# 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

Manon Erguy, Sébastien Morilhat, Guillaume Artigue, Julien Trincal, Anne Johannet, Séverin Pistre

# 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

**Manon Erguy [1,2], Sébastien Morilhat [2], Guillaume Artigue [1], Julien Trincal [2], Anne Johannet [1,*], Séverin Pistre [3]**

[1] HydroSciences Montpellier (Univ. Montpellier, IMT Mines Alès, IRD, CNRS), 6 av. de Clavières, 30100 Alès, France
[2] CEA, DES, IRESNE, Nuclear Technology Department, Cadarache, F-13108, Saint Paul les Durance, France
[3] HSM, Univ. Montpellier, CNRS, IMT, IRD, Montpellier, France, 163 rue Auguste Broussonnet – 34090 Montpellier
* Corresponding author: E-mail address: annne.johannet@mine-ales.fr

**Abstract**

Karst aquifers are complex and linked to numerous management issues, particularly concerning the the risk of flooding or the availability of water resources. In view of the associated stakes, it is necessary to have the best possible knowledge of the hydrodynamic behavior of these aquifers. In this context, a semi-automated unsupervised clustering of the hydrodynamic behavior at different piezometers has been developed. It was applied to the Cadarache site of the the French Alternative Energies & Atomic Energy Commission (CEA) in South-East of France; site which is subject to rapid groundwater floods. At the methodological level, this paper proposes a grid for the interpretation of sorted water levels in relation to the hydrogeological properties of the formations. The results of this study, achieved on 75 boreholes, allows identifying several clusters of boreholes. Three clusters appeared to be representative of specific areas. The method only uses sorted graphs of water levels. A relevant relation between each cluster and its geological structure is highlighted. This study revealed the role of preferential flow paths of NW-SE and NE-SW lineaments previously identified by regional tectonic or structural studies and with same directions than fractures observed at local scale (borehole core).

**Keywords** Karst aquifer, Groundwater floods, Hydrodynamics, Sorted groundwater levels, Cadarache karst aquifer.

## 1 Introduction

Karst is one of the most common types of aquifers in the world, with karstifiable carbonate rock formations occupying about 15% of the global ice-free continental lands (Goldscheider et al. 2020). These aquifers are associated with many issues related to resource management (drinking water supply, agriculture…) and risk management. This includes rapid transfers of pollutants and flooding. Indeed, karst hydrosystems may absorb a part of the flooding, or on the contrary constitute a partial or total groundwater contribution to the flood episodes (López-Chicano et al. 2002; Pinault et al. 2005; Jourde et al. 2007; Bailly-Comte 2008). A karst is a set of surface and underground features produced by the dissolution of carbonate rocks (Ford and Williams 2007). These very specific landscapes are anisotropic, heterogeneous and associated to hydrosystems with a non-linear behavior. The resulting flow conditions are therefore complex and linked to the structural properties of the physical object (Bakalowicz 1999). These characteristics make karst hydrosystems particularly difficult and challenging to study. They also complicate the understanding of the transfers that take place within them. Conceptual representations of karst are then developed describing karst as a physical object or compartments (generally: epikarst, unsaturated zone and saturated zone) crossed by flows. Because of the lack of physical knowledge about their structure and their hydrodynamic parameters, functional representations are also developed. In the systemic approach, karsts are then represented as a function transforming an input signal into an output signal. Many types of hydrogeological models are derived from this approach, such as reservoirs models (Jourde et al. 2015) or neural network models (Johannet et al. 1994; Kong a Siou et al. 2011; Kong-A-Siou et al. 2014; Taver 2014). However, because these models often take into account the global functioning of the aquifer, the spatialization of hydrodynamic properties remains difficult to determine (Kovács and Sauter 2007).

However, a spatial representation of the internal flow dynamics of the karst hydrosystem can be approached using statistical methods on pluviometric, piezometric or discharges data. For this purpose, several methods have been developed for hydrological time series. In a non-exhaustive way, these include: hierarchical clustering or K-means (Boukharouba et al. 2013; Naranjo-Fernández et al. 2020), Principal Component Analysis (PCA) (Haaf and Barthel 2018), Self-Organizing Maps (SOM) (Nourani et al. 2012; Han et al. 2016; Gholami et al. 2022; Wunsch et al. 2022). Two approaches are generally adopted: classification directly from the raw piezometry data (Raj 2004; Haaf and Barthel 2018) or via features describing the dynamics of the phenomenon (Nourani et al. 2012; Wunsch et al. 2022). This last approach is more generalizable because it is less dependent on the data series specificities (length, period...), and fairly independent from gaps or errors (Wang et al. 2006; Wunsch et al. 2022).

Specifically in the Mediterranean region, due to the intense rainy episodes, underground karst floods can affect many public, private and, in particular, industrial issues (Weng and Dörfliger 2002). It is therefore particularly important to predict them in order to be able to mitigate the consequences of these floods. In karst context, the first stage consists in improving the local understanding of the karst aquifer hydrodynamic behavior.

To this end, we propose in the current study a new simple method based on a simple and well-known statistical method: the data sorted curves. This statistical approach offers the advantage to visualize the changes in the process according to the state of the system (Mangin 1975). Applying this method to groundwater levels allows describing the local hydrodynamic behavior of the aquifer in order to better manage it (Wunsch et al. 2022). Validation of the method is done on the French Alternative & Atomic Energy Commission center of Cadarache, in the French Mediterranean region. The site is subject to karst groundwater floods with rapid kinetics and high amplitude (rising water tables about tens of meters, until it reaches the topography, in a few hours).

The main contributions of this paper are twofold: first, presenting a new simple clustering method based on the analysis of sorted groundwater level curves. Second, building an interpretation grid to link the obtained clusters with the geological knowledge of the study site.

The article is organized as follows: first, the material and methods section will present the Cadarache site and the clustering method. The results section will focus on the obtained clusters and their comparison with the geological knowledge. These results will then be discussed and their limitations analyzed in the last part.

# 2 Material and methods

## 2.1 Study area: the Cadarache center

### 2.1.1 Geological and hydrogeological settings

#### *2.1.1.1 Geological context*

The Cadarache center of the French Alternative Energies & Atomic Energy (CEA) is located at Saint Paul Lez Durance in the south of France. The center takes place on outcropping Hauterivian limestone (Lower Cretaceous) partly covered by Miocene sediments and / or Pliocene and Quaternary alluvium (Guerin 2001) (Fig. 1).

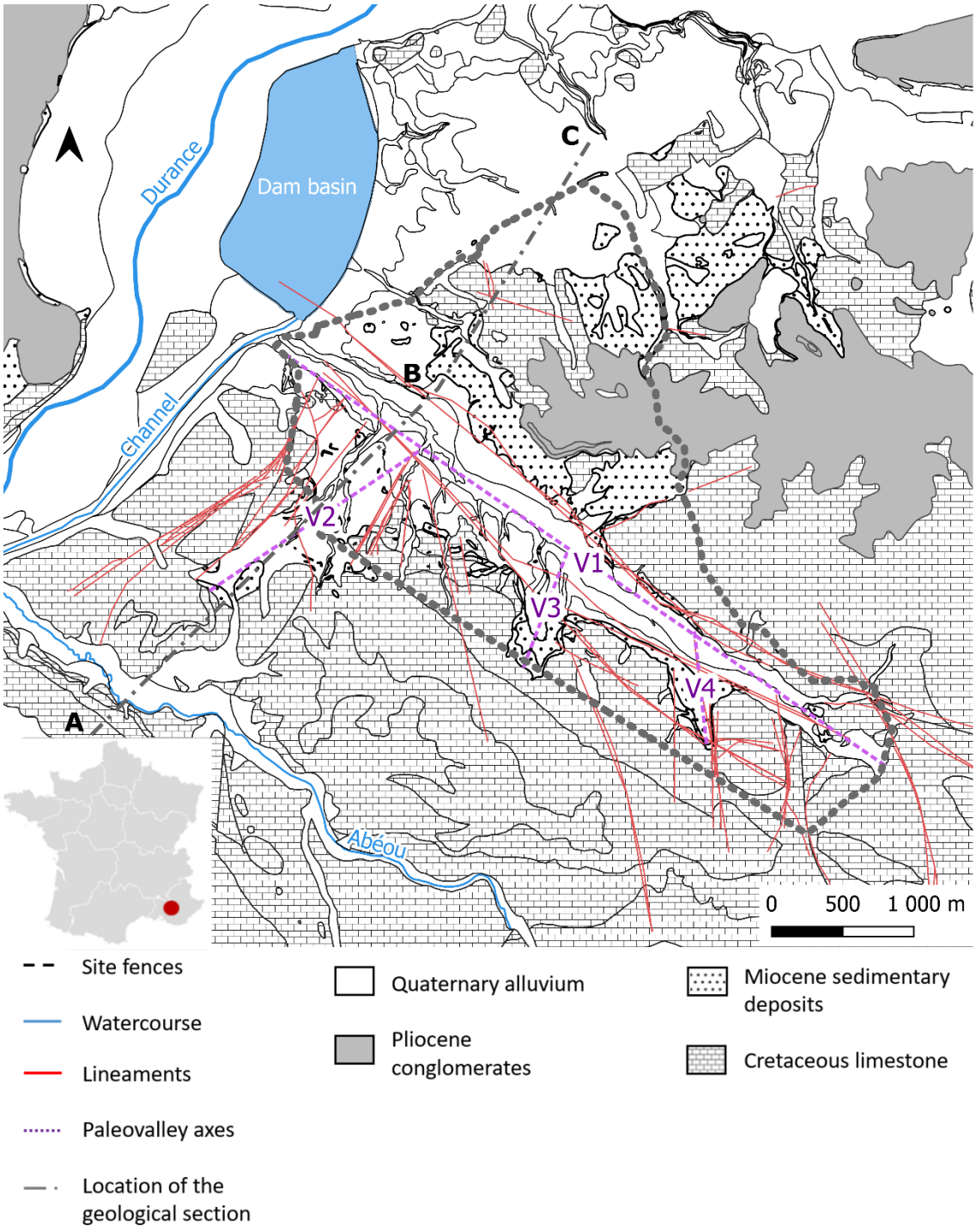


**Fig. 1** Simplified geological map of the Cadarache center

The area of the site is affected by numerous subvertical faults (Fig. 1 and Fig. 2) inherited from the successive tectonic regimes of the geological situation of South-East France, in particular linked to the Pyrenean and Alpine orogenies (Guerin 2001). At the site scale, the subhorizontal Cretaceous limestones are structurally characterized by:

*i.* plurikilometric NE-SW faulted structures (strike-slip type), framing Miocene-filled palaeovalleys, notably in the south of the site (e.g. V2, V3 and V4 in Fig. 1 and Fig. 2);

*ii.* a NW-SE paleovalley (V1 in Fig. 1 and Fig. 2) is present in the central axis of the site. Geological studies assume the presence of plurikilometric faults bordering this structure (Guerin 2001; Fenart 2008).

These fractures directions N120-140 and N040-060 are also found at the local (metric) scale (Guerin 2001; Fenart 2008).

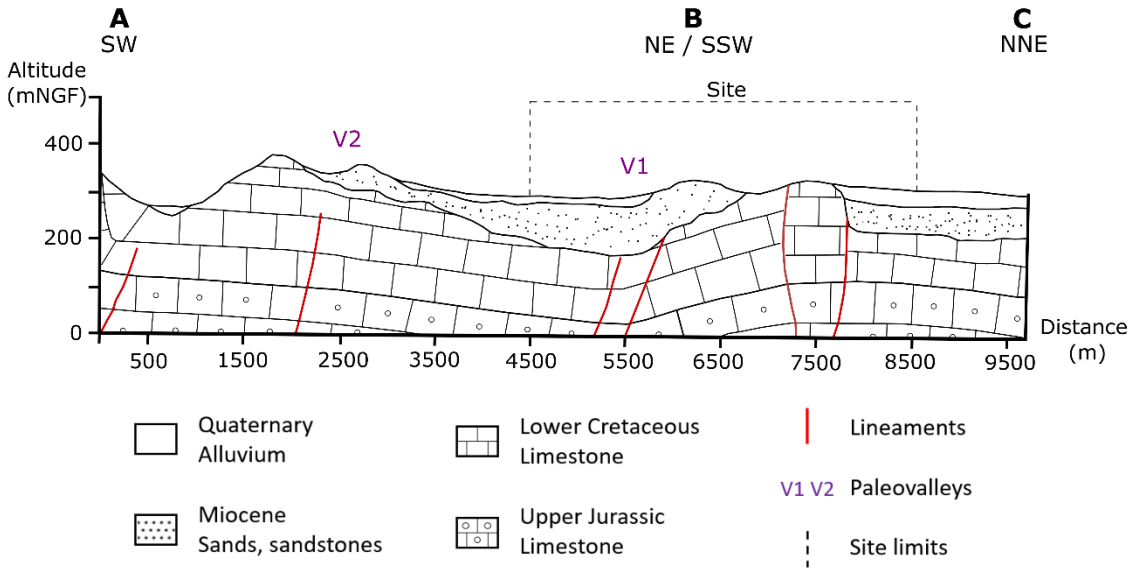


**Fig. 2** Geological section of the site (modified after Guerin (2001))

### *2.1.1.2 Hydrogeological context*

Each of the above-mentioned formations, Quaternary, Miocene and Cretaceous, constitutes an aquifer having a specific behavior. The main flow direction of the aquifers is towards the Durance river (NW of the center). The Miocene and Quaternary aquifers are not very reactive and their level variations are of low amplitudes (water level rising of few meters in several weeks, or several months). From outcrop and borehole observations, the Cretaceous aquifer shows moderate karstic features (Guerin, 2001) but it is characterized by an important reactivity, a hydrodynamic spatial heterogeneity and by the absence of highly developed karst conduits. It seems that the flow is organized via fractures and faults, more or less karstified (Guerin 2001). This fractured karst aquifer is subject to groundwater flooding. For example, the major rainfall events of November 2011 or November 2019 with a cumulative rainfall respectively of 225 mm in 72h, and 101 mm in 4 days (this episode followed an important rainfall event that took place earlier in October), both led to groundwater level increasing of nearby 40 m locally,

causing phenomena of artesianism. Important questions remain concerning the complex links and exchanges between the different aquifers.

### 2.1.2 Monitoring network

Given the flood challenges associated with this industrial site, the Cadarache center has an outstanding hydrogeological monitoring network. Rainfall was measured at the center: (*i*) manually since 1960 on a daily time step and, (*ii*) by automatic rain gauges (1 to 10 minutes time step) since 2006. Currently, the site has eight rain gauges spread over five zones (Fig. 3).

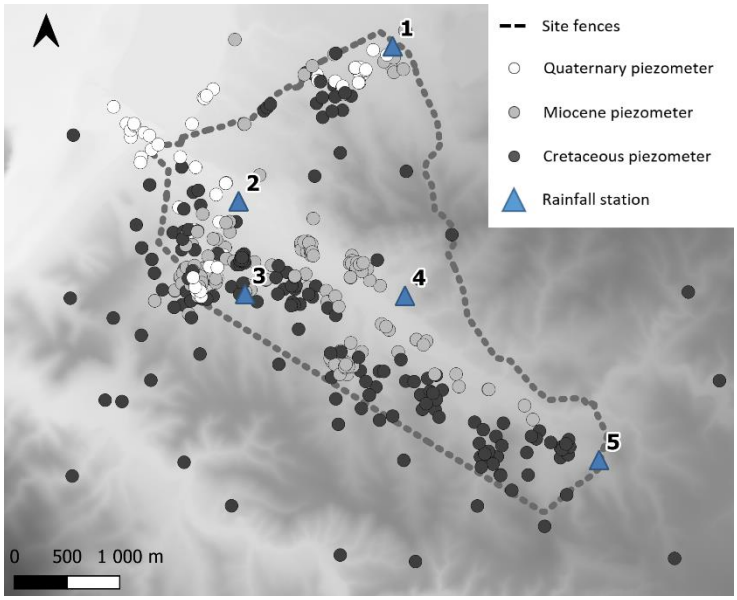


**Fig. 3** Hydrogeological monitoring network of the Cadarache site

Concerning the piezometric monitoring (Fig. 3), it consists of more than 400 piezometers for the different aquifers, of which 234 are currently equipped with pressure sensors (with 115 piezometers measuring the Cretaceous aquifer). This dense monitoring network has enabled the acquisition of a large database of rainfall and groundwater levels over more than 15 year.

## 2.2 Database

### 2.2.1 Rainfall time series

The rainfall data reliability and spatial variability were analyzed for all rain gauges using the double-mass curves, and compared with cumulative rainfall. The first method consists of comparing rainfall accumulations between two rain gauge stations. The aim is to identify potential malfunctions, or trends in the underestimation or overestimation of rainfall. All the 28 different stations combinations were compared on common acquisition periods. In order to study the spatial variability of rainfall at the site scale, in a more complete way, a second type of analysis was to compare the rainfall amounts over different periods to identify possible differences. The cumulative rainfalls of the eight stations were calculated: (*i*) for the 30 major rainy events since 2006, (*ii*) on a monthly scale since 2017 (this period contains the most available data), (*iii*) annually since 2006. The differences

observed on the reliable data (when no malfunction has been identified via the acquisition rate) do not exceed 10%. In addition to these differences, these methods revealed malfunctions and data gaps for some rain gauges. Whereas the water table is an integrated variable of rainfall at the scale of the hydrogeological basin, it was considered that 10% differences would not have significant impact on the aquifer hydrodynamic behavior.

Based on these concerns, a "reference rainfall chronicle" was built by averaging data, for each time step, from two stations, respectively located to the north (station 1 in Fig. 3) and to the south (station 5 in Fig. 3) of the site. This offers the advantage of providing a reliable and continuous rainfall chronicle since 2006. This chronicle was used to select the rainfall events studied in the rest of this paper. Based on this reference rainfall, the average annual cumulative precipitation calculated between 2007 and 2020 at the Cadarache center is 635 mm.

### 2.2.2 Groundwater time series

The automatic acquisition of the piezometric data of the center begins at the end of the 90s (1998/99) on some piezometers at the south of the site. Subsequently, other pressure/temperature sensors were installed and new piezometers were built until they resembled the current impressive network. The acquisition time step for these sensors was fixed at 30 minutes. The pressure sensors are read manually every 3 months. A manual measurement with the piezometric probe is made at this time in order to readjust the recorded measurements if necessary and correct any potential drift of the sensor.

### 2.2.3 Studied periods

Three events, emblematic of different high flood situations rainfall, where chosen for this study. Two major fall episodes were selected that generated artesianism phenomena: November 2011 and November 2019. These are the two events which caused the most important increases of the water table (40 m locally) recorded on the site since the automatic measurements (2000's). The difference between these two events is mainly related to the initial state of the aquifer. In the case of the 2011 event, the rain came after a period of low water (five months). For the 2019 event, the water level was high from the beginning because of a previous episode in October (cumulative rainfall of: 185 mm in 4 days). The March 2017 event, on the other hand, was a moderate rainfall in low/mid water level. The rainfall accumulation and the available number of piezometric records for each event are summarized in Table 1.

**Table 1** Overview of the studied rainfall episodes

| Rainfall episode | Cumulative rainfall | State of the aquifer | Number of piezometer records available |
|---|---|---|---|

| | | | |
|---|---|---|---|
| November 2011 | 212 mm in 72 h | Low | 61 |
| March 2017 | 53 mm in 72 h | Low/mid | 73 |
| November 2019 | 101 mm in 96 h | High | 92 |

To complete this overview, an analysis at the seasonal scale was conducted from fall 2008 to spring 2020 (**Erreur ! Source du renvoi introuvable.**). This analysis was limited to piezometers that have been identified using clustering (see section 3).

**Table 2** Means and standard deviations of rainfall per season between fall 2008 and spring 2020 on the Cadarache site

| **Season** | **Average cumulative rainfall (mm)** | **Standard deviation of seasonal rainfall totals in different years (mm)** |
|---|---|---|
| Winter | 117 | 61 |
| Spring | 179 | 45 |
| Summer | 110 | 66 |
| Fall | 265 | 136 |

According to **Erreur ! Source du renvoi introuvable.**, the rainfall of the site is marked by very variable and rainy falls (large standard deviation). The variability in cumulative rainfall for other seasons are smaller. The second season in terms of cumulative rainfall is spring, then winter and summer. The last one is marked by a strong variability because the rainfall is linked to stormy episodes, frequent during this season. This pattern corresponds well to a Mediterranean logic with rather dry summers and winters, a relatively rainy spring and a fall with significant rainfall totals. The significant standard deviations for all the seasons show the inter-annual rainfall variability, as well as the importance of intense events in the annual cumulative rainfall.

### 2.3 Sorted groundwater levels

The graphs of sorted values of a variable are used to represent the process associated with this variable. Applied to groundwater levels, the sorted groundwater level curves represent how the cumulative empirical probability (in %) evolve versus the piezometric values ranked by classes. As it is generally accepted that rain on a small time-scale behaves like white noise (Mangin 1975), the ruptures observed on sorted water levels are essentially representative of the hydrogeological properties or the underlying processes (compartmentalization, fracturing, connectivity or permeability), without rainfall influence. This allows identifying variations in the slope of the curve

as modifications in behavior, or particularities of the studied process (Mangin 1975). For example, in case of karst springs discharge, Marsaud (1997) described the kind of slopes ruptures that could correspond, for example, to a problem in the gauge station, an inflow of water from another catchment area, an overflow... This analysis, carried out on discharges, is not directly transposable to the groundwater level in a fractured and karstified aquifer. Another major difference with the study of discharges is that in the case of groundwater levels, instead of a river, the considered reservoir is complex and heterogeneous. Moreover, when studying piezometric data coming from a borehole, the sorted groundwater level is the result of the coupling between the geological heterogeneity and the heterogeneity created by the borehole itself, which corresponds to a vertical drain having an almost infinite permeability. In order to identify the patterns related to groundwater level slope changes, we proposed a new interpretation grid, with some examples of configurations encountered at the Cadarache site. Several main cases are illustrated in the Fig. 4.

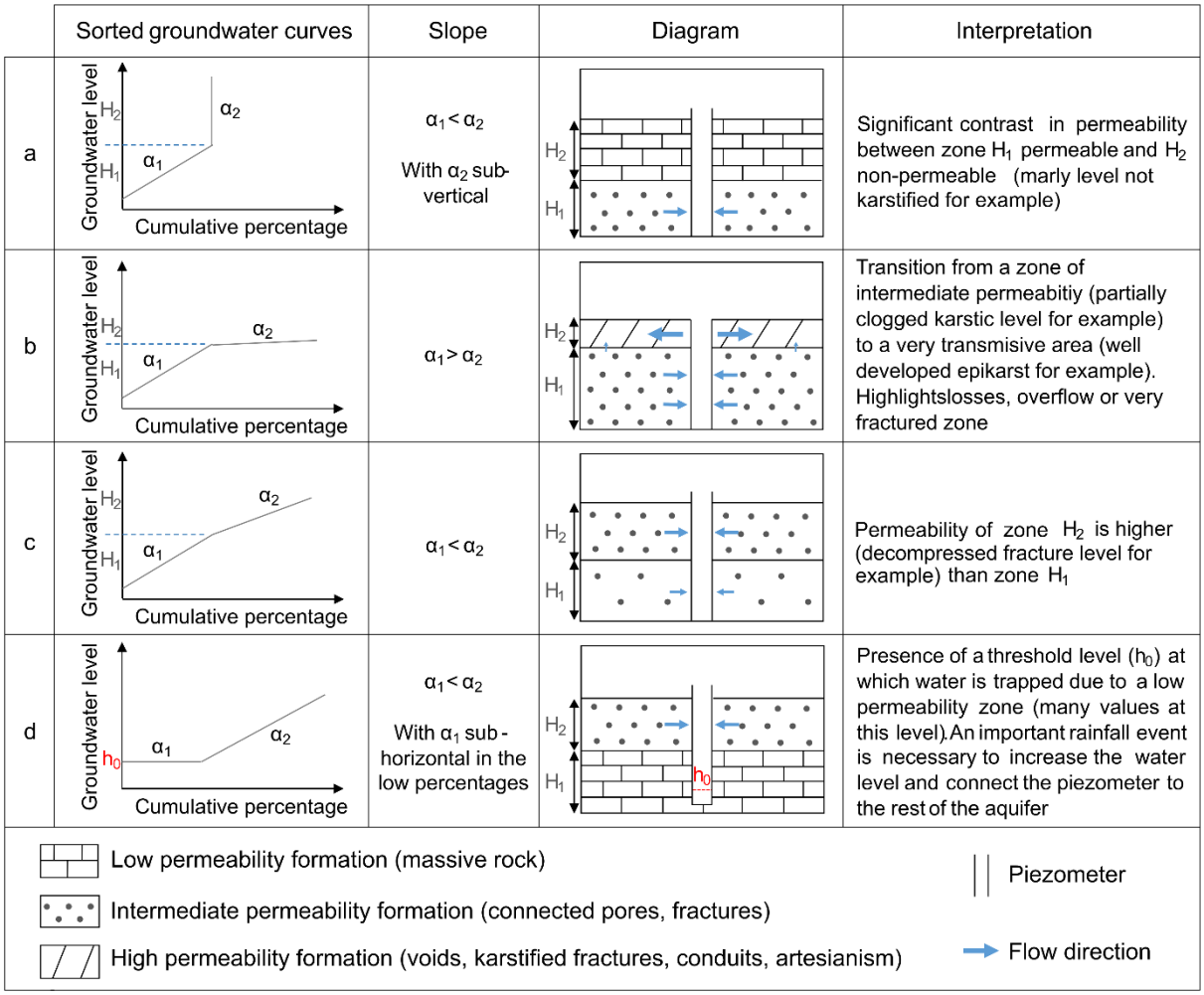


**Fig. 4** Interpretation grid for several cases of sorted groundwater levels in a fissured and/or karstified aquifer. Attention, the water level is represented in vertical axe to facilitate the interpretation, contrary to what is generally done for the sorted discharges, in particular in the work of Marsaud (1997).

Fig. 4 shows the evolution of the water level as a function of the cumulative probability for different configurations. In the Fig. 4.a, the upper layer is very little permeable compared to the one below. In this case, during recharge, when water come from layer $H_1$ or from layer upper layer $H_2$, the behavior is roughly the same. The water level in the borehole will rise with a moderate slope in the layer $H_1$, and will increase very quickly into the impermeable zone (no leak) showing a sub-vertical slope. Fig. 4.b corresponds to the superimposition of a horizontal draining layer, for example a stratification joint, on top of a layer with a median connected porosity. In this case, the layer $H_1$ fills up, which corresponds to a certain intermediate slope $\alpha_1$ on the graph of Fig. 4.b. Then, when the filling up reaches the draining layer, the water drains away as an overflow would. The slope is therefore almost horizontal on the graph in Fig. 4. However, if it is a case where the water comes from above, the results would be different. In this configuration the water would fill the borehole very quickly (almost vertical slope at the beginning), then it would stay at the same level, if layer $H_2$ can absorb it, until layer $H_1$ fills completely. In the third one (Fig. 4.c), the layer $H_2$ has a higher permeability than the lower. In this case, the slope $\alpha_1$ is greater than the slope $\alpha_2$ because, for the same ”flow”, the porosity will be filled faster in the lower layer; a greater slope corresponds to a greater permeability. Fig. 4.d corresponds to a particular case called: “bottle bottom”. In this case, a volume of water is trapped at the bottom of the borehole due to a low permeability zone like massive limestone without fractures for example: it is disconnected from the behavior of the water table. The piezometer then shows an "initial level" $h_0$. A significant rainfall event is necessary to increase the water level and reconnect the piezometer to the rest of the aquifer via fractures or faults above for example.

Interpretation of geological structures by means of sorted groundwater level curves is complicated by the dependence of piezometry on rainfall statistics, specifically for rare values. The cases presented above are valid with respect to slope breaks, but it is more difficult to make definitive statements about the interpretation of slope values since rainfall has a Gaussian distribution only at short time scales.

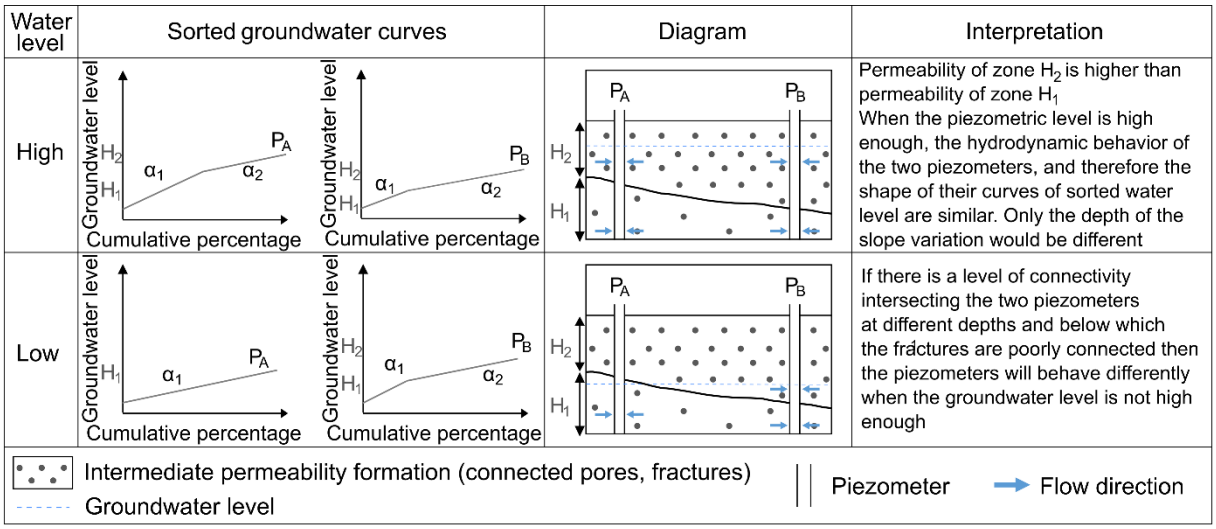


**Fig. 5** Identification of a connectivity threshold using event-driven analysis

When doing an event-based analysis of the sorted groundwater levels, the curves can be different depending on the water level reached. Thus in Figure 5, for weak events or in periods of average humidity, only one type of porosity is observed in the $H_{1A}$ layer (Borehole A), and two in the $H_{1B}$ layer (Borehole B). When the water level increases in Borehole A, due to a more intense event or to an event occurring in a humid period, the higher permeability of the $H_{2A}$ layer becomes visible. When several boreholes are close to each other, it is possible to describe the spatial variations of the transition from layer 1 to layer 2 by looking at the values of the rupture altitude for all the boreholes.

### 2.4 Clustering protocol

The Cretaceous aquifer of Cadarache is complex and heterogeneous. Although information is available on fracturing, its influence on hydrodynamic functioning is not known; finally, little information is available on the fracturing and its role in the flow. As presented above, the sorted groundwater level diagrams can provide information on the hydrodynamic processes involved, in relation to the associated depths. For this purpose, a simple semi-automated comparison protocol has been developed to group the sorted water levels curves based on their slope breaks (Fig. 6) (Erguy et al. 2022).

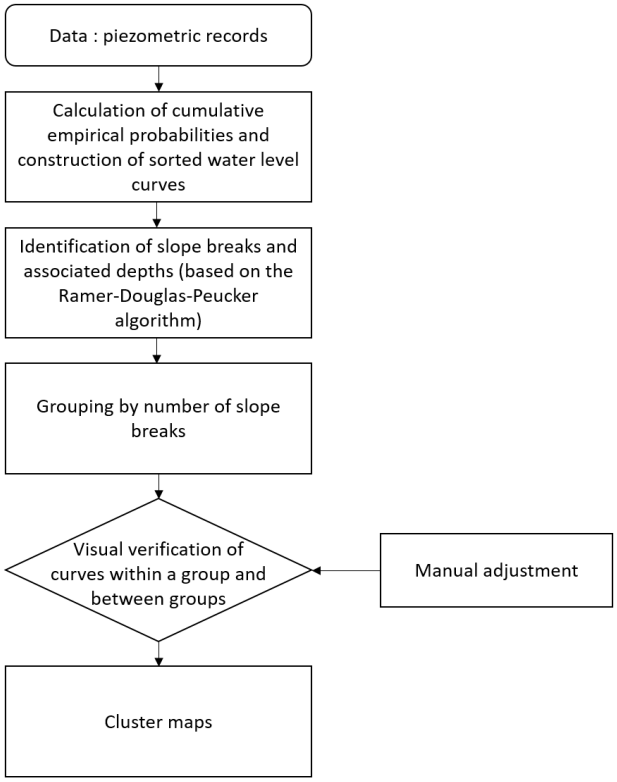


**Fig. 6** Flowchart of the clustering protocol

First, the curve variations and associated piezometric levels are automatically detected. The procedure is based on the Ramer-Douglas-Peucker algorithm; it simplifies the curve into straights segments and highlights the significant slope changes (Douglas and Peucker 1973). In a second step, an automatic clustering by number of slope breaks is made. This makes it possible to identify, in a first approach, the main groups with similar sorted groundwater level curves. After this preliminary clustering, the clustering assessment is done by a visual inter-comparison of the graphs curves, within each group, and between the different groups. This verification allows checking the groups formed and modify them if necessary. However, it should be noted that the need for a visual control is a limitation due to the risk of human bias, inconsistency and the problems of the time-consuming (Haaf and Barthel 2018).

## 3 Results

### 3.1 Piezometers clustering

The clustering protocol using the sorted groundwater level curves was applied to the events presented previously and to the available piezometric records (Table 1). The processed data and clustering results are presented in **Erreur ! Source du renvoi introuvable.**, and Fig.7.

Table 3 Piezometers hydrodynamics behavior clustering

| Flood episode | Group | Number of break slopes | Number of piezometers in the group |
|---|---|---|---|
| November 2011 | a | 4 | 6 |
| | b | 5 | 26 |
| | c | ≥ 6 | 7 |
| | Non-grouped | | 22 |
| March 2017 | d | 0 | 4 |
| | e | 1 | 7 |
| | f | 2 | 5 |
| | g | 3 – 4 | 27 |
| | h | 5 – 6 | 10 |
| | Non-grouped | | 20 |
| November 2019 | i | 3 | 9 |
| | j | 4 | 5 |
| | k | 5 – 6 | 33 |
| | l | 6 | 9 |
| | Non-grouped | | 36 |

Between three and five clusters were identified depending on the episodes studied and the data available for each period (Table 1). A cluster consists of more than two piezometers with similar sorted water level curves shapes (Fig. 7).

For the November 2011 episode, the analysis of the sorted water level curves allowed to identify three main clusters. For the March 2017 and November 2019 rain events, more piezometric records are available, respectively four and five clusters were identified. This proves that the Cretaceous aquifer shows different types of hydrodynamic behavior. Some of which are complex.

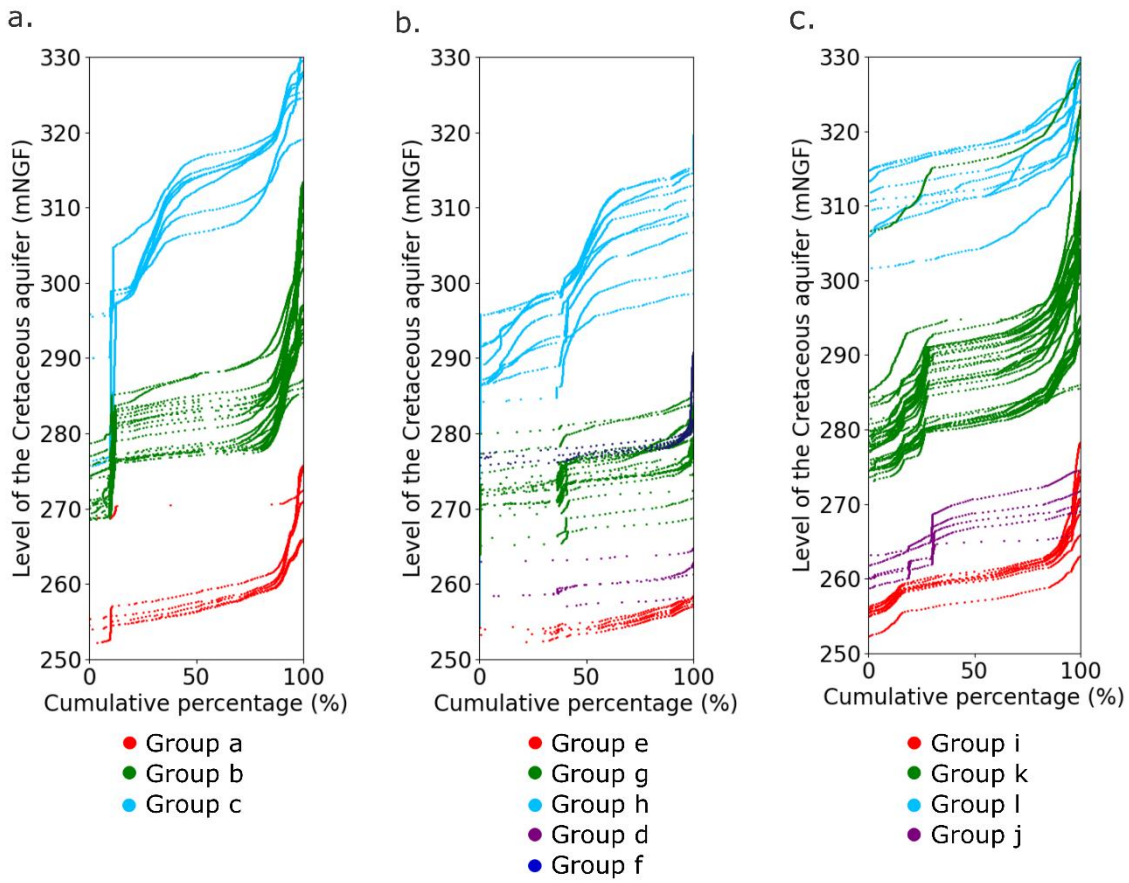


**Fig. 7** Sorted water level graphs and identified clusters. a. November 2011, b. March 2017, c. November 2019.

The spatial distribution of clusters have been represented in Fig. 8. It is remarkable to note that clusters (a, e, i) are superposed in one group (denoted 1), clusters b, g, k are superposed in another group (denoted 2) and clusters (c, h, j) are superposed in a third group (denoted 3). Group 4 corresponds to cluster d and j and group 5 to cluster f. Thus we can deduce that:

- each cluster {a,b,..,l} defined for each event is located in a physical area that can be identified (Fig. 8),

- the clusters defined by event are spatially overlapping and can thus be grouped spatially to define groups and geographical zones of coherent hydrodynamic behavior $\{Z_1, Z_2, .., Z_4\}$,

- the sorted water level curves are found in a coherent way by clusters according to the altitudes (Fig. 7),

- as clusters within the same zone do not have the same number of slope breaks across episodes, this indicates that the sorted water level curve is influenced by the event dynamics. As pointed out in the limit of the method in section 2, on a single event, the dynamics of the rainfall therefore influence the curve of the sorted water levels.

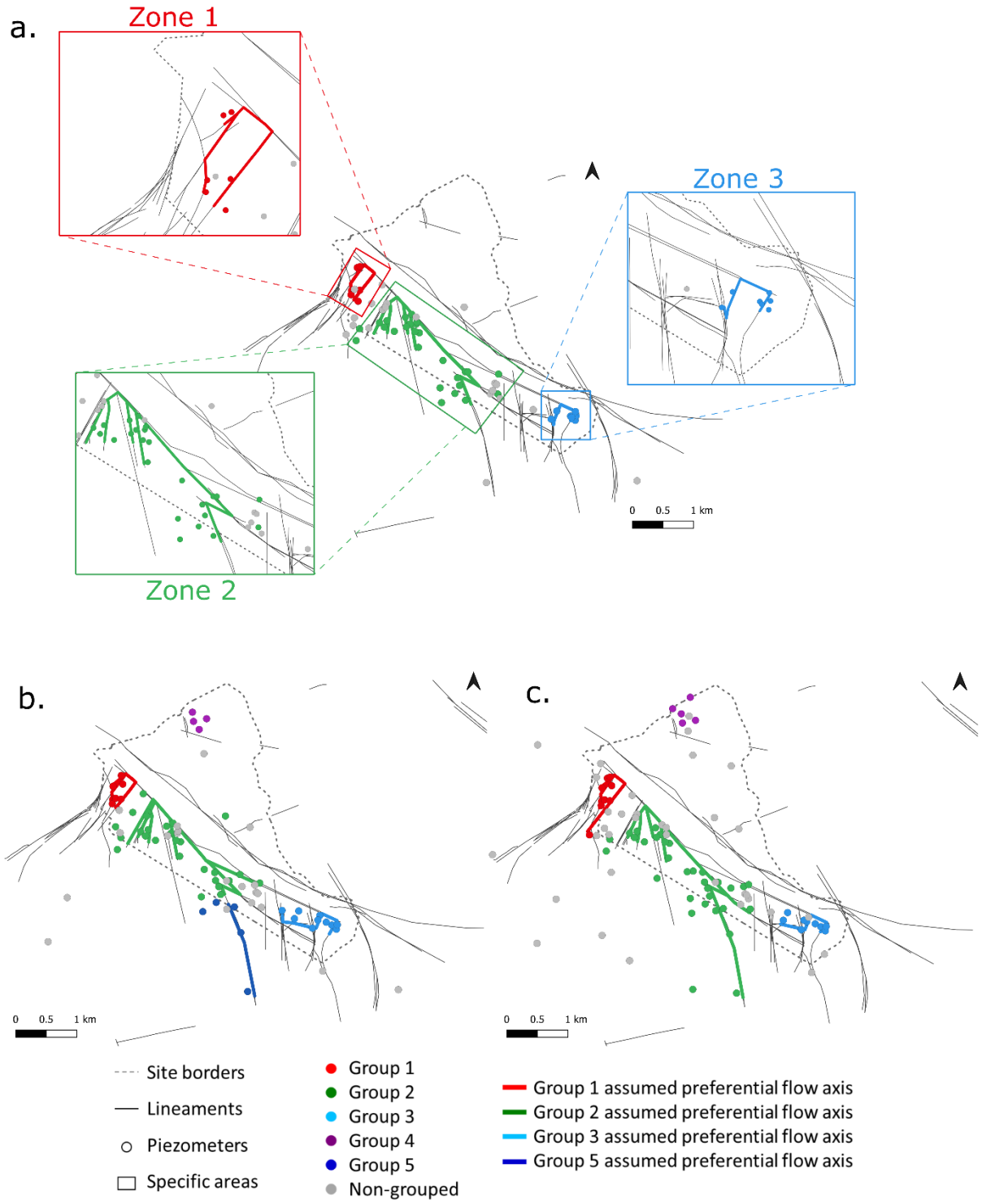


**Fig. 8** Spatial representation of clusters. a. November 2011, b. March 2017, c. November 2019, and assumed axes of connectivity (lineaments,faults).

### 3.2 Seasonality of piezometric variations

For the study of sorted groundwater levels on a seasonal scale, piezometers were selected according to the clustering made previously (Fig. 9). The objective is to represent the behavior by group of particular hydrodynamic behavior. It was thus retained: one piezometer from group 1 ($P_1$), one from group 2 ($P_2$), one for group 3 ($P_3$) and one non-grouped ($P_4$).

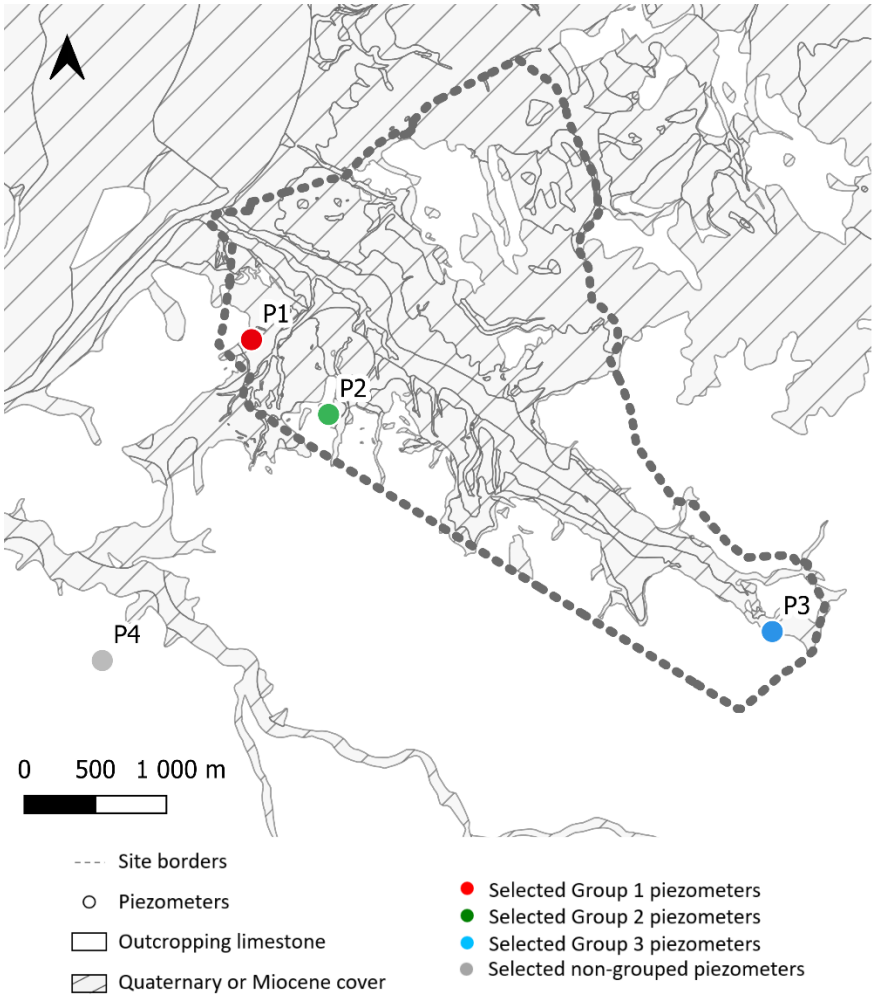


Fig. 9: Location of the piezometers selected for the study by season.

The water level curves classified by season are shown in Fig. 10. All the seasons (spring, summer, fall and winter) between fall 2008 and spring 2020 have been considered in order to minimize the effects of particular meteorological events and therefore to be more representative of the behavior of the Cretaceous aquifer.

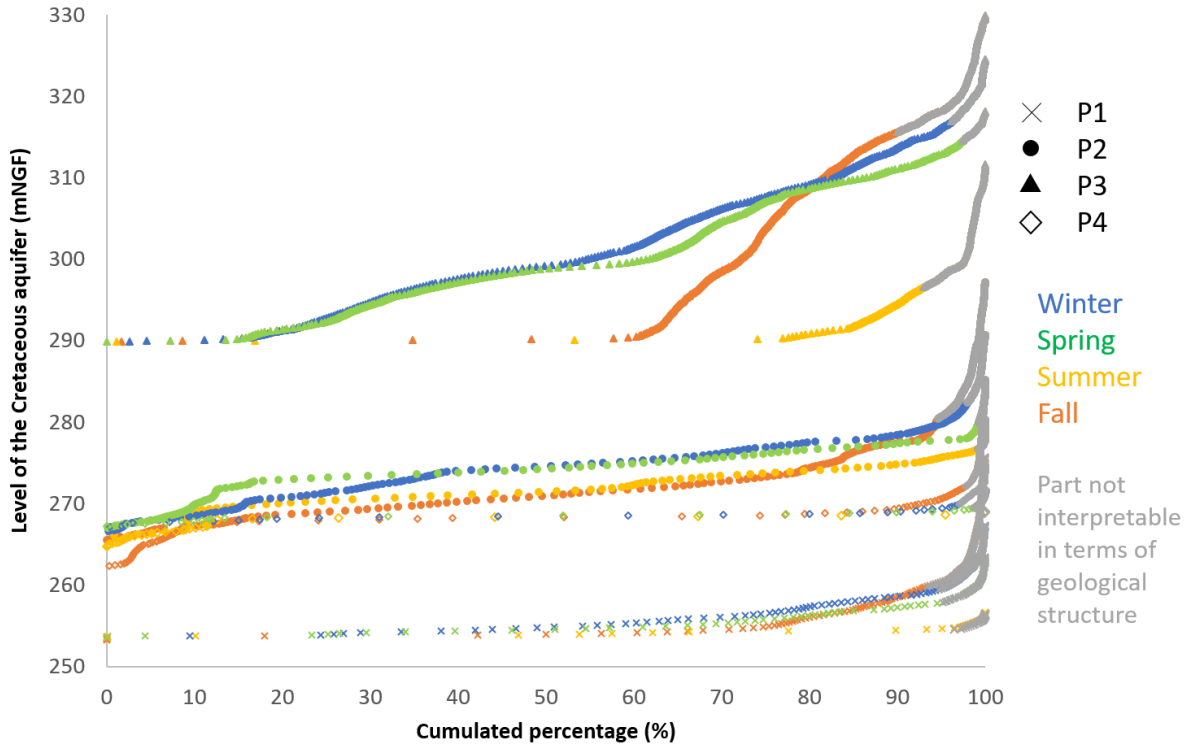


**Fig. 10** Graph of sorted groundwater levels by season

In Figure 10, we note that all curves become vertical for high percentages. These high values of water level are obtained for extreme events and are therefore rare, in small numbers. On the curves of sorted water levels, one thus notes a strong increase of the level for only some values (vertical slope). These high increases are therefore not interpretable in terms of geological structure, and we have represented them in grey in Figure 10.

These limits have been calculated as follows:

$$Limit(x) = med(x) + 2\sigma$$

With: $x$ the piezometric data, $med(x)$ the median and $\sigma$ the standard deviation of $x$.

According to Fig. 10, inside of each group, we find the same sequence: autumn presents the strongest amplitude of piezometric variations then winter, spring and summer. As mentioned in Table 2, on average over the period, winter had a lower cumulative rainfall than spring, so lower levels could be expected. However, in this region, intense rainfall events such as those of November 2011 and 2019 occur in autumn, so flood levels are still high in winter due to the discharge stage. Then, water levels are then also high in winter.

The piezometer $P_1$, has lower piezometric variations compared to $P_2$ and $P_3$ regardless of season. They can be linked to Miocene cover, corresponding to small variations of slopes of the sorted water levels. This indicates a less reactive zone than zones 2 and 3 although it is permeable on a large scale. The piezometer $P_2$ (central part of

the site), is the only with a lower level depending on the season. This shows that it has a more integrative property than the others. It depends more on the conditions of the season before, the initial level is lowest in autumn (after summer, usually dry in Mediterranean region); then in summer (after spring, a rainy period but also with important evapotranspiration), then in winter (after rainy autumn), then in spring (after winter when there is no evapotranspiration). The slope breaks indicate that the stucture of this zone is marked by a succession of more or less permeable zones, with low permeability layers of limestone matrix and more transmissive fracture/karstified layers. Concerning piezometer $P_3$, corresponding to group 3 (southern sector), the first element to note is the horizontal slope at 290 mNGF present for all seasons. This level corresponds to the "bottle bottom" level described in the interpretation grid of sorted groundwater levels (Fig. 4.c). This level is particularly frequent in summer and autumn, these two periods having the most marked low water levels. The slope breaks in the higher levels still indicate a succession of fractured zones and massive limestone. These observations confirm that, as in Zone 3, fracturing at the shallow depths plays an important role in the flow. $P_4$ is one of the non-clustered piezometers; it presents a low variability regardless of the season. This seems to indicate that it is not very connected to the rest of the aquifer. This may correspond to a piezometer that only intersects the matrix continuum.

## 4 Discussion

### 4.1 Links with the geological structures

After clustering the piezometers, the objective is to deepen the links between the type of sorted groundwater levels diagram and the geological structure of the reservoir. Once the areas of similar hydraulic behavior were identified (Fig. 8), we sought to understand how piezometers in the same cluster are connected, and to identify preferential flow paths. As the Cretaceous aquifer does not have well-developed karst conduits, the main hypothesis is that water flows are related to faults/lineaments (mainly subvertical on the site). In this karst, clay filling is considered marginal. The lineaments have been identified or supposedly determined from field observation, mapping and geological/hydrogeological models (GEOTER 2017). For this purpose, we have plotted the known lineaments on the map in Fig. 8.

Group 1 (Fig. 8), is located to the northern part of the site, at the northwest of the paleovalley $V_2$ (Fig. 1). This valley has a low permeability due to the Miocene sediments fillings. They act as a hydraulic barrier to the flows of the Cretaceous aquifer, especially during high groundwater level periods (Guerin 2001). The sorted water levels curves (Fig. 10) also shows the influence of the Miocene cover on the piezometric variations. In this area, the N20 faults seem to be preferential circulation paths. Pumping tests have been carried out on different boreholes in the

area and have shown synchronous piezometric evolutions, which attests to a good hydraulic continuity of the conductive fractures. This also assesses the relevance of the cluster 1.

The geological log of $P_1$ is available and can be associated to a borehole core. They are juxtaposed next to the curve of the sorted water levels (Fig. 11). A certain concordance between the information is apparent. The red dotted rectangle on Fig. 11 points to a zone (254.7 - 256.8 mNGF) where the limestone is massive and poorly fractured, and therefore has a low permeability. The underlying layer (254 - 254.7 mNGF) is more fractured and more permeable, as evidenced by the carrot. We can link to this succession a vertical increase of the sorted water levels of the borehole $P_1$. This corresponds to the interpretation proposed for the configuration of Figure 4.a: permeable layer underlying an impermeable layer. Similarly, the upper layer (256.8 - 258.8 mNGF) is permeable and explains that the slope of the water level decreases very strongly. Moreover, we can observe two small horizontal slopes around 258 mNGF and 258.8 mNGF (indicated by two dark blue dotted lines on Fig. 11). If we look at the log at these depths, it corresponds to the end of the karstified zone and the passage towards a more massive limestone. These transitions could correspond to high permeabilities explained by the stratification joint or an open karstified fracture. Recall, as shown in Figure 10, that the level above about 260 m is considered uninterpretable regarding the reservoir structure.

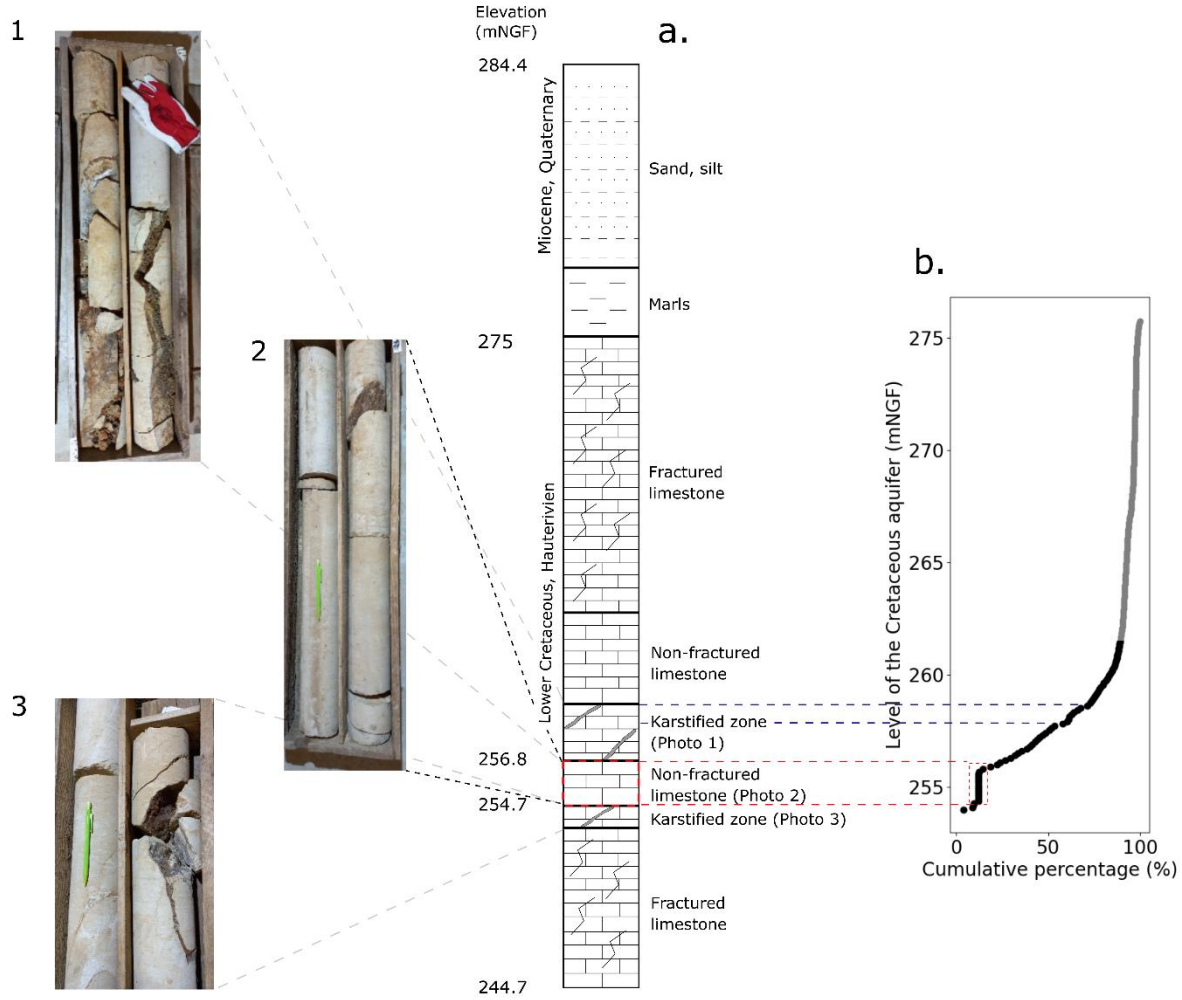


**Fig. 11** Multi-aspect analyze of a Group 1 piezometer. a. Simplified log of a borehole belonging to group 1 and photos of the borehole core: 1. Very fractured and karstified limestone 2. More massive limestone, 3. Karstified fracture. b. Sorted groundwater level curve of the piezometer for the 2011 event.

Group 2 corresponds to the central part that is the south of the paleovalley $V_1$, and presents N120-N150 plurikilometric and N40-N60 faults, locally more or less karstified (Guerin 2001). This can be related to the sorted water levels, which show slope heterogeneities at many elevations (Fig. 10). Group 3 to the southern part, is characterized by outcropping limestone and by the highest amplitude of rising water table recorded on the site (over 40 meters during the major floods). Fracturing is particularly heterogeneous in this zone, both in density and direction (Palasse et al. 1999). Group 4 is only present for the 2017 and 2019 events but not for the 2011 flood episode because the boreholes were drilled after 2011. The northern part of the $V_1$ valley (group 4) has a low density of piezometers. The behavior of the Cretaceous water table in this area is relatively unknown. Group 5, is only present for the 2017 event at the south of the site. For the other events, much intense, the piezometers were associated with the group 2. This difference can be explained by the existence of a fractures connectivity threshold in this area as shown in Fig. 5. During low water periods, the water table is not high enough to reach the depth were the fractures of the south and central part are connected. This highlights the major role of shallow fractures

on the flows circulation in this zone. This also shows the interest of carrying out this type of study at the event scale. The piezometers with non-groupable behaviors are spread over the site. They can correspond to non-representative piezometers (poor equipment, location...), very local matrix zones or geological specificities.

The study of the sorted groundwater levels, in addition to being used for clustering, also provides information on the role of geological structures with respect to the hydrodynamic behavior of the aquifer. In the Cadarache site, this highlights the role of fracturing and stratification joints in the flows. This methodology could be applied to other less well characterized sites in order to improve the knowledge of their hydrodynamic and geological properties.

## 5 Conclusion

As well as many karst aquifers, the faulted and karst aquifer of the Cadarache site (SE of France) is very sensitive to the rapid groundwater floods. The density of instrumented piezometers on this site allows to explore a statistic approach of spatial distribution of piezometry. The objective of this work is to improve the knowledge of the hydrodynamic functioning of this aquifer to better manage its flooding. Firstly, a semi-automatic method of clustering using sorted groundwater levels was developed and applied to a large amount of piezometric data while integrating a visual control of the observer. The need for operator control is a limitation of this protocol. Nevertheless, this protocol offers the advantage of being able to process a large amount of data thanks to a simple and understandable method providing hydrodynamic information in addition to the only clustering.

Secondly, a new interpretation grid of the sorted water levels is proposed. It allows identifying certain configurations where the heterogeneity created by the borehole interacts with the heterogeneity of the reservoir. Thirdly, based on the three majors flood events of the last 20 years the clustering provided several clusters. These groups can be related to three main distinct zones of hydrogeological behavior at the site scale: the northwest, central, and southeast zones. These areas are morphologically conditioned by geological structural elements. Finally, for the borehole where all three sources of information were available: stratigraphic log, core photography, and sorted groundwater levels, a concordance could be observed. This validates the richness of the information provided by the sorted curves and the interpretation grid.

For this aquifer, despite its complexity and the limited knowledge of its structure, the study of the link between the statistical approach and the geology of the site have allowed to identify preferential flow axes. Subvertical faults and fractures (NW-SE and NNE-SSW) appear to correspond to the preferential flow paths of the karst aquifer, especially at shallow depths.

The clustering method necessitates a good density of piezometers compared to the extent of the aquifer and scale of study. However, the sorted piezometric levels method should be an efficient and easy-to-use tool to understand the hydrodynamic behavior of karst aquifer based on few piezometers. In the continuity of the project, it is planned to improve the comparison protocol towards a more automatic clustering methodology integrating classical methods such as k-means, PCA, and to explore the potential of the analysis grid on other aquifers. .

**References**


Bailly-Comte V (2008) Interactions hydrodynamiques surface/souterrain en milieu karstique- Approche descriptive, analyse fonctionnelle et modélisation hydrologique appliquées au bassin versant expérimental du Coulazou, Causse d'Aumelas, France. These de doctorat, Hydrogéologie, Université Montpellier II

Bakalowicz M (1999) Connaissance et gestion des ressources en eaux souterraines dans les régions karstiques. SDAGE Rhône Méditerranée Corse

Boukharouba K, Roussel P, Dreyfus G, Johannet A (2013) Flash flood forecasting using Support Vector Regression: An event clustering based approach. In: 2013 IEEE International Workshop on Machine Learning for Signal Processing (MLSP). pp 1–6

Douglas DH, Peucker TK (1973) Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. Cartographica 10:112–122. https://doi.org/10.3138/FM57-6770-U75U-7727

Erguy M, Morilhat S, Artigue A, et al (2022) Application of statistical approaches to piezometry to improve the understanding of the karst aquifer hydrodynamic behaviours at the Cadarache CEA centre (France). Eurokarst 2022, Malaga

Fenart P (2008) Analyse préliminaire à l'étude de traçage. Centre CEA Cadarache - Sources de l'Abéou. Rapport CEA Cadarache

Ford D, Williams P (2007) Karst Hydrogeology and Geomorphology, John Wiley&Sons

GEOTER (2017) Bilan et hiérarchisation des failles cartographiées sur le site de Cadarache. CEA

Gholami V, Khaleghi MR, Pirasteh S, Booij MJ (2022) Comparison of Self-Organizing Map, Artificial Neural Network, and Co-Active Neuro-Fuzzy Inference System Methods in Simulating Groundwater Quality: Geospatial Artificial Intelligence. Water Resources Management: An International Journal, Published for the European Water Resources Association (EWRA) 36:451–469

Goldscheider N, Chen Z, Auler AS, et al (2020) Global distribution of carbonate rocks and karst water resources. Hydrogeol J 28:1661–1677. https://doi.org/10.1007/s10040-020-02139-5

Guerin R (2001) Synthèse des connaissances géologiques et hydrogéologiques. Rapport CEA Cadarache

Haaf E, Barthel R (2018) An inter-comparison of similarity-based methods for organisation and classification of groundwater hydrographs. Journal of Hydrology 559:222–237. https://doi.org/10.1016/j.jhydrol.2018.02.035

Han J-C, Huang Y, Li Z, et al (2016) Groundwater level prediction using a SOM-aided stepwise cluster inference model. Journal of Environmental Management 182:308–321. https://doi.org/10.1016/j.jenvman.2016.07.069

Johannet A, Mangin A, D'Hulst D (1994) Subterranean Water Infiltration Modelling by Neural Networks : Use of Water Source Flow. In Proc. of ICANN 1994. M. Marinaro and P.G. Morasso eds, Springer, pp 1033-1036

Jourde H, Mazzilli N, Lecoq N, et al (2015) KARSTMOD: A Generic Modular Reservoir Model Dedicated to Spring Discharge Modeling and Hydrodynamic Analysis in Karst. In: Andreo B, Carrasco F, Durán JJ, et al. (eds) Hydrogeological and Environmental Investigations in Karst Systems. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 339–344

Jourde H, Roesch A, Guinot V, Bailly-Comte V (2007) Dynamics and contribution of karst groundwater to surface flow during Mediterranean flood. Environ Geol 51:725–730. https://doi.org/10.1007/s00254-006-0386-y

Kong a Siou L, Johannet A, Borrell V, Pistre S (2011) Complexity selection of a neural network model for karst flood forecasting: The case of the Lez Basin (southern France). Journal of Hydrology 403:367–380. https://doi.org/10.1016/j.jhydrol.2011.04.015

Kong-A-Siou L, Fleury P, Johannet A, et al (2014) Performance and complementarity of two systemic models (reservoir and neural networks) used to simulate spring discharge and piezometry for a karst aquifer. Journal of Hydrology 519:3178–3192. https://doi.org/10.1016/j.jhydrol.2014.10.041

Kovács A, Sauter M (2007) Modelling karst hydrodynamics. In: Methods in Karst Hydrogeology. CRC Press

López-Chicano M, Calvache ML, Martín-Rosales W, Gisbert J (2002) Conditioning factors in flooding of karstic poljes—the case of the Zafarraya polje (South Spain). CATENA 49:331–352. https://doi.org/10.1016/S0341-8162(02)00053-X

Mangin A (1975) Contribution à l'étude hydrodynamique des aquifères karstiques. Thèse de doctorat, Sciences de la Terre, Université de Dijon

Marsaud B (1997) Structure et fonctionnement de la zone noyée des karsts à partir des résultats expérimentaux. Thèse de doctorat, BRGM, Université Paris XI Orsay

Naranjo-Fernández N, Guardiola-Albert C, Aguilera H, et al (2020) Clustering Groundwater Level Time Series of the Exploited Almonte-Marismas Aquifer in Southwest Spain. Water 12:1063. https://doi.org/10.3390/w12041063

Nourani V, Baghanam AH, Vousoughi FD, Alami MT (2012) Classification of Groundwater Level Data Using SOM to Develop ANN-Based Forecasting Model. International Journal of Computer and Information Engineering 2:464–469

Palasse J-R, Tauveron N, Guerin R, Hillel E (1999) Caractérisation géologique et hydrogéologique de la colline du Medecin : bilan des actions au 01/09/99. CEA Cadarache

Pinault J-L, Amraoui N, Golaz C (2005) Groundwater-induced flooding in macropore-dominated hydrological system in the context of climate changes. Water Resources Research 41:. https://doi.org/10.1029/2004WR003169

Raj P (2004) Classification and interpretation of piezometer well hydrographs in parts of southeastern peninsular India. Env Geol 46:808–819. https://doi.org/10.1007/s00254-004-1031-2

Taver V (2014) Caractérisation et modélisation hydrodynamique des karsts par réseaux de neurones: application à l'hydrosystème du Lez. Thèse de doctorat, Eaux Continentales et Société, Université Montpellier II

Wang X, Smith K, Hyndman R (2006) Characteristic-Based Clustering for Time Series Data. Data Min Knowl Disc 13:335–364. https://doi.org/10.1007/s10618-005-0039-x

Weng P, Dörfliger N (2002) Projet PACTES. Module : contribution des eaux souterraines aux crues et inondations ; site de l'Hérault. Rapport BRGM

Wunsch A, Liesch T, Broda S (2022) Feature-based Groundwater Hydrograph Clustering Using Unsupervised Self-Organizing Map-Ensembles. Water Resour. Manage. 36:39–54. https://doi.org/10.1007/s11269-021-03006-y

**Acknowledgments**

The authors wish to thank Mr. Patrick Lachassagne and Mr. David Labat for their contributions to some very interesting discussions.

**Author contributions**

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Manon Erguy, Anne Johannet, Séverin Pistre, Sébastien Morilhat, Julien Trincal and Guillaume Artigue . The first draft of the manuscript was written by Manon Erguy and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.”