien Trincal – julien.trincal@cea.fr

Anne Johannet – anne.johannet@mines-ales.fr

Séverin Pistre – severin.pistre@umontpellier.fr





Annexes

Annexes

Annexe 1 : Scores de validation croisée et critères de performance sur le test

Méthode 1 :

Piézomètres	Zone	Distance au forage pilote (m)	Critères de performance sur le test (40 initialisations)			
			<i>NSE</i>	<i>NSE</i> mises en charge	C_p	C_p mises en charge
Pilote	3	0	0,980	0,89	0,58	0,52
P _{c1}	3	75	0,976	0,89	0,50	0,51
P _{c2}	3	102	0,971	0,91	0,47	0,52
P _{c3}	3	394	0,967	0,91	0,57	0,61
P _{c4}	3	708	0,951	0,88	0,15	0,10
P _{c5}	3	338	0,924	0,70	-0,34	-0,23
P _{c6}	3	353	0,94	0,77	-2,49	-2,42
P _{c7}	2	2016	0,74	0,45	-0,03	0,01
P _{c8}	2	3040	0,19	-0,56	-6,43	-3,26
P _{c9}	2	3298	0,47	0,10	-1,47	-0,93
P _{c10}	2	3477	0,47	0,09	-1,47	-0,89
P _{c11}	2	3627	-4,74	-2,77	-65,82	-13,49
P _{c12}	1	4117	-1,51	-0,55	-14,75	-3,24

Fig. X : Localisation des piézomètres cibles et du piézomètre pilote

Annexes

Annexe 1 : Scores de validation croisée et critères de performance sur le test

Méthode 2 :

Scores de validation croisée (10 suites d'initialisation)				
Piézomètres	NSE	NSE mises en charge	C _p	C _p mises en charge
Pc5	0,97	0,90	0,63	0,70
Pc7	0,81	0,72	0,58	0,64
Pc8	0,97	0,93	0,75	0,82
Pc9	0,78	0,67	0,33	0,47
Pc10	0,80	0,69	0,50	0,55
Pc11	0,92	0,85	0,62	0,68
Pc12	0,94	0,84	0,54	0,59
Critères de performance sur le test oct-nov 2019 (20 initialisations)				
Piézomètres	NSE	NSE mises en charge	C _p	C _p mises en charge
Pc5	0,97	0,88	0,50	0,51
Pc7	0,91	0,81	0,64	0,66
Pc8	0,97	0,91	0,68	0,75
Pc9	0,86	0,75	0,46	0,47
Pc10	0,93	0,86	0,68	0,71
Pc11	0,97	0,92	0,67	0,70
Pc12	0,90	0,75	0,37	0,31

Fig. X : Localisation des piézomètres cibles et du piézomètre pilote

Annexes

Annexe 2 : Validation croisée

- **Objectif** : estimer l'**erreur de généralisation** sur l'ensemble de la base de données (Stone, 1974)
- **Principe** : division de l'ensemble d'apprentissage en D **sous-ensembles** e , l'apprentissage se fait donc sur $D-1$ sous-ensembles et chacun **passé en validation tour à tour** → calcul du critère de validation C_e

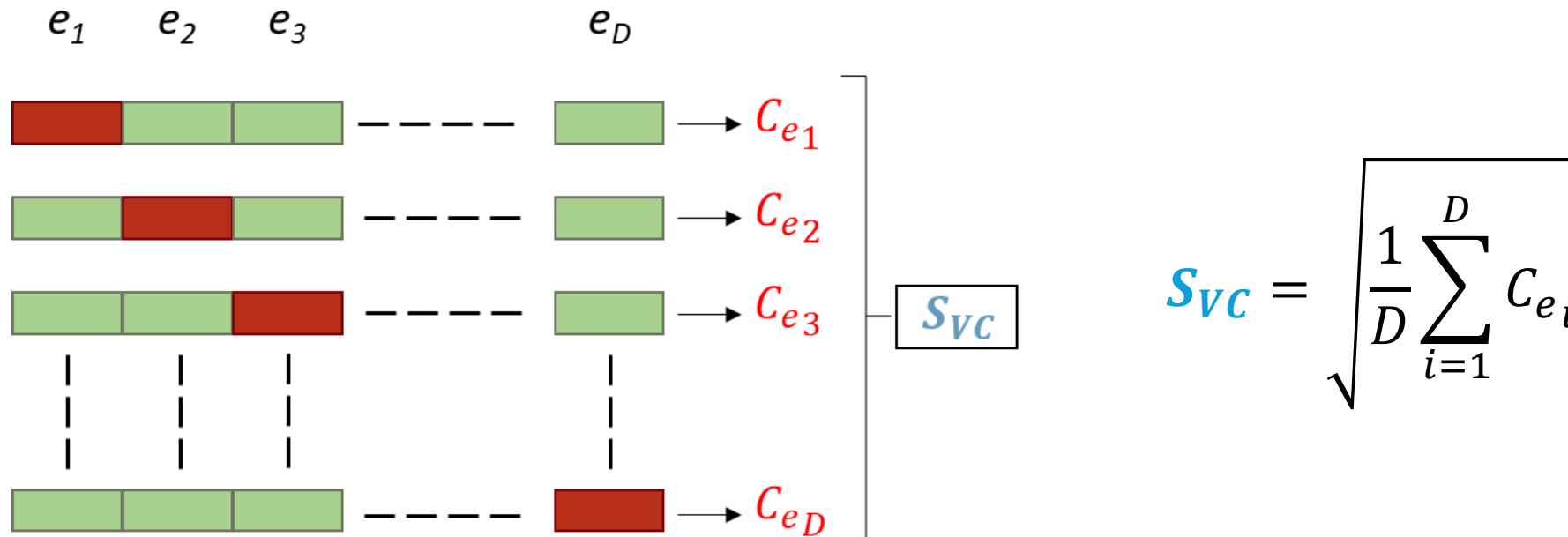


Schéma de la méthode de validation croisée

Annexes

Annexe 3 : Arrêt précoce

Méthode : consiste à arrêter l'apprentissage avant la convergence de l'algorithme pour éviter le **surapprentissage** (Sjöberg *et al.* 1995)

→ Choix important de l'ensemble d'arrêt

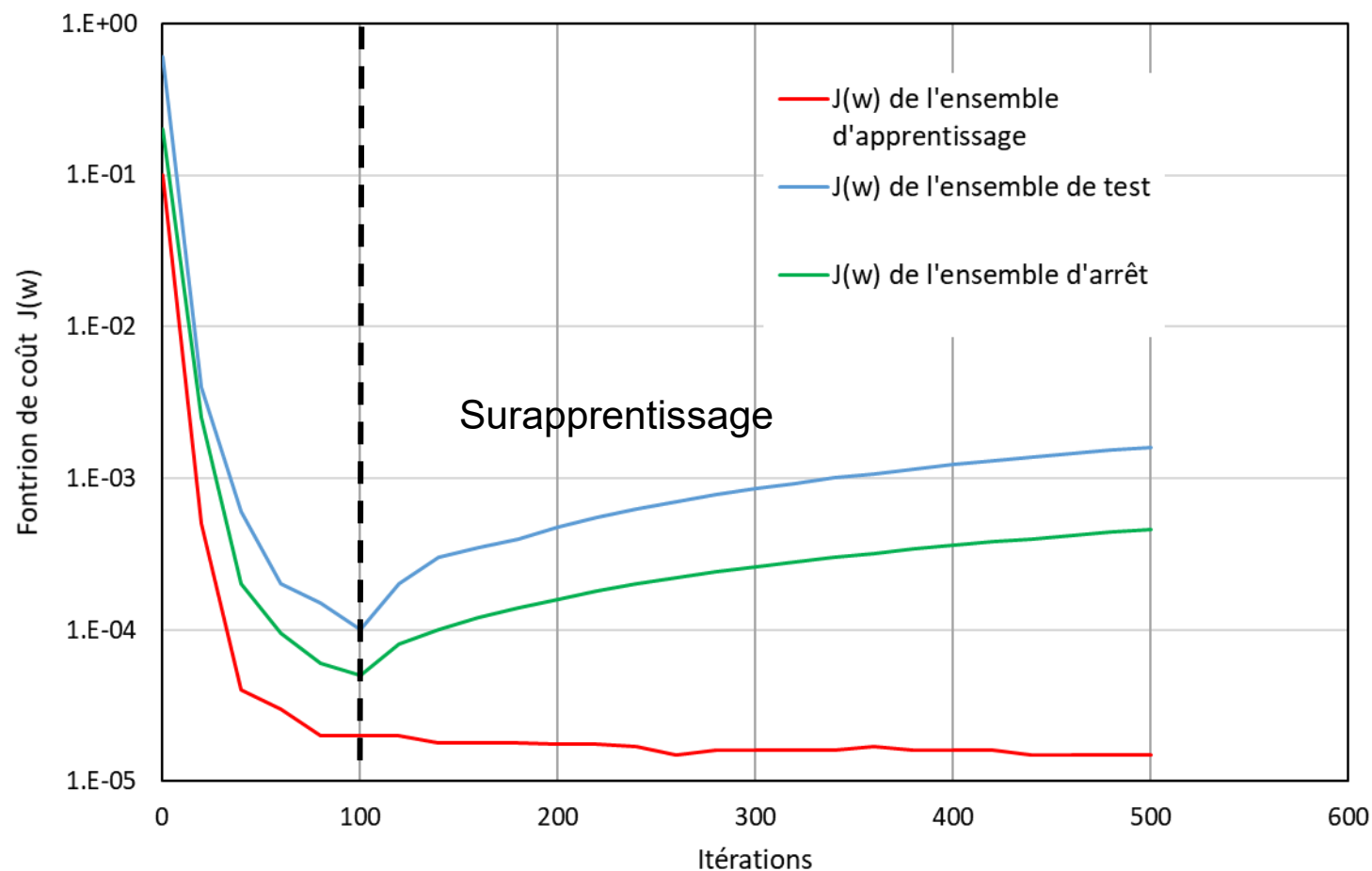


Fig. 27 : Illustration de la méthode d'arrêt précoce

Annexe 4 : Modèles d'ensemble

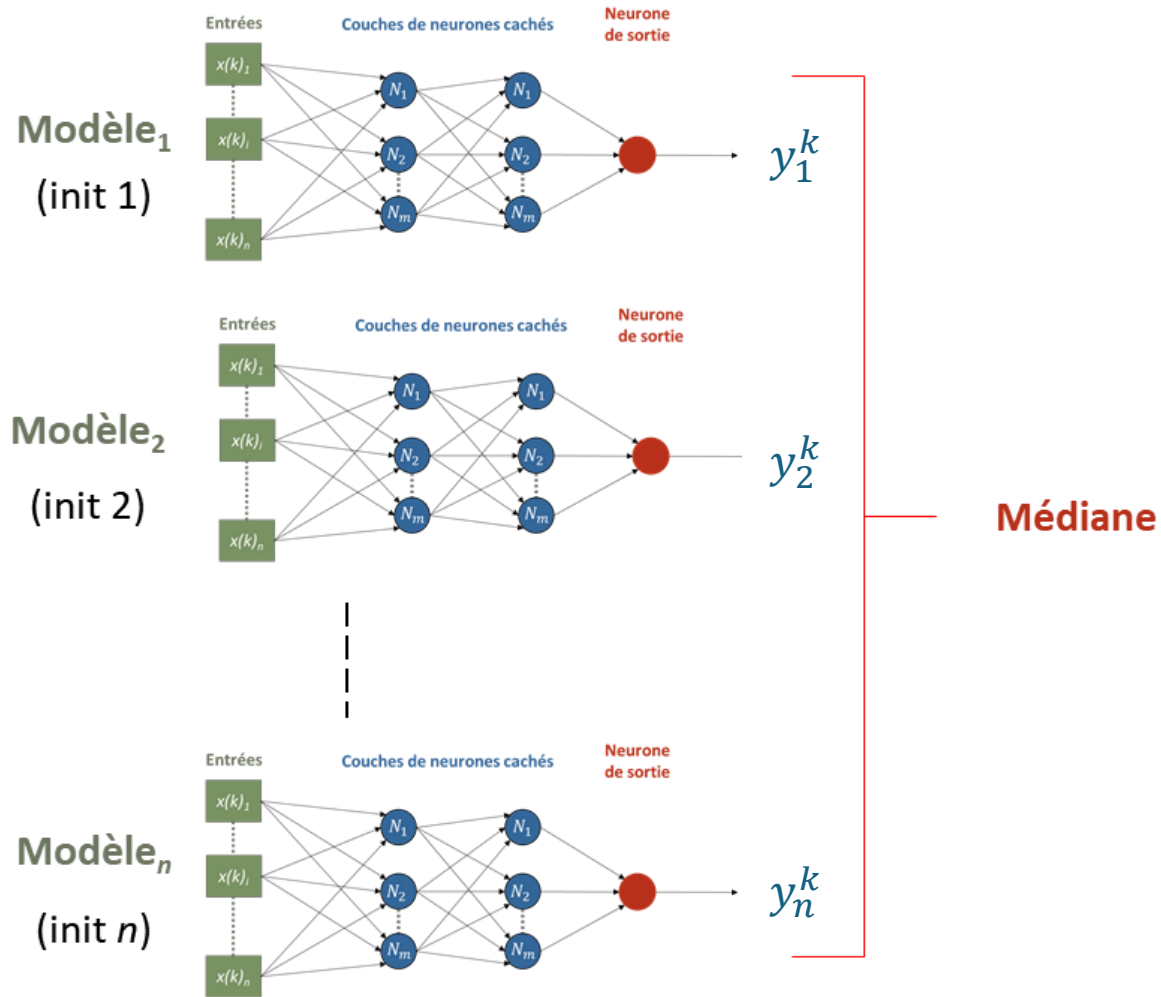


Fig. 28 : Illustration de la méthode des modèles d'ensemble

Méthode :

- Minimiser l'impact de l'initialisation du modèle (Darras *et al.*, 2018)
- Consiste à calculer à chaque pas de temps la médiane des sorties des n modèles (avec n initialisations)

$$\hat{y}_n^k = \text{Médiane}_n(y_i^k)$$

- Diminue la variabilité des sorties