

Article

Multivariate Time Series Clustering of Groundwater Quality Data to Develop Data-Driven Monitoring Strategies in a Historically Contaminated Urban Area

Chiara Zanotti ^{1,*}, Marco Rotiroti ¹, Agnese Redaelli ¹, Mariachiara Caschetto ¹, Letizia Fumagalli ¹, Camilla Stano ², Davide Sartirana ¹ and Tullia Bonomi ¹

¹ Department of Earth and Environmental Sciences, University of Milano—Bicocca, 20126 Milan, Italy

² A2A Ciclo Idrico S.p.A, 25124 Brescia, Italy

* Correspondence: chiara.zanotti@unimib.it

Abstract: As groundwater quality monitoring networks have been expanded over the last decades, significant time series are now available. Therefore, a scientific effort is needed to explore innovative techniques for groundwater quality time series exploitation. In this work, time series exploratory analysis and time series cluster analysis are applied to groundwater contamination data with the aim of developing data-driven monitoring strategies. The study area is an urban area characterized by several superimposing historical contamination sources and a complex hydrogeological setting. A multivariate time series cluster analysis was performed on PCE and TCE concentrations data over a 10 years time span. The time series clustering was performed based on the Dynamic Time Warping method. The results of the clustering identified 3 clusters associated with diffuse background contamination and 7 clusters associated with local hotspots, characterized by specific time profiles. Similarly, a univariate time series cluster analysis was applied to Cr(VI) data, identifying 3 background clusters and 7 hotspots, including 4 singletons. The clustering outputs provided the basis for the implementation of data-driven monitoring strategies and early warning systems. For the clusters associated with diffuse background contaminations and those with constant trends, trigger levels were calculated with the 95° percentile, constituting future threshold values for early warnings. For the clusters with pluriannual trends, either oscillatory or monotonous, specific monitoring strategies were proposed based on trends' directions. Results show that the spatio-temporal overview of the data variability obtained from the time series cluster analysis helped to extract relevant information from the data while neglecting measurements noise and uncertainty, supporting the implementation of a more efficient groundwater quality monitoring.

Keywords: groundwater contamination; diffuse contamination; dynamic time warping; anthropic background level; time series analysis; early warning system



Citation: Zanotti, C.; Rotiroti, M.; Redaelli, A.; Caschetto, M.; Fumagalli, L.; Stano, C.; Sartirana, D.; Bonomi, T. Multivariate Time Series Clustering of Groundwater Quality Data to Develop Data-Driven Monitoring Strategies in a Historically Contaminated Urban Area. *Water* **2023**, *15*, 148. <https://doi.org/10.3390/w15010148>

Academic Editor: Fernando António Leal Pacheco

Received: 11 November 2022

Revised: 23 December 2022

Accepted: 26 December 2022

Published: 30 December 2022



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Groundwater is a crucial resource, providing social, economic, and environmental benefits and opportunities, currently constituting half of the volume of water withdrawn for domestic purposes by the global population [1]. On the other hand, groundwater pollution reduces the suitability of abstracted groundwater for drinking purposes and affects groundwater-dependent ecosystems [2]. Groundwater quality management, especially for drinking purposes, involves monitoring through regular collection and analysis of water samples over prolonged periods, and a considerable effort is needed to collect high-quality data with sufficient frequency [3]. In this regard, a scientific effort is needed to explore new techniques to exploit and interpret the time series resulting from the prolonged effort for water quality monitoring of the last decades.

Currently, monitoring data of raw groundwater quality are mostly used by water suppliers and agencies at every survey as a picture of the current state with respect to the

regulatory limit. On the other hand, if those data are compared with a longer-term trend, they can help to identify the processes that may lead to an improvement or aggravation of the current situation, facilitating wells management and further monitoring. Particularly, observing single time-pictures or averages can be useful for major ions or physico-chemical parameters mostly associated with natural, and therefore stable, processes. However, when studying anthropic contamination, evaluating pluriannual trends can provide valuable additional information.

Time series analysis is increasingly adopted to treat, understand and forecast water availability and quality data and earthquake-induced hydrogeochemical changes [4–13]. Indeed, interpreting trends of several monitoring stations can become confusing and time demanding when working with a large set of wells. In this regard, clustering techniques can help to group data with similar trends. Clustering is a data mining technique that arranges similar data into related or homogeneous groups without previous knowledge of the groups' definitions [14]. It has been proven to be a useful methodology for exploratory data analysis as it recognizes structure(s) in datasets by objectively organizing data into similar groups (clusters) [15]. A special type of clustering is time-series clustering. While each time series consists of multiple data, it can also be seen as a single object [16], and clustering these kinds of complex objects can be advantageous, allowing for the discovery of relevant patterns in time series datasets [17,18]. Time series clustering allows for the identification of homogeneous groups of monitoring stations with similar behaviors over time [19–22].

In the last few years, successful attempts have been made to apply time series clustering to water data. Among those, some works applied time series clustering to water availability datasets related to surface water discharges [23–25] and groundwater level data [26–29]. On the other hand, only a few works applied time series clustering to water quality data: Huang et al. [30] and Lee et al. [31] analyzed water quality parameters in river monitoring stations while, to the best of our knowledge, time series cluster analysis application to groundwater quality data is yet to be explored.

The main aim of this work is to explore the application of univariate and multivariate time series cluster analysis for the interpretation of groundwater quality data. This work constitutes a novel approach for groundwater quality data investigation through time series cluster analysis, which can become a valuable tool for monitoring and investigating any kind of physico-chemical data. In this work, time series cluster analysis is applied to groundwater contamination data, to support water quality monitoring and wells management in a historically contaminated urban area.

Urban and industrial sprawl is often associated with the presence of wide areas with contaminated groundwater, where it becomes difficult to discriminate and manage different local sources and plumes of contamination [32]. These areas are often characterized by multiple superimposing point sources of contamination, generating diffuse contamination over wide areas, together with local high-concentration plumes [33]. In these cases, for proper use and monitoring of water, the need emerges for scientific-based tools to discriminate the areas affected by diffuse anthropic contamination from local hotspots linked with different processes [34] since the associated risks can be different.

Here, time series exploratory analysis and time series cluster analysis are applied to contamination data of raw water samples from an urban area characterized by several superimposing historical contamination sources and a complex hydrogeological setting. The results of the cluster analysis are exploited to investigate the wells' temporal profiles of the contamination, discriminating between diffuse and local contaminations with the aim of designing data-driven monitoring strategies.

Implementing data-driven monitoring strategies can lead to more efficient monitoring networks tailored for specific territories and focused on relevant ongoing processes. Furthermore, data-driven monitoring strategies can help to avoid redundant analyses while targeting relevant trends by implementing early warning systems.

The proposed approach constitutes a widely applicable tool, easily reproducible by researchers, water suppliers, practitioners and environmental protection agencies.

2. Materials and Methods

2.1. Study Area

The study area is the municipality of Brescia (N Italy, Figure 1, cross-section in Figure S1), the second largest municipality in the Lombardy region by population, hosting ca. 200,000 inhabitants and one of the 20 most populous municipalities in Italy. According to Kottek et al. [35], the climate in the area is classified as Cfa (Humid subtropical climate): the average annual temperature is 13.1 °C, whereas the average annual rainfall is 1091 mm.

The municipality of Brescia lies downstream of the Trompia valley. From the hydrological point of view, the study area includes the Mella river fan in the northern part of the study area and a Higher Plain area in the southern part [36–40].

In the Mella fan area, northern zone, the structure of the aquifers was strongly influenced by the incision of the bedrock, in correspondence with the Mella River, flowing from the upstream Trompia valley, which caused a large depression subsequently filled by river deposits. The Mella river runs through the valley, diverging and creating considerable lithological variations with zones with different permeabilities laterally and vertically. The water circulation mainly occurs among overlaying levels generating a multilayer aquifers system consisting of a shallower aquifer hosted in the more gravel-sandy deposits (gravel-sandy unit) and a deeper aquifer within conglomerate deposits (conglomerate units). This structure limits, but does not prevent, exchanges and intercommunications between the aquifers, especially on a local level. The groundwater flow inside the gravelly-sandy material encounters lower permeability layers, consisting of fine-grained levels and compact conglomerates. The shallower aquifer, with greater permeability and transmissivity, is also highly vulnerable to anthropogenic contamination, while the deeper aquifer, hosted in conglomerates, is less vulnerable to pollution from the surface.

Moving southward toward the higher plain area, the conglomerate gradually disappears, and the higher plain area hosts a multilayer aquifer system with several superimposed aquifers separated by low permeability aquicludes.

Due to the hydrogeological setting and the high-density urbanization and industrialization, the piezometric map shows a complex situation. The main flow direction is North-South, with a first significant piezometric depression on the left bank of the Mella River, due to water withdrawals and a second one in the southern area. Since the shape of the water table is mainly driven by anthropic withdrawals, it can also rapidly vary over time, determining wide and rapid changes in the flow directions, which can hamper the interpretation of chemical data.

The industrial growth in the Trompia valley and Brescia city goes back to the early 1900s and includes mainly metallurgic activities, foundries, weapons and ammunition manufacturing, tanning, paints, and varnish production.

The monitoring network of the environmental authorities showed evidence of widespread qualitative degradation of the groundwater tapped by public and private wells since the 1980s, with the presence of anthropic contaminants such as organo-halogen compounds and hexavalent Chromium [41]. The local environmental Protection Agency observed local and widespread contaminations throughout the territory, especially in the northern and central parts of Brescia municipality, and associated them with inadequate waste and discharge management [41].

In 2002 part of the Brescia municipality (ca. 262 hectares, Figure 1) was included in the National Priority List of contaminated sites according to Italian law, based on its extension and quantity and concentrations of pollutants, including chlorinated compounds and heavy metals [41].

The Trompia valley as well, upstream to the city center, has been known for a general deterioration of the aquifer from the 90s, related to various outbreaks of hexavalent Chromium and chlorinated solvents. The environmental protection agency described such

degradation as mainly related to point sources of pollution, which are responsible for the most relevant pollution phenomena, as well as to the persistent or occasional discharge of effluents or waste on the ground or in the surface water [42].

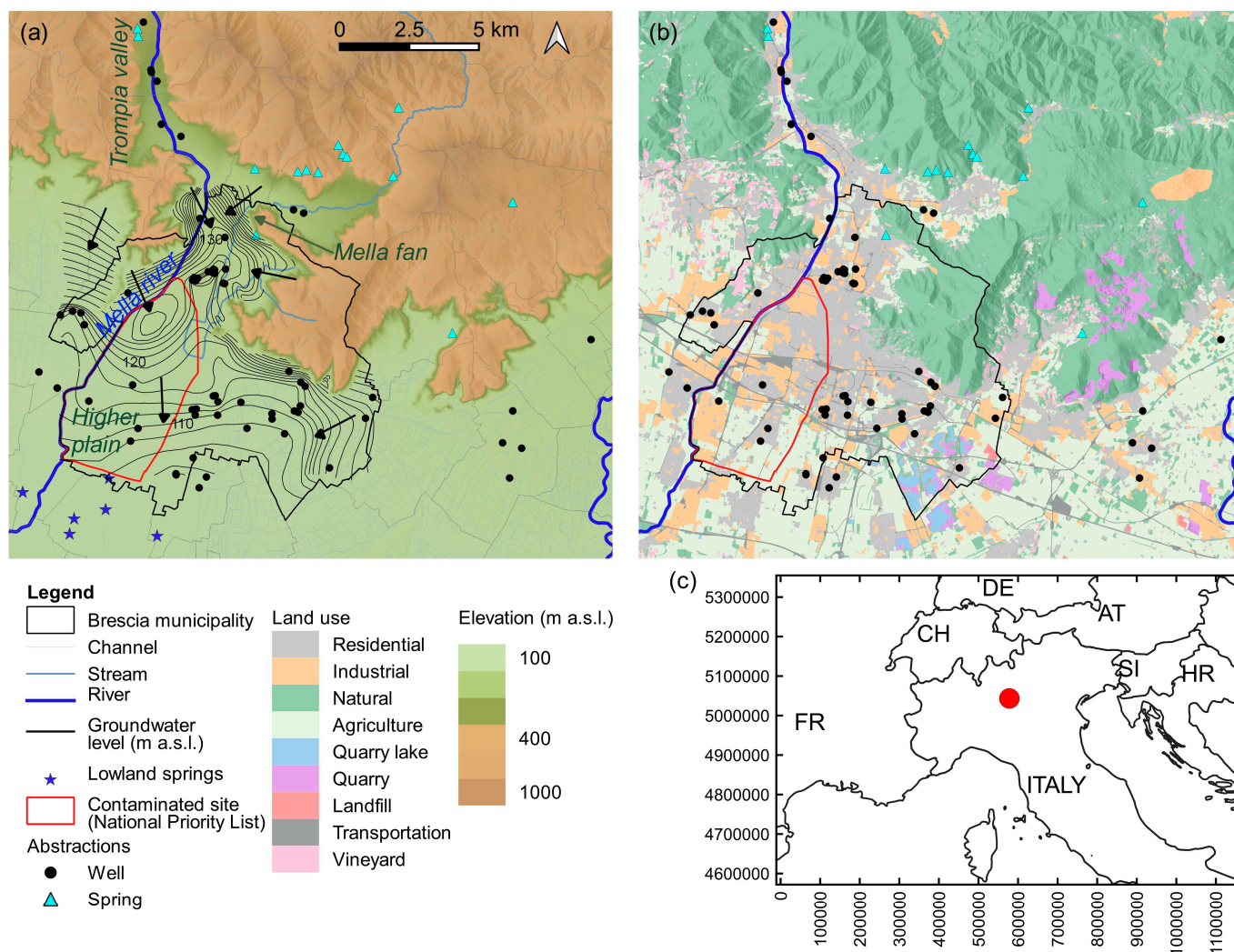


Figure 1. Study area: (a) elevation, and piezometric map by Osservatorio Acqua Bene Comune [40] and (b) land use, (c) study area location (WGS 84 UTM 32N).

2.2. Available Data

Water quality data were made available by A2A Ciclo Idrico SpA, water supplier of several municipalities in the province of Brescia. Data refer to raw water prior to any potabilization treatment. For a broader understanding of the ongoing processes, available data concerning the municipalities surrounding Brescia city were considered in this study, for a total of 68 wells and 16 springs in the dataset.

The wells have different depths and different screens number and distributions. Therefore, even wells close in space can be subjected to different contamination sources and processes. For each well and spring, data are available for the 2009–2020 time window for the following dissolved contaminants: tetrachloroethylene (PCE), trichloroethylene (TCE), and hexavalent Chromium (Cr(VI)), for a total of ca. 3000 data points.

The WHO Guideline Value for drinking water for PCE and TCE are, respectively, 40 µg/L and 20 µg/L while a value of 50 µg/L is set for the total Chromium value [43]. The Italian regulation for drinking water (D Lgs 31/01) specified a more restrictive 10 µg/L value for the sum of PCE and TCE, similarly to other countries and territories world-

wide [43]. As for the Chromium, the Italian regulation for drinking water adopted the WHO guide value of 50 µg/L for the total Chromium.

To perform the statistical analysis, data below the Limit of Detection (<LOD) were substituted with the LOD/2 value. When considering wide periods, the LOD value for a single parameter can vary over time in the dataset, which can cause fictitious variability [44]. Therefore, <LOD data were substituted with the minimum LOD/2 of the data available for each parameter.

For most of the wells, one sample per year is available, but in several cases, sampling and analyses were intensified for specific wells in specific periods; therefore, more data are available. The wells' time series were homogenized for the subsequent statistical analysis by considering yearly averages. The use of yearly averages also reduces the effect of seasonality and noise [10], which could affect the interpretation of pluriannual trends.

2.3. Data Analysis

The work was carried through three successive work phases. First, an exploratory data analysis was carried out to assess the data variability. In the second phase, the time series clustering technique was applied and the results were explored and interpreted. Lastly, the results of the time series clustering were exploited for the development of data-driven monitoring strategies.

2.3.1. Exploratory Analysis

In a first phase, an exploratory analysis of the dataset was performed through the Mann-Kendall test and Sen's slope estimator. The Mann-Kendall test [45,46] is a popular test aimed at statistically assessing the presence of monotonic upward or downward trends. It is a non-parametric test (i.e., no assumptions about the probability distribution of the dataset are required) and robust to potential outliers. Hence, it is particularly suitable for environmental data. Similarly, the Sen' Slope estimator is a non-parametric and robust statistic for calculating the trend's slope [47,48].

2.3.2. Time Series Clustering

Time series clustering was applied on the time series of PCE, TCE and Cr(VI) to identify homogeneous groups of wells with similar dynamic responses to anthropic pressures. The 'dtw' package [49] was used in the R environment version 4.1.0 (18 May 2021). As for the clustering technique, in this study, the Ward hierarchical method was used, which has been previously applied to hydrological data, resulting in more homogeneous and consistent clusters than other methods [7,50].

The most common distance used for cluster analysis is the Euclidean distance, but when working with time series objects the Euclidean distance may fail to produce an intuitively correct measure of similarity between two sequences, being very sensitive to small distortions in the time axis [51]. This is because the Euclidean distance would only compare feature values at the same time-step without considering adjacent measurements (Figure 2). Therefore, if two time series are very similar but even slightly shifted in time, they will be classified as very different when applying straight Euclidean distance. To overcome this issue, the Dynamic Time Warping (DTW) methodology has been proposed [52,53], which is an elastic, shape-based similarity measure created to deal with temporal drift. The DTW method matches time series considering all the directions that minimize the distance between two time series and allowing data at different time steps to be compared.

The DTW method explores potential associations within the whole time series, searching for similarities even in data points that can be far in time (Figure 2) and allowing for sweeping shift and warping of the time series. This approach can be useful for several applications but might become counterproductive for certain fields of work. Indeed, when working with hydrological data, researchers can be interested in evidencing the significant time shifts that can represent specific processes or geological structures. For this purpose, a distance measure should neglect the time shifts which are considered insignificant in

relation to the field of study and the time-resolution of the data, but it should also highlight, with higher distance values, the time shifts that could have an informative meaning. In this regard, Dau et al. [54] highlights the importance of tuning the maximum amount of warping through the window size parameter (w). The parameter w constraints the DTW algorithm allowing for time warping only on a specific time window around each data point. In this work, considering the yearly time resolution and the hydrogeological structure and extension of the area, a 2-year window size was set, which is considered a relevant temporal shift for the purpose of this study with respect to the distances of the wells, the study area extension and the high geological and hydrodynamic variability.

By fixing the warping window size w , the DTW searches for similarities around each data point considering w precedent and w successive time steps (Figure 2). Therefore, when working with hydrological data, the choice of the parameter w should be based on the time-resolution of the data (seconds, days, months, years) and the hydrogeological context of the studied system to highlight temporal shift that could be meaningful and neglect those that are not.

The optimal number of clusters was selected by comparing the results of different solutions by means of the internal clustering validity indices (CVIs). The CVIs are a set of statistics developed to quantify and compare different clustering solutions properties such as compactness, separation between clusters, etc. A large number of CVIs have been proposed, and reports comparing different CVIs suggest that no single CVI can always outperform the others [55,56] and that the performance of single CVIs can decrease when working on small datasets or noisy data. Therefore, it is common practice to adopt and compare several CVIs, using a majority vote as a supporting tool for choosing the number of clusters. Here, the seven most used CVIs were calculated and compared: Silhouette [57] (Sil), to be maximized; Dunn index [56] (D), to be maximized; COP index [56] (COP), to be minimized; Davies-Bouldin index [56] (DB), to be minimized; Modified Davies-Bouldin index [58] (DB*), to be minimized; Calinski-Harabasz index [56] (CH), to be maximized; Score Function [59] (SF), to be minimized.

In this work, two different cluster analyses were carried: (a) a multivariate analysis on PCE and TCE data and (b) a univariate analysis on Cr(VI). The two datasets were analyzed separately since they are associated with different kinds of anthropic sources, they have different spatial distributions and different environmental processes. Hence, a single multivariate analysis would only lead to known information about the different distributions of the two types of compounds, while treating them separately allows for a more specific investigation of the time variability of each type of compound. On the other hand, PCE and TCE in a single well can be equally associated with a single source or come from different sources [60]; therefore, analyzing them simultaneously can help reveal useful information about different hotspots.

For the purpose of the time series clustering, highly incomplete time series (i.e., time series with less than 7 years of data) and wells in which the analyzed compound was never detected were excluded in each analysis.

In most cases, it is appropriate to standardize the data for clustering purposes and, for time series clustering, it is sometimes appropriate to subtract the mean from each time series (e.g., when working with piezometric data) [61–63]. Since this work aims to cluster chemical data, it was important to separate wells with higher concentrations from those with lower concentrations. Therefore, no mean subtraction was performed on the time series. For the multivariate analysis, the two entire PCE and TCE datasets were standardized as z-score so that they could have the same weight within the multivariate analysis, but the relative differences among wells were preserved.

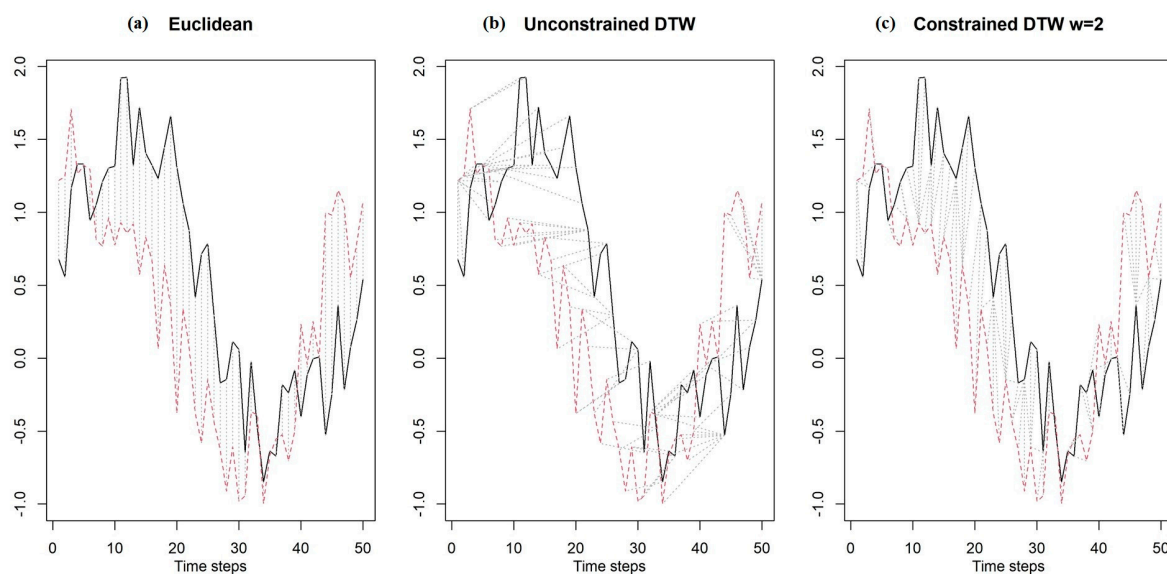


Figure 2. Graphic representation on synthetic data of (a) Euclidean distance, (b) unconstrained Dynamic Time Warping, (c) Dynamic Time Warping with reduced window size ($w = 2$). Black and red lines are the synthetic time series generated through a sin function plus a random component, while the dotted grey lines indicate the association among the data of the two time series calculated with the three methods. Euclidean distance only allows for the comparison of the data at the same time step without any warping of the series, and even the slightest shift would increase the distance value. The DTW searches for similarities in the time series allowing for sweeping shift and warping of the time series, coupling also data which can be far in time. In this example, DTW couples data that are more than 10–15 time-steps away.

2.3.3. Development of Data-Driven Monitoring Strategies

The output of time series cluster analysis allows for a meaningful synthesis of the chemical variability of the wells over space and time. Here, the characteristics of the different clusters' time profiles were analyzed to support the monitoring by designing specific early warning systems tailored for each time profile.

First, the clusters interpretation, based on the time profile and spatial distribution of the wells, is used for discriminating among diffuse background contamination and local hotspots. Subsequently, based on the time profile, the early warning systems indications and, when appropriate, the trigger levels are provided for each cluster. The trigger levels are intended to be threshold values for the implementation of early warning systems: if future data are above this threshold value, the monitoring should be intensified to check or rule out the presence of new contamination processes.

In particular, for clusters associated with diffuse background contaminations, trigger levels are here calculated as the 95° percentile of the data in the cluster. The 95° percentile is the most used method for identifying threshold values for specific groups of wells in the scope of both, natural and anthropic contamination [64–66].

On the other hand, using the 95° percentile is only appropriate for static conditions as it would lead to an overestimation of possible threshold values in case of decreasing trends or an underestimation in case of increasing trends.

Hence, calculating a trigger level through a percentile approach would hardly be useful for the clusters that present peculiar time profiles with increasing or decreasing trends or significant pluriannual oscillations. Instead, successive monitoring should aim to verify that specific trend's persistence or reversal over time. Here three likely cases are considered:

- Increasing trends, which can be immediately considered as a warning situation. Therefore, immediate actions must be taken, and no indications for future warnings can be given.

- Decreasing trends, which indicate ongoing attenuation processes. For these wells any threshold value would be overestimated if based on the entire time series, while the early warning should be triggered if a peak or a new uptrend were to occur.
- Trends characterized by at least one changing point determining the trend inversion. For these cases, it is necessary to focus on the most recent part of the time series, identifying the current trend, which leads back to cases a and b.

3. Results and Discussion

3.1. Exploratory Analysis

Results of the exploratory analysis are reported in Figure 3. For each well, the time series of PCE (Figure 3a), TCE (Figure 3b) and Cr(VI) (Figure 3c) were analyzed, searching for potential trends, with regard to the most recent situation. In Figure 3 the symbol indicates whether the time series showed a non-significant trend or a significant upward or downward trend. The wells are color-coded based on the concentration of the last available measurement.

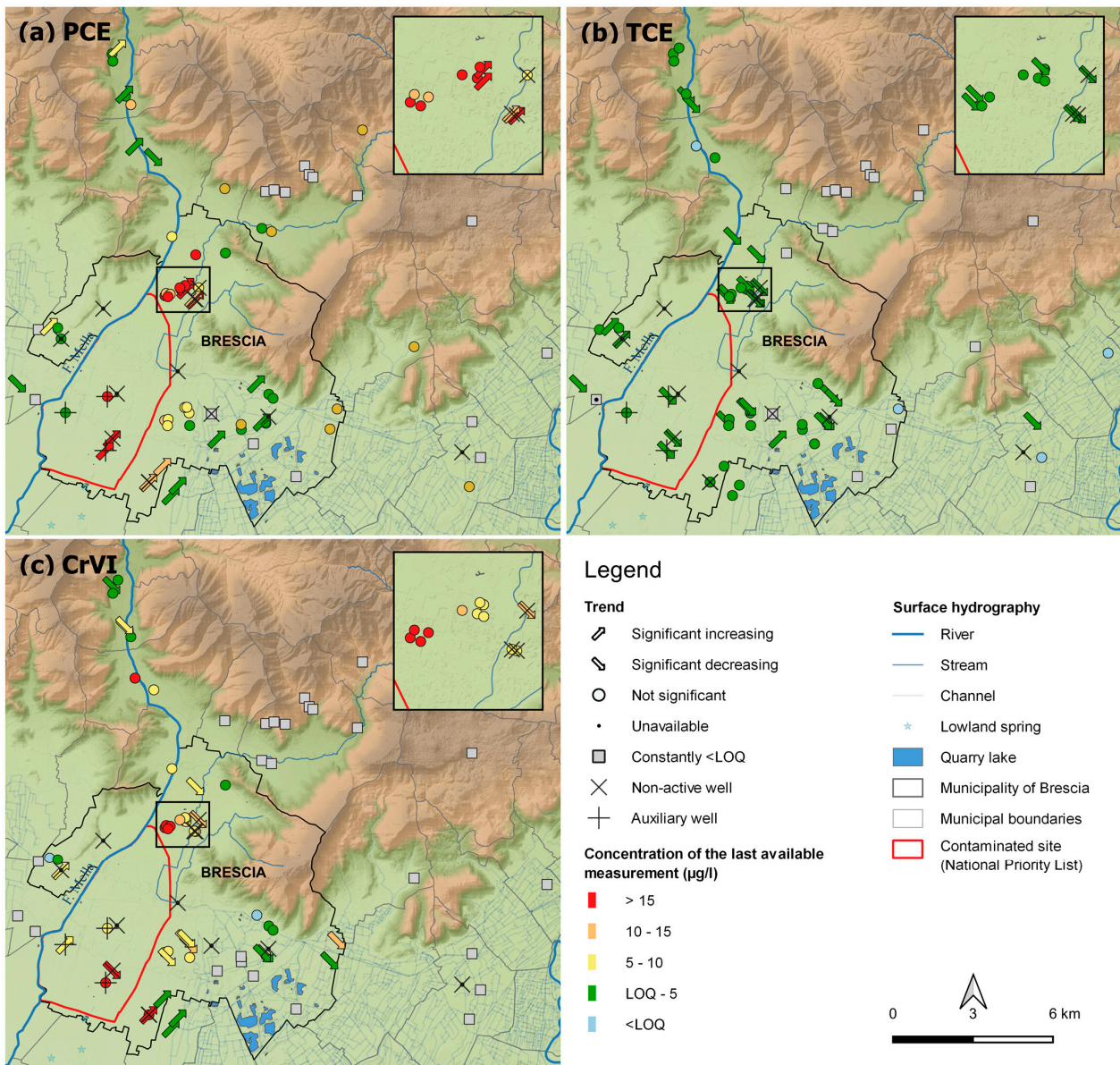


Figure 3. Exploratory trend analysis for (a) PCE, (b) TCE and (c) Cr(VI).

The PCE shows a heterogeneous situation, with several zones of higher concentration with a static or increasing trend, mainly in the city center, and surrounding areas with lower concentrations and static, increasing or decreasing trends. TCE, on the other hand, shows a more homogeneous pattern, with generally lower concentrations and mostly non-significant or decreasing trends.

The Cr(VI) data also show a heterogeneous situation, with high concentration and static trends in the city center and increasing trends with both low and high concentrations in the southern part.

3.2. Time Series Clustering

3.2.1. Multivariate Time Series Clustering of PCE and TCE

Multivariate time series clustering was performed on PCE and TCE data. For the purpose of this analysis, one well was excluded, due to its extreme PCE concentrations, out of the range of the remaining wells, potentially masking the variability of the other wells.

The solutions from 5 to 10 clusters were compared, and the CVI results (Table 1) suggested that the 10-cluster solution was the best performing one. Furthermore, the analysis of the results confirmed that it was an environmentally interpretable solution.

Table 1. CVIs results for the multivariate time series analysis on PCE and TCE data: Silhouette (Sil), to be maximized; Dunn index (D), to be maximized; COP index (COP), to be minimized; Davies-Bouldin index (DB), to be minimized; Modified Davies-Bouldin index (DB*), to be minimized; Calinski-Harabasz index (CH), to be maximized; Score Function (SF), to be minimized. Bold font indicates the best solution according to each CVI.

No. of Clusters	Sil↑	SF↑	CH↑	DB↓	DB*↓	D↑	COP↓
5	0.44	4.89 × 10⁻⁶	35.48	0.85	1.05	0.14	0.15
6	0.43	1.41 × 10 ⁻⁶	30.93	0.87	1.18	0.14	0.13
7	0.40	1.83 × 10 ⁻⁷	27.55	0.99	1.27	0.18	0.12
8	0.40	1.35 × 10 ⁻⁷	25.50	0.88	1.20	0.18	0.11
9	0.40	2.26 × 10 ⁻⁷	23.11	0.81	1.12	0.20	0.11
10	0.39	3.67 × 10 ⁻⁷	21.11	0.77	0.98	0.24	0.10

In Figure 4 the temporal profiles of the ten clusters are shown, together with their spatial distribution.

The cluster PTA groups 23 abstractions, including wells and springs. The PCE values range below 6 µg/L and TCE values are below 1 µg/L. The wells are scattered, mostly in peripheral areas (Figure 4). No relevant pluriannual trend is evident from the data regarding both, PCE and TCE.

The cluster PTB includes 7 wells. Concentrations of PCE range from 1 to 7 µg/L while TCE mostly ranges from 0.5 to 1 µg/L with two peaks above 2 µg/L in 2010. No relevant pluriannual trend is evident from the data regarding PCE, while TCE shows slightly higher concentrations around 2010 (Figure 4). As in the case of the PTA cluster, the PTB wells are scattered in marginal areas.

The cluster PTC includes 10 abstractions, including wells and springs, with PCE values below 8 µg/L and TCE mostly ranging between 1 and 2 µg/L. No relevant pluriannual trend is evident from the data regarding PCE. TCE shows a wider pluriannual variability, but no environmentally relevant trend emerges (Figure 4). Also in this case, the wells are mostly scattered in peripheral areas. The position of the wells, the constantly low concentrations over time, and the absence of relevant trends, suggest the association of these three clusters with the diffuse background contamination connected with the multiple historical sources in the upstream valley and the Brescia or surrounding municipalities.

The cluster PTD represents a group of 7 neighboring wells in the plain area with a peculiar temporal profile for PCE and TCE and a well in the northern valley area (Figure 4). The PCE concentrations in PTD show an increasing trend up to 2015, followed by a static/decreasing behavior. TCE instead shows an oscillatory behavior from 2009 to 2016,

followed by a slightly decreasing trend. The specific time profile and the geographical distributions led to the association of this cluster's wells with two local hotspot contaminations: one in the northern part and one in the southern part of the study area.

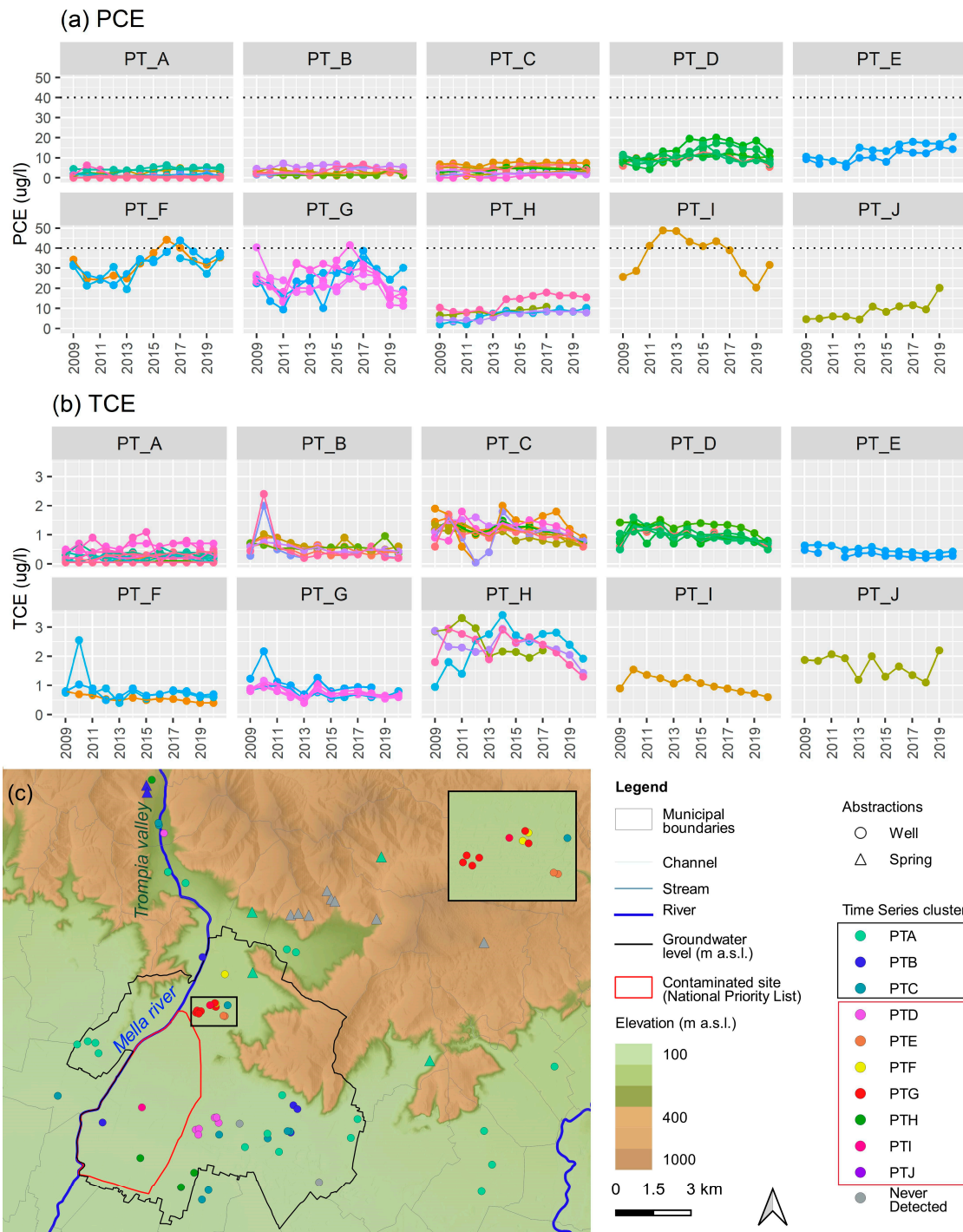


Figure 4. Results of multivariate time series cluster analysis on PCE and TCE: (a) PCE time profile of the clusters (dashed lines indicate WHO guide value for PCE in drinking water), (b) TCE time profile of the clusters (the WHO guide value for TCE is 20 µg/L, outside the range of the graphs) and (c) spatial distribution of the clusters, black rectangle in the legend indicates clusters associated with diffuse background contamination, while red rectangle indicates clusters associated with local hotspots.

The cluster PTE includes two neighboring wells in the city center, with an evident monotonous increasing trend for PCE and a decreasing trend for TCE (Figure 4). This distinct temporal profile marks these two wells as a hotspot but differentiates them from the surrounding wells in clusters PTF and PTG.

The cluster PTF includes three neighboring wells in the city center, with a distinct time profile for PCE, characterized by a minimum in 2010–2012, and a maximum in 2016–2017, while the TCE shows consistently low concentrations (Figure 4). The grouped position and the peculiar time profile indicate that this cluster represents a local hotspot.

The cluster PTG presents a time profile close to PTF, but it presents a lower minimum in 2010–2012 and reaches lower values in the recent years, while TCE is still constantly low, with a marked oscillation in 2013–2014 (Figure 4). Therefore, PTG can be associated with a local hotspot.

The cluster PTH groups wells in the southernmost part of the area, with a well in the northern area. These wells share a PCE time profile characterized by an increase around 2014, followed by a more stationary condition. On the other hand, the TCE shows concentrations slightly higher than the other clusters with decreasing values in the last years (Figure 4).

The cluster PTI is a singleton, far from other monitoring wells, and presents a unique profile for PCE with a maximum around 2013, reaching the highest values in the dataset and a second peak around 2016 (Figure 4).

The cluster PTJ is another singleton, representing a well very close to one of the wells in the PTH cluster, from which it differentiates due to a maximum in 2019 (Figure 4).

3.2.2. Univariate Time Series Clustering of Cr(VI)

For the Cr(VI) time series, a univariate time series clustering has been performed. As for the PCE and TCE analysis, five solutions were compared, with 5 to 10 clusters. The CVI output (Table 2) indicated the 5-clusters solution, which on examination are not sufficiently environmentally interpretable and explicative. Therefore, the second-best option was chosen, which was the 10-cluster solution.

Table 2. CVIs results for the univariate time series analysis on Cr(VI) data: Silhouette (Sil), to be maximized; Dunn index (D), to be maximized; COP index (COP), to be minimized; Davies-Bouldin index (DB), to be minimized; Modified Davies-Bouldin index (DB*), to be minimized; Calinski-Harabasz index (CH), to be maximized; Score Function (SF), to be minimized. Bold font indicates the best solution according to each CVI.

No of Clusters	Sil \uparrow	SF \uparrow	CH \uparrow	DB \downarrow	DB* \downarrow	D \uparrow	COP \downarrow
5	0.54	2.24×10^{-12}	20.52	0.39	0.45	0.13	0.06
6	0.38	3.63×10^{-13}	20.97	0.44	0.60	0.09	0.04
7	0.38	9.33×10^{-15}	22.05	0.46	0.55	0.16	0.04
8	0.34	2.22×10^{-16}	24.66	0.58	0.84	0.11	0.03
9	0.37	0.00	23.51	0.60	0.74	0.15	0.03
10	0.37	2.22×10^{-16}	21.93	0.52	0.64	0.16	0.02

Out of the ten clusters, three were associated with diffuse background contaminations. These three clusters are CrA, which includes 5 wells and 3 springs, CrB, which includes 16 wells and CrC, which includes 10 wells. In terms of Cr(VI) concentrations, CrA, CrB and CrC show consistently low values (i.e., 0–5 $\mu\text{g/L}$ for CrA, 3–10 $\mu\text{g/L}$ for CrB and 5–15 $\mu\text{g/L}$ for CrC), and their distribution appear scattered over the study area, including upstream and lateral areas (Figure 5).

The cluster CrD groups three neighboring wells, which present an oscillatory pluri-annual trend, with a maximum around 2012, a minimum around 2017 and increasing values in the 2017–2020 time span (Figure 5). The specific time profile, together with their proximity in space support the identification of these wells as a local hotspot.

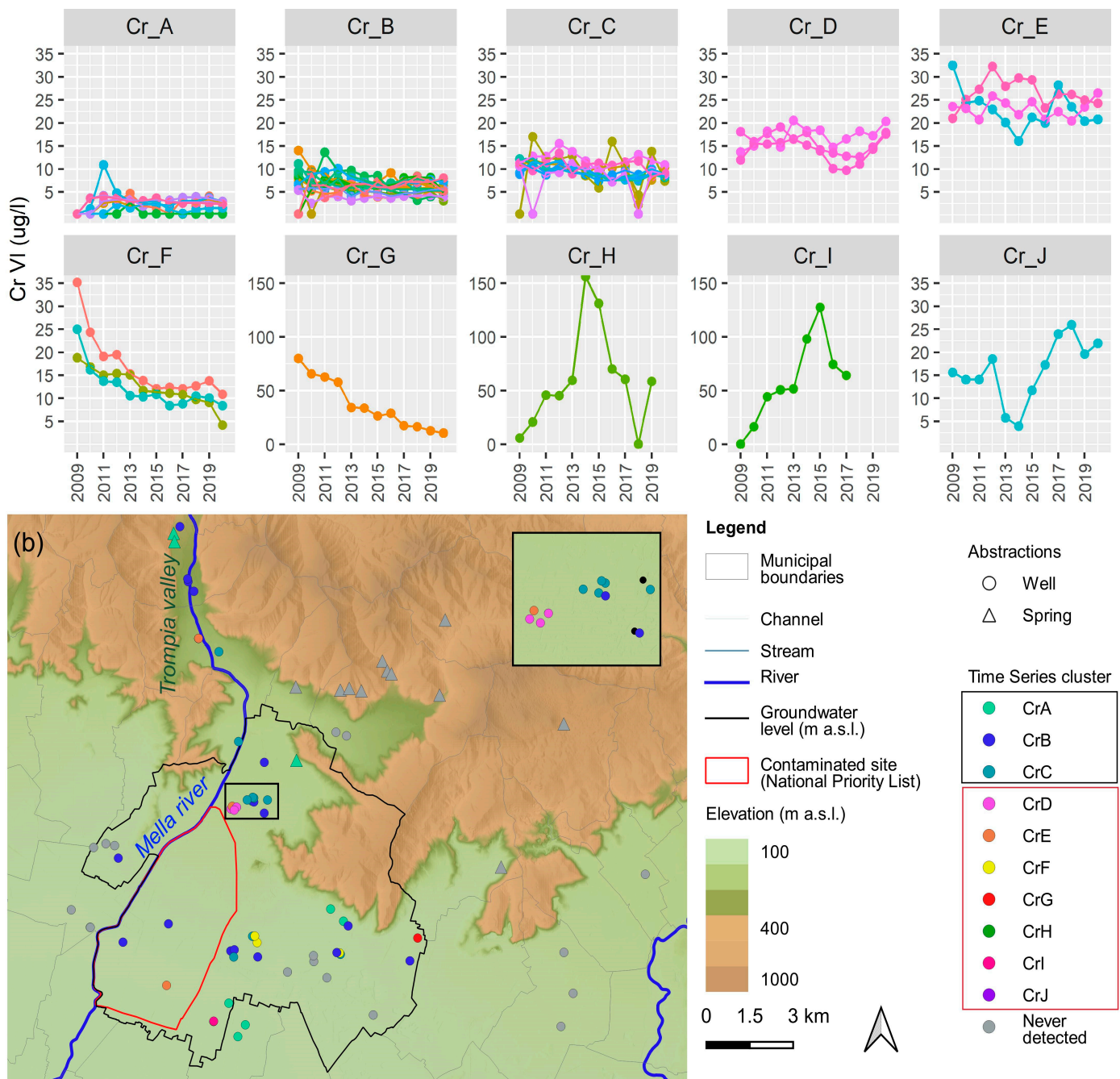


Figure 5. Results of univariate time series cluster analysis on Cr(VI): (a) Cr(VI) time profile of the clusters, (b) spatial distribution of the clusters, black rectangle in the legend indicates clusters associated with diffuse background contamination, while red rectangle indicates clusters associated with local hotspots.

CrE groups 3 wells with constant concentrations in the range 20–30 $\mu\text{g/L}$. Two of these wells are close to each other, in the south-eastern part of the area, while the third one is separated, in a more eastern area (Figure 5). All the wells in CrE show values consistently above the range of the background clusters (i.e., CrA, CrB and CrC) with stronger interannual variability which supports the identification of these three wells as local hotspots.

CrF groups 3 wells, scattered in different positions, with an evident decreasing trend (Figure 5). CrH–CrJ are singletons, with time profiles different from the remaining wells of

the dataset. CrH and CrI though, which are close in space, show similar trends with the highest concentrations of the dataset reached respectively in 2014 and 2015 (Figure 5).

3.3. Data-Driven Monitoring Strategies

Here, specific monitoring indications are proposed for each cluster based on a detailed interpretation of the characterizing time profile.

As regards the PCE and TCE monitoring, the PTA, PTB and PTC clusters were associated with diffuse background contaminations without significant trends. Therefore, trigger levels were calculated, through the 95° percentile, leading to the values 4.1, 5.9 and 7.3 µg/L for PCE and 0.6, 0.9 AND 1.7 µg/L for TCE for respectively PTA, PTB and PTC.

Clusters PTD, PTF, PTG and PTI were associated with local hotspots, and they show an oscillatory behavior in terms of PCE with at least 2 reversals of the trend directions during the considered period. In all these clusters, the last part of the time profile of PCE appears to be decreasing or stationary, starting from different years. Therefore, the PCE monitoring could be intensified, evaluating the results with respect to the previous data, and the early warning should be triggered if future data appear to be higher than the previous data. In these clusters, the TCE shows constantly low concentrations. Even if slight trends are visible, the range of values is very narrow, therefore it is appropriate to calculate a trigger level through the 95° percentile method which results in the TCE values of: 1.4 µg/L for PTD, 0.9 µg/L for PTF, 1.1 µg/L for PTG.

Cluster PTH was associated with local hotspots. Since 2014, the PCE has shown a stationary behavior while the TCE shows decreasing values. In this case, the monitoring should be aimed at verifying the persistency of these conditions and the early warning should be triggered if new measurements give higher results than previous data.

Clusters PTE and PTJ were associated with local hotspots, and show an increasing trend for PCE and stationary/oscillatory trend for TCE. Here, monitoring should be intensified for a more detailed observation of the ongoing processes that could lead to an increase in PCE, which could also be associated with future increase in TCE.

As regards Cr(VI) monitoring, results highlighted that CrA, CrB and CrC were associated to diffuse contamination, with no pluriannual trends and small variability. Therefore, trigger levels were calculated, through the 95° percentile, leading to the values 4 µg/L, 9.3 µg/L and 13.1 µg/L for respectively the CrA, CrB and CrC clusters.

CrE was associated to local hotspots, and its wells show wider variability, but around a static average without any evident pluriannual trend. Therefore, it is appropriate to elaborate a trigger level with the 95° percentile, which result in 30.4 µg/L.

CrD, CrH, CrI and CrJ show clear pluriannual oscillation, with different behaviors. In these cases, the monitoring should be aimed at evaluating the progress of the ongoing processes, with regard to the most recent trends which are increasing for CrD and CrJ and decreasing for CrH and CrI. Therefore, for CrD and CrJ the monitoring could be intensified, and warning should arise if a significant positive trend takes place considering data from 2017 for CrD and 2014 for CrJ. While For CrH and CrI the warning should be activated for new peaks or increasing trends, therefore for values higher than the previous data.

CrF and CrG show decreasing trends; in these cases, early warning should arise in case of peaks and trend reversal, i.e., for concentration values higher than the previous data.

3.4. Methodological Approach Pros and Cons

Currently, monitoring data of raw groundwater quality are mostly analysed by environmental agencies and water suppliers and at every survey to check the compliance with regulatory limits [67,68].

On the other hand, if time series of data are available, analysing longer-term trend can help to identify the processes that may lead to an improvement or aggravation of the current situation, facilitating wells management and further monitoring. Particularly, observing single time-pictures or averages can be useful for major ions or physico-chemical parameters mostly associated with natural, and therefore stable, processes. However, when

studying anthropic contamination, evaluating pluriannual trends can provide valuable additional information.

In this work, an exploratory data analysis was performed, through the most common methods for trend detection and quantification and, subsequently, a time series cluster analysis was applied for a more detailed trend overview.

The results of the exploratory analysis reported a highly variable situation (Figure 3), and no evident information about specific hotspots emerged. Coupling in the same visualization the concentration and the trend information is informative, but interpreting different trends against concentrations requires a high level of interpretation and analysis. Indeed, the most common statistical methods for trend analysis applied here, are valuable and well standardized methods [10,69], mostly applied on groundwater level data [11,12] but the results have shown how these methods present severe limitations when applied on highly variable chemical data. Particularly, the Mann-Kendall test for trends detection works properly only with monotonous trends, while its performances decrease when coping with oscillatory behaviors. Furthermore, the Mann-Kendall test mostly focuses on the sign of the differences among data neglecting the amplitude of possible trends while, for environmental applications, it can be useful to distinguish slight or dramatic trends. For this last application, Sen's slope estimator is applied, which provides information about the slope amplitude. Nevertheless, the Sen's slope estimator also performs better on monotonous trends, while it can struggle with strong oscillations. Furthermore, for environmental applications, it could be useful to discriminate between cases of similar trend slopes but different ranges of concentrations (e.g., lower values or values closer or higher to a regulatory limit).

Therefore, if working with anthropic contamination data, it becomes useful to assess all these abovementioned aspects: the presence or absence of trends, their direction and amplitude, the concentration ranges, and the presence of oscillatory behaviors or trend reversals. Assessing all these aspects for each well for a significant number of wells and parameters can become time demanding and confusing.

In this regard, the time series cluster analysis allowed for a concise representation of the time profiles of different wells (Figures 4 and 5), by grouping them based on the most relevant features of their time profiles.

Time series clustering has been increasingly adopted for the analysis of groundwater level data [26,27,63,70], and recently its application was extended to the analysis of surface water quality data [30].

The results of the study highlighted that time series clustering could become a valuable tool for the analysis, exploration and exploitation of groundwater quality data, especially in the scope of anthropic contamination.

Particularly, the time series cluster analysis, performed with the DTW method, with a window size tuned to the hydrogeological characteristics of the study area and length and resolution of the time series, provided a meaningful and environmentally interpretable grouping of the different temporal profiles of the wells. There is a wide range of applications for these kinds of results, and two main applications were explored in this work. As a first application, the time series cluster analysis helped to discriminate diffuse background contamination and different local hotspots. If looking, for example, at Figure 3 the information of the concentrations at a single time step could have led to a flawed interpretation, especially in the context of the neighboring wells in the zoom rectangle, which have a similar range of concentration. The time profile analysis, instead, clearly showed different behaviors for some of these wells. Indeed, by observing a single concentration or an average it is not possible to distinguish whether a value results from an increasing or decreasing trend or is static over time, while this information is crucial for a proper water resource management.

These results are particularly valuable when considering the complex hydrogeology of the study area, with a wide geological variability over a narrow territory and complex flow paths which are highly variable over time.

In this work, the second application of time series cluster analysis highlighted that it could be a valuable support for future monitoring. Also in this case, being able to deal with groups of wells instead of single wells allows for a more immediate, efficient, and smart design of successive monitoring standards and early warning system implementation.

On the other hand, a limitation of this application is that it can mix up different hotspots with similar time profiles (e.g., PTD) since the information about the location of the wells with respect to the flow path is not entered into the analysis. For this reason, the interpretation of the results should not neglect all the hydrogeological information concerning the structure of the aquifers, the flow directions, and the nature of the investigated contaminants.

Another limitation of this type of analysis is associated with being unsupervised techniques. Since it is a data-mining technique, it does not support an absolute classification of the wells, for a validation of the grouping obtained by the cluster analysis. As for every data-mining application, the main validation is provided by the environmental interpretability of the solutions and their information content which are also driven by the dataset quality. In this regard, the choice of the w parameter was here performed by exploring possible solutions, and it was based on the hydrogeological knowledge of the study area dynamics. On the other hand, as mentioned, validation is not supported for unsupervised applications. The CVI index could help to compare different solutions. However, each CVI measures specific cluster characteristics such as intra-cluster variability, compactness, and separation among clusters, but none of these characteristics are exhaustive for determining the environmental interpretability and the solution's usefulness in highlighting relevant information for the specific purpose of the study. Furthermore, there is no standardized use of the CVIs: different scientific works use different CVIs, and comparison works highlighted that single CVIs have low performance, further decreased when the structure of the analyzed dataset encompasses noisy data, overlapping clusters or cluster which are closer in the variables space [56].

The specific application of the present study is out of the scope of source identification and apportionment or attenuation processes assessment, mainly because of the data availability. Nevertheless, the proposed method could easily be integrated with multivariate source apportionment techniques, isotopic analyses and attenuation processes investigations.

4. Conclusions

In this work, time series analysis of contamination data in a historically contaminated urban area was undertaken: first, an exploratory analysis was performed through Mann-Kendall and Sen's Slope estimator and then univariate and multivariate time series cluster analysis were carried out.

The main conclusions of this work can be summarized in the following points:

- Time series analysis of contamination data provides deep insights on the processes governing water quality, which would not be provided by the analysis of single field surveys
- Results of the exploratory analysis highlighted that the most common methods for trend analysis such as Mann-Kendall and Sen's Slope could be non-exhaustive when dealing with highly variable groundwater chemical data since (a) they only work with monotonous trends and struggle with oscillatory behaviors and (b) they do not discriminate between lower and higher concentrations, focusing only on the trend's shape while even increasing trends, over very low concentrations, can have scarce environmental relevance.
- Time series clustering overcame these issues and demonstrated to be an efficient tool for summarizing spatio-temporal variability of contamination data, allowing for an easier interpretation, and supporting the implementation of data-driven monitoring strategies

- The implementation of data-driven monitoring strategies can lead to more efficient, site-specific monitoring networks, able to avoid redundant analysis to focus on relevant or alarming trends.

Future lines of research based on the results presented in this work may deal with a validation of the cluster interpretations, through e.g., a backward approach thus investigating a contaminated site chemically and isotopically. Furthermore, possible future step of this work involve the widening of the analysed contaminants set and the validation of the developed monitoring strategies, evaluating their future effectiveness in identifying critical situations and in monitoring ongoing processes.

The approach proposed in this work represents an easily reproducible methodology, ready-to-use for researchers, water suppliers, practitioners and environmental protection agencies. This methodology could indeed serve for several different applications in the different fields of groundwater quality assessment, monitoring, and management.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/w15010148/s1>, Figure S1: (a) cross section, modified from: Carloni, A. Studio Di Fattibilità per La Realizzazione Degli Interventi Di Messa in Sicurezza e Bonifica Delle Acque Di Falda Del “SIN Brescia Caffaro”.; 2013; (b) study area, cross section and lithologs.

Author Contributions: Conceptualization, C.Z.; methodology, C.Z.; validation, T.B., M.R., L.F. and C.Z.; formal analysis, C.Z., A.R. and M.C.; resources, C.S.; data curation, C.Z., A.R. and M.C.; writing—original draft preparation, C.Z.; writing—review and editing, C.Z., M.R., A.R., M.C., L.F., C.S., D.S. and T.B.; visualization, C.Z., A.R. and D.S.; supervision, C.Z., M.R., and T.B.; project administration, C.Z. and T.B.; funding acquisition, C.Z. and T.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by A2A Ciclo Idrico Spa, contract number 2020-ECO-0025.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Restrictions apply to the availability of these data. Data was obtained from A2A Ciclo Idrico Spa and are available from the authors with the permission of A2A Ciclo Idrico Spa.

Acknowledgments: We thank Daniel T. Feinstein of USGS for providing valuable suggestions and English revisions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. UNESCO. *The Role of Sound Groundwater Resource Management and Governance to Achieve Water Security*; UNESCO: Paris, France; i-WSSM: Daejeon, Republic of Korea, 2021; ISBN 9789231004681.
2. UNESCO. *Water Security and the Sustainable Development Goals*; UNESCO: Paris, France; i-WSSM: Daejeon, Republic of Korea, 2019; ISBN 9789231003233.
3. United Nations. *Water Development Report 2022: Groundwater: Making the Invisible Visible*; United Nations: New York, NY, USA, 2022; ISBN 9789231005077.
4. Zanotti, C.; Rotiroti, M.; Sterlacchini, S.; Cappellini, G.; Fumagalli, L.; Stefania, G.A.; Nannucci, M.S.; Leoni, B.; Bonomi, T. Choosing between Linear and Nonlinear Models and Avoiding Overfitting for Short and Long Term Groundwater Level Forecasting in a Linear System. *J. Hydrol.* **2019**, *578*, 124015. [[CrossRef](#)]
5. Wunsch, A.; Liesch, T.; Broda, S. Forecasting Groundwater Levels Using Nonlinear Autoregressive Networks with Exogenous Input (NARX). *J. Hydrol.* **2018**, *567*, 743–758. [[CrossRef](#)]
6. Bakker, M.; Schaars, F. Solving Groundwater Flow Problems with Time Series Analysis: You May Not Even Need Another Model. *Groundwater* **2019**, *57*, 826–833. [[CrossRef](#)] [[PubMed](#)]
7. Giese, M.; Haaf, E.; Heudorfer, B.; Barthel, R. Comparative Hydrogeology–Reference Analysis of Groundwater Dynamics from Neighbouring Observation Wells. *Hydrol. Sci. J.* **2020**, *65*, 1685–1706. [[CrossRef](#)]
8. Kayhomayoon, Z.; Milan, S.G.; Azar, N.A.; Kardan, H. A New Approach for Regional Groundwater Level Simulation: Clustering, Simulation, and Optimization. *Nat. Resour. Res.* **2021**, *30*, 4165–4185. [[CrossRef](#)]

9. De Luca, D.A.; Destefanis, E.; Forno, M.G.; Lasagna, M.; Masciocco, L. The Genesis and the Hydrogeological Features of the Turin Po Plain Fontanili, Typical Lowland Springs in Northern Italy. *Bull. Eng. Geol. Environ.* **2014**, *73*, 409–427. [[CrossRef](#)]
10. Frollini, E.; Preziosi, E.; Calace, N.; Guerra, M.; Guyennon, N.; Marcaccio, M.; Menichetti, S.; Romano, E.; Ghergo, S. Groundwater Quality Trend and Trend Reversal Assessment in the European Water Framework Directive Context: An Example with Nitrates in Italy. *Environ. Sci. Pollut. Res.* **2021**, *28*, 22092–22104. [[CrossRef](#)]
11. Meggiorin, M.; Passadore, G.; Bertoldo, S.; Sottani, A.; Rinaldo, A. 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 Sotter. Ital. J. Groundw.* **2021**, *10*, 35–48. [[CrossRef](#)]
12. Egidio, E.; Lasagna, M.; Mancini, S.; De Luca, D.A. Climate Impact Assessment to the Groundwater Levels Based on Long Time-Series Analysis in a Paddy Field Area (Piedmont Region, NW Italy): Preliminary Results. *Acque Sotter. Ital. J. Groundw.* **2022**, *11*, 21–29. [[CrossRef](#)]
13. Barbieri, M.; Franchini, S.; Barberio, M.D.; Billi, A.; Boschetti, T.; Giansante, L.; Gori, F.; Jónsson, S.; Petitta, M.; Skelton, A.; et al. Changes in Groundwater Trace Element Concentrations before Seismic and Volcanic Activities in Iceland during 2010–2018. *Sci. Total Environ.* **2021**, *793*, 148635. [[CrossRef](#)]
14. Rai, P.; Singh, S. A Survey of Clustering Techniques. *Int. J. Comput. Appl.* **2010**, *7*, 1–5. [[CrossRef](#)]
15. Aghabozorgi, S.; Seyed Shirkhorshidi, A.; Ying Wah, T. Time-Series Clustering—A Decade Review. *Inf. Syst.* **2015**, *53*, 16–38. [[CrossRef](#)]
16. Kumar, R.; Nagabhushan, P. Time Series as a Point—A Novel Approach for Time Series Cluster Visualization. *Conf. Data Min.* **2006**, 24–29. Available online: <https://www.semanticscholar.org/paper/Time-Series-as-a-Point-A-Novel-Approach-for-Time-Kumar-Nagabhushan/507cc47a5d0954fd87591929c50974d96c93ad24> (accessed on 10 November 2022).
17. Li, L.; Prakash, B.A. Time Series Clustering: Complex Is Simpler! 2011. Available online: <https://www.pdl.cmu.edu/PDL-FTP/associated/li-icml11-time.pdf> (accessed on 10 November 2022).
18. Rani, S. Recent Techniques of Clustering of Time Series Data: A Survey. *Int. J. Comput. Appl.* **2012**, *52*, 1–9. [[CrossRef](#)]
19. Caiado, J.; Maharaj, E.A.; D’Urso, P. Time Series Clustering. In *Handbook of Cluster Analysis*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2015; pp. 241–263.
20. Akay, Ö. Examination of the 21 European Countries and Turkey in Terms of Water Resources along with the Effect of Climate Change by Time Series Clustering. *Environ. Earth Sci.* **2021**, *80*, 784. [[CrossRef](#)]
21. Utimula, K.; Hunkao, R.; Yano, M.; Kimoto, H.; Hongo, K.; Kawaguchi, S.; Suwanna, S.; Maezono, R. Machine-Learning Clustering Technique Applied to Powder X-Ray Diffraction Patterns to Distinguish Compositions of ThMn12-Type Alloys. *Adv. Theory Simul.* **2020**, *3*, 2000039. [[CrossRef](#)]
22. Warren Liao, T. Clustering of Time Series Data—A Survey. *Pattern Recognit.* **2005**, *38*, 1857–1874. [[CrossRef](#)]
23. Prakaisak, I.; Wongchaisuwat, P. Hydrological Time Series Clustering: A Case Study of Telemetry Stations in Thailand. *Water* **2022**, *14*, 2095. [[CrossRef](#)]
24. Lee, W.; Zeyar, W.; Catalina, A.; Stuart, F.; Eds, M.; Goebel, R.; Arslan, Y.; Küçük, D.; Eren, S.; Birturk, A. Clustering River Basins Using Time-Series Data Mining on Hydroelectric Energy Generation. In Proceedings of the International Workshop on Data Analytics for Renewable Energy Integration, Dublin, Ireland, 10 September 2018; pp. 103–115.
25. Mishra, S.; Saravanan, C.; Dwivedi, V.K.; Shukla, J.P. Rainfall-Runoff Modeling Using Clustering and Regression Analysis for the River Brahmaputra Basin. *J. Geol. Soc. India* **2018**, *92*, 305–312. [[CrossRef](#)]
26. Sartirana, D.; Rotiroti, M.; Bonomi, T.; De Amicis, M.; Nava, V.; Fumagalli, L.; Zanotti, C. Data-Driven Decision Management of Urban Underground Infrastructure through Groundwater-Level Time-Series Cluster Analysis: The Case of Milan (Italy). *Hydrogeol. J.* **2022**, *30*, 1157–1177. [[CrossRef](#)]
27. Naranjo-Fernández, N.; Guardiola-Albert, C.; Aguilera, H.; Serrano-Hidalgo, C.; Montero-González, E. Clustering Groundwater Level Time Series of the Exploited Almonte-Marismas Aquifer in Southwest Spain. *Water* **2020**, *12*, 1063. [[CrossRef](#)]
28. Rinderer, M.; van Meerveld, H.J.; McGlynn, B.L. From Points to Patterns: Using Groundwater Time Series Clustering to Investigate Subsurface Hydrological Connectivity and Runoff Source Area Dynamics. *Water Resour. Res.* **2019**, *55*, 5784–5806. [[CrossRef](#)]
29. Moghaddam, H.K.; Milan, S.G.; Kayhomayoon, Z.; Kivi, Z.R.; Azar, N.A. The Prediction of Aquifer Groundwater Level Based on Spatial Clustering Approach Using Machine Learning. *Environ. Monit. Assess.* **2021**, *193*, 173. [[CrossRef](#)] [[PubMed](#)]
30. Huang, L.; Feng, H.; Le, Y. Finding Water Quality Trend Patterns Using Time Series Clustering: A Case Study. In Proceedings of the IEEE Fourth International Conference on Data Science in Cyberspace (DSC), Hangzhou, China, 23–25 June 2019; pp. 330–337. [[CrossRef](#)]
31. Lee, S.; Kim, J.; Hwang, J.; Lee, E.J.; Lee, K.J.; Oh, J.; Park, J.; Heo, T.Y. Clustering of Time Series Water Quality Data Using Dynamic Time Warping: A Case Study from the Bukhan River Water Quality Monitoring Network. *Water* **2020**, *12*, 2411. [[CrossRef](#)]
32. Pollicino, L.C.; Masetti, M.; Stevenazzi, S.; Colombo, L.; Alberti, L. Spatial Statistical Assessment of Groundwater PCE (Tetrachloroethylene) Diffuse Contamination in Urban Areas. *Water* **2019**, *11*, 1211. [[CrossRef](#)]
33. Alberti, L.; Azzellino, A.; Colombo, L.; Lombi, S. Cluster Analysis to Identify Tetrachloroethylene Pollution Hotspots for Transport Numerical Model Implementation in Urban Functional Area of Milan, Italy. In Proceedings of the 16th International Multidisciplinary Scientific Conference SGEM2016, Albena, Bulgaria, 28 June–7 July 2016; Book 1.

34. Azzellino, A.; Colombo, L.; Lombi, S.; Marchesi, V.; Piana, A.; Merri, A.; Alberti, L. Groundwater Diffuse Pollution in Functional Urban Areas: The Need to Define Anthropogenic Diffuse Pollution Background Levels. *Sci. Total Environ.* **2019**, *656*, 1207–1222. [CrossRef]
35. Kottek, M.; Grieser, J.; Beck, C.; Rudolf, B.; Rubel, F. World Map of the Köppen-Geiger Climate Classification Updated. *Meteorol. Z.* **2006**, *15*, 259–263. [CrossRef]
36. Francani, V. La Stato Di Inquinamento Delle Risorse Idriche Della Pianura Padana e Gli Interventi Possibili. In *Studi Idrogeologici Sulla Pianura Padana*; 1987. Available online: http://wwwdb.gndci.cnr.it/php2/gndci/gndci_f_regione.php?®ione=Italia+Settentrionale&inizio=50&formato=&lingua=en (accessed on 10 November 2022).
37. Vercesi, P.L. Aspetti Quali-Quantitativi Delle Risorse Idriche Sotterranee Del Bresciano. *Nat. Brescia* **1994**, *29*, 21–52.
38. Denti, E.; Lauzi, S.; Sala, P.; Scesi, L. Studio Idrogeologico Della Pianura Bresciana Tra i Fiumi Oglio e Chiese. In *Studi Idrogeologici Sulla Pianura Padana*; ERSAL: Milano, Italy, 1998.
39. Gasparetti, D.; Tribani, M.; Ribolla, G.; Gavazzi, F.; Treccani, L. *Adeguamento Della Componente Geologica, Idrogeologica e Sismica Del PGT Al Piano Di Gestione Del Rischio Alluvioni*; 2009. Available online: https://www.comune.brescia.it/servizi/urbanistica/PGT/Pagine/pgt_approvazione_%20variante_idrogeologica.aspx (accessed on 10 November 2022).
40. Osservatorio Acqua Bene Comune. *Primo Rapporto*; Osservatorio Acqua Bene Comune: Comune di Brescia, Brescia, 2015.
41. ARPA—Lombardia. *Attività Di Affinamento Delle Conoscenze Sulla Contaminazione delle Acque Sotterranee in Cinque Aree Della Provincia Di Brescia Con Definizione Dei Plumes Di Contaminanti Ed Individuazione Delle Potenziali Fonti Di Contaminazione—Area BS002—Brescia—C*; ARPA: Milano, Italy, 2016.
42. ARPA—Lombardia. *Attività Di Affinamento Delle Conoscenze Sulla Contaminazione delle Acque Sotterranee in Cinque Aree Della Provincia Di Brescia Con Definizione Dei Plumes Di Contaminanti Ed Individuazione Delle Potenziali Fonti Di Contaminazione—Lotto A—Area BS001—F*; ARPA: Milano, Italy, 2015.
43. WHO. *A Global Overview of National Regulations and Standards for Drinking-Water Quality. Second Edition*; WHO: Geneva, Switzerland, 2021; ISBN 978-92-4-151376-0.
44. European Commission. *Guidance Document No. 18 Guidance on Groundwater Status and Trend Assessment*; European Commission: Brussels, Belgium, 2009; ISBN 9789279113741.
45. Mann, H.B. Nonparametric Tests Against Trend. *Econometrica* **1945**, *13*, 245–259. [CrossRef]
46. Kendall, M.G. *Rank Correlation Methods*; Charles Griffin: London, UK, 1975.
47. Sen, P.K. Estimates of the Regression Coefficient Based on Kendall’s Tau. *J. Am. Stat. Assoc.* **1968**, *63*, 1379–1389. [CrossRef]
48. Almazroui, M.; Şen, Z. Trend Analyses Methodologies in Hydro-Meteorological Records. *Earth Syst. Environ.* **2020**, *4*, 713–738. [CrossRef]
49. Giorgino, T. Computing and Visualizing Dynamic Time Warping Alignments in R: The Dtw Package. *J. Stat. Softw.* **2009**, *31*, 1–24. [CrossRef]
50. Haaf, E.; Barthel, R. An Inter-Comparison of Similarity-Based Methods for Organisation and Classification of Groundwater Hydrographs. *J. Hydrol.* **2018**, *559*, 222–237. [CrossRef]
51. Chu, S.; Keogh, E.; Hart, D.; Pazzani, M. Iterative Deepening Dynamic Time Warping for Time Series. In Proceedings of the 2002 SIAM International Conference on Data Mining, Arlington, VA, USA, 11–13 April 2002; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2002; pp. 195–212. [CrossRef]
52. Sakoe, H. Dynamic-Programming Approach to Continuous Speech Recognition. In *1971 Proceedings of the International Congress of Acoustics*; Akademiai Kiado: Budapest, Hungary, 1971.
53. Sakoe, H.; Chiba, S. Dynamic Programming Algorithm Optimization for Spoken Word Recognition. *IEEE Trans. Acoust.* **1978**, *26*, 43–49. [CrossRef]
54. Dau, H.A.; Silva, D.F.; Petitjean, F.; Forestier, G.; Bagnall, A.; Mueen, A.; Keogh, E. Optimizing Dynamic Time Warping’s Window Width for Time Series Data Mining Applications. *Data Min. Knowl. Discov.* **2018**, *32*, 1074–1120. [CrossRef]
55. Kryszczuk, K.; Hurley, P. Estimation of the Number of Clusters Using Multiple Clustering Validity Indices. In *Multiple Classifier Systems. MCS 2010*; Springer: Berlin/Heidelberg, Germany, 2010; Volume 3590, pp. 114–123.
56. Arbelaitz, O.; Gurrutxaga, I.; Muguerza, J. An Extensive Comparative Study of Cluster Validity Indices. *Pattern Recognit.* **2013**, *46*, 243–256. [CrossRef]
57. Rousseeuw, P.J. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. [CrossRef]
58. Kim, M.; Ramakrishna, R.S. New Indices for Cluster Validity Assessment. *Pattern Recognit. Lett.* **2005**, *26*, 2353–2363. [CrossRef]
59. Saitta, S.; Raphael, B.; Smith, I.F.C. A Bounded Index for Cluster Validity. In *Machine Learning and Data Mining in Pattern Recognition. MLDM 2007*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 174–187.
60. Nijenhuis, I.; Schmidt, M.; Pellegatti, E.; Paramatti, E.; Richnow, H.H.; Gargini, A. A Stable Isotope Approach for Source Apportionment of Chlorinated Ethene Plumes at a Complex Multi-Contamination Events Urban Site. *J. Contam. Hydrol.* **2013**, *153*, 92–105. [CrossRef]
61. Colyer, A.; Butler, A.; Peach, D.; Hughes, A. How Groundwater Time Series and Aquifer Property Data Explain Heterogeneity in the Permo-Triassic Sandstone Aquifers of the Eden Valley, Cumbria, UK. *Hydrogeol. J.* **2022**, *30*, 445–462. [CrossRef]
62. Shen, S.; Chi, M. Clustering Student Sequential Trajectories Using Dynamic Time Warping. In Proceedings of the 10th International Conference on Educational Data Mining (EDM 2017), Wuhan, China, 25–28 June 2017; pp. 266–271.

63. Lafare, A.E.A.; Peach, D.W.; Hughes, A.G. Use of Seasonal Trend Decomposition to Understand Groundwater Behaviour in the Permo-Triassic Sandstone Aquifer, Eden Valley, UK. *Hydrogeol. J.* **2016**, *24*, 141–158. [[CrossRef](#)]
64. Zanutti, C.; Caschetto, M.; Bonomi, T.; Parini, M.; Cipriano, G.; Fumagalli, L.; Rotiroti, M. Linking Local Natural Background Levels in Groundwater to Their Generating Hydrogeochemical Processes in Quaternary Alluvial Aquifers. *Sci. Total Environ.* **2021**, *805*, 150259. [[CrossRef](#)]
65. Stefania, G.A.; Zanutti, C.; Bonomi, T.; Fumagalli, L.; Rotiroti, M. Determination of Trigger Levels for Groundwater Quality in Landfills Located in Historically Human-Impacted Areas. *Waste Manag.* **2018**, *75*, 400–406. [[CrossRef](#)]
66. Parrone, D.; Frollini, E.; Preziosi, E.; Ghergo, S. ENaBLE, an On-Line Tool to Evaluate Natural Background Levels in Groundwater Bodies. *Water* **2021**, *13*, 74. [[CrossRef](#)]
67. Bouteraa, O.; Mebarki, A.; Bouaicha, F.; Nouaceur, Z.; Laignel, B. Groundwater Quality Assessment Using Multivariate Analysis, Geostatistical Modeling, and Water Quality Index (WQI): A Case of Study in the Boumerzoug-El Khroub Valley of Northeast Algeria. *Acta Geochim.* **2019**, *38*, 796–814. [[CrossRef](#)]
68. Zolekar, R.B.; Todmal, R.S.; Bhagat, V.S.; Bhailume, S.A. Hydro—Chemical Characterization and Geospatial Analysis of Groundwater for Drinking and Agricultural Usage in Nashik District in Maharashtra, India. *Environ. Dev. Sustain.* **2021**, *23*, 4433–4452. [[CrossRef](#)]
69. Egidio, E.; Mancini, S.; De Luca, D.A.; Lasagna, M. The Impact of Climate Change on Groundwater Temperature of the Piedmont Po Plain (NW Italy). *Water* **2022**, *14*, 2797. [[CrossRef](#)]
70. Li, J.; Hassan, D.; Brewer, S.; Sitzenfrei, R. Is Clustering Time-Series Water Depth Useful? An Exploratory Study for Flooding Detection in Urban Drainage Systems. *Water* **2020**, *12*, 2433. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.