



Syndicat Mixte de gestion
de la nappe phréatique de la Crau

Dossier technique n° 5

-

Analyse statistiques des chroniques piézométriques



Titre : Analyse statistiques des chroniques piézométriques

Date : Février 2018

Auteurs : Baillieux A., Alcazar C.

**Avec le soutien
technique et
financier de :**



Agence de l'eau
RMC



Région PACA

Cadre et objectifs de l'étude

Analyser la corrélation existante entre les données piézométriques recueillies entre 1955 et 1969 (122 points), puis entre 2013 et 2018 (32 points). L'analyse de corrélation permet de délimiter des secteurs ayant une dynamique piézométrique similaire. Ce type d'analyse est déjà décrit dans la littérature scientifique¹ (Prinos, Scott T., 2005).

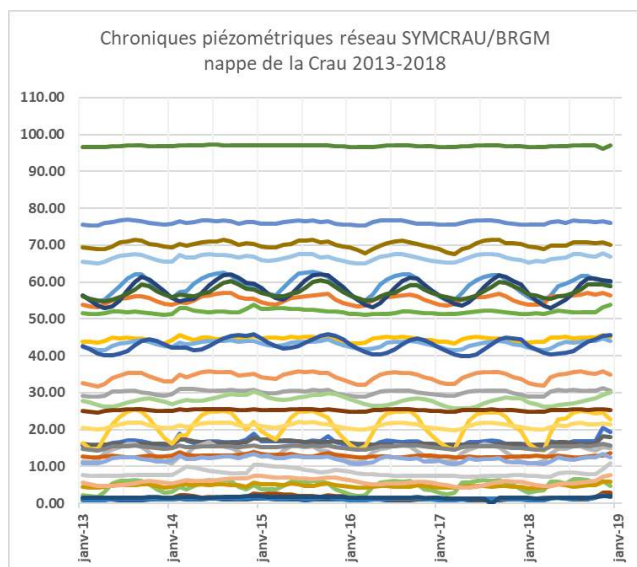


Figure 1 Chroniques piézométriques 2013 - 2018

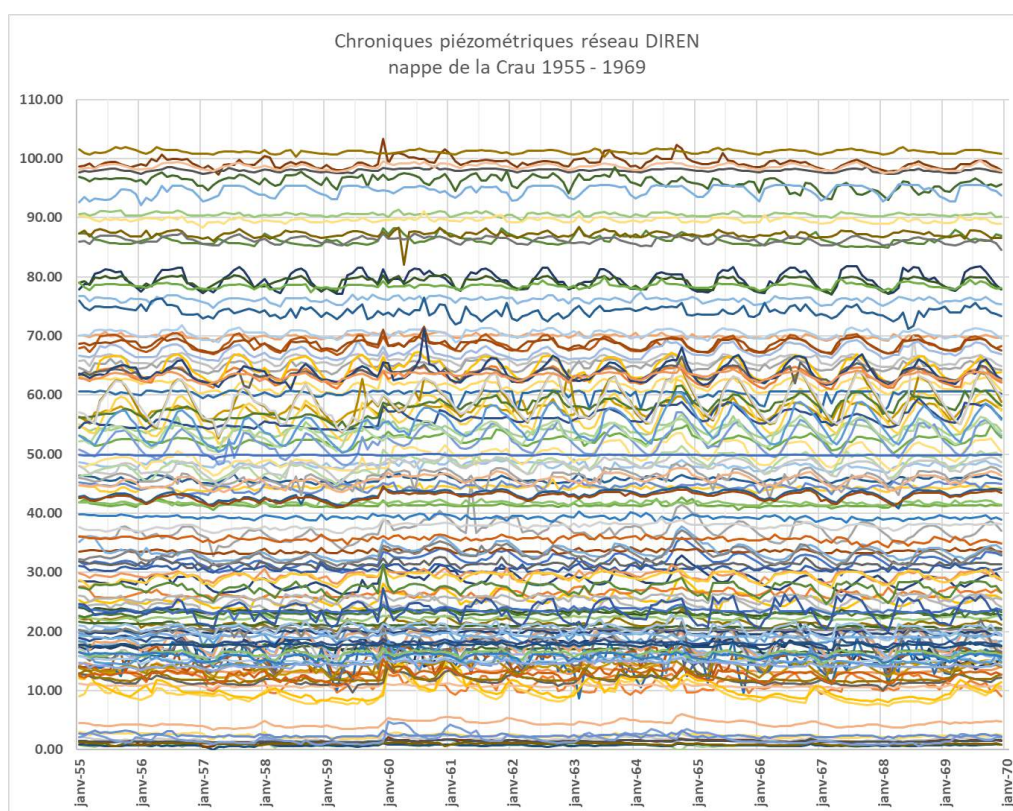


Figure 2 Chroniques piézométriques 1955-1969

¹ Prinos, Scott T., 2005), Correlation Analysis of a Ground-Water Level Monitoring Network, Miami-Dade County, Florida: U.S. Geological Survey Open-File Report 2004-1412, <https://pubs.usgs.gov/of/2004/1412/correlation.html>

Méthode

L'analyse de la corrélation est effectuée sur les niveaux piézométriques au pas de temps mensuel. L'indice de similarité est estimé par le calcul du coefficient de corrélation de Pearson (recherche de la corrélation linéaire entre les variables) et celui de Spearman (recherche d'une corrélation non linéaire). Les théories sur ces analyse sont rappelés dans les encarts ci-après ([source](#)²).

<p>Le coefficient de corrélation linéaire de Bravais-Pearson</p> <p>Ce coefficient permet de détecter la présence ou l'absence d'une relation linéaire entre deux caractères quantitatifs continus.</p> <p>Pour calculer ce coefficient il faut tout d'abord calculer la covariance. La covariance est la moyenne du produit des écarts à la moyenne.</p> $\text{Cov}(X, Y) = \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})$ <p>ou</p> $\text{Cov}(X, Y) = \left(\frac{1}{N} \sum_{i=1}^N X_i \cdot Y_i \right) - (\bar{X} \cdot \bar{Y})$ <p>Le coefficient de corrélation linéaire de deux caractères X et Y est égal à la covariance de X et Y divisée par le produit des écarts-types de X et Y</p> $r(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$ <p>Remarque : lorsque deux caractères sont standardisés, leur coefficient de corrélation est égal à leur covariance puisque leurs écarts-types sont égaux à 1.</p>	<p>Le coefficient de corrélation de rang de Spearman.</p> <p>Le coefficient de corrélation de rang (appelé coefficient de Spearman) examine s'il existe une relation entre le rang des observations pour deux caractères X et Y, ce qui permet de détecter l'existence de relations monotones (croissante ou décroissante), quelle que soit leur forme précise (linéaire, exponentiel, puissance, ...). Ce coefficient est donc très utile lorsque l'analyse du nuage de point révèle une forme curviligne dans une relation qui semble mal s'ajuster à une droite.</p> <p>On notera également qu'il est préférable au coefficient de Pearson lorsque les distributions X et Y sont dissymétriques et/ou comportent des valeurs exceptionnelles.</p> <p>Le coefficient de Spearman est fondé sur l'étude de la différence des rangs entre les attributs des individus pour les deux caractères X et Y :</p> $r(X, Y) = 1 - \frac{6 \sum_{i=1}^N [r(X_i) - r(Y_i)]^2}{N^3 - N}$ <p>avec</p> <p>$r(X_i)$ rang de X_i dans la distribution X_1, X_2, \dots, X_N</p> <p>$r(Y_i)$ rang de Y_i dans la distribution Y_1, Y_2, \dots, Y_N</p> <p>Ce coefficient varie entre -1 et +1. Son interprétation est la même que celui de Pearson, mais il permet de mettre en évidence des relations non-linéaires lorsqu'elles sont positives ou négatives.</p>
--	--

Une fois la matrice de similarité produite, il est effectué une classification ascendante hiérarchique³ (CAH) des données en fonction de la matrice similarité des individus calculée selon la méthode précédente. La CAH est représenté sous la forme d'un dendrogramme.

² http://grasland.script.univ-paris-diderot.fr/STAT98/stat98_6/stat98_6.htm#LE%20CALCUL%20DES%20COEFFICIENTS%20DE%20CORRELATION

³ Cf : <https://perso.univ-rennes1.fr/valerie.monbet/ExposesM2/2013/Classification2.pdf>

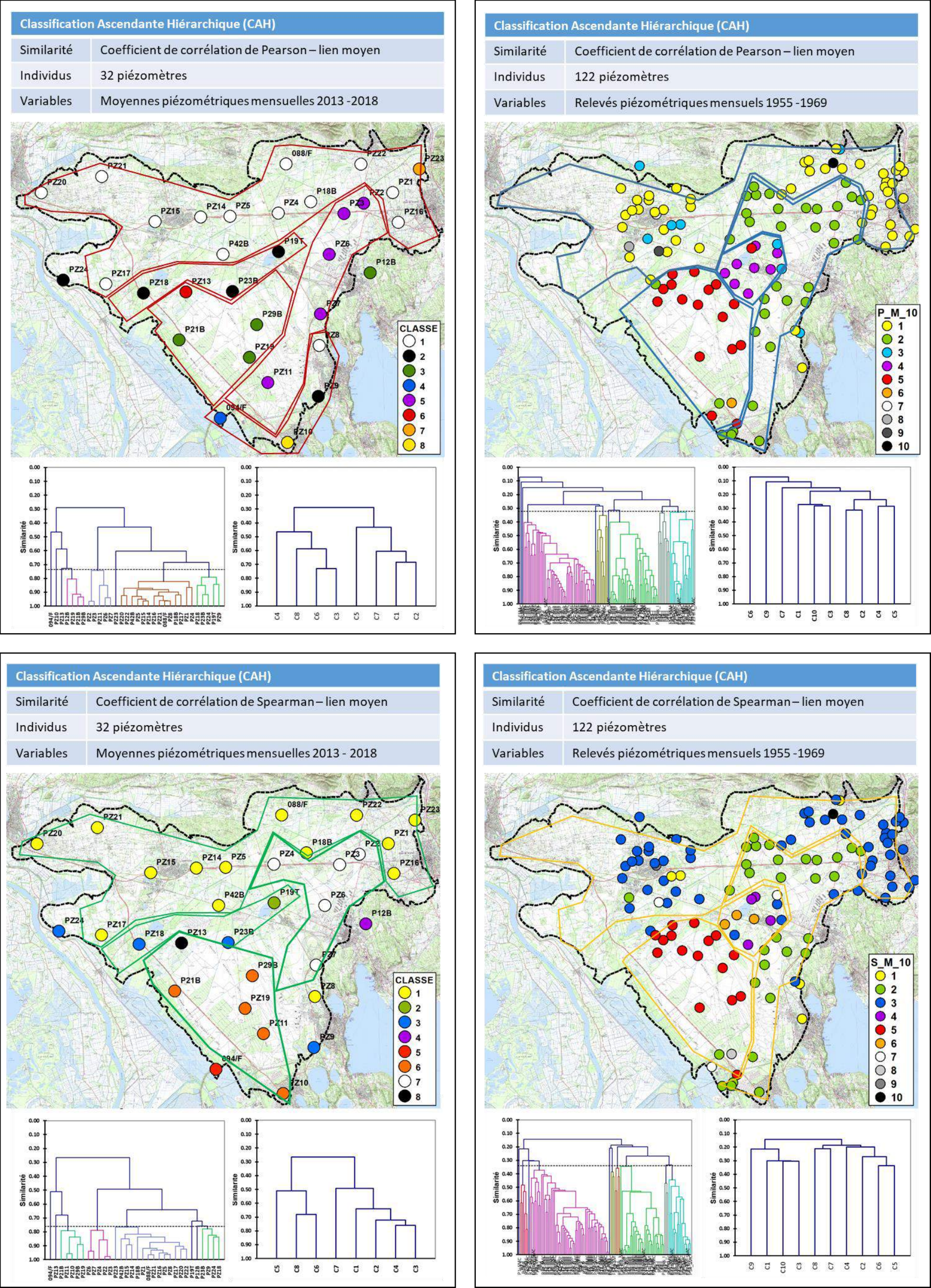


Figure 3 Répartition spatiale des groupes constitués par l'analyse ascendante hiérarchique

Les différentes analyses statistiques proposent des regroupements assez cohérents, variant légèrement d'une méthode à l'autre (Pearson/Spearman) et de manière légèrement plus significative d'une période à l'autre (1955-1969 / 2013/2018).

Les variations entre méthodes sont explicables par la nature du lien de corrélation estimé : pour Pearson il est recherché une relation linéaire, c'est-à-dire, des variations piézométriques qui évoluent dans le même sens avec un coefficient de corrélation linéaire. Dans le cas de Spearman, il est recherché les piézomètres qui ont tendances à se trouver dans la même fréquence d'apparition des événements. Les deux méthodes sont donc complémentaires.

Les variations d'une période à l'autre peuvent s'expliquer d'une part par la différence entre la taille de l'échantillon (nombre d'individus et variables piézométriques plus longues), et d'autre part par les conditions de recharge et de prélèvements qui ont pu évoluer entre ces deux périodes.

Cependant, en cumulant les groupes constitués (figure 4), on retrouve une certaine cohérence des résultats, permettant de distinguer 10 zones à l'intérieur desquelles les variations piézométriques présentent une cohérence significative (Figure 5).

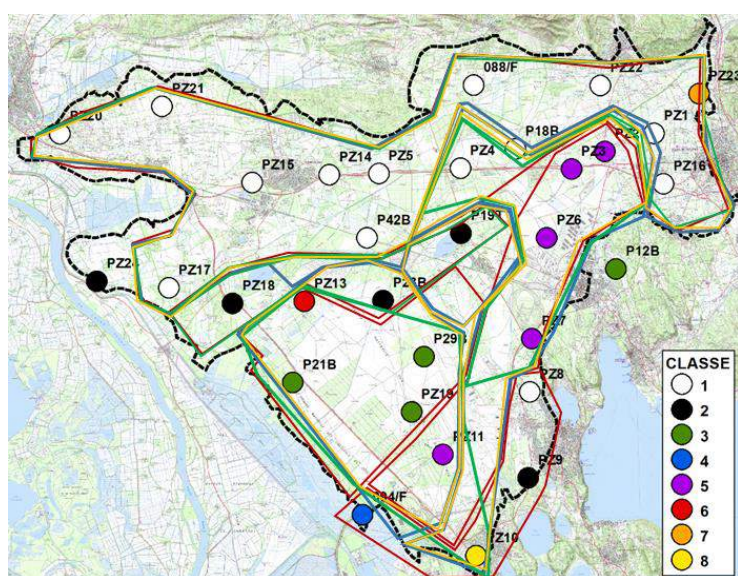


Figure 4 Cumul des groupes constitués par les CAH

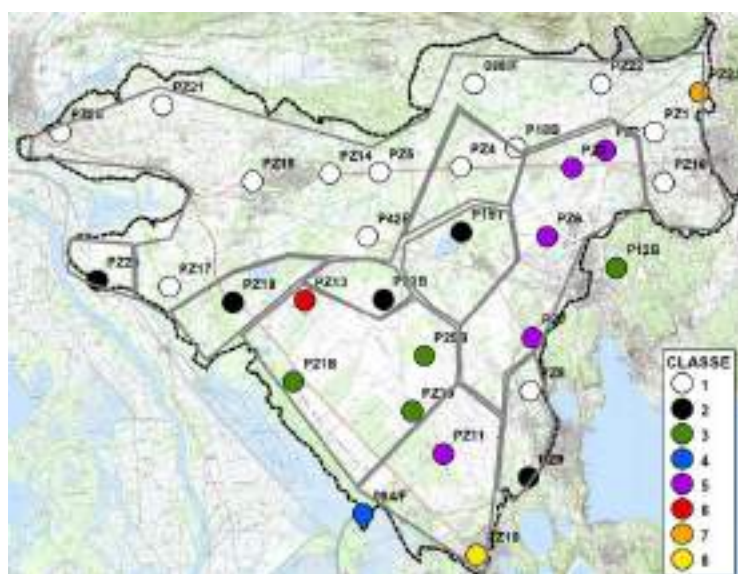


Figure 5 Zones se distinguant après analyses statistiques

Le regroupement de ces zones, tenant compte de la proximité des groupes, permet de mettre en évidence 5 grandes zones (Figure 6). En tenant compte des connaissances du terrain nous pouvant proposer les interprétations suivantes :

- Une zone amont (vert) caractérisée par la présence des prairies irriguées et dont la recharge est principalement contrôlée par la pratique de l'irrigation gravitaire
- Une zone de transition (violet) correspondant au sillon de Miramas, dont la piézométrie est influencée par la fonction de drain exercé par ce sillon
- Une zone aval (bleu) dont la dynamique piézométrique semble assez dissociée des pratiques d'irrigation et semble contrôlé par les précipitations efficaces
- Une zone intermédiaire (orange) caractériser par un substratum proche de la surface. Cette zone a une dynamique piézométrique plus inhomogène, contrôlée à la fois par les pluies, l'irrigation et des phénomènes de drainage (canaux et étangs).
- Une zone à l'est (jaune), avec un comportement piézométrique plus inhomogène influencé par des effets de bordure. Le piézomètre PZ23, en bordure de l'aquifère au nord de Salon et présentant un comportement singulier, peut être rattaché à cette zone.

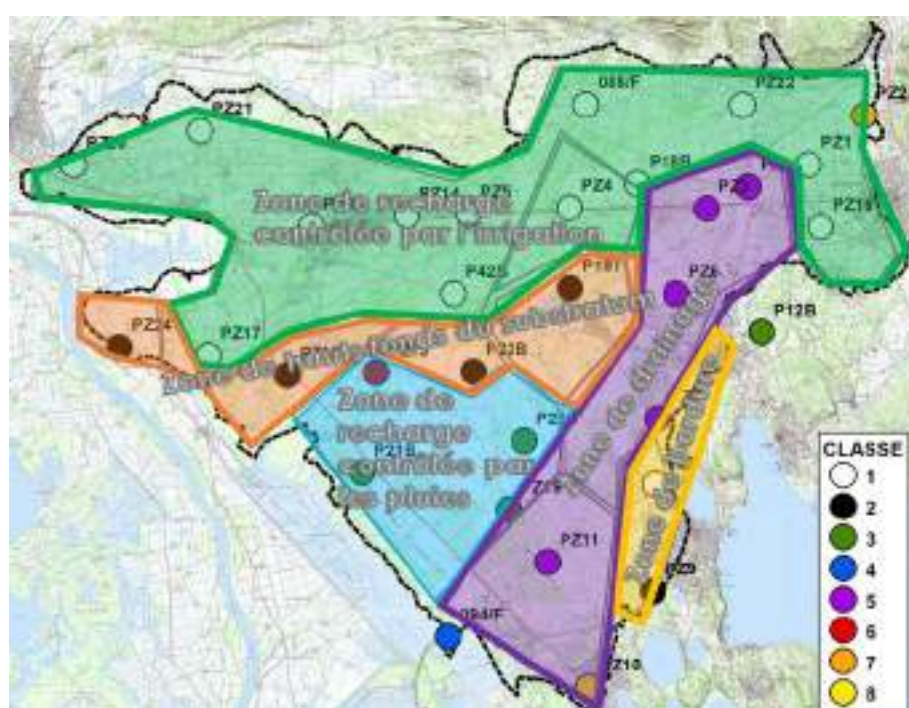


Figure 6 Grandes zones de cohérence piézométriques

Conclusions et perspectives

Le calcul de corrélation statistique sur les variations piézométriques, appliquée au réseau de surveillance 1955-1969 et 2013-2018, permet de distinguer 5 grandes zones. Théoriquement l'état global de la nappe pourrait être caractérisée par 5 piézomètres répartis parmi ces zones. Pour SINERGI, cela signifie que le modèle doit être capable de simuler les variations piézométriques sur ces cinq points caractéristiques.

La prochaine étape consiste donc à définir les indicateurs piézométriques selon les enjeux des usages et services écosystémiques de la nappe.