

Applying the Principal Component Analysis for a deeper understanding of the groundwater system: case study of the Bacchiglione Basin (Veneto, Italy)

Applicazione della Principal Component Analysis per una maggiore comprensione del sistema acquifero monitorato: caso studio del bacino del Bacchiglione (Veneto, Italia)

Mara Meggiorin^a, Pierluigi Bullo^b, Valentina Accoto^b, Giulia Passadore^c, Andrea Sottani^b, Andrea Rinaldo^{c,d}

^a Ramboll Italy Srl - Vicenza, Italy - email: mara.meggiorin@dicea.unipd.it

^b Sinergeo Srl - Vicenza, Italy - email: pbullo@sinergeo.it; vaccoto@sinergeo.it; asottani@sinergeo.it

^c Dipartimento di Ingegneria Civile, Edile e Ambientale, Università di Padova - email: giulia.passadore@dicea.unipd.it

^d Laboratory of Echohydrology (ECHO/IIE/ENAC), École Polytechnique Fédérale de Lausanne, Switzerland - email: andrea.rinaldo@epfl.ch

ARTICLE INFO

Ricevuto/Received: 02 May 2022

Accettato/Accepted: 03 June 2022

Publicato online/Published online:
30 June 2022

Handling Editor:
Daniele Pedretti

Publication note:

This contribution has been selected from Flowpath 2021 congress held in Naples 1-3 December 2021

Citation:

Meggiorin M, Bullo P, Accoto V, Passadore G, Sottani A, Rinaldo A (2022) Applying the Principal Component Analysis for a deeper understanding of the groundwater system: case study of the Bacchiglione Basin (Veneto, Italy). *Acque Sotterranee - Italian Journal of Groundwater*, 11(2), 07 - 17
<https://doi.org/10.7343/as-2022-573>

Correspondence to:

Mara Meggiorin ✉
mara.meggiorin@dicea.unipd.it

Keywords: Groundwater, Principal Component Analysis, Conceptual model, Bacchiglione.

Parole chiave: acque sotterranee, Principal Component Analysis, modello concettuale, Bacchiglione.

Copyright: © 2022 by the authors. License Associazione Acque Sotterranee. This is an open access article under the CC BY-NC-ND license: <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Riassunto

In idrogeologia è spesso difficile comprendere appieno le componenti della ricarica delle idrostrutture sotterranee. In particolare, il sistema idrico sotterraneo può essere soggetto a diversi meccanismi di alimentazione e migrazione, quali, ad esempio: la permeabilità del suolo, la presenza di paleoalvei e le precipitazioni efficaci. Alcuni traccianti, tra i quali la temperatura delle acque di falda o composti chimici specifici ivi disciolti, possono migliorare la comprensione delle dinamiche in gioco. Tuttavia, su scala regionale le informazioni chimico fisiche sono generalmente scarse, mentre il monitoraggio piezometrico è il dato di più frequente acquisizione. Questo studio quindi si prefigge lo scopo di sfruttare le misurazioni dei trend idrogeologici per validare ed eventualmente approfondire i modelli geologico strutturali dei sistemi acquiferi locali. Lo studio è consistito nel raggruppare i punti di controllo sulla base delle variazioni medie annue della quota piezometrica in ordine a: (i) identificare tramite i cluster aree con un comportamento idrogeologico simile e (ii) studiare i fattori di ricarica analizzando la variazione annuale della quota piezometrica del cluster. Poiché il raggruppamento visivo in cluster può essere una procedura lunga e soggettiva, lo studio propone l'uso dell'analisi delle componenti principali dei regimi piezometrici. Nel dettaglio lo studio è stato applicato ai dati del monitoraggio piezometrico di una serie di pozzi distribuiti nella pianura alluvionale del Fiume Bacchiglione ottenendo come principali risultati (i) cluster definiti in base al numero, momento e lunghezza dei massimi e minimi delle variazioni medie annue della quota piezometrica, (ii) cluster geograficamente ben raggruppati, confermando la dipendenza dell'idrogramma delle acque sotterranee da fattori idrogeologici locali. Inoltre, l'analisi delle anomalie dei risultati ha evidenziato l'effetto della presenza di serie temporali con periodi di monitoraggio differenti, osservando però anche la capacità del metodo di individuare un eventuale cambiamento dell'idrogramma caratteristico del cluster nel corso degli anni.

Abstract

In hydrogeology, it is often difficult to fully understand the hydraulic factors affecting the recharge of groundwater systems. Particularly, at a regional scale, the groundwater system can have different drivers depending on the considered area, i.e., soil permeability, paleochannels, and precipitation. Chemical-physical (i.e. temperature) or hydrogeochemical data can help such understanding. However, this type of information is usually sparse at the regional scale, whereas extended groundwater piezometric head monitoring is more common. This study aims at exploiting these longitudinal observations of the hydraulic head to validate (and possibly bring more insights into) the geological structural model of aquifer systems. Clustering control points based on the piezometric head average annual variations can help the system conceptualization in two ways: (i) clusters can geographically identify areas with similar hydrogeological behavior; and (ii) the typical cluster annual variation with its ups and downs can bring insights on the recharge component of an aquifer system. Nevertheless, visual clustering can be a long and subjective procedure, thus this study suggests the use of the Principal Component Analysis to cluster the control points with a similar average annual variation of their recorded time series. This study supports the proposed analysis by applying it to the monitoring data of the Bacchiglione basin resulting in (i) clusters identified based on the number, moment, and lengths of groundwater level peaks and minima, (ii) well-gathered clusters in space, underpinning the groundwater hydrograph dependence on local driving factors. Furthermore, the investigation of clustering anomalies highlighted the relevance of the presence of time series with different recording periods pinpointing, however, the method's capacity to spot a change in the hydrogeological cycle over the years.

Introduction

Worldwide, groundwater resources are a major player in supporting surface ecosystems and economic and human development (Freeze and Cherry 1979; Giordano 2009), but their study and understanding are often complicated. Thus, resource exploitation often comes without a robust comprehension of the groundwater system increasing the risk of over-exploitation.

It is generally difficult to sustainably manage the groundwater resource (Giordano 2009) because of the complexity of fully understanding the main stresses and how to counteract them: conceptual models and hydraulic factors affecting the recharge of groundwater systems are not always obvious (Earman 2011). At a regional scale, groundwater systems can be highly complicated due, for example, to the complex distribution of lithologies and the transient flows connected to the hydrologic cycle (Fitts 2002): different areas of the system can be affected by different hydrogeological structures (e.g., impermeable layers, paleochannels, tectonic features) and hydraulic conditions (e.g., precipitation, river dispersion, snow melting, irrigation). A deeper understanding is sometimes achievable with tracers, studying chemical transport or temperature (as, for example, in Vettorello and Sottani 2019). However, this type of information is usually rare and sparse at the regional scale, whereas extended groundwater piezometric head monitoring is more common. Therefore, this study aims at exploiting this common widespread type of information, gathering insights into groundwater systems.

Specifically, this study suggests applying the Principal Component Analysis (PCA) (Pearson 1901) to analyze these longitudinal observations of available groundwater heads (hereafter termed 'groundwater levels' for simplicity) in order to validate and bring new insights into the geological structural model and into the hydraulic conditions of the aquifer systems. Clustering the control points depending on the piezometric head average annual variations can help the system conceptualization in two ways: (i) clusters can geographically identify areas with similar hydrogeological behavior; and (ii) the typical cluster annual variation with its ups and downs can bring insights on the groundwater recharge component of an aquifer system. As clustering method, the PCA is preferred because visual clustering can be a long and subjective procedure (Barthel et al. 2022), and other statistical clustering methods often require the subjective definition of different parameters that influence the whole process (Aghabozorgi et al. 2015).

This study supports the use of PCA for achieving insights into the groundwater system by applying it to the monitoring data of the Bacchiglione basin within the Veneto Plain (IT): the studied area includes a complex hydrogeological setting with an upstream unconfined aquifer and a downstream multi-layered confined aquifer system. Additionally, the presence of paleochannels and a pinch-out structure of aquitards complicate the local groundwater flow (Passadore et al. 2012). This domain has been often studied because of

the high exploitation of the groundwater resource and the recorded decreasing groundwater levels (Passadore et al. 2012; Meggiorin et al. 2021) and, consequently, some Managed Aquifer Recharge experiences have been there implemented (e.g., Altissimo et al. 2014; Mezzalira et al. 2014; Meggiorin et al. 2019).

Therein, the groundwater piezometric head is continuously monitored by a spread information-rich monitoring network. Nevertheless, the longitudinal data has to be preprocessed by removing outliers and by extrapolating normalized average annual variations of the groundwater elevation, a necessary step to highlight local seasonality and avoid clouding the analysis with the mean or the oscillation range of the piezometric head. Following, identified clusters are studied and associated with the local conceptual model.

Study Area

The selected study area is the so-called Bacchiglione Basin (shown in Figure 1), located in the Veneto plain (IT) and covering the alluvial plain north of Vicenza. This hydrogeological basin has been already studied in several previous studies (e.g., Dal Prà et al. 1976, 1977; Sottani et al. 1982; Altissimo et al. 1999; Passadore 2008; Marcolongo 2012; Passadore et al. 2012; Sottani and Vielmo 2014; Rinaldo and Passadore 2018; Meggiorin et al. 2021) because of (i) the complex structural system, (ii) its historical water abundance, (iii) its quantitative depletion over the years, and (iv) the widespread monitoring network that continuously measures the groundwater elevation in several control points. Thus, within this study area, a sizeable dataset is available, particularly suitable for statistical analyses.

The central area of the Veneto Region has always been characterized by an abundance of groundwater resources due mainly to (i) the permeable subsoil and (ii) the great rivers' recharge to the aquifer system. Indeed, in the upper plain close to the mountains there is a continuous undifferentiated graveled subsoil hosting an unconfined aquifer system. The underlying bedrock is located at a depth of a few meters in the upper Veneto plain and it deepens to several hundred meters going downstream, towards South-East. Additionally, going downstream, the coarse material thickness decreases, and layers of fine materials appear and thicken in the lower plain (Dal Prà et al. 1976, 1977). This downstream stratified complex hosts a layered multi-aquifer system.

Data and Methods

The proprietary dataset used for the following analysis corresponds with the one used in Meggiorin et al. (2021), where more details on the monitoring network are available.

As a brief summary of the available data, the piezometric head is measured in 102 monitoring points with a sampling interval of 1 or 3 hours and available recordings have different lengths: ranging from 3 months to almost 14 years. In addition to this heterogeneous recorded time period, the dataset is affected by missing values and outliers, as usual for real case studies. The firsts are present because of (i)

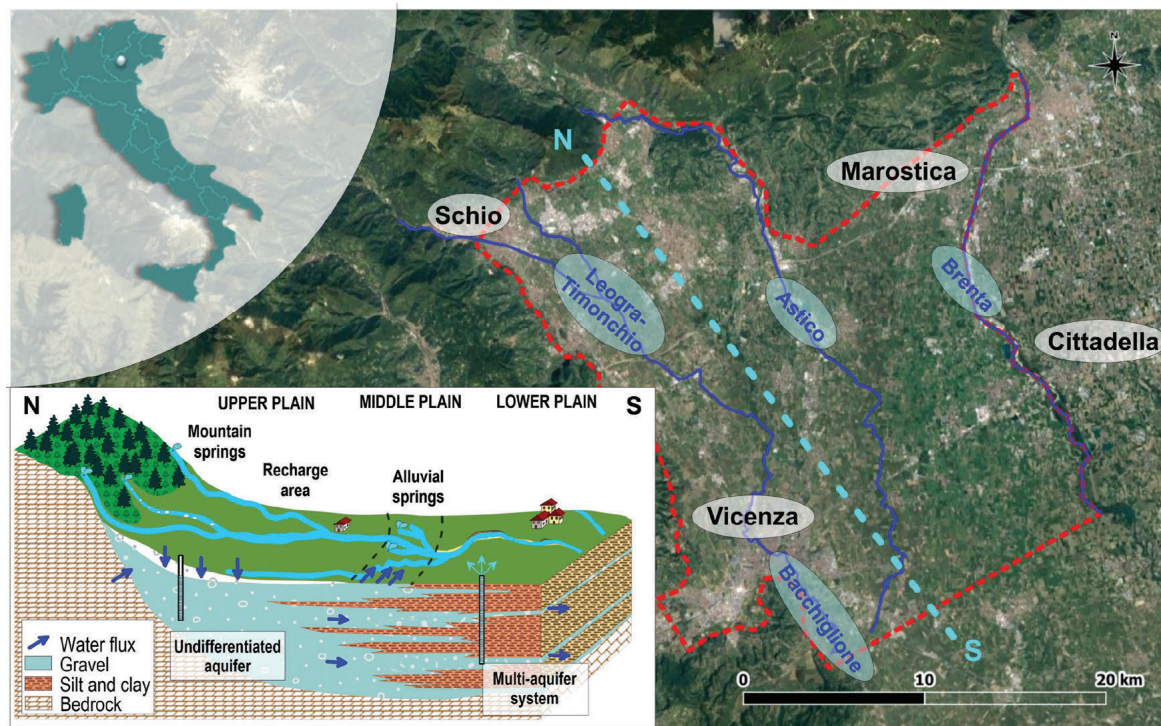


Fig. 1 - Plan view of the study area with main cities, main river tracks and the reference geological scheme of the Veneto plain. Image by Meggiorin et al. (2021).

Fig. 1 - Vista in pianta dell'area di studio con riportate le città principali, i fiumi più importanti nel territorio e lo schema geologico di riferimento della pianura veneta. Immagine di Meggiorin et al. (2021).

different sensors' installation times, (ii) sensor breakdowns, or (iii) particularly dry periods for which the sensor is not located deep enough, thus being suspended above the water table. The latter can be present due to sensor malfunctions or to the pumping effect of nearby wells. Outliers have to be removed for applying the statistical analyses and, in the present study, the same procedure as in Meggiorin et al. (2021) has been applied. Starting from that dataset, a further sensor-specific documental analysis led to removing 17 points because too close to pumping wells, likely affecting the recorded piezometric head. Indeed, this study aims to strengthen the conceptual model of the groundwater system, therefore focusing on the natural groundwater level trend. Records potentially affected by nearby pumping wells could mislead the analysis. Furthermore, to achieve robust results, records shorter than two years or with a majority of missing values within the recorded period have been removed from the dataset. In conclusion, the dataset includes 63 time series.

Finally, the dataset preparation for the PCA involves that time series have been processed as follows: (i) original time series are centered around the 0 value, (ii) records have been monthly averaged in order to have a uniform monthly timestep, (iii) monthly time series are normalized to range between -1 and 1, and (iv) typical Yearly Hydrographs are calculated by averaging records for each month (see Figure 2).

The two preprocessing steps of centering and normalizing the time series between -1 and 1 are necessary for the purpose of this analysis, otherwise, the average groundwater elevation and its oscillation range would mislead the PCA because those

aspects would influence the whole clustering. Even if those are certainly relevant aspects, the present study prefers to focus on the oscillations' seasonality in order to achieve the desired insights into the conceptual model of the groundwater system.

The PCA is applied in the present study because multivariate statistics can bring insights into the system by simultaneously looking at relationships between multiple variables and revealing hidden patterns (Bro and Smilde 2014). The PCA is a powerful statistical procedure that does not require any assumption on data structure, even if it uses the Pearson coefficient, thus considering only linear correlation.

Briefly, PCA aims to orthogonally transform the dataset by identifying the smallest number of principal components (PCs) that are able to account for most of the original dataset variability (Gunasekaran and Kasirajan 2017). The first PC accounts for most of the data variability and the following PCs explain the highest left variability, being orthogonal to all previous components. Thus, first few evaluated PCs explain most of the variability of the original dataset: hopefully first two or three components entail more than 70-80% of the original dataset variability. If this is the case, useful insights can be observed by PCs scatter plots, called factorial plans, showing how and if observations group along with first PCs. The remaining PCs contain much less information, and they can be neglected (Ribeiro et al. 2015). PCA is a well-known statistical method that has been already widely applied in many different fields, including the hydrogeochemical field analyzing physico-chemical parameters (Palmucci et al. 2016; Barbagli et al. 2019; Di Curzio 2019) and the hydrological field for investigating

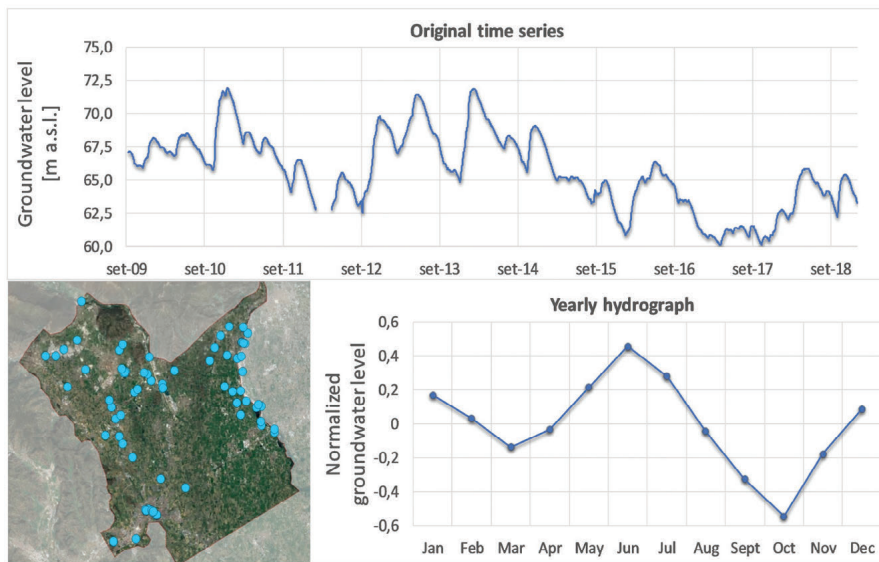


Fig. 2 - Map of the present monitoring network and example of preprocessed time series for control point 169. The example time series is plotted as the original time series with a total length of almost 10 years and as its yearly hydrograph.

Fig. 2 - Mappa della rete di monitoraggio esistente nell'area di studio e la serie storica al punto 169, di esempio per la fase di preprocessing del dataset. La serie storica di esempio è rappresentata sia come la serie originale a disposizione, lunga quasi 10 anni, che come il suo corrispettivo idrogramma annuale.

spatio-temporal relationships between monitoring stations of an existing network (e.g., Lins 1985; Winter et al. 2000; Menciò and Mas-Pla 2008; Ribeiro et al. 2015).

In the present study, the identified PCs represent reference groundwater yearly hydrographs, their linear combination explains most of the variability of the field hydrographs, and the factorial plan shows how the field hydrographs relate to the reference ones. Particularly, clusters are identified by observing records' distribution in the factorial plan, where hydrographs scores are plotted against PCs. The clusters have been hand-selected on the factorial plan, thus the split has been defined subjectively by observing the data spread and without defining a priori the number of clusters to identify.

Results

The PCA is applied to the dataset of yearly hydrographs. The first two identified PCs explain more than 68% of the dataset variability and they represent 'Reference Hydrographs' showing peaks and minima of the groundwater elevation. Each field hydrograph is a linear combination of the two Reference Hydrographs and their weights are visible in the factorial plan.

The PCA results are reported in Figures 3 and 4 respectively as Reference Hydrographs and as Factorial plan. The first PC, explaining most of the dataset variability, is a groundwater hydrograph showing maximum groundwater levels between December and March while minimum levels are between July and September. On the contrary, the second PC has the highest values between April and June and the lowest levels between October and November. These two Reference Hydrographs explain most of the original dataset and their weights for each field hydrograph are shown in the Factorial Plan, Figure 4. In this graph, observations can be grouped in clusters being similarly explained by main PCs, therefore being similarly characterized by the Reference Hydrographs. Nevertheless, in this specific case, the factorial plan is not particularly enlightening because clusters are not easily identifiable.

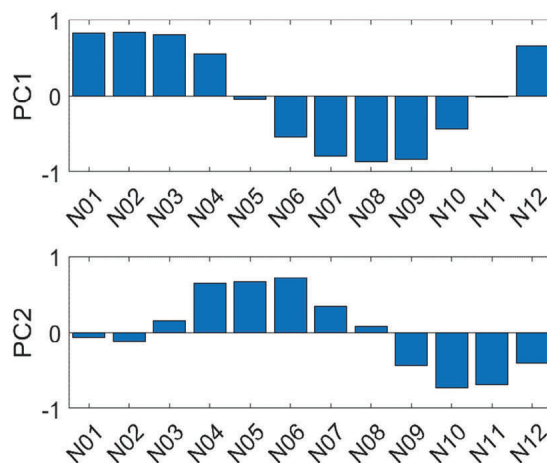


Fig. 3 - Identified principal components for the whole available dataset.

Fig. 3 - Componenti principali individuate dall'analisi dell'intero dataset disponibile.

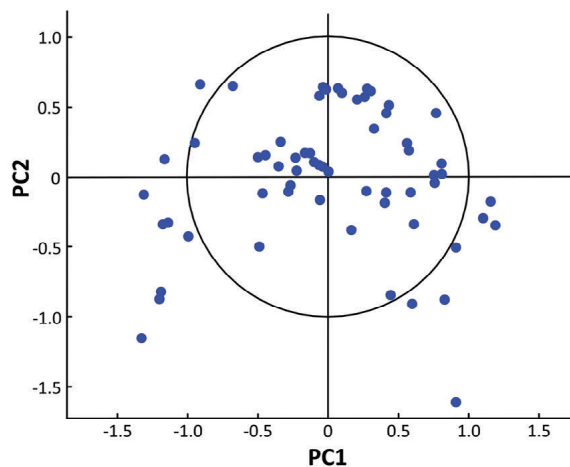


Fig. 4 - Factorial plan of the whole available dataset.

Fig. 4 - Piano fattoriale elaborato sui risultati dell'analisi dell'intero dataset disponibile.

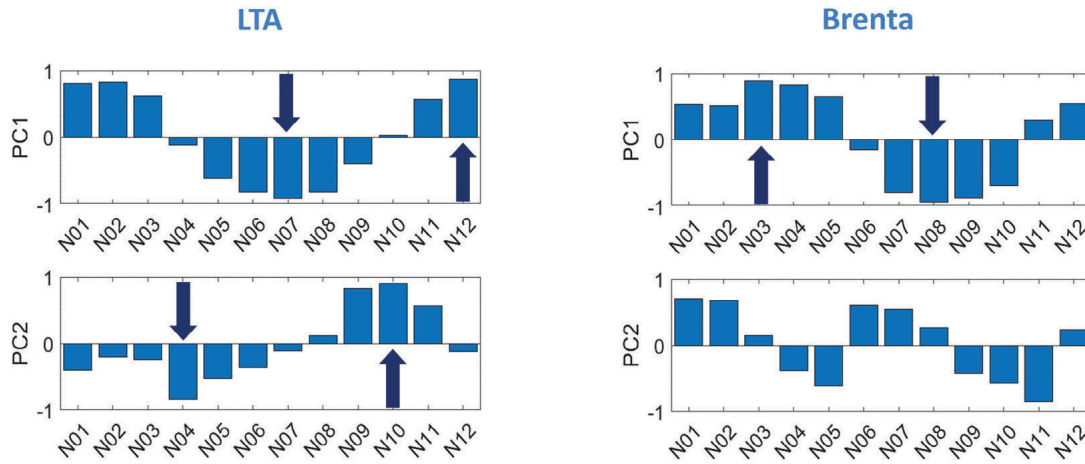


Fig. 5 - Identified principal components for the sub-datasets. The blue arrows highlight the identified hydraulic peaks: minimum and maximum.

Fig. 5 - Componenti principali individuate dall'analisi dei due sub-dataset. Le frecce blu sottolineano i picchi idraulici individuati: massimi e minimi.

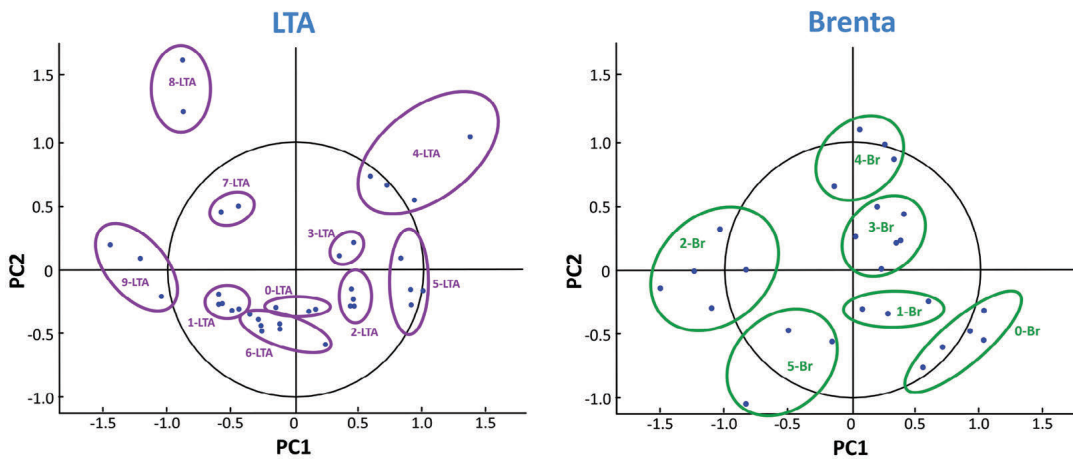


Fig. 6 - Factorial plans of the sub-datasets.

Fig. 6 - Piano fattoriale elaborato sui risultati dell'analisi dei due sub-dataset.

Consequently, the dataset has been split into two sub-domains depending on the hydrological basin, which in the study area is strongly connected with the underground hydrogeological basin: the Leogra-Timonchio-Astico rivers (LTA) at West and the Brenta river at East. Figures 5 and 6 show the results of the PCA for the two sub-datasets. Particularly, it is evident how the two domains have different Reference Hydrographs: the first PC has the maximum and minimum respectively in December and July for the LTA domain, in March and August for the Brenta domain. Therefore, in both datasets, the period with minimum groundwater levels is within the summer season, while the maximum is in early winter for the LTA while it is in late winter for the Brenta domain. The second PC also differs in the two datasets: the LTA Reference Hydrograph shows only one peak in October and one minimum in April, on the contrary, the Brenta hydrograph shows two minima and maxima along the year.

Factorial plans for the two sub-datasets are shown in Figure 6 and, in this case, the visual clustering comes relatively easy because of the great spread of the observations within the graph. From these results, 6 clusters are identified within the Brenta domain and 10 within the LTA domain. The same clusters are plotted above the original factorial plan calculated

in the first overall analysis (visible in Figure 7) showing how the clusters were already present in such analysis, even if it was more difficult to identify them. Additionally, the two sub-domains are relatively departed in the overall factorial plan: only two small LTA clusters overlap with Brenta clusters.

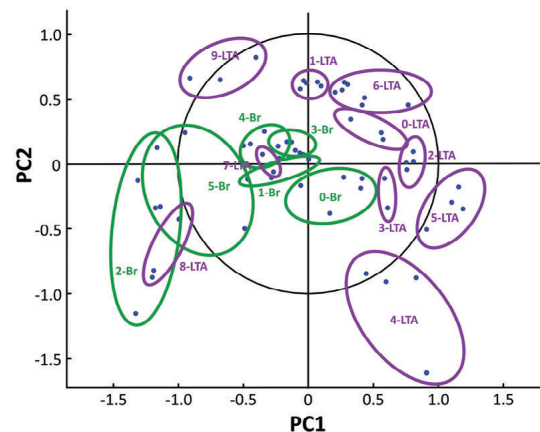


Fig. 7 - Factorial plan of the whole available dataset with highlighted the clusters identified by the sub-basins analysis.

Fig. 7 - Piano fattoriale elaborato sui risultati dell'analisi dell'intero dataset disponibile con evidenziati i cluster identificati dall'analisi dei due sottobacini.

A visual check of the clustering has been carried out by observing grouped hydrographs, results are visible in Figures 8 and 9. The overall comprehension is that the PCA well identifies the clusters depending on the number, location, and lengths of groundwater level peaks and minima. For example, for the LTA sub-dataset, the two hydrograph aspects

that are most relevant in differentiating the clusters are (i) the location and length of the maximum groundwater level and (ii) the groundwater trend during the fall rise, between October and December. On the contrary, for the Brenta sub-dataset, the most relevant hydrographs' characteristics are (i) the maximum magnitude and location and (ii) the groundwater trend during the early fall season.

Legend:

Different cluster trends during

- the winter-spring rise
- the fall rise

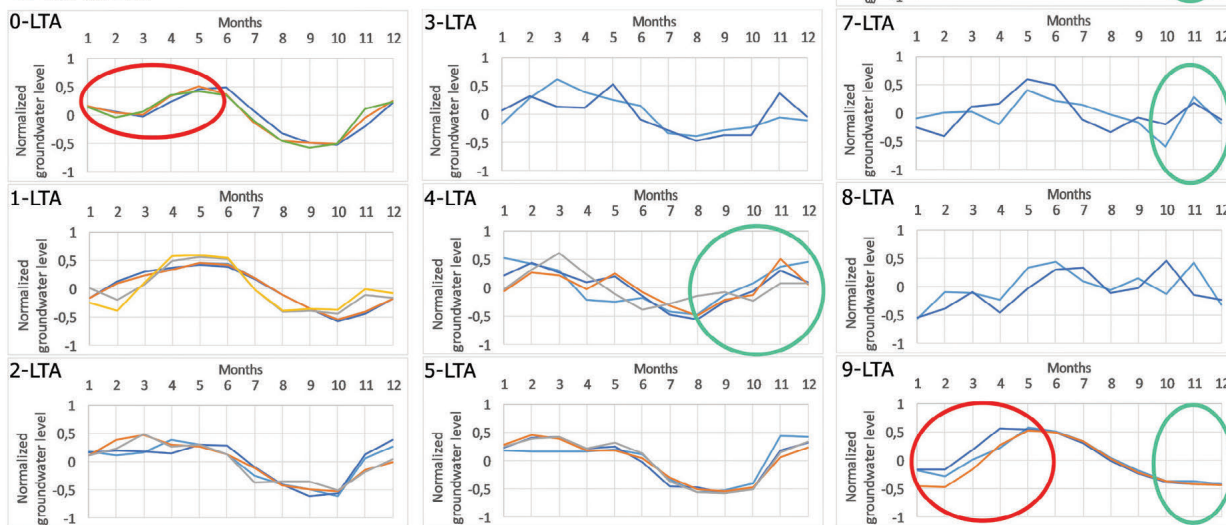
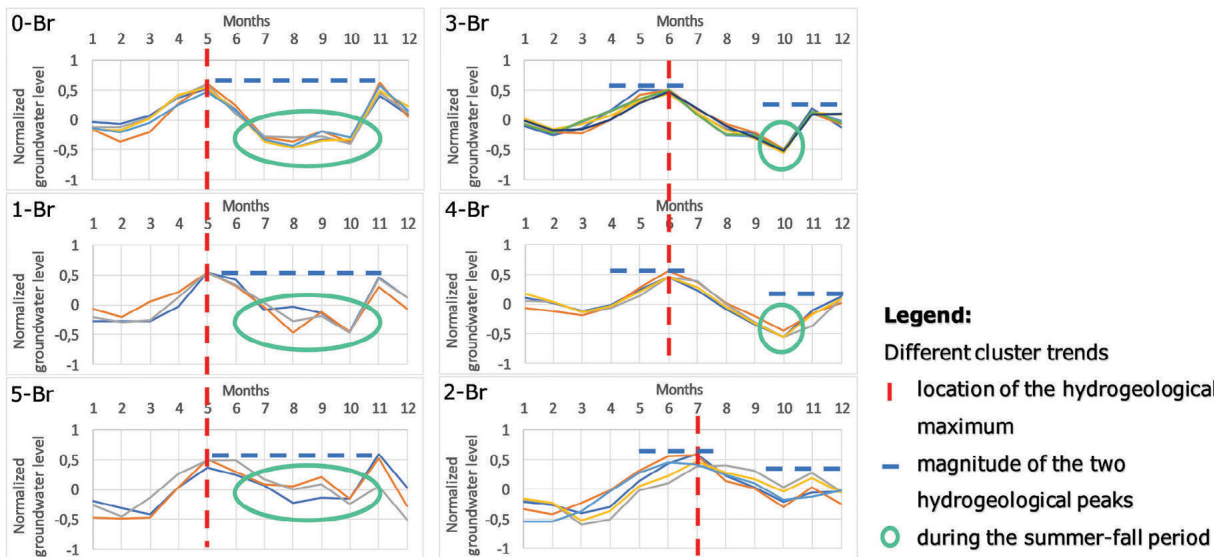


Fig. 8 - Clustered groundwater normalized yearly hydrographs for the LTA sub-basin: highlighted in red and green the characteristics mostly differentiating the clusters, the winter-spring rise in red and the fall rise in green.

Fig. 8 - Cluster degli idrogrammi annuali del livello normalizzato delle acque sotterranee per il sottobacino LTA: nella figura sono evidenziati i comportamenti osservati che maggiormente differenziano i cluster, in rosso i differenti trend invernali-primaverili di risalita del livello, mentre in verde sono evidenziati i differenti trend autunnali.



Legend:

- Different cluster trends
- | location of the hydrogeological maximum
 - magnitude of the two hydrogeological peaks
 - during the summer-fall period

Fig. 9 - Clustered groundwater normalized yearly hydrographs for the Brenta sub-basin: highlighted the characteristics mostly differentiating the clusters, in red different locations of the hydrogeological maximum, in blue the different magnitude of spring and fall peaks, and in green different summer-fall behaviors.

Fig. 9 - Cluster degli idrogrammi annuali del livello normalizzato delle acque sotterranee per il sottobacino Brenta: nella figura sono evidenziati i comportamenti osservati che maggiormente differenziano i cluster, in rosso sono evidenziati i differenti momenti del massimo idrologico, in blu si sottolinea la differente entità dei picchi primaverile e autunnale, mentre in verde sono evidenziati i differenti comportamenti estivi-autunnali.

Discussion

This study applied the PCA to analyze the available groundwater level dataset to identify hydrogeologically homogeneous areas. Specifically, the PCA examines the average yearly hydrograph of all control points within the study area in order to cluster them and identify areas characterized by similar hydrographs. Normalized average yearly hydrographs were preferred to the original records (with observations roughly between 2005 and 2019) because they allow having a uniform dataset without missing values and focused on the average groundwater level seasonality.

The main hydrogeological results of this analysis are reported in Figures 8 and 9. Clustered yearly hydrographs show how the analysis is able to differentiate such hydrographs depending on their peculiar trend. On a few occasions, a later visual inspection would lead to joining clusters or moving a time series from one cluster to another. Nevertheless, looking at Figures 8 and 9, the overall clustering is satisfying.

Within the LTA domain, the analyzed yearly hydrographs are grouped in 10 different clusters. Figure 8 highlights how some clusters group well-aligned hydrographs (e.g., clusters 0-LTA, 6-LTA, and 9-LTA) while others group couples of hydrographs that are visibly less aligned (e.g., clusters 3-LTA, 7-LTA, and 8-LTA). Additionally, a visual inspection could also join a few clusters, such as 0-LTA with 6-LTA and 2-LTA with 5-LTA, even if they have been split by the PCA. However, the split has been defined subjectively by observing the factorial plan in Figure 6, where indeed clusters 0-LTA, 1-LTA, 6-LTA and 2-LTA, 3-LTA, 5-LTA are respectively

closely placed. Another analysis operator could have decided to join them in two larger clusters in the first place.

On the contrary, within the Brenta domain, in Figure 6 the observations in the factorial plan are more spread, clusters defined with no uncertainty, and, indeed, all analyzed yearly hydrographs are well grouped and aligned. Figure 9 highlights how the location and magnitude of the hydrographs' peaks are differentiated as well as the early fall behavior. Specifically, the difference in magnitude of the two peaks can be insightful for understanding local recharging factors. Indeed, the greater response to recharge in the spring period would be attributable to the contribution of the Brenta water dispersion with the contribution of rain from outside the study area. Furthermore, the contribution of the snow melting that reaches the Brenta river before flowing into the plain must be considered. On the other hand, a similar value of spring and fall peaks suggests the absence of the external contribution related to the snow melting. The contemporaneity of spring peaks in clusters 0-Br, 1-Br, and 5-Br is observed as they are directly connected to the river flow. On the contrary, the peak delay in clusters 3-Br and 4-Br would be attributable to a deep feeding of the sub-river aquifer, while the further delay in cluster 2-Br is linked to the slower path towards the actual aquifer.

Observing the location of the clustered yearly hydrographs within the Bacchiglione basin, Figure 10 highlights how most of the clusters well gather the control points in space, underpinning the groundwater hydrograph dependence on the local underground context and local water balance.

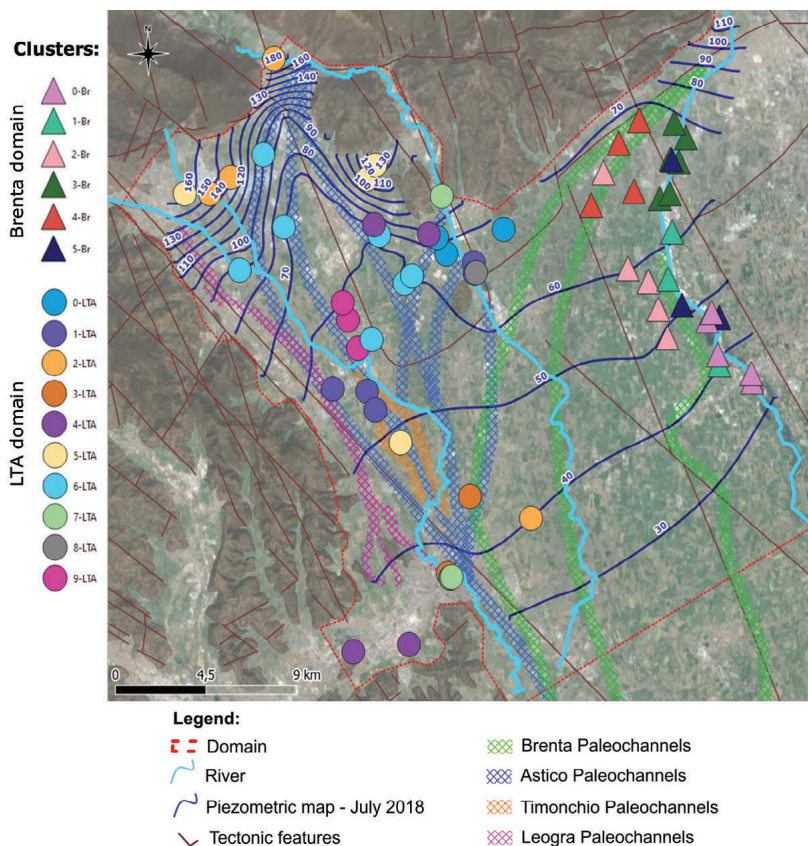


Fig. 10 - Location of clustered groundwater normalized yearly hydrographs for the whole Bacchiglione basin and local hydrogeological driving factors: tectonic features in brown, paleochannels in colored diamond patterns, the local piezometric map in blue, and main rivers in light blue.

Fig. 10 - Posizione degli idrogrammi annuali, raggruppati in cluster, per il bacino Bacchiglione e i drivers idrogeologici principali: strutture tettoniche in marrone, paleovalvei rappresentati con pattern a rombi colorati, la mappa piezometrica locale in blue e i principali fiumi in celeste.

Indeed, clusters are usually geographically well gathered, especially within the Brenta domain, even if a few exceptions are present. Additionally, it is evident how the Brenta domain is more orderly while the LTA domain is multifaceted.

Going into more detail on the reasons why these clusters are identified and differentiated by the PCA, the authors made an effort to pinpoint the hydrogeological driving factors of each cluster based on the geological and hydrogeological available knowledge. A summary is reported in Table 1. The driving factors can be generally classified as (i) tectonic features (i.e., faults) that determine the saturated thickness, (ii) geomorphological features such as impermeable layers position and the paleochannels presence, (iii) hydrogeological features such as the presence of graveled channels below the riverbeds, (iv) recharge proximity (i.e., river dispersion and drainage).

Tab. 1 - Summary of the main pinpointed contributions to recharge and driving factors for each cluster.

Tab. 1 - Riassunto dei principali contributi di ricarica e di altri fattori rilevanti identificati per ogni cluster.

Area	Group Cluster	Main contribution to recharge and driving factor
LTA	0 - 1 - 6	River dispersion, secondary rain Unconfined aquifer with high gravel thickness
	2 - 3 - 4 - 5	Local rainfall Shallow aquifer
	7 - 8	Anomalous
	9	River dispersion, no rain evidence Confined aquifer by an impermeable layer
Brenta	0 - 1	Local rainfall and riverbed contribution. Riverbed aquifer
	2 - 3 - 4	River dispersion and rain Unconfined aquifer with high gravel thickness
	5	Anomalous

Overall, it is evident that the Brenta domain is more orderly while the LTA domain is multifaceted. Clusters within the Brenta area are well defined either by: (i) the proximity to the mountain area and the southern springs belt (differentiation from clusters 3-4 and 0-1-2), (ii) the presence of only one main river and the proximity to it and to its riverbed aquifer (differentiation from cluster 0-1-3 and 2-4), (iii) presence of one main Paleochannel (cluster 4).

Clusters in the LTA area (Leogra-Timonchio-Astico river system) are not so clearly defined due to the major area complexity: (i) presence of multiple paleochannels (mainly related to the Astico river), (ii) presence of deposits from three different rivers (Leogra, Timonchio, and Astico), (iii) the nearness with an important tectonic dislocation (Schio-Vicenza fault, "Flessura Pedemontana"), (iv) the long North-South points distribution that covers the high plain area and the southern shallow aquifer zone (Vicenza area). For the reasons above, the identified clusters are partially overlapping, and

their hydrogeological characterization is complex. Observing the PCA results and the hydrogeological context, clusters in the LTA area could be grouped into just two main clusters: one cluster comprehending 0-LTA, 1-LTA, and 6-LTA associated with the main deep unconfined aquifer, a second cluster comprehending the 2-LTA, 3-LTA, 5-LTA associated with the local presence of a shallow unconfined aquifer.

The first relevant hydrogeological insight coming out from this study regards the complexity of the northern area of the LTA domain. Indeed, it is usually considered by literature as a unicum homogeneous undifferentiated aquifer (e.g., Dal Prà et al. 1976; Altissimo et al. 1999) but this study instead observes a much more complex system with different characteristic hydrographs. Indeed, the presence of clusters 2-LTA, 4-LTA, and 5-LTA in the high plain shows that it is not homogeneous as expected.

The examination of anomalies is considered an essential step to understanding the main limits of the proposed approach. Indeed, even if the PCA results are overall satisfying, the analysis not always identifies clusters well-gathered in space. Some exceptions stand out and are following deeply investigated: single points in unexpected positions (i.e., far away from the main cluster) or clusters grouping points in completely different geological settings. The main identified clustering anomalies are (i) cluster 5-Br across the Brenta domain, (ii) 2-LTA 4-LTA, 5-LTA, 7-LTA, 8-LTA grouping some points in the southern confined aquifer system (shallow aquifer) with the northern unconfined aquifer, and (iii) one single point of cluster 2-Br enclosed in cluster 4-Br.

Starting from the first anomaly, the 5-Br cluster gathers large diameter wells across the Brenta domain. The further anomaly investigation led to a deeper documental analysis of such monitoring points, and they have been found to be equipped with a pump for agricultural use even if their time series are not visibly affected by pumping activities. Nevertheless, their trend is not related to the local natural groundwater trend but is governed by the withdrawal for agricultural use. For this reason, they differ from the neighboring points and, despite being distant, they are grouped in one single cluster as the agricultural needs are the same within the area. Regarding the second identified anomaly, the 2-LTA 4-LTA, 5-LTA, 7-LTA, and 8-LTA clusters have a similar hydrogeological context, although sometimes located in opposite positions (i.e., intercepting the northern 'unconfined' aquifer, close to the mountains, or intercepting the shallow aquifer in the southern multi-layer aquifer system). In fact, further documental analysis of northern points identified the presence of local lenses of colluvial sediments, thus these points are also intercepting a shallow aquifer that responds to immediate stresses of the rain (as for the 3-LTA). Focusing then on the last anomaly within the Brenta domain, the PCA grouped the control point 168 within the cluster 2-Br but this point is geographically enclosed in cluster 4-Br. Figure 11a shows how the PCA returns clear results without clustering uncertainties in the factorial plan. Moreover, comparing the clustered hydrographs with the one of control point 168 in Figure 11b,

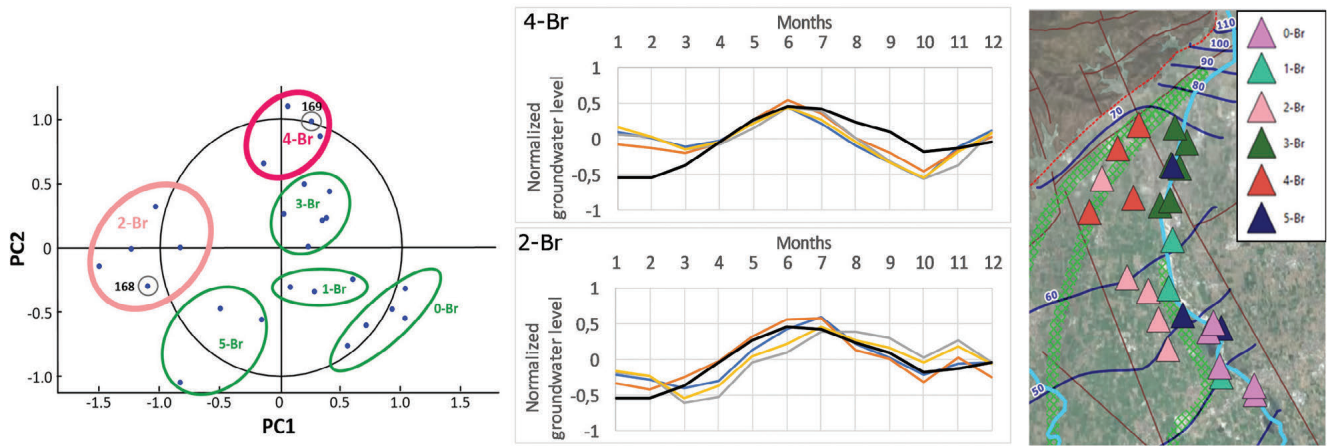


Fig. 11 - Analysis of the control point 168 and the two overlapping clusters (2-Br and 4-Br) within the Brenta domain. a) Factorial plan where the two investigated clusters are differently colored, b) Clustered yearly hydrographs and the hydrograph of control point 168 in black, c) zoom of the clusters within the Brenta domain.

Fig. 11 - Analisi del punto di controllo 168 ed i due cluster sovrapposti (2-Br e 4-Br) all'interno del dominio del Brenta. A) Piano fattoriale dove i due cluster sono colorati con un colore differente, b) gli idrogrammi annuali raggruppati a confronto con l'idrogramma del punto 168 in nero, c) zoom dei cluster all'interno del dominio del Brenta.

the clustering is confirmed beyond any doubt: the hydrograph of control point 168 is actually similar to ones of cluster 2-Br and different from those of cluster 4-Br. Nevertheless, this anomaly has been further investigated in order to understand the hydrogeological reason for such differentiation.

A visual comparison of the original time series (see Figure 12) highlights how the investigated control point recorded a trend like the time series grouped in cluster 4-Br. However, the different recorded time periods are evident, much shorter for point 168. Indeed, the investigated control point monitored the groundwater level between 2016 and 2018 but even in this period the groundwater level recorded in 168 is more associable to cluster 4-Br than to 2-Br. Thus,

a test has been applied to the time series recorded in control point 169, grouped in cluster 4-Br: a second hydrograph has been calculated considering only the time period 2016-2018, coincident with the recording period of point 168. Figure 13 clearly shows how the calculated hydrograph for 169 differs from its original hydrograph (considering the whole available time series between 2009 and 2018) and how it is similar to the hydrograph of 168. Thus, during the last years, the typical hydrograph of cluster 4-Br changed converging to the typical hydrograph of cluster 2-Br.

This anomaly shows a limit of the PCA for studying the groundwater system. Indeed, even if the data has been preprocessed to have a homogeneous dataset, the different

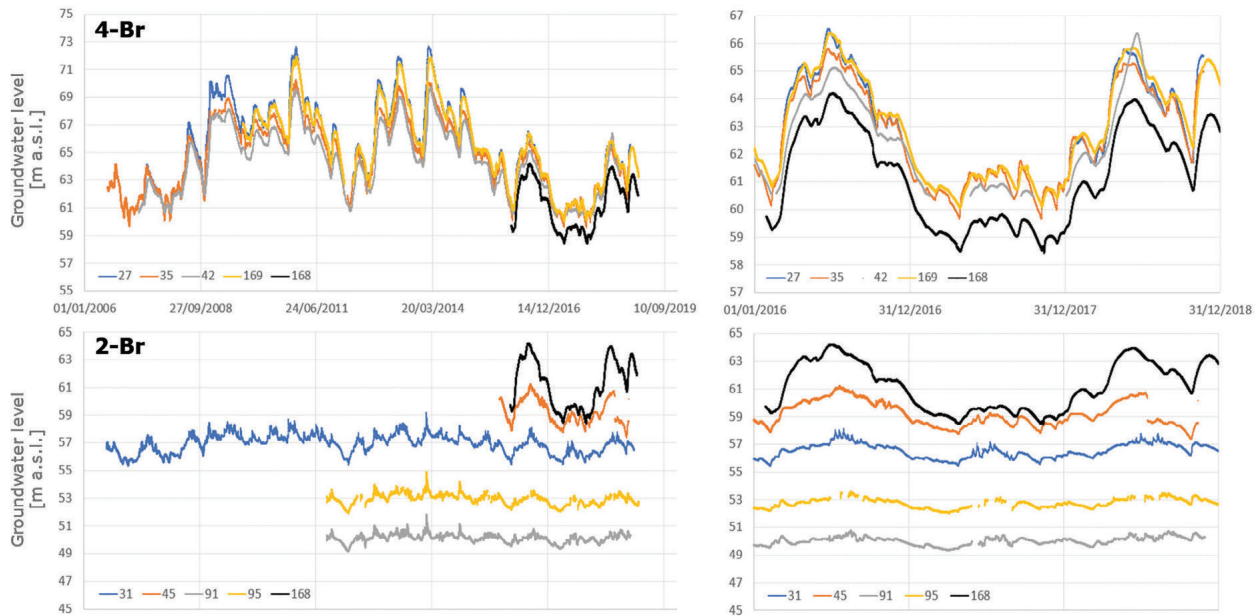


Fig. 12 - Analysis of the control point 168. Original time series clustered within the two overlapping clusters in space (2-Br and 4-Br) and the record of control point 168 in black: original time series on the left, last 3 years of recording on the right.

Fig. 12 - Analisi del punto di controllo 168. Serie temporali originali raggruppate secondo i due cluster sovrapposti (2-Br e 4-Br) a confronto con la serie storica del punto 168 in nero: serie storiche originali a sinistra, ultimi tre anni di monitoraggio a destra.

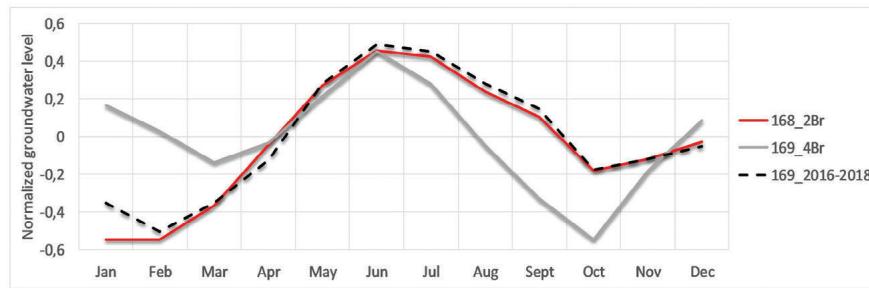


Fig. 13 - Analysis of the control point 168. Comparison of three hydrographs: control point 168 included in cluster 2-Br; control point 169, representative of cluster 4-Br which hydrograph is calculated based on the original available time series (2013-2018); control point 169 but the hydrograph has been calculated considering only the period 2016-2018.

Fig. 13 - Analisi del punto di controllo 168. Confronto di tre idrogrammi annuali: punto di controllo 168 raggruppato nel cluster 2-Br; punto di controllo 169, rappresentativo del cluster 4-Br il cui idrogramma è stato calcolato considerando l'intera serie storica disponibile (2013-2018); punto di controllo 169 ma il cui idrogramma è stato calcolato considerando solo il periodo 2016-2018.

lengths of the original time series could affect the clustering due to potential hydrological changes over the years (e.g., Climate Change effects). If such anomalies are missed, this issue could mislead the achievable insights on the conceptual model but, if such anomalies are observed and investigated, this analysis could be able to pinpoint such exceptions, potentially leading to insightful comprehensions of the variations of the hydrogeological cycle.

Conclusions

The present study proposed and tested the application of PCA to a regional scale dataset of the groundwater elevation in order to achieve insights into the conceptual model of the monitored aquifer system. The results clearly show how the analysis made it possible to highlight the correlation between the groundwater regime and some hydrogeological and environmental factors. The comparison between two different hydrogeological systems (Brenta and LTA) gave the opportunity to verify the sensitivity of the method with respect to the hydrogeological complexity of the area. The PCA returned clear results, with neat driving factors, for the orderly Brenta domain while results are more multifaceted for the complex LTA domain. This analysis highlighted indeed how the LTA system is far from the homogeneous unconfined system described in the literature and how it hides local complexities. Furthermore, the study of a few anomalies in the clustering results highlighted (i) the peculiarities of some points, (ii) the relevance of records' different lengths that can affect the statistical analysis. Nevertheless, in this study, the investigation of such clustering anomalies spotted a change in the hydrogeological cycle over the years (affecting the groundwater yearly hydrograph) and this is an interesting side result of the PCA application.

As for the future prospects, the achievable insights applying the PCA to many different large-scale monitoring networks of the groundwater level could unveil relationships between catchment properties (i.e., paleochannels and rivers dispersion) and the dynamic behavior of groundwater. The possibility of building a very large database of such relationships is foreseen to quickly classify each new observation point on the basis of

a limited series of data (as suggested in Barthel et al. (2021), a study in Central and Northern Europe, considering almost 1200 time series).

Acknowledgement

We wish to thank the FLOWPATH session's conveners (Tullia Bonomi, Antonio Menghini and Maria Filippini in session "Groundwater Resource and Sustainability") for their interest in this study and the two anonymous reviewers for their precious comments and suggestions.

Funding

This work was self-funded by the authors.

Competing interest

The authors declare no competing interest.

Author contributions

Collection of data, Sottani A, Bullo P, Accoto V; data processing, Meggiorin M; interpretation of results, all authors; writing-original draft preparation, Meggiorin M, Bullo P, Accoto V; writing-review and editing, Passadore G, Sottani A, Rinaldo A; graphical editing, Meggiorin M; supervision, Rinaldo A; project administration, Sottani A, Rinaldo A. All authors have read and agreed to the published version of the manuscript.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.7343/as-2022-573>

Reprint and permission information are available writing to acquessotterranee@anipapozzi.it

Publisher's note Associazione Acque Sotterranee remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

- Altissimo L, Bertoldo S, Campagnolo F, Gusmaroli G, Muraro T, Sottani A (2014) Managed aquifer recharge experiences with shallow wells: first analysis of the experimental activities in the high Vicenza plain (Northern Italy). *Acque Sotterranee - Italian Journal of Groundwater* 3(3), 31-39. <https://doi.org/10.7343/as-082-14-0108>.
- Altissimo L, Dal Prà A, Scaltriti G (1999) Relazione conclusiva "Final report". Osservatorio interprovinciale per la tutela delle falde acquifere, Vicenza.
- ARPAV (2016) Stato quantitativo dei corpi idrici sotterranei "Quantitative status of groundwater systems" - ALLEGATO A alla Dgr n. 552 del 26 aprile 2016. Regione del Veneto.
- Aghabozorgi S, Shirkhorshidi AS, Wah TY (2015) Time-series clustering - a decade review. *Information Systems* 53, 16–38.
- Barbagli A, Jensen BN, Raza M, Schüth C, Rossetto R (2019) Assessment of soil buffer capacity on nutrients and pharmaceuticals in nature-based solution applications. *Environmental Science and Pollution Research* 26(1):759-774.
- Barthel R, Haaf E, Nygren M, Giese M (2022) Systematic visual analysis of groundwater hydrographs: potential benefits and challenges. *Hydrogeology Journal*, 1-20.
- Barthel, R, Haaf, E, Giese, M, Nygren, M, Heudorfer, B, Stahl, K (2021) Similarity-based approaches in hydrogeology: proposal of a new concept for data-scarce groundwater resource characterization and prediction. *Hydrogeology Journal*, 29(5), 1693-1709.
- Bro R, Smilde AK (2014) Principal component analysis. *Analytical methods*, 6(9), 2812-2831
- Dal Prà A, Bellati R, Antonelli A, Costacurta R, Sbettega G (1977) Distribuzione dei materiali limoso-argillosi nel sottosuolo della pianura veneta "Distribution of silty-clayey materials in the subsoil of the Venetian plain". Quaderni IRSA e del Consiglio Nazionale delle Ricerche, IRSA, Rome.
- Dal Prà A, Bellati R, Costacurta R, Sbettega G (1976) Distribuzione delle ghiaie nel sottosuolo della pianura veneta "Distribution of gravaled materials in the subsoil of the Venetian plain". Quaderni IRSA e del Consiglio Nazionale delle Ricerche, IRSA, Rome.
- Di Curzio D (2019) Hydrogeochemical and hydrodynamic features affecting redox processes in groundwater. *Acque Sotterranee-Italian Journal of Groundwater*, 8(3).
- Earman S, Dettinger M (2011) Potential impacts of climate change on groundwater resources - a global review. *Journal of water and climate change*, 2(4), 213-229.
- Fitts CR (2002) *Groundwater science*. Elsevier.
- Freeze RA, Cherry JA (1979) *Groundwater*. Prentice-Hall.
- Giordano M (2009) Global groundwater? Issues and solutions. *Annual review of Environment and Resources*, 34, 153-178.
- Gunasekaran R, Kasirajan T (2017) Principal Component Analysis (PCA) for Beginners. *Int. J. Adv. Sci. Res. Manag*, 2, 9-11.
- Lins HF (1985) Interannual streamflow variability in the United States 1143 based on principal components. *Water Resources Research* 21 (5), 691-701.
- Marcolongo B (2012) Foto interpretazione sull'alta pianura vicentina per una ricostruzione paleo-idrografica dei corsi del Leogra e del Timonchio "Photo interpretation on the upper Vicenza plain for a paleo-hydrographic reconstruction of the Leogra and Timonchio rivers". in N. Sottani "Antica Idrografia Vicentina", pp. 265-275, 4, Tavv. f.t., Accademia Olimpica, Vicenza
- Meggiorin M, Passadore G, Bertoldo S, Sottani A, Rinaldo A (2019) Stochastic analysis of groundwater temperature timeseries for characterizing check dams efficiency: case study on a Managed Aquifer Recharge site (Veneto, Italy). *Acque Sotterranee - Italian Journal of Groundwater*, 8(4), 7-13. <https://doi.org/10.7343/as-2019-410>
- Meggiorin M, Passadore G, Bertoldo S, Sottani A, Rinaldo A (2021) Assessing the long-term sustainability of the groundwater resources in the Bacchiglione basin (Veneto, Italy) with the Mann-Kendall test: suggestions for higher reliability. *Acque Sotterranee-Italian Journal of Groundwater*, 10(1), 35-48. <https://doi.org/10.7343/as-2021-499>
- Menciò A, Mas-Pla J (2008) Assessment by multivariate analysis of groundwater-surface water interactions in urbanized mediterranean streams. *Journal of hydrology* 352 (3-4), 355-366.
- Mezzalana G, Niceforo U, Gusmaroli G (2014) Aree forestali di infiltrazione (AFI): principi, esperienze, prospettive "Forested Infiltration Areas: principles, experiences and prospective". *Acque Sotterranee-Italian Journal of Groundwater*, 3(3/137).
- Palmucci W, Rusi S, Di Curzio D (2016) Mobilisation processes responsible for iron and manganese contamination of groundwater in Central Adriatic Italy. *Environmental Science and Pollution Research* 23:11790 -11805.
- Passadore G (2008) Modello matematico di flusso nei sistemi acquiferi del veneto centrale "Mathematical flow model of the aquifer systems in Central Veneto". PhD Thesis, Università degli Studi di Padova.
- Passadore G, Monogo M, Altissimo L, Sottani A, Putti M, Rinaldo A (2012) Alternative conceptual models and the robustness of groundwater management scenarios in the multi-aquifer system of the Central Veneto Basin, Italy. *Hydrogeology Journal*, 20, 419-433
- Pearson K (1901) LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, 2(11), 559-572.
- Ribeiro L, Kretschmer N, Nascimento J, Buxo A, Rötting T, Soto G, Señoret M, Oyarzún J, Maturana H, Oyarzún R (2015) Evaluating piezometric trends using the Mann-Kendall test on the alluvial aquifers of the Elqui River basin, Chile. *Hydrological Sciences Journal*, 60(10), 1840-1852.
- Rinaldo A, Passadore G (2018) Modelli matematici per la tutela e la gestione delle risorse idriche sotterranee nell'ambito del territorio del consiglio di Bacino Bacchiglione "Mathematical models for the protection and management of groundwater resources within the Bacchiglione Basin". Seconda relazione tecnica intermedia (2°stralcio) - Centro Internazionale di Idrologia Dino Tonini-Università degli Studi di Padova.
- Sottani A, Vielmo A (2014) Groundwater conservation and monitoring activities in the middle Brenta River plain (Veneto Region, Northern Italy): preliminary results about aquifer recharge. *Acque Sotterranee - Italian Journal of Groundwater*, 3(3), 25-29. <https://doi.org/AS-081-14-0107>
- Sottani N, Pretto L, Viero C, Marcolongo B (1982) Gli acquiferi nella pianura a nord di Vicenza "Aquifers within the plain at North of Vicenza". Tipolitografica Sociale Artegrafica Cittadella di Padova.
- Vettorello L, Sottani A (2019) Using water level and temperature time series to improve hydrogeological parameterization in a complex alluvial system. *Acque Sotterranee - Italian Journal of Groundwater*, 8(4), 47-53. <https://doi.org/as-2019-413>
- Winter T, Mallory S, Allen T, Rosenberry D (2000) The use of principal component analysis for interpreting ground water hydrographs. *Groundwater* 38 (2), 234-246.