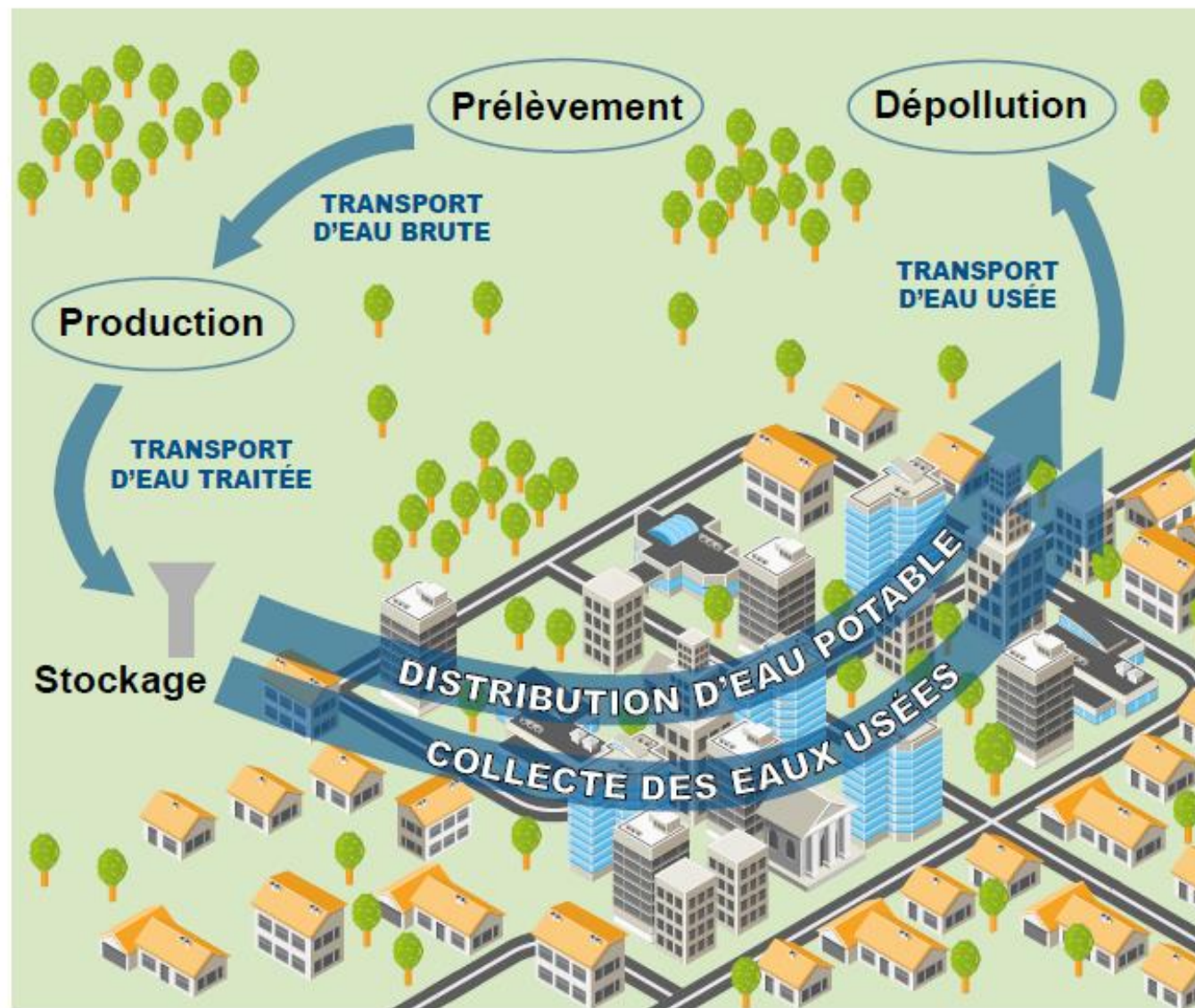




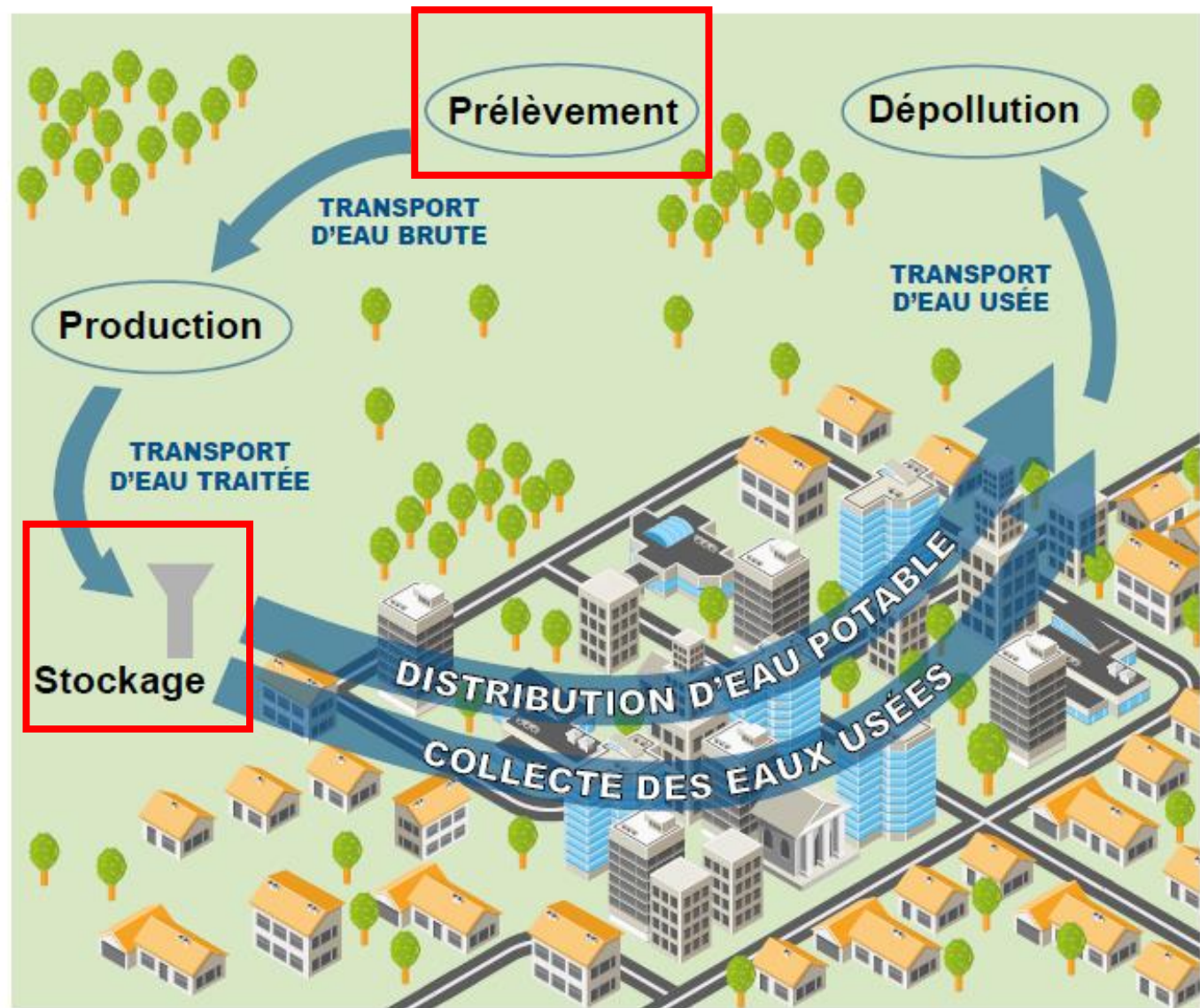
# Modèles mathématiques pour la prévision des demandes en eau des usagers et du niveau des nappes phréatiques

- Université : M. Chabaud, H. Munive, **R. Catellier**, S. Rubenthaler, T. Laloë
- Eau d'Azur :, A. Maupetit

# Niveau des nappes



# Niveau des nappes



# Niveau des nappes : contexte

Eau d'Azur : *distribution continue* d'eau en qualité et quantité  
Nécessite *gestion de la ressource* (prélèvement → distribution)

Changements climatiques :

- Problématiques de tension sur la ressource
  - Actions de sécurisation des prélèvements souterrains
  - Mise en place de système d'alimentation de secours

## → Anticiper le niveau de nappe phréatique

- à 10 jours
- À 3 mois

Contexte géographique : basse vallée du Var



Rivière de récolte : La Vésubie



Vésubie et station électrique EDF



Station de pompage



St-Jeanet

AMI  
VENCE  
CT 43.6

LA GAUDE

AMI  
St-Paul  
CT 43.7

LA COLLE-  
SUR-LOUP

ST-LAURENT-  
DU-Var

AMI  
NICE  
IPE 342.7

AMI  
VILLEFRANCHE-  
SUR-MER

AMI  
Beaulieu-sur-Mer  
CT 43.5

AMI  
St-Jean-Cap-Ferrat  
CT 43.9

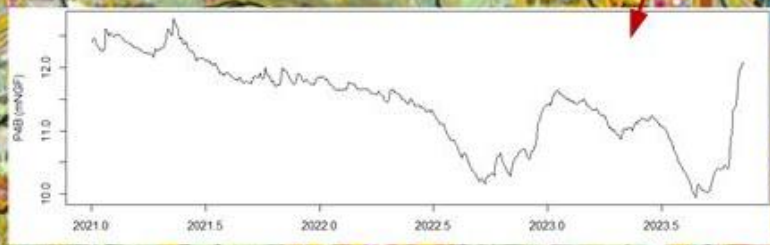
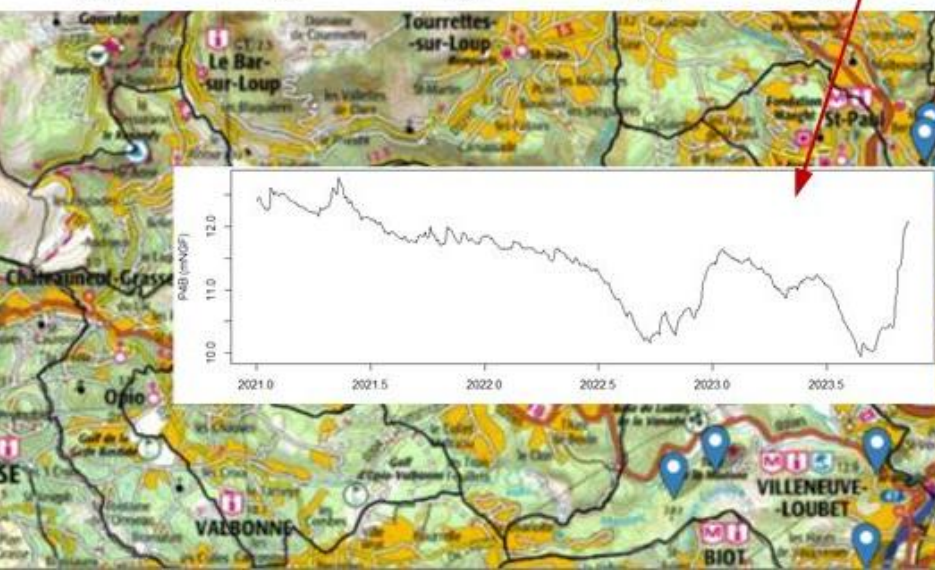
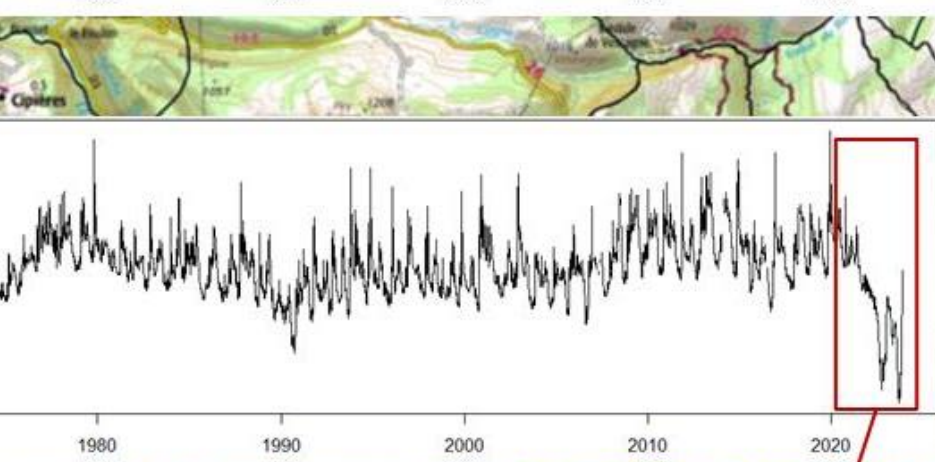
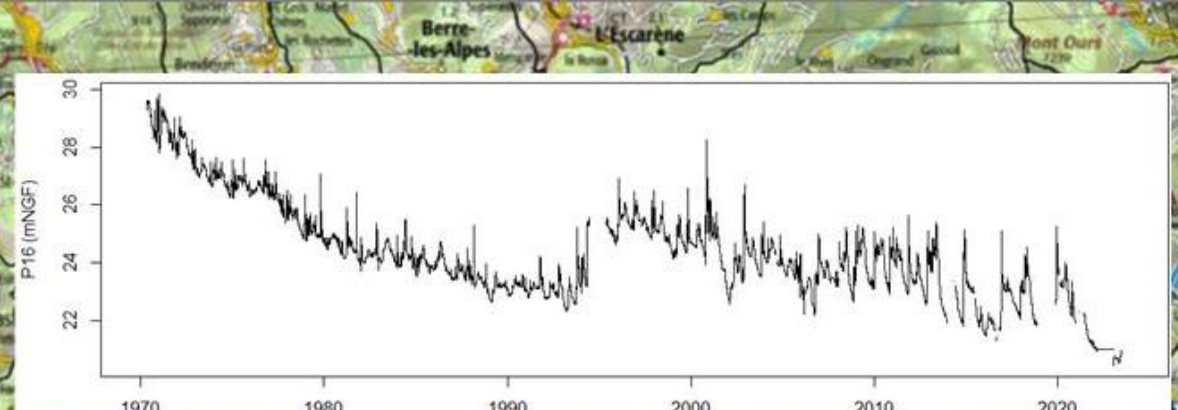
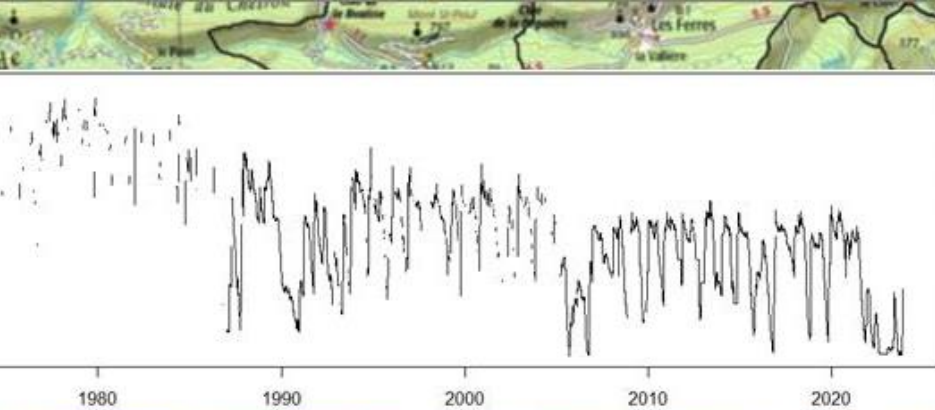
AMI  
VILLENEUVE-  
CT 43.8

AMI  
CAGNES SUR-MER  
CT 43.9

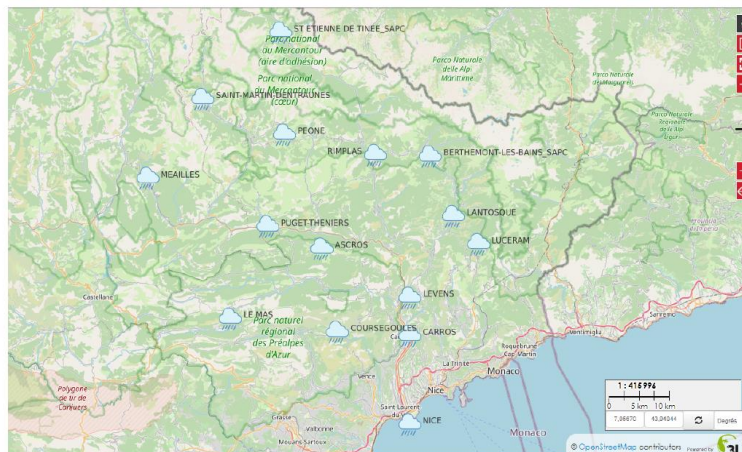
AEROPORT  
DE NICE-CÔTE D'AZUR

ANGES

CAP FERRAT



# Données : Description des valeurs manquantes



| Variable              | Première date | NA restantes    |
|-----------------------|---------------|-----------------|
| X09724X0023.P2        | 2007-01-01    | 248 (3.923 %)   |
| X09728X0032.P15       | 2008-06-25    | 1064 (18.408 %) |
| X09994X0399.P.35      | 2008-06-25    | 1036 (17.924 %) |
| X09728X0099.P38       | 2008-06-26    | 1258 (21.768 %) |
| X09728X0168.PZS10A    | 2009-03-10    | 605 (10.956 %)  |
| X09994X0521.P4B       | 2007-01-16    | 488 (7.739 %)   |
| X09735X0161.P33BIS    | 2008-06-25    | 2671 (46.211 %) |
| X09735X0149.P34BIS    | 2008-06-25    | 3924 (67.889 %) |
| X09735X0214.PZS9AM    | 2021-09-02    | 559 (58.048 %)  |
| PZ_Roguez             | 2021-04-14    | 500 (45.290 %)  |
| PZ_Drap               | 2014-07-11    | 515 (14.414 %)  |
| PZ_Roguez_REA         | 2022-10-05    | 5 (0.885 %)     |
| PZ_P6_Les_Plans_REA   | 2023-02-01    | 5 (1.121 %)     |
| PZ_Pugets_REA         | 2022-10-05    | 5 (0.885 %)     |
| PZ_Plan_du_Var_REA    | 2023-02-01    | 5 (1.121 %)     |
| PZ2_Roquebilliere_REA | 2022-10-05    | 5 (0.885 %)     |
| PZ_Bastion_REA        | 2023-02-01    | 19 (4.260 %)    |
| PZ_Gattieres_REA      | 2023-11-28    | 37 (25.342 %)   |
| PZ_Bec_Esteron_REA    | 2023-04-21    | 303 (82.561 %)  |

| Variable                         | Première date | NA restantes    |
|----------------------------------|---------------|-----------------|
| cumul_ASCROS                     | 1987-08-14    | 6 (0.046 %)     |
| cumul_CARROS_SAPC                | 1997-03-10    | 690 (7.009 %)   |
| cumul_COURSEGOULES_SAPC          | 1997-03-10    | 170 (1.727 %)   |
| cumul_LANTOSQUE_SAPC             | 2006-12-27    | 6 (0.095 %)     |
| cumul_LEVENS                     | 2008-05-16    | 8 (0.137 %)     |
| cumul_PEIRA.CAVA                 | 1991-02-08    | 747 (6.220 %)   |
| cumul_LE.MAS                     | 2006-04-20    | 20 (0.304 %)    |
| cumul_NICE                       | 1987-08-14    | 6 (0.046 %)     |
| cumul_PEONE                      | 2002-11-20    | 23 (0.295 %)    |
| cumul_PUGET.THENIERS_SAPC        | 1997-03-09    | 2590 (26.308 %) |
| cumul_RIMPLAS_SAPC               | 2007-02-14    | 7 (0.112 %)     |
| cumul_BERTHEMONT.LES.BAINS_SAPC  | 2023-07-04    | 4 (1.365 %)     |
| cumul_ST.ETIENNE.DE.TINEE_SAPC   | 2007-03-29    | 17 (0.273 %)    |
| cumul_ST.MARTIN.D.ENTRAUNES_SAPC | 2007-05-07    | 65 (1.049 %)    |

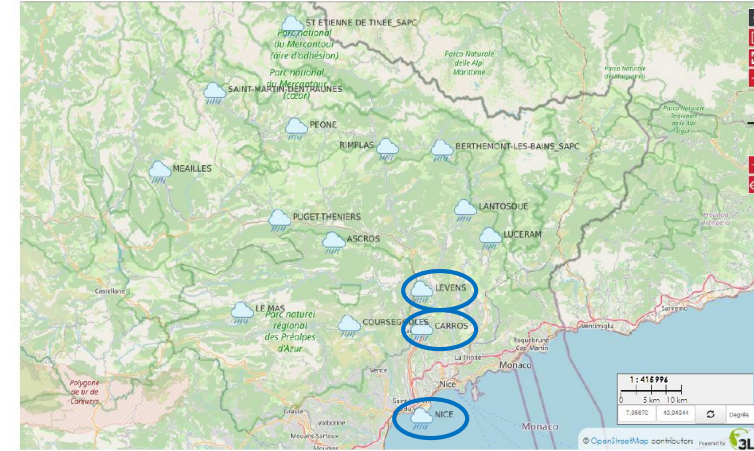
# Données :



## Focus P4B :

- Historique long (2007)
- 8% de valeurs manquantes (488 sur 6306)
- Représentatif du comportement global de la nappe (corrélations)

| Variable              | Première date | NA restantes    |
|-----------------------|---------------|-----------------|
| X09724X0023.P2        | 2007-01-01    | 248 (3.923 %)   |
| X09728X0032.P15       | 2008-06-25    | 1064 (18.408 %) |
| X09994X0399.P.35      | 2008-06-25    | 1036 (17.924 %) |
| X09728X0099.P38       | 2008-06-26    | 1258 (21.768 %) |
| X09728X0168.PZS10A    | 2009-03-10    | 605 (10.956 %)  |
| X09994X0521.P4B       | 2007-01-16    | 488 (7.739 %)   |
| X09735X0161.P33BIS    | 2008-06-25    | 2671 (46.211 %) |
| X09735X0149.P34BIS    | 2008-06-25    | 3924 (67.889 %) |
| X09735X0214.PZS9AM    | 2021-09-02    | 559 (58.048 %)  |
| PZ_Roguez             | 2021-04-14    | 500 (45.290 %)  |
| PZ_Drap               | 2014-07-11    | 515 (14.414 %)  |
| PZ_Roguez_REA         | 2022-10-05    | 5 (0.885 %)     |
| PZ_P6_Les_Plans_REA   | 2023-02-01    | 5 (1.121 %)     |
| PZ_Pugets_REA         | 2022-10-05    | 5 (0.885 %)     |
| PZ_Plan_du_Var_REA    | 2023-02-01    | 5 (1.121 %)     |
| PZ2_Roquebilliere_REA | 2022-10-05    | 5 (0.885 %)     |
| PZ_Bastion_REA        | 2023-02-01    | 19 (4.260 %)    |
| PZ_Gattieres_REA      | 2023-11-28    | 37 (25.342 %)   |
| PZ_Bec_Esteron_REA    | 2023-04-21    | 303 (82.561 %)  |

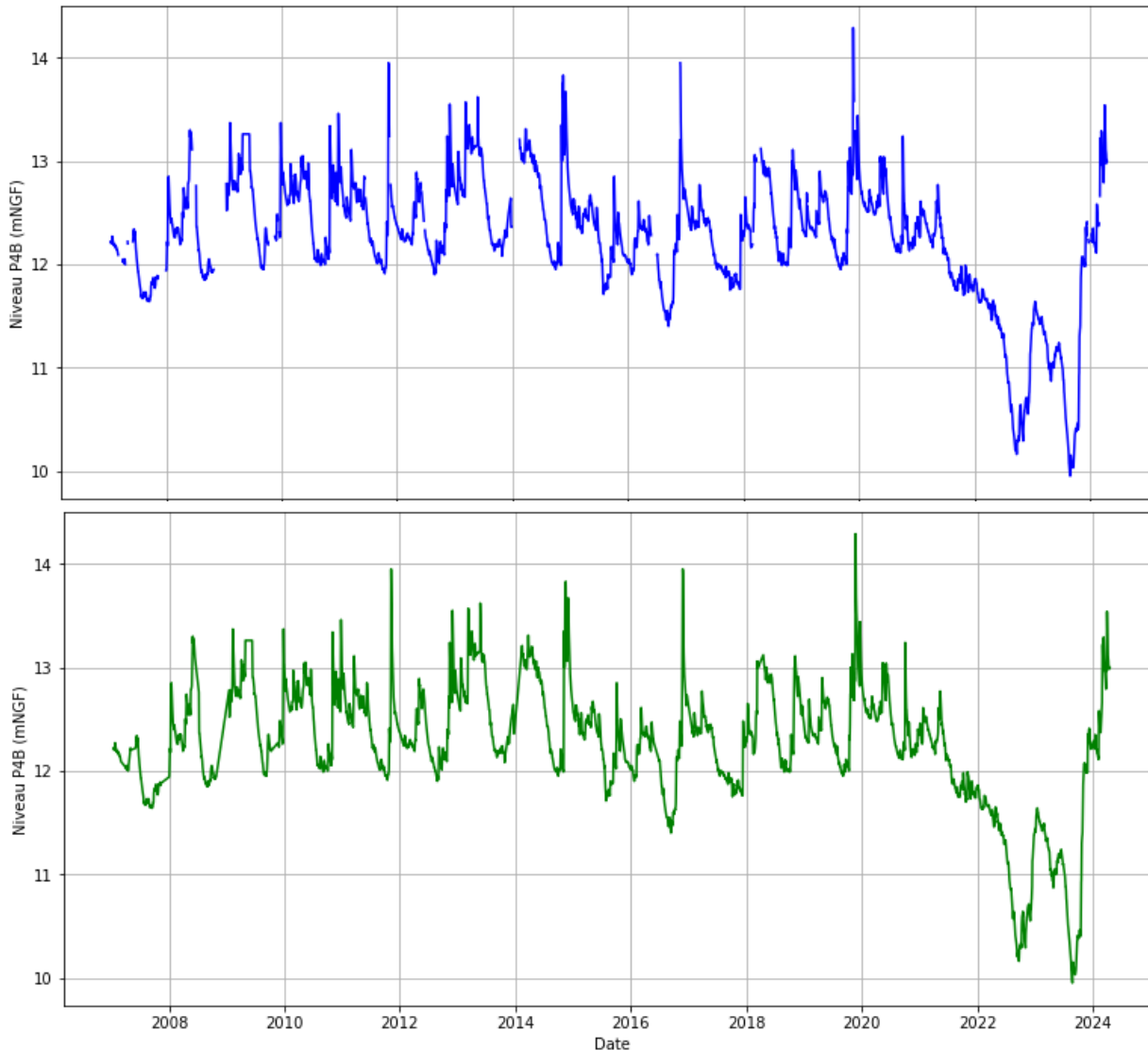


Focus Nice, Carros et Levens :  
→ Stations géographiquement les plus proches

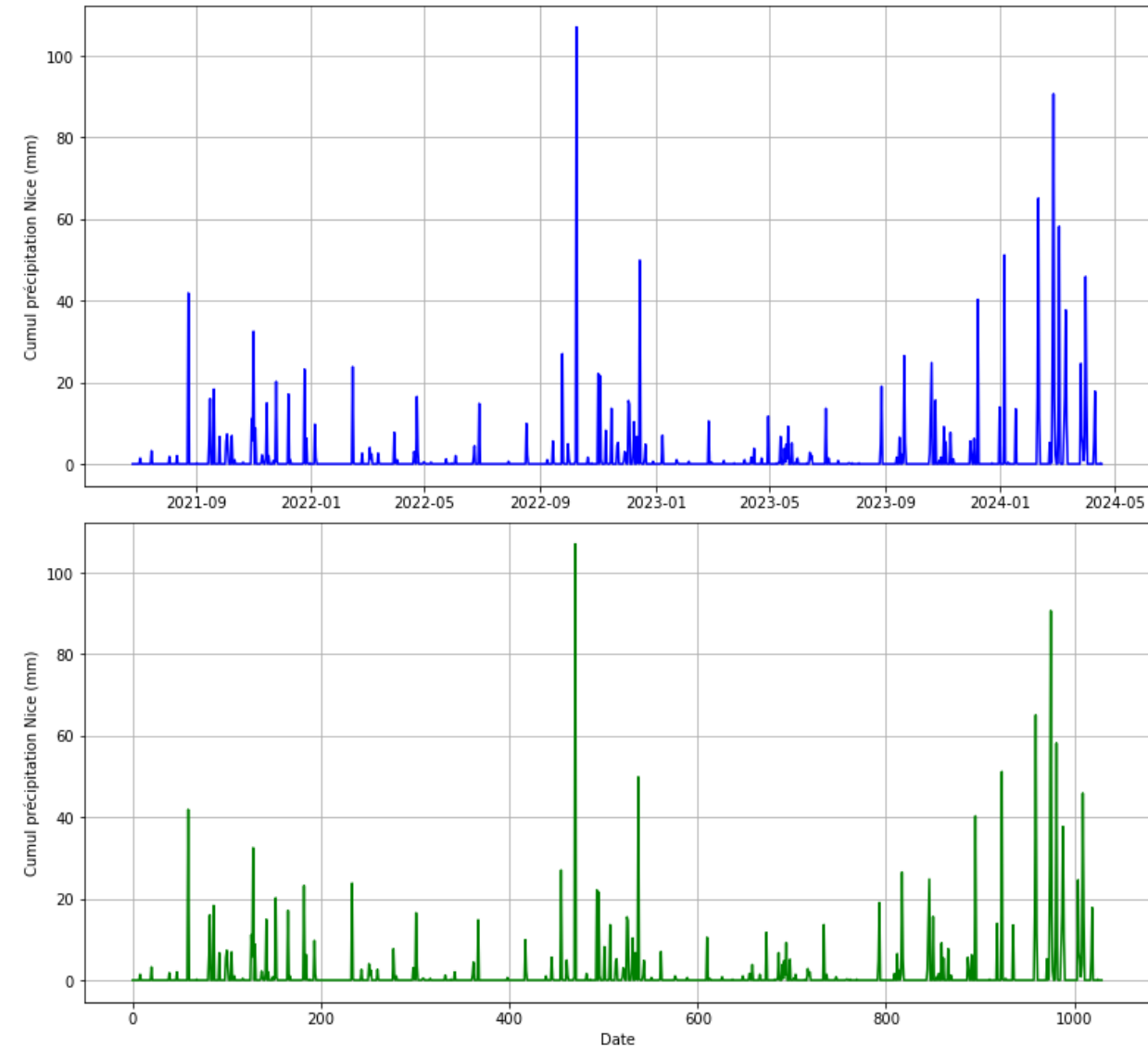
| Variable                         | Première date | NA restantes    |
|----------------------------------|---------------|-----------------|
| cumul_ASCROS                     | 1987-08-14    | 6 (0.046 %)     |
| cumul_CARROS_SAPC                | 1997-03-10    | 690 (7.009 %)   |
| cumul_COURSEGOULES_SAPC          | 1997-03-10    | 170 (1.727 %)   |
| cumul_LANTOSQUE_SAPC             | 2006-12-27    | 6 (0.095 %)     |
| cumul_LEVENS                     | 2008-05-16    | 8 (0.137 %)     |
| cumul_PEIRA.CAVA                 | 1991-02-08    | 747 (6.220 %)   |
| cumul_LE.MAS                     | 2006-04-20    | 20 (0.304 %)    |
| cumul_NICE                       | 1987-08-14    | 6 (0.046 %)     |
| cumul_PEONE                      | 2002-11-20    | 23 (0.295 %)    |
| cumul_PUGET.THENIERS_SAPC        | 1997-03-09    | 2590 (26.308 %) |
| cumul_RIMPLAS_SAPC               | 2007-02-14    | 7 (0.112 %)     |
| cumul_BERTHEMONT.LES.BAINS_SAPC  | 2023-07-04    | 4 (1.365 %)     |
| cumul_ST.ETIENNE.DE.TINEE_SAPC   | 2007-03-29    | 17 (0.273 %)    |
| cumul_ST.MARTIN.D.ENTRAUNES_SAPC | 2007-05-07    | 65 (1.049 %)    |

# Données : Interpolation des valeurs manquantes

Piezo P4B : Interpolation linéaire



Cumuls : Interpolation par la médiane de la station

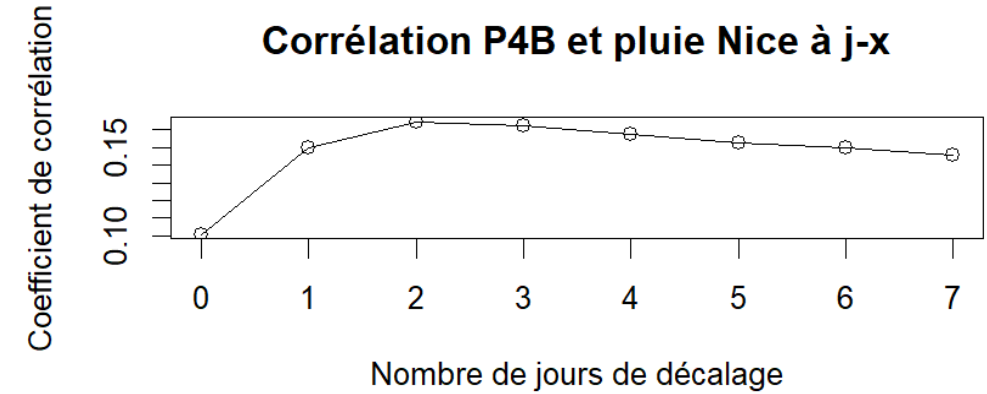


# Données : Corrélations entre niveau de nappe et précipitations

Corrélation entre niveau de nappe et précipitations

Coefficient de corrélation max à J+2  $\sim 0,16$

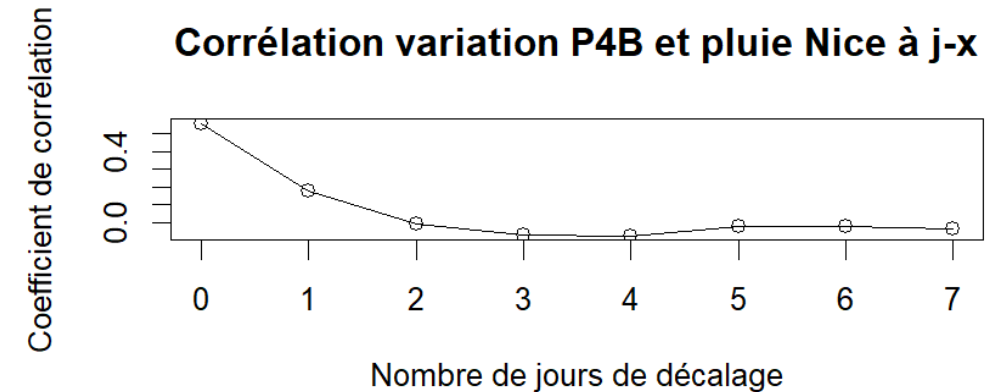
→ Niveau de nappe faiblement corrélé aux précipitations de l'avant-veille.



Corrélation entre **variation** du niveau de nappe et précipitations

Coefficient de corrélation max à J  $\sim 0,60$

→ Variation du niveau de nappe corrélée aux précipitations de la veille.



Coefficient de corrélation avec les précipitations de

Carros et Levens > Nice

→ Variation du niveau de nappe d'avantage corrélées aux précipitations en amont du piézomètre.

# Méthode 1 : SARIMA

Seasonal AutoRegressive Integrated Moving Average

Particulièrement efficace pour identifier les tendances et les motifs saisonniers

- Niveau de nappe :  $(X_t)_t$ ,
- Lien entre  $X_{t-1}, \dots, X_{t-p}$ , des valeurs de bruit  $\epsilon_t, \epsilon_{t-1}, \dots, \epsilon_{t-q}$ .
- Le *Seasonal* veut dire que  $X_t$  est également fonction de  $X_{t-52}, \dots, X_{t-(2 \times 52)}, \dots, X_{t-P \times 52}$ , et de  $\epsilon_{t-52}, \dots, \epsilon_{t-Q \times 52}$ .
- La valeur significative  $\Delta X_t = X_t - X_{t-1}$ ,
- 6 paramètres :  $(p, d, q)(P, D, Q)_{52}$ .

L'opérateur de shift  $B$  est défini par  $BX_t = X_{t-1}$ , alors  $(X_t)$  est un  $ARIMA(p, d, q)(P, D, Q)_{52}$  s'il existe  $(\theta_i)_{i \in \{1, \dots, p\}}$ ,  $(\Theta_j)_{j \in \{1, \dots, P\}}$ ,  $(\phi_k)_{k \in \{0, \dots, q\}}$  et  $(\Phi_l)_{l \in \{0, \dots, Q\}}$  tels que

$$(1 - B)^d (1 - B^s)^D X_t = c + \frac{(\phi_0 + \phi_1 B + \dots + \phi_q B^q)(\Phi_0 + \Phi_1 B^s + \dots + \Phi_Q B^{sQ})}{(1 - \theta_1 B - \dots - \theta_p B^p)(1 - \Theta_1 B^s - \dots - \Theta_P B^{sP})} \epsilon_t.$$

## Inconvénients

Choix des paramètres par tâtonnement

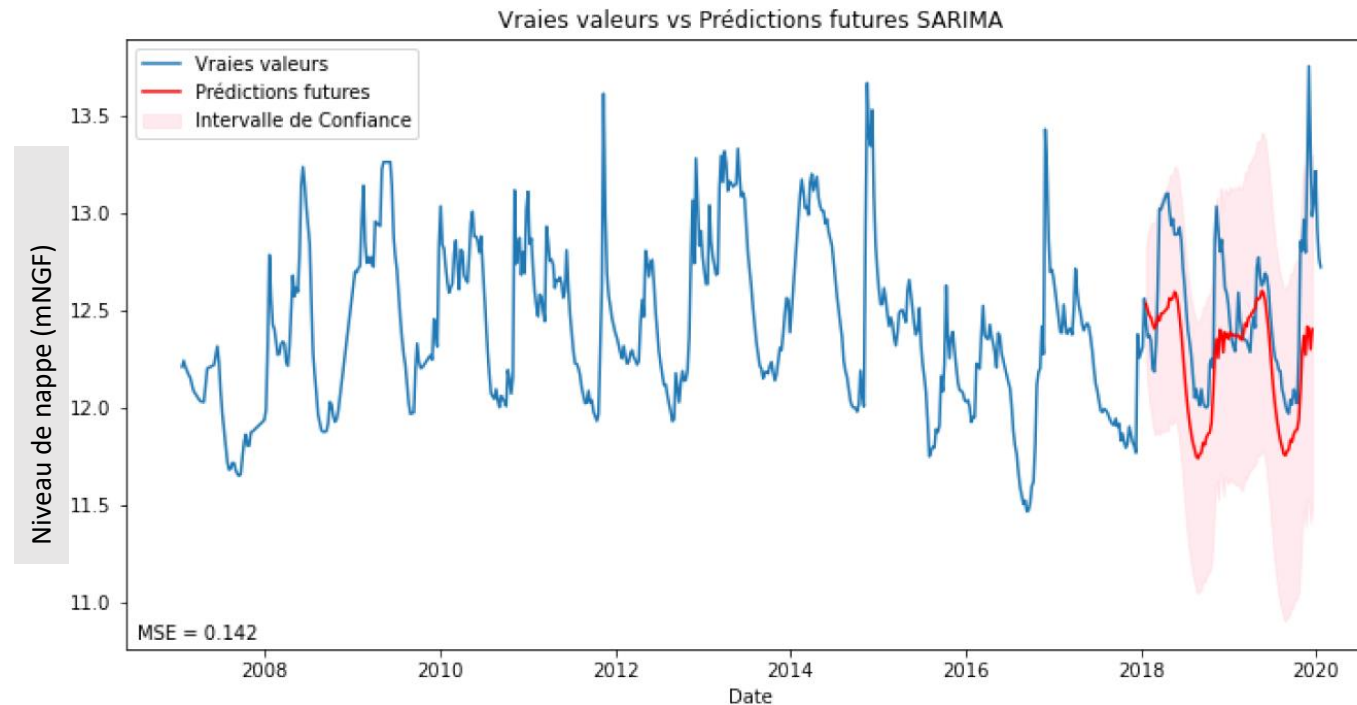
# Modéliser le niveau de nappe

SARIMA(X)

Données agrégées par semaine, Entraînement : 2007-2018; test : 2018-2020

Métrique :  $\text{MSE}(\hat{\theta}) \stackrel{\text{def}}{=} \mathbb{E} [(\hat{\theta} - \theta)^2]$

→ Meilleur modèle : SARIMA(1,1,1)(0,1,1)<sub>52</sub>



# Méthode 1 : SARIMAX

Seasonal AutoRegressive Integrated Moving Average

Particulièrement efficace pour identifier les tendances et les motifs saisonniers

- Niveau de nappe :  $(X_t)_t$ ,
- Lien entre  $X_{t-1}, \dots, X_{t-p}$ , des valeurs de bruit  $\epsilon_t, \epsilon_{t-1}, \dots, \epsilon_{t-q}$ .
- Le *Seasonal* veut dire que  $X_t$  est également fonction de  $X_{t-52}, \dots, X_{t-(2 \times 52)}, \dots, X_{t-P \times 52}$ , et de  $\epsilon_{t-52}, \dots, \epsilon_{t-Q \times 52}$ .
- La valeur significative  $\Delta X_t = X_t - X_{t-1}$ ,
- 6 paramètres :  $(p, d, q)(P, D, Q)_{52}$ .

L'opérateur de shift  $B$  est défini par  $BX_t = X_{t-1}$ , alors  $(X_t)$  est un  $ARIMA(p, d, q)(P, D, Q)_{52}$  s'il existe  $(\theta_i)_{i \in \{1, \dots, p\}}$ ,  $(\Theta_j)_{j \in \{1, \dots, P\}}$ ,

$(\phi_k)_{k \in \{0, \dots, q\}}$  et  $(\Phi_l)_{l \in \{0, \dots, Q\}}$  tels que

$$(1 - B)^d (1 - B^s)^D X_t = c + \frac{(\phi_0 + \phi_1 B + \dots + \phi_q B^q)(\Phi_0 + \Phi_1 B^s + \dots + \Phi_Q B^{sQ})}{(1 - \theta_1 B - \dots - \theta_p B^p)(1 - \Theta_1 B^s - \dots - \Theta_P B^{sP})} \epsilon_t + (\beta_0 Y_0 + \beta_1 Y_1 + \dots + \beta_N Y_N).$$

- $Y$  représente le niveau des précipitations

Inconvénients

Choix des paramètres par tâtonnement

# Méthode 3 : Filtre de Kalman

Estimateur récursif

Pour estimer l'état courant, seules l'estimation de l'état précédent et les mesures actuelles sont nécessaires.

Estime les variations du niveau de nappe à partir des volumes d'eau entrants et sortants de la nappe

## Étapes

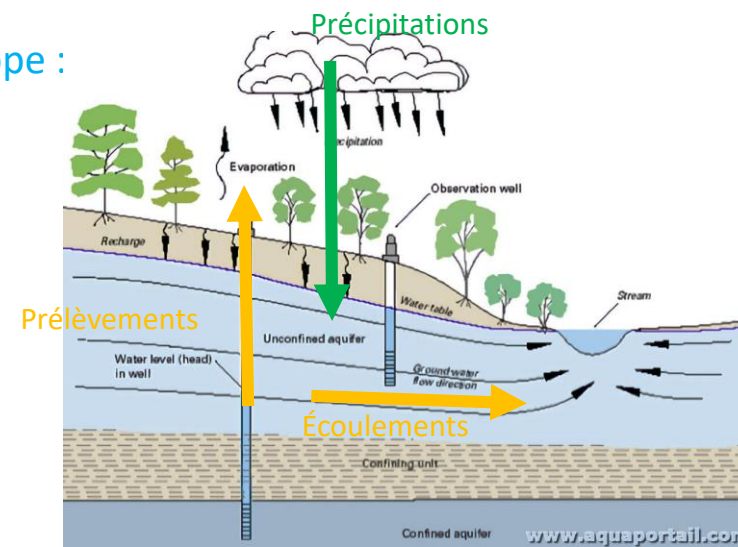
- 1- Prédiction
- 2- Mise à jour du modèle
- 3- Prévisions au pas de temps suivant

$$x_k = F_k x_{k-1} + B_k u_k + w_k$$

matrice de transition      commande en entrée      bruit d'évolution

Données supplémentaires de volumes sortants de la nappe :

- Prélèvements faits par Eau d'Azur
- Écoulement naturel souterrain moyen





## Méthode 3 : Filtre de Kalman

Données supplémentaires de volumes sortants de la nappe :

- Écoulement naturel souterrain moyen (constante)

$$\Delta X_t = \sum_{i=0}^{q_{rain}} a_i^T Y_{t-i} + \sum_{j=0}^{q_{withdrawals}} b_j^T Z_{t-j} + \sum_{k=0}^{q_{noise}} b_k \epsilon_{t-k}$$

Débit volumique  $Q$  (m<sup>3</sup>/s) avec la loi de Darcy :

$$Q = k \times A \times \frac{\Delta H}{L}$$

Avec :

- $A$  : la surface de la section étudiée (m<sup>2</sup>) ;
- $k$  : la conductivité hydraulique du milieu poreux (m/s) ;
- $\Delta H / L$  : le gradient hydraulique, où  $\Delta H$  est la différence des hauteurs piézométriques en amont et en aval de l'échantillon,  $L$  est la longueur de l'échantillon.

# Méthode 3 : Filtre de Kalman

Données supplémentaires de volumes sortants de la nappe :

- Écoulement naturel souterrain moyen (constante)
- Prélèvements faits par Eau d'Azur (07/2021 – 05/2024)

Modélisation des prélèvements sur la période étudiée (2007-2020) :

Hypothèse : prélèvements identiques d'une année sur l'autre

$\text{Prélèvement}(\text{Jn}20\text{xx}) = \text{moyenne} ( \text{Prélèvement}(\text{Jn}2021) \text{ à } \text{Prélèvement}(\text{Jn}2024) )$



# Résultat 3 : Modéliser le niveau de nappe

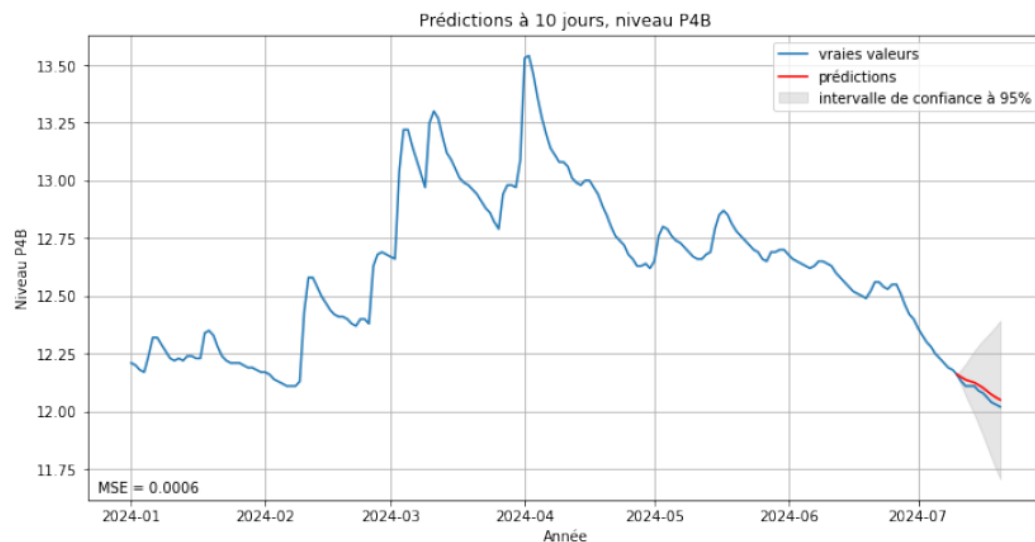
Prévisions à 10 jours avec les prévisions météo de précipitations

Comparaison des prédictions effectuées avec les observations météo et celle effectuées avec les prévisions météo

Modèle de prévision utilisé : GFS (Global Forecast System) haute résolution

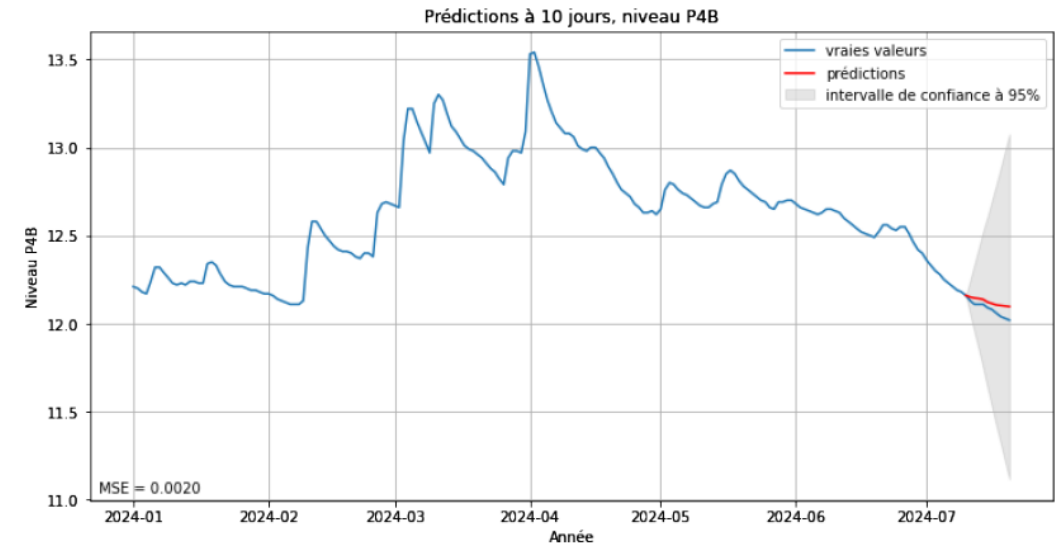
Mais maillage large (~28km) non représentatif des microclimats potentiels de la région

## Régime « sécheresse »



→ Bonne prédiction à 10 jours

## Régime « pluvieux »



# Comparaison des méthodes

|                          | SARIMA         | SARIMAX        | Kalman                           |
|--------------------------|----------------|----------------|----------------------------------|
| Historique considéré     | 11 ans         | 11 ans         | ~3 ans                           |
| Prévisions faites        | ~2 ans (700 j) | ~2 ans (700 j) | ~3 ans                           |
| Données exogènes         | non            | météo          | météo, prélèvements, écoulements |
| MSE                      | 0,142          | 0,073          | $10^{-3}$ à $10^{-4}$            |
| Intervalles de confiance | ~ constant     | croissant      | croissant                        |

➔ Difficulté de comparaison des résultats

Modèles incluant les prévisions météo

**Nécessaire à la prévision du niveau de nappe**

Dépendent de la qualité des prévisions météo ➔ limite à 3 ou 10 jours de prévision

➔ Quelles stations météo inclure ?

Modèle « filtre de Kalman » avec deux régimes

Choix du régime à utiliser au moment de la prédiction pour les jours suivants ?

Intervalle de confiance exponentiel.

➔ Les variables incluses dans le modèle sont-elles pertinentes ?

# Analyses supplémentaires et perspectives

## Ecoulements souterrains

Ne pas inclure les écoulements souterrains dans le modèle ou calculer un débit volumique à plus petite échelle spatiale et temporelle

## Prélèvements Eau d'Azur

Inclure les prélèvements de  $j$  et  $j-1$  dans le modèle. Enlever les prélèvements antérieurs.

## Stations météo

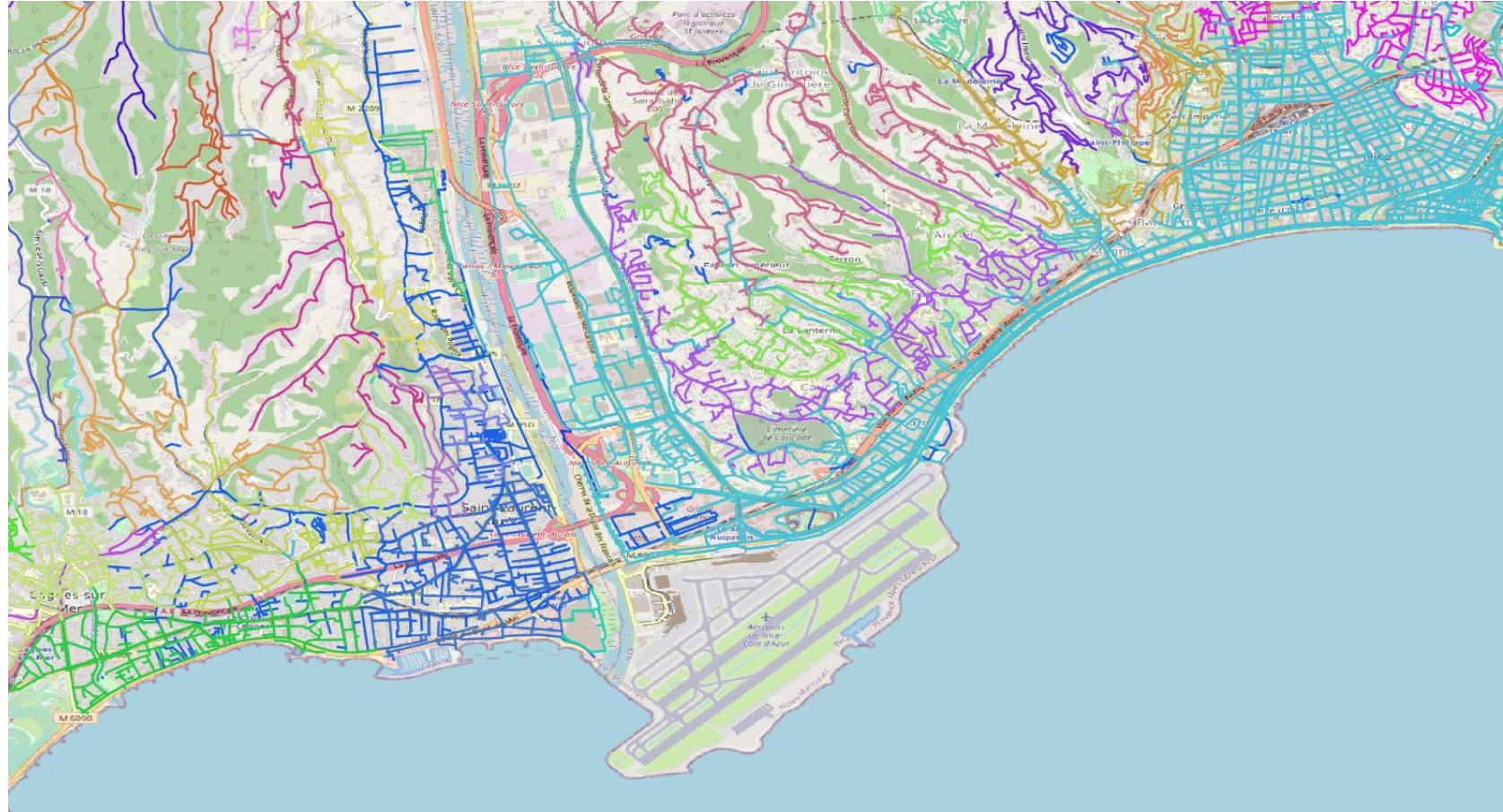
Inclure les stations météo des cercles 1, 2 et 3 (enlever Peone, Saint Étienne de Tinée et Saint Martin d'Entraunes)

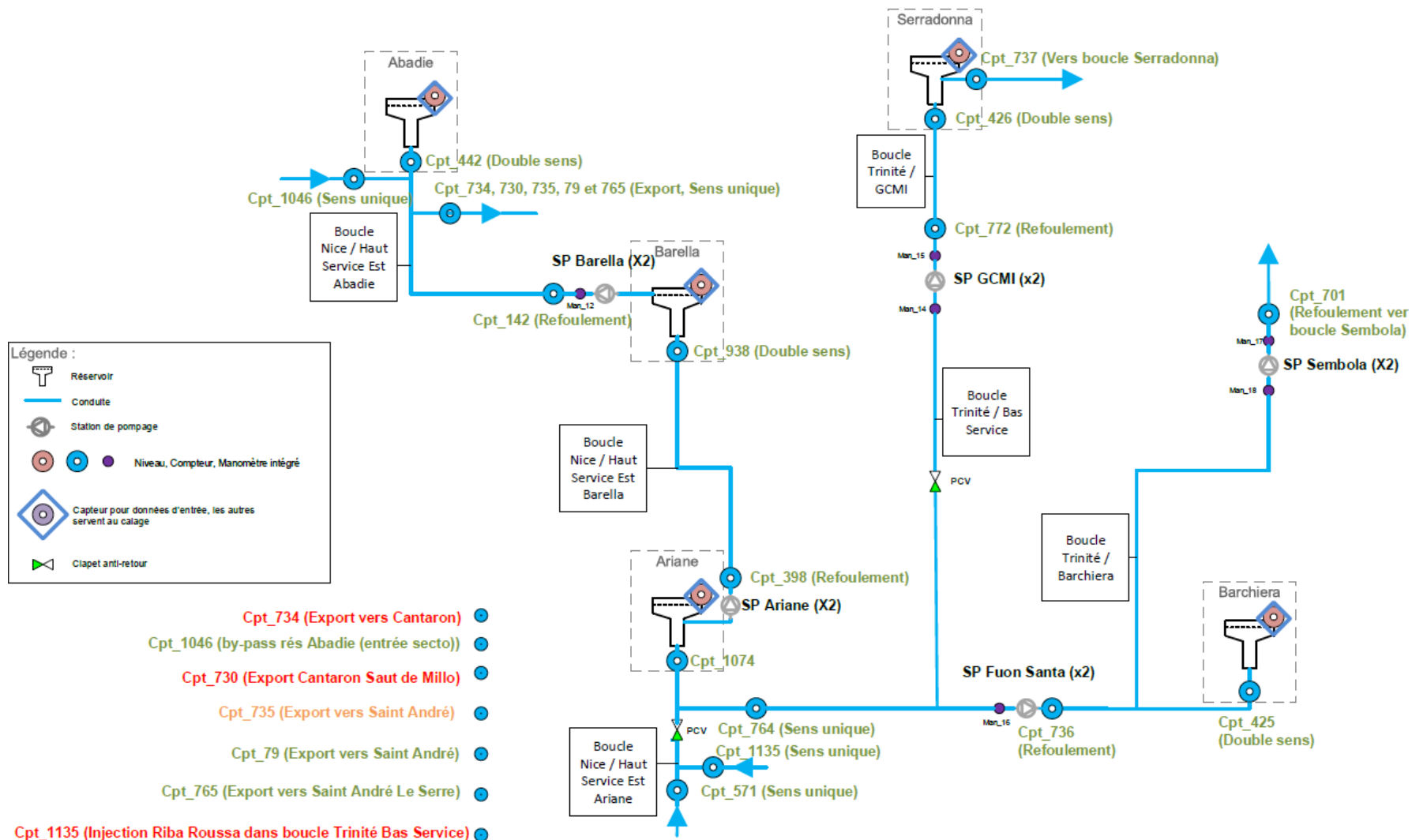
## Débits du Var et affluents ??

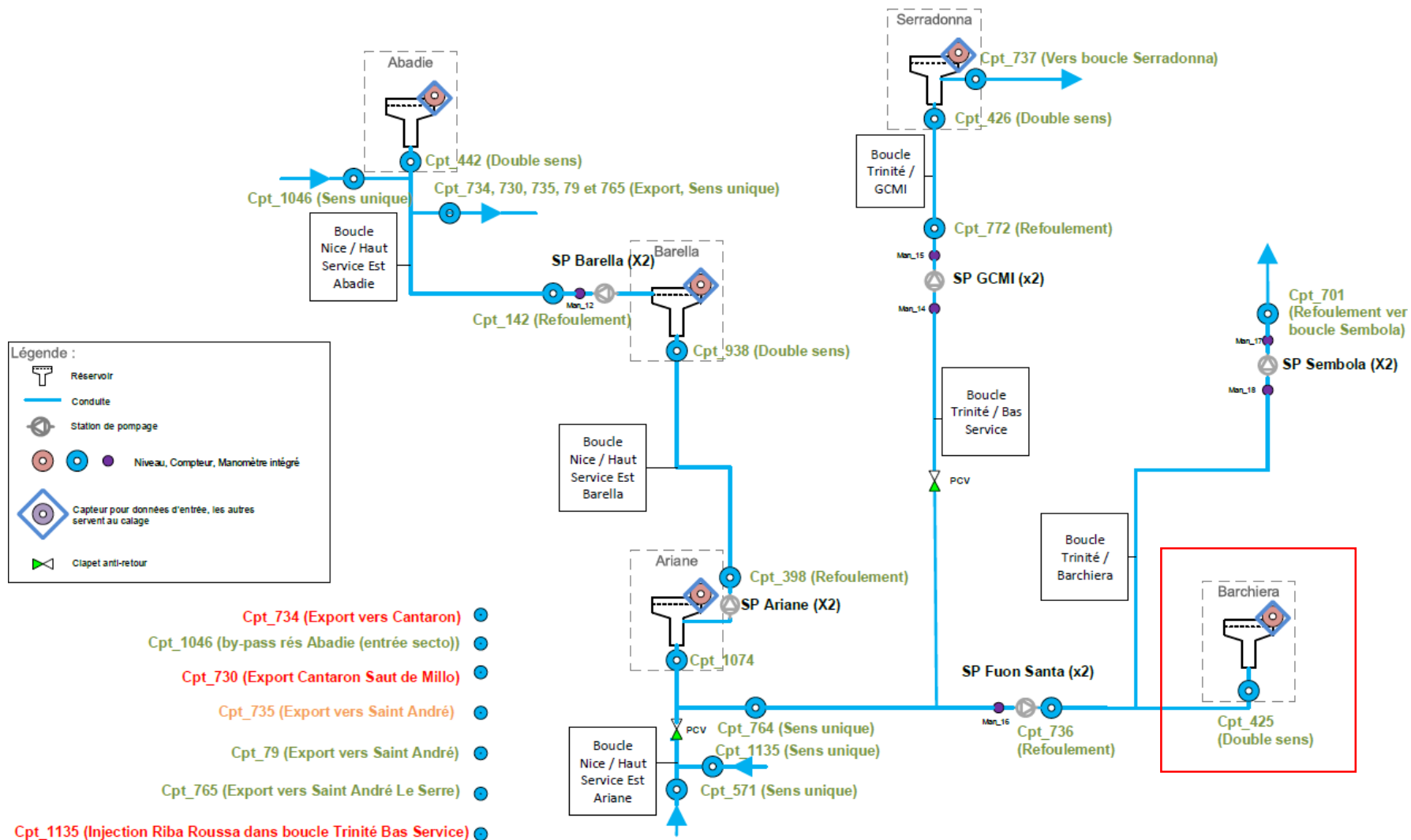
## ➔ Prévisions à 3 mois

Classification d'année type avec des méthodes s'appuyant sur la signature des données

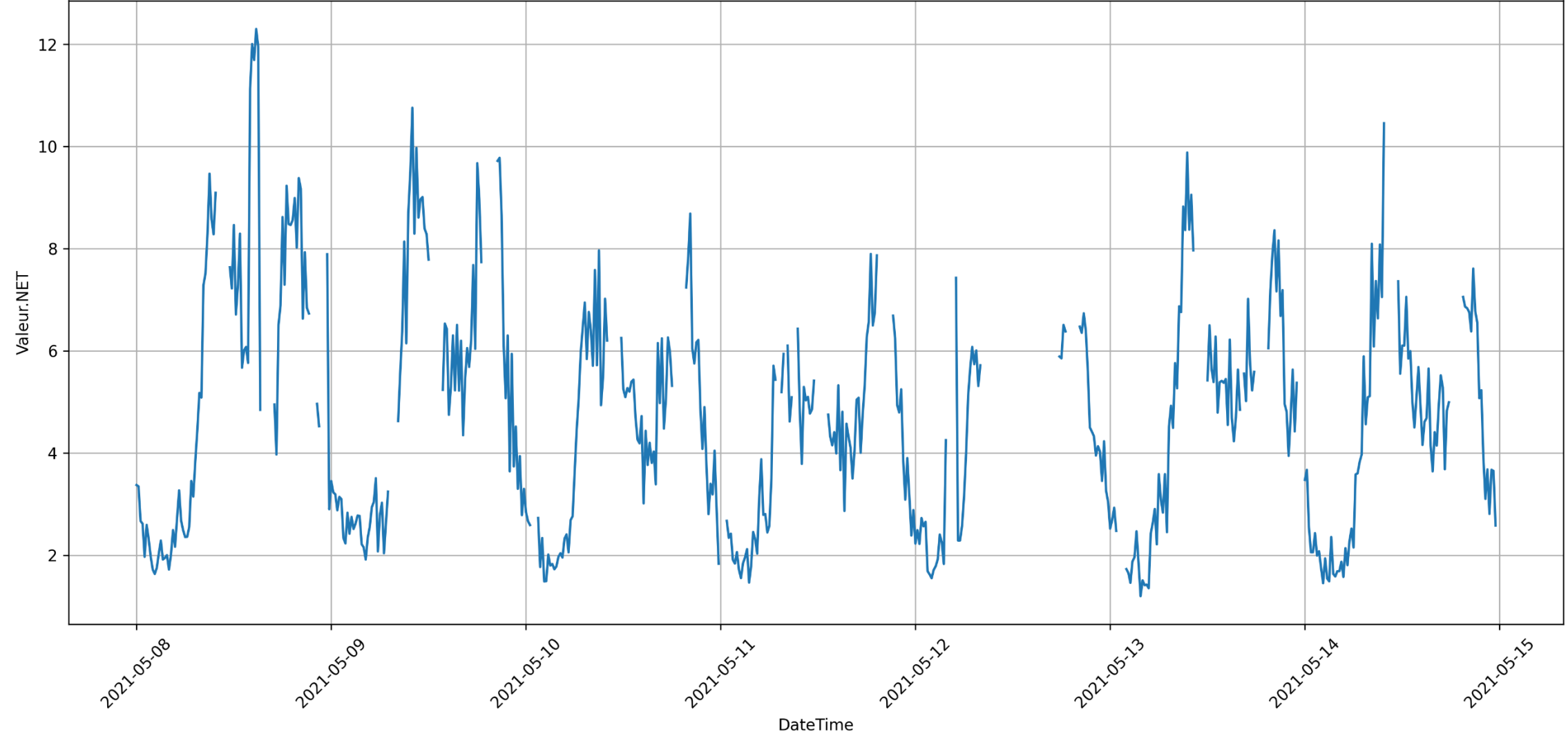
# Consommation



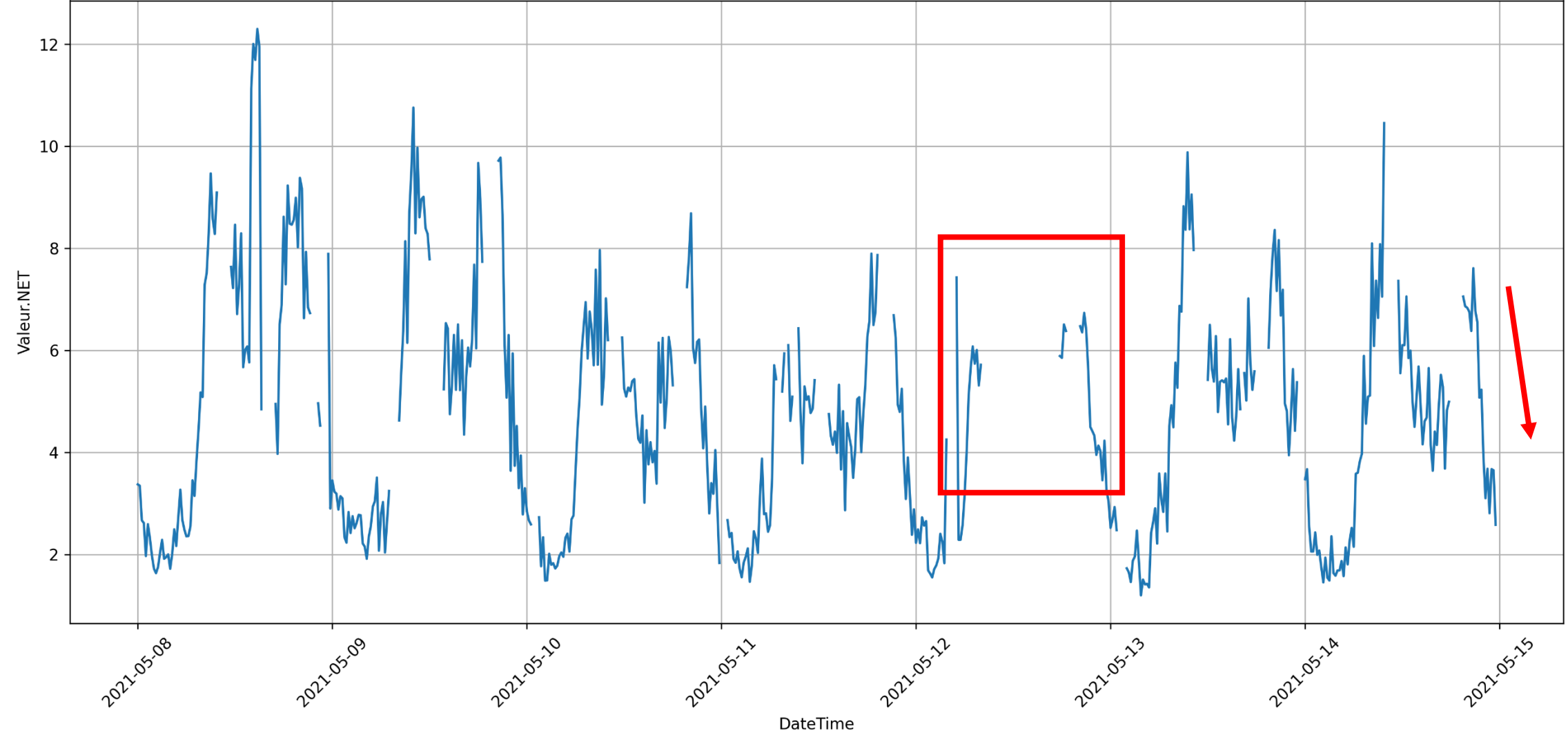




Série temporelle des valeurs de MPQ02\_LITT\_BARCHIERA\_DIST\_DEB avec NaNs

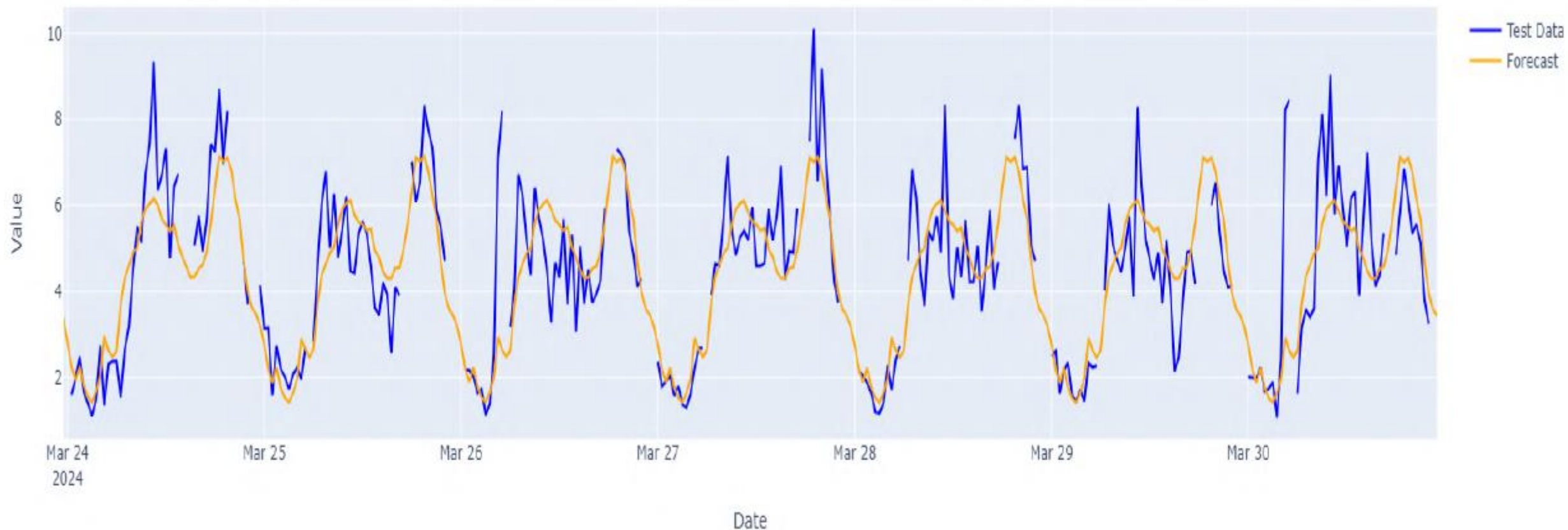


Série temporelle des valeurs de MPQ02\_LITT\_BARCHIERA\_DIST\_DEB avec NaNs

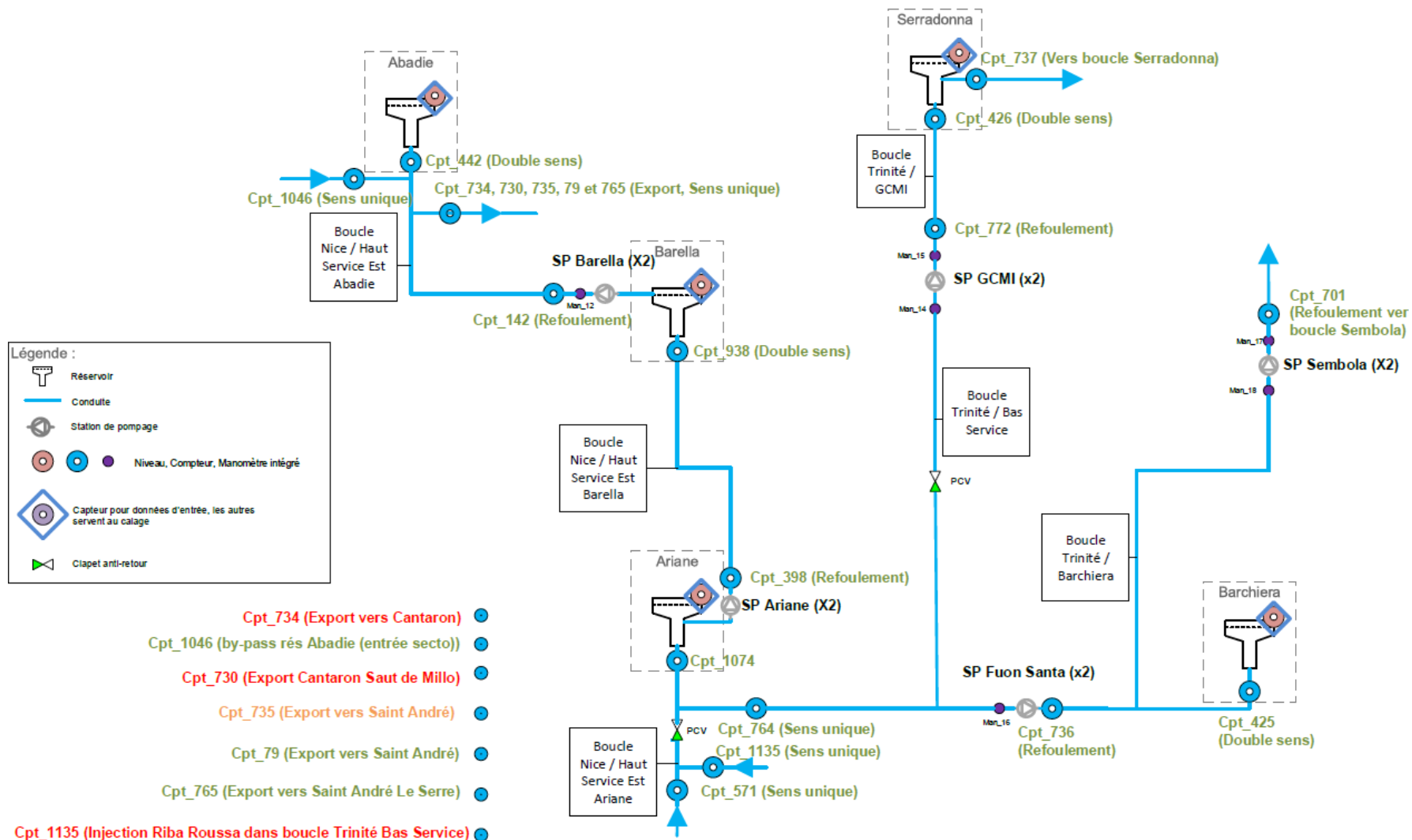


# Méthodologie

- But : **Prévoir la consommation (24h)**
- Jeux de données : 15min, sur plusieurs années
- Méthodologie :
  - Exploration des données
  - Complétion les données manquantes (interpolation via segmentation et régression)
  - Prédiction (via régression)
  - Stratégie :
    - méthode d'apprentissage/régression pour l'interpolation des données (OSSPI)
    - Méthode de régression (SARIMA) pour la prédiction



| Métrique                | Valeurs réelles (NaN) | Valeurs interpolées (OSSPI) |
|-------------------------|-----------------------|-----------------------------|
| Temps d'exécution moyen |                       | 26m12s                      |
| MSE moyen               | 2.1880                | 1.7275                      |
| MAPE moyen              | 22.4383%              | 18.7228%                    |
| Accuracy moyenne        | 77.8950%              | 81.2772%                    |



Merci pour votre attention