hydrology

MDPI

Article

Multivariate and Spatial Study and Monitoring Strategies of Groundwater Quality for Human Consumption in Corsica

Hajar Lazar [1], Meryem Ayach [1], Abderrahim Bousouis [2], Frederic Huneau [3], Christophe Mori [3], Emilie Garel [3], Ilias Kacimi [1], Vincent Valles [4,5] and Laurent Barbiero [6,*]

[1] Geosciences, Water and Environment Laboratory, Faculty of Sciences Rabat, Mohammed V University, venue Ibn Batouta, Rabat 10100, Morocco; hajar_lazar2@um5.ac.ma (H.L.); meryem_ayach@um5.ac.ma (M.A.); i.kacimi@um5r.ac.ma (I.K.)
[2] Laboratoire de Géosciences, Faculté des Sciences, Université Ibn Tofaïl, BP 133, Kénitra 14000, Morocco; abderrahim.bousouis@uit.ac.ma
[3] Université de Corse Pascal Paoli, CNRS UMR 6134 SPE, Campus Grimaldi, BP52, 20250 Corte, France; huneau@univ-corse.fr (F.H.); mori@univ-corse.fr (C.M.); garel_e@univ-corse.fr (E.G.)
[4] Mixed Research Unit EMMAH (Environnement Méditerranéen et Modélisation des Agro-Hydrosystèmes), Hydrogeology Laboratory, Avignon University, 84916 Avignon, France; vincent.valles@outlook.fr
[5] Faculté des Sciences et Techniques (FSTBM), BP 523, Beni Mellal 23000, Morocco
[6] Institut de Recherche pour le Développement, Géoscience Environnement Toulouse, CNRS, University of Toulouse, Observatoire Midi-Pyrénées, UMR 5563, 14 Avenue Edouard Belin, 31400 Toulouse, France
* Correspondence: laurent.barbiero@get.omp.eu

Abstract: Groundwater, widely used for supplying drinking water to populations, is a vital resource that must be managed sustainably, which requires a thorough understanding of its diverse physico-chemical and bacteriological characteristics. This study, based on a 27-year extraction from the Sise-Eaux database (1993–2020), focused on the island of Corsica (72,000 km^2), which is diverse in terms of altitude and slopes and features a strong lithological contrast between crystalline Corsica and metamorphic and sedimentary Corsica. Following logarithmic conditioning of the data (662 water catchments, 2830 samples, and 15 parameters) and distinguishing between spatial and spatiotemporal variances, a principal component analysis was conducted to achieve dimensionality reduction and to identify the processes driving water diversity. In addition, the spatial structure of the parameters was studied. The analysis notably distinguishes a seasonal determinism for bacterial contamination (rain, runoff, bacterial transport, and contamination of catchments) and a more strictly spatial determinism (geographic, lithological, and land use factors). The behavior of each parameter allowed for their classification into seven distinct groups based on their average coordinates on the factorial axes, accounting for 95% of the dataset's total variance. Several strategies can be considered for the inventory and mapping of groundwater, namely, (1) establishing quality parameter distribution maps, (2) dimensionality reduction through principal component analysis followed by two sub-options: (2a) mapping factorial axes or (2b) establishing a typology of parameters based on their behavior and mapping a representative for each group. The advantages and disadvantages of each of these strategies are discussed.

Keywords: water quality monitoring; groundwater database; bacteriological composition; chemical composition; cluster analysis; principal component analysis; Corsica

Citation: Lazar, H.; Ayach, M.; Bousouis, A.; Huneau, F.; Mori, C.; Garel, E.; Kacimi, I.; Valles, V.; Barbiero, L. Multivariate and Spatial Study and Monitoring Strategies of Groundwater Quality for Human Consumption in Corsica. *Hydrology* **2024**, *11*, 197. https://doi.org/10.3390/hydrology11110197

Academic Editor: Tammo Steenhuis

Received: 28 October 2024
Revised: 18 November 2024
Accepted: 19 November 2024
Published: 20 November 2024

1. Introduction

The definition of drinking water quality is based on the comparison of various bacteriological, chemical, organoleptic, and radiological criteria on the one hand, and the analysis results of these same criteria in water on the other [1,2]. With the continuous development of analytical techniques, the list of quality criteria is becoming increasingly long, making it complex to provide a concise evaluation of water quality for human consumption. This list now includes hundreds of pesticides and their metabolites, heavy

metals, bacteriological parameters, as well as the physico-chemistry of major ions, metals, and other trace elements [3–8]. In this context, synthetic mapping of drinking water quality requires strategic choices. Often, the most critical parameters are selected, but this approach does not always clearly reveal the mechanisms responsible for the situation or account for their spatial and temporal variations. Groundwater plays a crucial role among the available water resources for human consumption, primarily due to its lower vulnerability to surface pollution [9–11]. Consequently, several databases on groundwater quality are now emerging such as the Global Groundwater Information System (GGIS), the World Water Assessment Program (WWAP) Groundwater Database, the Water Data of the United States Geological Survey (USGS), the European Environment Agency (EEA) Groundwater Database, and the French Groundwater Information and Management System, known under the acronym "Sise-Eaux" [12–16]. The establishment of such databases is essential for monitoring groundwater quality, particularly within the framework of sustainable water resource management [17–20]. These databases enable researchers, government authorities, and water managers to track quality trends, assess risks, and implement strategies for aquifer protection. However, a comprehensive understanding of aquifers requires long-term monitoring and a high density of sampling points, which entails significant costs for the organizations responsible for this oversight.

The aim of this study is precisely to extract and interpret the information contained in a large groundwater database for Corsica, a Mediterranean island with significant geological and altitudinal diversity, in order to propose rational approaches for cost-effective monitoring and surveillance. To achieve this, the study will compare the spatial structure of various quality criteria, examine their potential multiple correlations, and propose a synthetic, spatially referenced approach to groundwater quality assessment.

2. Materials and Methods

2.1. Study Area, the Corsica Island

Corsica is a mountainous French island in the Mediterranean, stretching 180 km from north to south and 82 km at its widest point, covering a total area of 8722 km^2 (Figure 1). The average altitude is 568 m, with numerous peaks exceeding 2000 m, and the highest point being Monte Cinto at 2706 m. The significant elevation changes create considerable variations in the landscape depending on the watersheds. Two major geological regions can be distinguished: on one hand, the Hercynian Corsica, consisting of crystalline rocks in the west and south (granodiorites, monzogranites, alkaline granites, volcanic formations in the northwest, and a basic tholeiitic complex [21]); on the other hand, Alpine Corsica in the northeast, composed of metamorphic formations, mainly schists. Carbonate-quartz sedimentary formations are found in the extreme south of the island.

The climate is primarily Mediterranean but varies depending on the location on the island and the altitude. On the coast, summers are hot and dry, with temperatures exceeding 30 °C. Winters are mild (with an average temperature of around 10 °C) and humid, with moderate rainfall. Inland, due to the increasing altitude, summers are cooler, with temperatures rarely exceeding 25 °C. Precipitation is higher, often falling as snow in winter from altitudes of 600 to 800 m. The island is exposed to strong winds, such as the Mistral (from the northwest) and the Libeccio (from the southwest).

Under the Water Framework Directive (WFD) [22–24], the French Geological Survey (BRGM) mapped the island's aquifers, identifying 40 distinct groundwater bodies. These aquifers are highly variable due to the island's complex geology [25], with significant differences in thickness, depth, and hydraulic conductivity. The majority consist of compartmentalized and fractured aquifers (over 20 groundwater bodies) within the island's granitic and metamorphic bedrock, which provide approximately 60% of the water extracted for drinking supplies in local communities. Sedimentary aquifers include the karstified Bonifacio molasses in the island's far south and the Miocene aquifer in the eastern plain, which remains relatively unexplored [26]. Coastal aquifers alongside rivers are distributed around the island and significantly contribute to water supply for populations [27].

These groundwater resources are fragile due to the risks of saline intrusions in aquifers near the coast and, more generally, due to limited surface protection. In the context of climate change, authorities anticipate a decrease in aquifer recharge, increased evapotranspiration, and a reduction in effective rainfall. The flow rates of island springs, often modest and highly dependent on this recharge, are expected to decline, potentially leading to the drying up of some springs during low-flow periods [28].

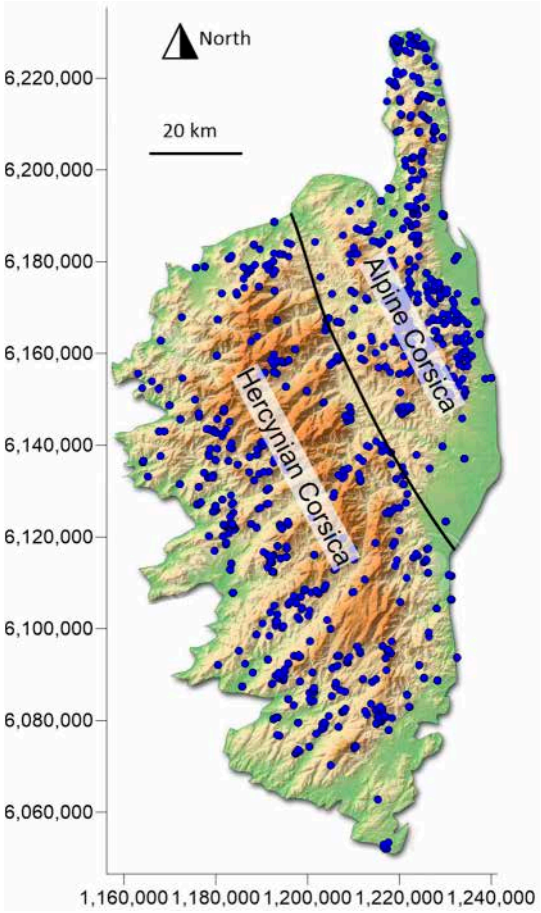


Figure 1. Relief, distribution of groundwater sampling points, and distinction between Hercynian Corsica and Alpine Corsica.

2.2. The Sise-Eaux Database

Sise-Eaux is a database that centralizes the archiving of surface and groundwater quality data intended for human consumption [16]. This database is managed nationally but administered regionally by Health Agencies (ARS). The water samples were collected and analyzed by the ARS service provider laboratories, approved by the Ministry of Health, and have all the international certifications for analytical quality. The database is primarily fed by the results of sanitary controls on water sources that supply municipalities. It includes both untreated waters, directly from the source, and treated waters following disinfection by chlorination, filtration, or decantation. In this study, only raw groundwater was considered, and where several aquifers are superimposed, only the most superficial aquifer has been retained. For more details on data extraction, manual error correction, coordinate retrieval, and the creation of a final georeferenced dataset consisting of 2830 observations and 15 parameters, readers can refer to the previous work [25]. This extraction, limited to unconfined aquifers of the island of Corsica, covers a 27-year period (from April 1993 to September 2020), and the selected parameters include major ions (Ca, Mg, Na, SO_4, Cl, HCO_3), electrical conductivity (EC), bacteriological parameters (total Coliforms (Col.), revivable aerobic bacteria at 22 °C and 37 °C (Aer.22, Aer.37)), as well as fecal con-

tamination parameters *Enterococci* and *Escherichia coli* (Ent., *E.coli*), nitrate ions (NO_3), and two trace metals (Fe and Mn). In the following text, a distinction will be made between parameters and major ions, for example, SO_4 representing the analyzed ion parameter SO_4^{2-}. Ultimately, these 2830 water samples were collected from 662 sampling points (Figure 1), with an average of 4.3 samples per sampling point. Given the size of the study area, the samples do not all come from the same aquifer but from 40 groundwater bodies, averaging 71 samples per groundwater body [25]. It should be noted, moreover, that for the crystalline part of the island (representing two-thirds of the territory), this concept of groundwater body has a geographic meaning but lacks a clear hydrogeological significance, as it involves flow in a fractured environment.

2.3. Mathematical Tools

2.3.1. Normality Tests, Q-Q Plots, and Data Conditioning

Before analysis, the dataset underwent Kolmogorov–Smirnov normality tests, which are suitable for high-dimensional statistical distributions [26]. Additionally, a visual comparison of the distribution residuals for each parameter against a normal distribution with the same mean and standard deviation was made using quantile–quantile (Q-Q) plots. In these graphs, the diagonal represents the normal distribution, and the closer the data points are to this diagonal, the closer the distribution is to normality [27]. Based on these tests, logarithmic conditioning was applied to all parameters using the formula $y = \log_{10}(x + DL)$, where x represents the value of parameter X (whether physico-chemical or bacteriological), and DL is the determination limit. This conditioning was applied to reduce the impact of extreme values without removing them from the dataset, by dilating the gaps between low values and contracting those between high values [28]. Indeed, these extreme values can obscure certain processes responsible for the variation in water quality during analysis [29,30].

2.3.2. Principal Component Analysis

In order to reduce the dimensionality of the data space while minimizing the loss of information contained in the dataset, a principal component analysis (PCA) based on the correlation matrix was performed on all parameters [31]. This procedure, by diagonalizing the correlation matrix, based on standardized and centered data, ensures that each variable carries the same weight in the analysis, regardless of the unit used. This method, frequently employed due to its robustness, is based on the principle that the resulting factorial axes (principal components) are orthogonal to each other and thus carry concise information related to independent processes [32,33]. This helps identify and prioritize sources of variability. Dimensionality reduction was assessed using Bartlett's sphericity test [34].

In the context of the dataset, sampling was conducted at various points and dates, leading to a combination of spatial and temporal variability. To distinguish between these two aspects, two PCAs were calculated: one including all data, capturing both spatial and temporal variability, and the other based on the mean values of each parameter at each sampling point. This second approach minimizes temporal variance. However, the variability studied this way is not purely spatial since the samples were not all collected on the same date.

2.3.3. Agglomerative Hierarchical Clustering

Agglomerative hierarchical clustering (AHC) [35,36] was then performed based on the mean values of each parameter of the principal components obtained from the PCA, with the aim of grouping parameters by degree of similarity across the majority of the information contained in the dataset. For this analysis, 95% of the initial information was retained, with the remaining 5% considered statistical and analytical noise and eliminated. The relative similarities between the parameters were quantified using Euclidean distance, and the similarity levels at which the parameters were merged were used to construct a dendrogram.

2.3.4. Variograms and Map Calculation Method

The spatial structure of the parameters, and, subsequently, of the factorial axes, was studied through the construction and analysis of variograms, representing the evolution of semivariance between pairs of points as a function of the distance between them. As with the PCA processing, two variogram calculation methods were used [37]:

- Using the entire dataset. The variability measured by the semivariance is both spatial and temporal;
- Using the mean values for each sampling point to minimize temporal variance.

Experimental variograms obtained under similar conditions to allow for comparison (same number of points and same number of distance classes) were fitted with a model including a nugget effect and a spherical structure. This latter was used for developing parameter distribution maps for the island. The comparison of spatial structure is based on variogram characteristics such as range, sill, and nugget effect. To detect possible anisotropy in the distribution of parameter values, directional variograms were calculated with a 15° increment.

3. Results

The descriptive statistics of the 15 parameters are summarized in Table 1. The greatest variations concerned fecal bacteria, as well as the Ca and HCO_3 parameters.

Table 1. Descriptive statistics for the 15 log-transformed parameters.

Parameter (2830 Values)	Unit	Min.	Max.	Mean	Standard Deviation
Ent.	n/100 mL	0	2.44	0.32	0.50
E.coli	n/100 mL	0	2.55	0.23	0.44
Col.	n/100 mL	0	0.52	0.01	0.05
Aer.22	n/100 mL	0	2.00	0.02	0.15
Aer.37	n/100 mL	0	1.69	0.01	0.10
EC	mS cm^{-1}	1.49	3.12	2.38	0.30
Ca	mg L^{-1}	0.02	2.19	1.30	0.47
Mg	mg L^{-1}	−0.15	1.91	0.78	0.34
Cl	mg L^{-1}	0.46	2.34	1.22	0.32
SO_4	mg L^{-1}	0.16	2.13	0.96	0.31
Na	mg L^{-1}	0.27	2.09	1.04	0.30
HCO_3	mg L^{-1}	0.55	2.70	1.92	0.42
NO_3	mg L^{-1}	−1	1.55	0.23	0.30
Fe	µg L^{-1}	0.13	2.27	1.12	0.25
Mn	µg L^{-1}	−1	2.74	1.04	0.19

The values correspond to the calculation $y = \log_{10}(x + DL)$ with DL = 1 for bacteriological parameters, 10^{-3} mg L^{-1} for major ions and nitrate, and 10^{-3} µg L^{-1} for metals.

The normality tests showed that the statistical distributions of the parameters do not follow a normal distribution. However, the Q-Q plot of the statistical distribution of electrical conductivity, which is representative of major ions, presented as an example in Figure 2, highlights that logarithmic transformation significantly brought this distribution closer to normality (Figure 2a,b). The comparison of the first factorial plane calculated on raw and log-transformed data (Figure 2c,d) showed that the cloud representing the majority of observations was more spread out due to the reduction of the undesirable effect of extreme values. Finally, for this same parameter, while the variations in semivariance as a function of the distance between pairs of points were relatively similar (Figure 2e,f, all data), the distribution map of log-transformed values across the island was more informative, once again due to the reduced influence of extreme values (Figure 2g,h). Consequently, from here on, only the log-transformed data will be considered.

Several examples of spatio-temporal and spatial variograms obtained are presented in Figures 2f and 3. The difference between the two methods of calculating the variogram,

namely, using all data (spatio-temporal variance, black curve) or averaging each parameter at each sampling point (spatial variance, red curve), indicated, for each parameter, the significance of the strictly temporal variance. Electrical conductivity and major ions exhibited a high range of around 30 to 40 km depending on the parameters. The nugget effect was low, approximately 20% of the sill. The difference between the two types of variograms was small for major ions and conductivity, and slightly higher for chlorides (Figure 3d) and sodium Directional variograms revealed, again for electrical conductivity and major ions, a significant difference in semivariance beyond 30 km (Figure 3l), with higher semivariance in the N45° direction and lower in the N150° direction. For fecal contamination parameters (*E. coli* and *Enterococci*), a significant nugget effect was observed, ranging from about one-half to one-third of the sill, along with a low range of just a few kilometers at most (Figure 3g,h). The difference between the two methods of calculating the variogram was substantial, indicating high temporal variance relative to total variance. Similar characteristics (low range, high temporal variance) were also observed for coliforms (in this case, essentially temporal variance, Figure 3f) and indigenous revivable bacteria (Figure 3i). Metals (Figure 3j,k) exhibited a nugget effect equivalent to the sill, suggesting that variability is very high, even over very short distances, with a significant difference between the two variograms as well.

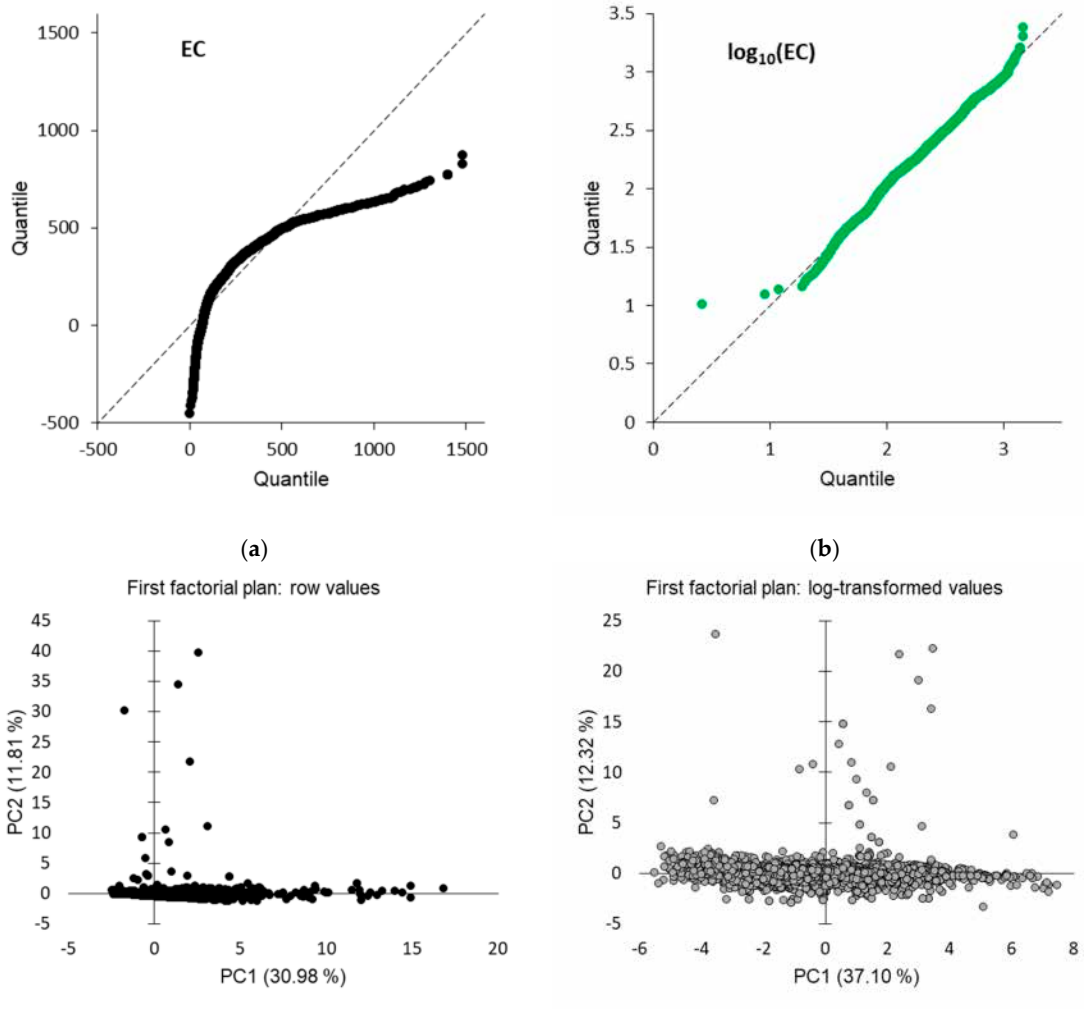


Figure 2. *Cont.*

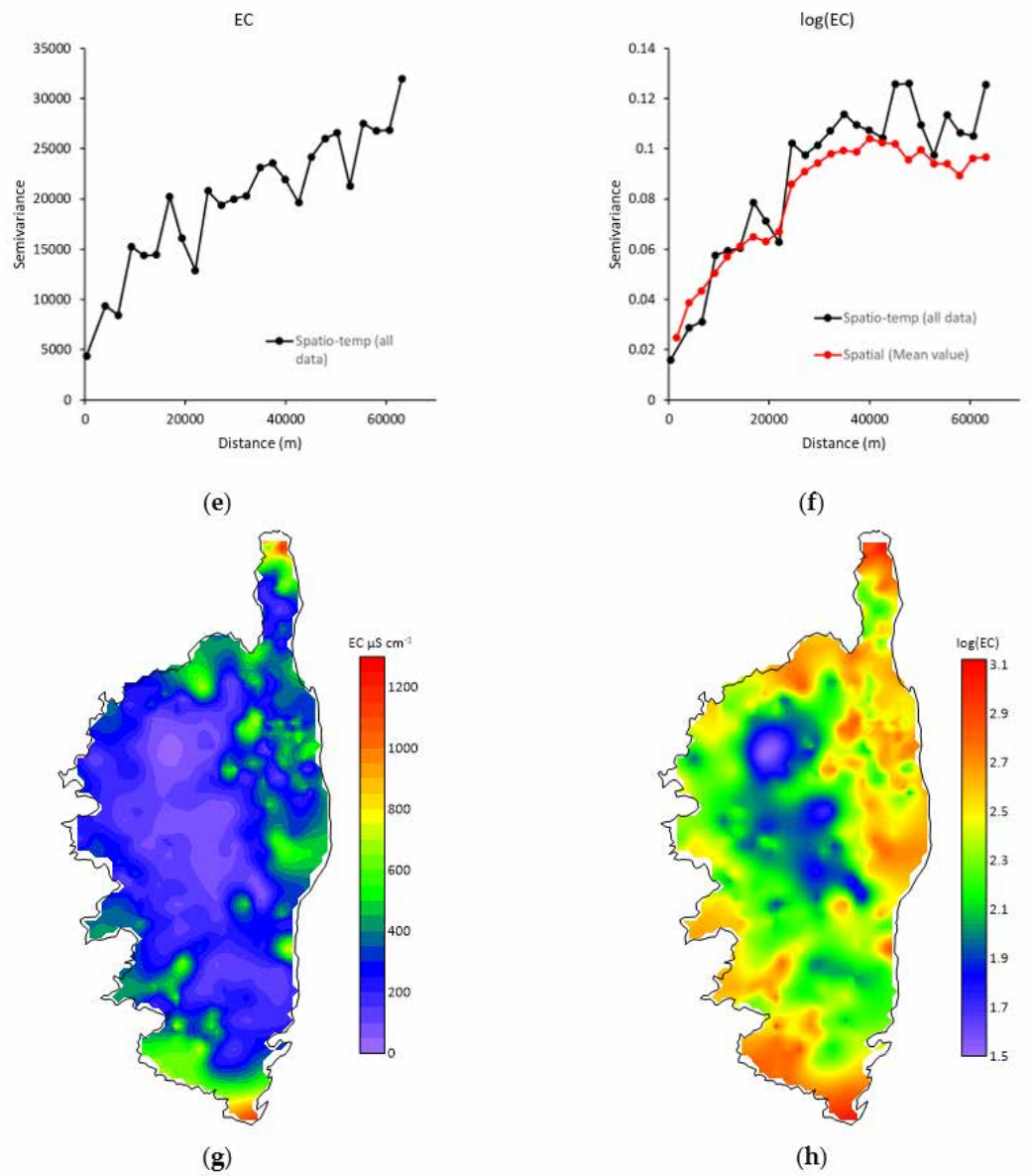


Figure 2. Effects of logarithmic data transformation on (**a**,**b**) the approximation to a normal distribution, (**c**,**d**) the spread of sampling points on the first factorial plane, (**e**,**f**) the variograms, and (**g**,**h**) the distribution maps of electrical conductivity in Corsica.

The distribution maps of several parameters are presented in Figure 4. These maps, developed from several thousand pairs of sampling points, revealed contrasting regions with varying distributions depending on the parameters considered. Certain parameters, such as major ions Mg, Ca, and SO_4, clearly distinguish the flows within the fractured rocks of Hercynian crystalline Corsica from the porous metamorphic environments of Alpine and sedimentary Corsica (Figure 4a,b,e). However, we also observed many similarities in the distributions, for example, between log(EC) (Figure 2h) and log(SO_4) or log(Mg) (Figure 4a,b), between log(Na) and log(Cl) (Figure 4c,d), and between log(*E.coli*) and log(Ent.) (Figure 4g,h). The similarities in the distribution of parameters across Corsica highlighted the redundancy of the information provided by the parameters, which justified dimensional reduction through PCA.

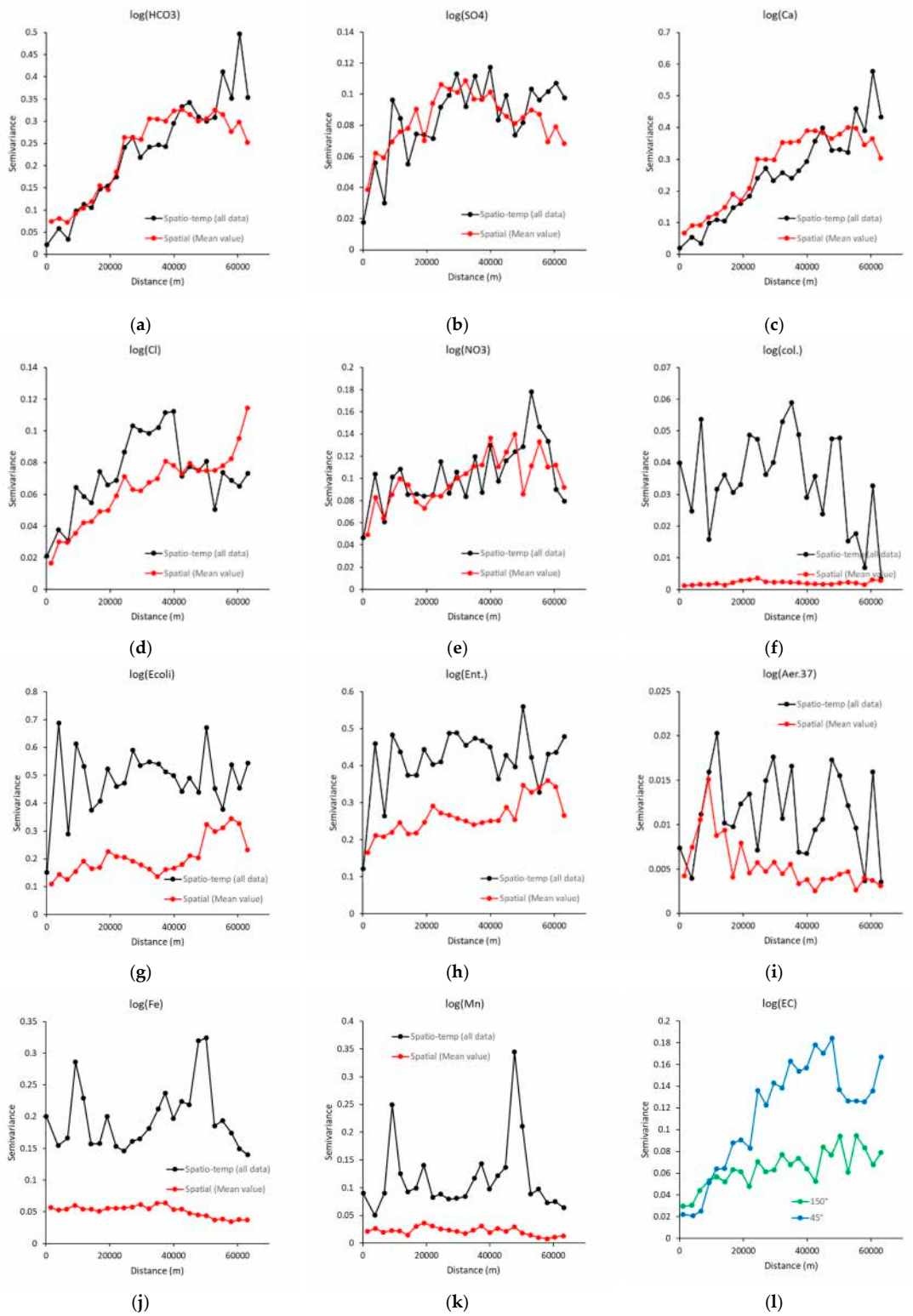


Figure 3. Experimental variograms for parameters (**a**) log(HCO_3), (**b**) log(SO_4), (**c**) log(Ca), (**d**) log(Cl), (**e**) log(NO_3), (**f**) log(Col.), (**g**) log(*E.coli*), (**h**) log(Ent.), (**i**) log(Aer.37), (**j**) log(Fe), (**k**) log(Mn), and (**l**) directional variogram for log(EC) at direction N45° and N150°.

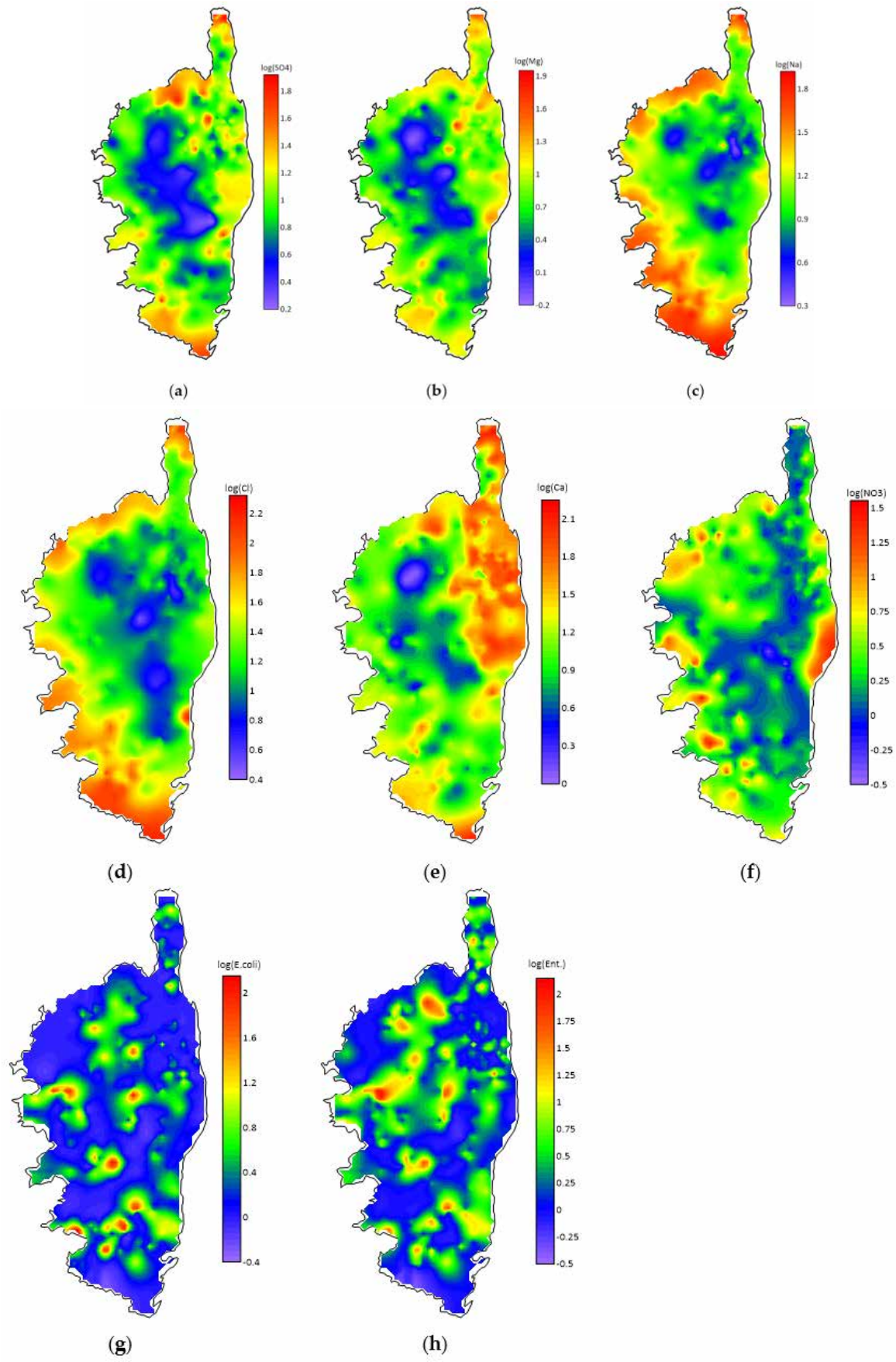
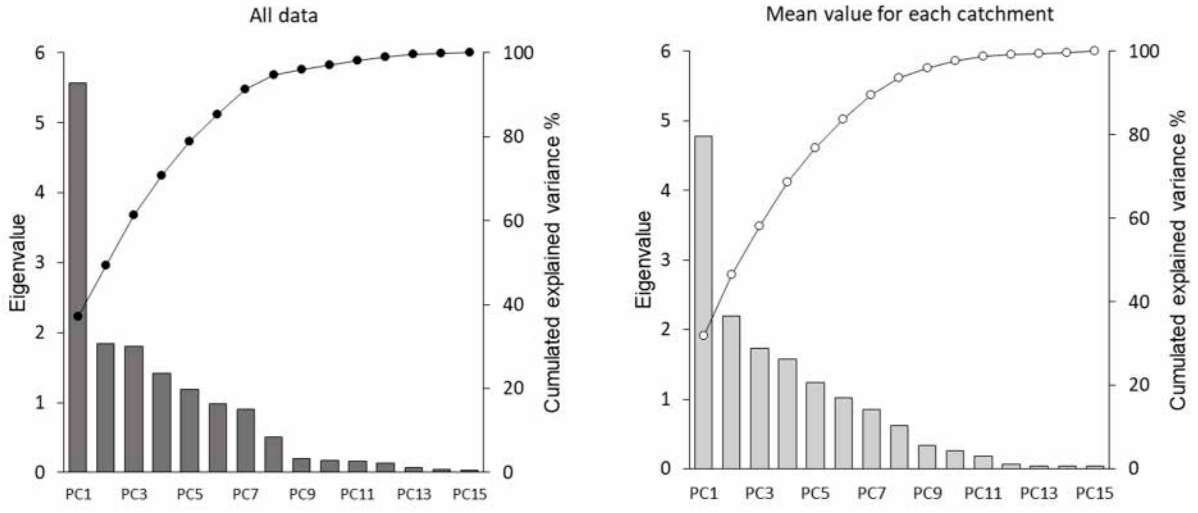


Figure 4. Distribution maps of several parameters across Corsica: (**a**) log(SO_4), (**b**) log(Mg), (**c**) log(Na), (**d**) log(Cl), (**e**) log(Ca), (**f**) log(NO_3), (**g**) log(*E.coli*), (**h**) log(Ent.).

The distribution of inertia of the factorial axes resulting from the PCA performed on the entire dataset is shown in Figure 5a. It revealed that the PCA allowed for significant dimensionality reduction, with the first seven axes representing over 90% of the information conveyed by the 15 initial parameters. The first 5 principal components (PCs) had eigenvalues greater than one, meaning they each contained more information than the initial parameters, with PC1 alone accounting for the information carried by 5.5 of these initial parameters. This dataset, therefore, showed a high degree of redundancy, which is significantly reduced by focusing only on the first PCs. Bartlett's sphericity test gave a value of $\chi^2 = 40063$, far exceeding the critical value of 82 (significance level of 0.05 and p-value < 0.0001), confirming the effectiveness of the dimensionality reduction. When considering the average value of each parameter at each sampling point, thereby minimizing temporal variations, the PCA still achieved significant dimensionality reduction (Figure 5b, $\chi^2 = 9035$). The eigenvalue of the first axis decreased from 5.5 to 4.8, while the second axis slightly increased from 1.85 to 2.2. The eigenvalues of the subsequent PCs remained largely unchanged, and as in the calculation with the entire dataset, the first seven factorial axes still carried 90% of the information in the dataset.



(**a**) (**b**)

Figure 5. Inertia of the factorial axes for the PCAs conducted on (**a**) the entire dataset (spatio-temporal variance, adapted from Lazar et al. [25]), and (**b**) using the average of each parameter for each sampling point (spatial variance).

The position of the parameters in the main factorial planes PC1-PC2 and PC3-PC4 is shown in Figure 6. The results of the spatio-temporal variance analysis (Figure 6a,c) were presented in a previous study [25], and are reiterated here. The first factorial axis explained 37.1% of the total variance and contrasted mineralized waters, with positive coordinates for electrical conductivity and major ions (Ca, Mg, Na, Cl, SO_4, HCO_3), with poorly mineralized waters characterized by fecal contamination. The second PC contrasted waters were marked by revivable bacteria with waters displaying a predominantly chloride-sodium chemical profile, influenced by the presence of metals and fecal contamination. The third PC (12.1% of the variance) was also influenced by the presence of revivable bacteria, combined with fecal contamination and the presence of metals, again within a predominantly chloride-sodium chemical context, while the fourth factorial axis showed positive coordinates for fecal contamination but negative ones for metals. The PCA conducted by minimizing temporal variance (Figure 6b,d) presented similar factorial axes to the description above, with some notable distinctions. The first PC still represented water mineralization, but the weight of fecal contamination, negatively correlated with this axis, was significantly reduced. The second factorial axis was marked by revivable bacteria in a chloride-sodium geochemical context, but also positively correlated with metals and fecal contamination,

corresponding to the third factorial axis in the analysis with the full dataset. The third axis was similarly driven by revivable bacteria, but also negatively scored by the chloride-sodium chemical profile, fecal contamination, and nitrates. The fourth PC represented fecal contamination in a calcium carbonate context.

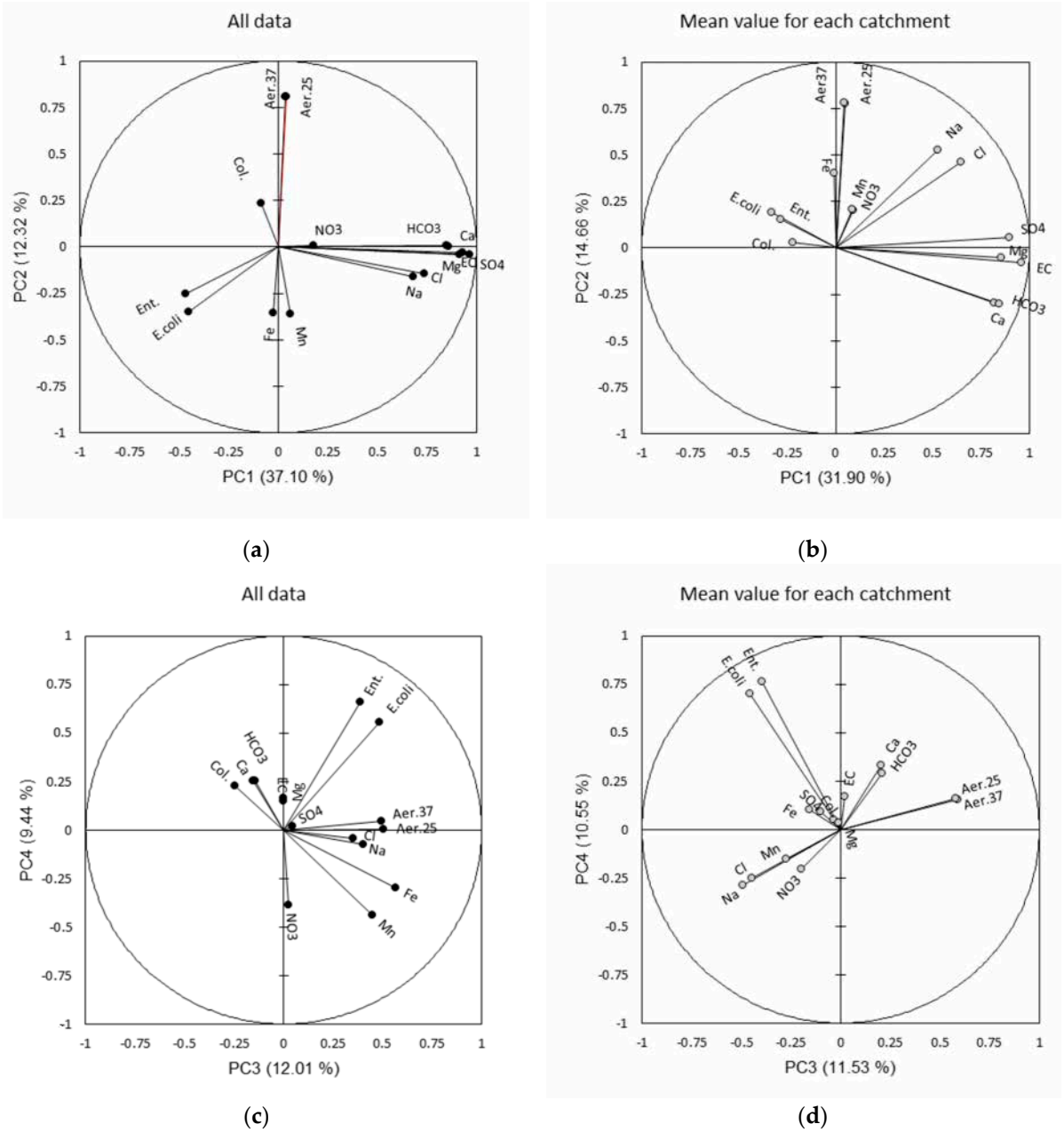


Figure 6. Factorial planes PC1–PC2 and PC3–PC4 obtained from the entire dataset ((**a**,**c**), adapted from Lazar et al. [25]) and from (**b**,**d**) the averages of each parameter for each sampling point.

The maps of the first four factorial axes (all data) are presented in Figure 7. These four axes represent more than 70% of the information contained in the dataset. The classification of parameters based on their similarity in position across the first eight factorial axes of the PCA, i.e., their correlational similarities, is shown in Figure 8. This is presented both for the full dataset (spatio-temporal variance, Figure 8a) and based on the average of each parameter for each sampling point (spatial variance, Figure 8b). There were minimal differences between the two classification methods with seven groups. Overall, there were strong similarities within each group, and strong dissimilarities between groups (dissimilarity > 13). Bacteriological parameters were correlated by origin, with Enterococcus

and Escherichia coli on one hand, indigenous revivable bacteria on the other, and finally, coliforms correlated with nitrates in the analysis using all data (Figure 8a), whereas they were separated when temporal variance was minimized (Figure 8b).

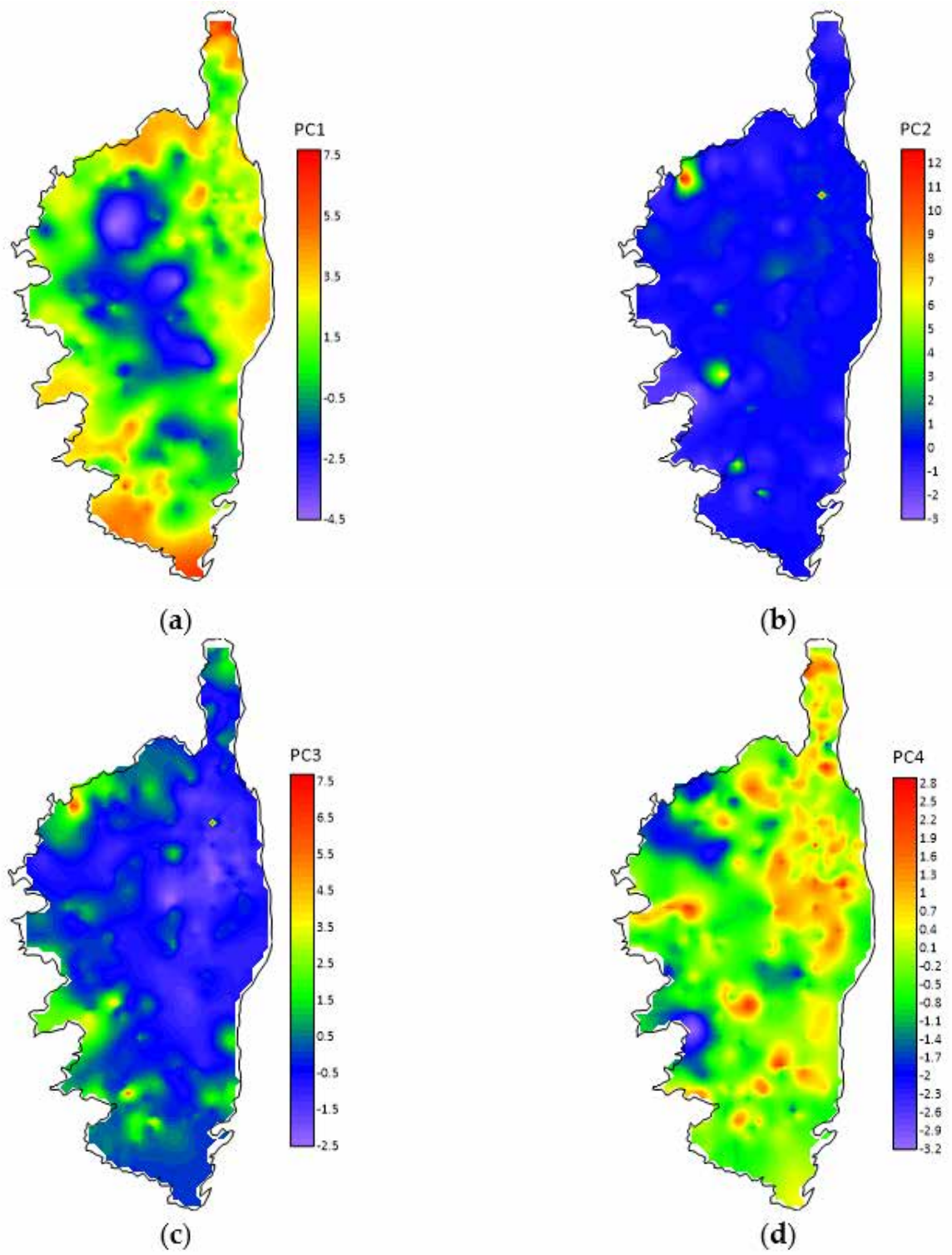
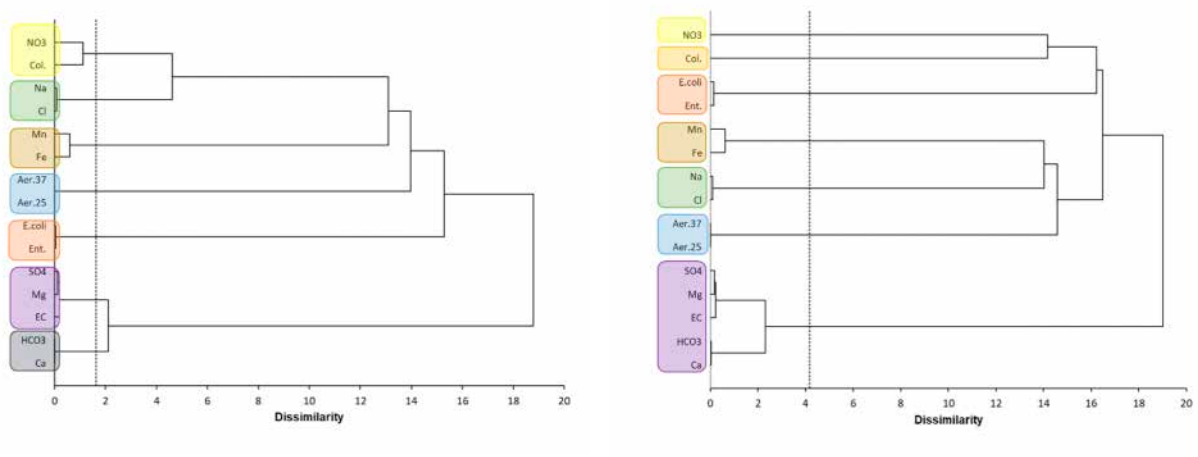


Figure 7. Distribution of the first four factorial axes across the island of Corsica. ((**a**–**d**) = PC1 to PC4).



(**a**) (**b**)

Figure 8. Clustering of parameters based on (**a**) the full dataset and (**b**) the average values of each parameter for each catchment point. The dashed line is the phenom line chosen to select number of clusters.

4. Discussion

4.1. Redundancy Within Multifactorial Information

The similarity between the distribution maps of parameters within the island's groundwater reflects not only the similar behaviors of certain parameters distinguishing distinct environments (fractured medium/porous medium), but also a redundancy of information, which is condensed after performing a principal component analysis (PCA), significantly reducing the dimensionality of the data hyperspace [32]. By arbitrarily setting a 10% loss of information, the dimensionality decreases from the initial 15 parameters to 7 PCs, while for a 20% loss, it drops to 5 PCs. This redundancy of information, quantified by the PCA, is evident in the similar distributions of parameters across Corsica (Figure 4), such as Na and Cl, *E.coli* and Ent., etc. This concretely means that the information from one variable can, in part, be inferred from the other variable. This information redundancy is also reflected in the clustering of parameters (Figure 8), based on the first eight principal components, which capture 95% of the information. This clustering leads to seven distinct clusters with low similarity. The mapping of the principal components synthesizes the information: for instance, PC1 largely groups the information carried by EC, SO_4, Mg, but also a part by Na and Cl. The remaining information carried by Na and Cl, associated with fecal contamination in low-altitude coastal waters, is captured by PC3.

However, several findings demonstrate that the information within the dataset is multifactorial, and despite significant redundancy among parameters, this information cannot be summarized by just one or two parameters. This is mainly evidenced by the high eigenvalues of the first principal components and the necessity to consider the first seven PCs to explain 90% of the total variance. The analysis of the position of parameters in the main factorial planes (Figure 6) further shows that the relationship between parameter groups and principal components is not straightforward. For example, some parameters like *E.coli* and Ent. load onto multiple principal components, indicating that their spatial and temporal distributions depend on several independent determinants. The maps of the principal components represent the spatial distribution of each of these determinants. Thus, the principal component maps do not exactly synthesize several parameters but rather illustrate the key mechanisms responsible for their distribution, which is notably different.

4.2. Local or Regional and Spatial or Temporal Determinants of Water Quality

The cartographic representation based on the logarithmic transformation of electrical conductivity (Figure 2h), along with the lower semivariance in the N150° direction (Figure 2f), reveals a structure that coincides with the distinction between Hercynian and

Alpine Corsica, the island's two major structural units [25]. The least mineralized waters are found in the crystalline bedrock at the center of the island, while the most mineralized waters are observed in the metamorphic terrains of Alpine Corsica, the sedimentary formations in the far south, and along some coastal rivers. This geographical and lithological coincidence between major ions/conductivity and the main structural units is further supported by the low nugget effect and the variogram ranges for these parameters, which are about 30 to 40 km, matching the size of the structural units in the region. Thus, there is a regional (geographical and lithological) determinism in the distribution of major ions. This finding is consistent with results obtained from Sise-Eaux database extractions for other French regions, such as Provence-Alpes-Côte d'Azur, Occitanie, Bourgogne-Franche-Comté, and Auvergne-Rhône-Alpes [37–41]. The higher electrical conductivity of waters associated with coastal rivers, primarily on the island's western side, aligns with the general observation of longer flow paths and prolonged water–rock interaction. Additionally, a sodium-chloride chemical profile reflects the influence of strong northwest (Mistral) to southwest (Libeccio) winds, which can carry salty sea spray inland for several kilometers [42,43]. The difference between the two variogram calculations for Cl (Figure 3d) and Na suggests some temporal variance, possibly indicating a seasonal phenomenon.

In contrast, bacteriological parameters exhibit a high nugget effect relative to total semivariance and a short range, no more than a few kilometers, indicating a local determinism for fecal contamination. Again, this result is consistent with observations from other regions. The extraction of the Sise-Eaux database for Corsica highlights the complexity of fecal contamination determinants, which contribute to several principal components, particularly PC1 and PC4 for the analysis with all data, and primarily PC4 for the analysis based on parameter averages at each catchment. Since the principal components are orthogonal to each other, they reflect independent processes, including two distinct determinants of contamination: one with significant temporal variations (PC1), which diminishes when temporal variance is minimized, and another with low temporal variability (PC4), whose trace on the PCA remains unaffected by different processing methods. One could suggest contamination during late-summer storms, which are violent in this Mediterranean climate, combining runoff, water turbidity, and bacterial transport [44–48] to poorly protected catchments, as well as strictly spatial contamination in environments with a predominantly carbonate-calcium chemical facies, reflecting the weathering of metamorphic rocks. Soils in these environments are slightly richer than soils on crystalline rocks and are used for livestock farming, which maintains pollution pressure year-round. Methodologically, this result parallels previous studies in other French regions [37–41] and in Corsica [25], where the spatial and temporal components of variance were not distinguished. Such a distinction allows for more detailed analysis and interpretations of fecal contamination determinants. Our results also confirm observations and conclusions made in the Bourgogne-Franche-Comté region, showing that the factors favoring fecal contamination are multiple and interdependent. They cannot be reduced solely to runoff caused by heavy rainfall but also depend on environmental factors that influence the presence or absence of extensive livestock farming (cattle and pigs). The two approaches used in our work—statistical (PCA and AHC) and geostatistical (variogram analysis)—are independent but converge in terms of interpretation. To conclude on fecal contamination, which remains the main cause of non-compliance with drinking water standards, the vulnerability of catchments has both a temporal (predominant) and a strictly spatial dimension, reflecting several aspects such as the following:

- A diversity of soil types and their varying degrees of flocculating power for particles, the main carriers of bacteria;
- A diversity of soil textures, impacting the intrusion of contaminated surface water;
- A diversity of contamination pressure, with livestock farming not evenly distributed across the territory, being more prominent in mountainous areas and less so in the plains, with differences in livestock type (cattle vs. pigs) depending on lithology.

4.3. Strategies for Water Quality Inventory and Mapping

Several strategies can be considered for the inventory and mapping of water resources. The methodology traditionally adopted by regional health agencies involves creating distribution maps for individual parameters, with the need to select the most limiting parameter to estimate the quality of water intended for human consumption. In this study, we have shown that mapping parameters one by one results in significant redundancy and complicates water quality monitoring.

A second option is to reduce dimensionality through principal component analysis (PCA), then map the factorial axes across the studied territory. This option provides a precise, synthetic view of each independent source of variability in water quality. Mapping the principal components effectively eliminates redundancies, but its interpretation is less intuitive and might be more suitable for those familiar with mathematical analysis.

A third option involves, after dimensionality reduction, grouping the parameters into categories that show strong intra-group similarity (meaning that the parameters within the same group vary similarly) and strong inter-group dissimilarity. In the case of Corsica, we observed that the seven groups exhibit these characteristics (Figure 8), but this may not always be the case, necessitating clustering for each region studied. Choosing to monitor one representative parameter from each group is more intuitive than the previous option and offers a more familiar approach for water quality monitoring agents, who may not necessarily be experts in mathematical processing. However, this option does not reveal independent sources of information, concealing redundancies.

A fourth option is to cross-reference the information from the Sise-Eaux database with the groundwater body framework, which introduces an independent physical constraint in the analysis of water diversity. This option involves grouping groundwater bodies that exhibit similar behavior in terms of the processes driving diversity. The information loss associated with this grouping must be quantified on a case-by-case basis. This approach to mapping, monitoring, and surveillance of water resources was developed in a previous study on the island of Corsica [25].

It is important to keep in mind that the total variance in an extraction from a database like Sise-Eaux (or its equivalent in other countries) includes both spatial and temporal variance. The fact that samples are collected over an extended period raises the question of the impact of temporal variability on the reliability of spatial distribution characterization. Here, a preliminary step to partially distinguish between the two variance components has been undertaken, allowing for a more detailed analysis and a better identification and distinction of the processes responsible for water quality.

4.4. Consequences for Sustainable Management of Groundwater in Corsica

Comprehensive monitoring is costly, especially given the large number of parameters involved. The proposal to select a representative from each group of parameters allows for a significant reduction in these costs without substantial loss of information. The savings made should enable an expansion of the water quality monitoring scheme, making this monitoring more effective. Most issues of water non-compliance are related to fecal contamination, with a limit set at zero cells per 100 mL [49]. Our approach provides a clear view of the most vulnerable sectors or sources, which can lead to increased sampling frequency for better monitoring in these areas, while allowing for less frequent sampling in less vulnerable regions. Additionally, the principal component analysis offers a more precise understanding of contamination mechanisms, particularly those related to rainfall events. These contaminations indicate a vulnerability of the catchments or the overall water resource, which should guide targeted monitoring in these specific areas.

5. Conclusions

This study of groundwater in Corsica allows for advancements in the handling of large databases such as Sise-Eaux, or equivalent databases around the world. This study confirms the need for data conditioning (in this case, log transformation) aimed at reducing

the impact of extreme values without eliminating or artificially limiting them in the dataset. The risk of such extreme values is high, especially for metals and bacteriological parameters, which can obscure certain processes responsible for water diversity and make the analysis more challenging. The extraction of the database, covering a 27-year period and 662 sampling points distributed within 40 groundwater bodies across the island, contains both temporal and spatial variability, which have been (partially) distinguished. This distinction reveals two components of fecal contamination that had not been clearly identified in previous studies: on one hand, a strong temporal component likely due to the impact of violent late-summer storms, and on the other hand, a weaker, strictly spatial component, which we have linked to the presence of a permanent pollution pressure, driven by several factors (topography, lithology, land use, etc.). This work represents a new step in the analysis and understanding of groundwater diversity, although further efforts are needed in multivariate analysis after conditioning the dataset to specifically preserve only temporal or spatial variance. This will allow for a more rigorous refinement and better distinction of the processes responsible for water quality diversity.

Author Contributions: Conceptualization, V.V., I.K., F.H., C.M. and E.G.; methodology, V.V., H.L. and M.A.; software, A.B.; validation, H.L., M.A., F.H., E.G. and A.B.; formal analysis, H.L.; investigation, H.L. and L.B.; resources, L.B. and V.V.; data curation, H.L.; writing—original draft preparation, H.L. and V.V.; writing—review and editing, L.B.; visualization, H.L. and L.B.; supervision, V.V. and I.K.; project administration, I.K.; funding acquisition, L.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The dataset in this study is the property of Corsica Regional Health Agency and is included in the SISE-Eaux French database. The datasets presented in this article are not easily accessible for reasons of sensitivity to possible malicious acts. Requests should be addressed to the Corsica Regional Health Agency (ARS).

Acknowledgments: The authors would like to thank Patrice Grandjean (Corsica Regional Health Agency) for his help with data extraction.

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

References

1. WHO. *Guidelines for Drinking-Water Quality: Incorporating 1st and 2nd Addenda, Volume 1, Recommendations*, 3rd ed.; World Health Organization: Geneva, Switzerland, 2008; ISBN 978-92-4-154761-1.
2. Saeedi, M.; Abessi, O.; Sharifi, F.; Meraji, H. Development of groundwater quality index. *Environ. Monit. Assess.* **2010**, *163*, 327–335. [CrossRef] [PubMed]
3. Lapworth, D.J.; Baran, N.; Stuart, M.E.; Ward, R.S. Emerging organic contaminants in groundwater: A review of sources, fate and occurrence. *Environ. Pollut.* **2012**, *163*, 287–303. [CrossRef] [PubMed]
4. Ng, J.C. Environmental Contamination of Arsenic and its Toxicological Impact on Humans. *Environ. Chem.* **2005**, *2*, 146–160. [CrossRef]
5. Morales-Simfors, N.; Bundschuh, J. Arsenic-rich geothermal fluids as environmentally hazardous materials—A global assessment. *Sci. Total Environ.* **2022**, *817*, 152669. [CrossRef]
6. Hinsby, K.; Condesso de Melo, M.T.; Dahl, M. European case studies supporting the derivation of natural background levels and groundwater threshold values for the protection of dependent ecosystems and human health. *Sci. Total Environ.* **2008**, *401*, 1–20. [CrossRef]
7. Bradford, S.A.; Harvey, R.W. Future research needs involving pathogens in groundwater. *Hydrogeol. J.* **2017**, *25*, 931–938. [CrossRef]
8. Bunting, S.Y.; Lapworth, D.J.; Crane, E.J.; Grima-Olmedo, J.; Koroša, A.; Kuczyńska, A.; Mali, N.; Rosenqvist, L.; van Vliet, M.E.; Togola, A.; et al. Emerging organic compounds in European groundwater. *Environ. Pollut.* **2021**, *269*, 115945. [CrossRef]
9. Kemper, K.E. Groundwater—From development to management. *Hydrogeol. J.* **2004**, *12*, 3–5. [CrossRef]
10. Koundouri, P. Current Issues in the Economics of Groundwater Resource Management. *J. Econ. Surv.* **2004**, *18*, 703–740. [CrossRef]
11. Bhunia, G.S.; Shit, P.K.; Brahma, S. Chapter 19—Groundwater conservation and management: Recent trends and future prospects. In *Case Studies in Geospatial Applications to Groundwater Resources*; Shit, P., Bhunia, G., Eds.; Elsevier: Amsterdam, The Netherlands, 2023; pp. 371–385. ISBN 978-0-323-99963-2.

12. Qiu, W.; Ma, T.; Wang, Y.; Cheng, J.; Su, C.; Li, J. Review on status of groundwater database and application prospect in deep-time digital earth plan. *Geosci. Front.* **2022**, *13*, 101383. [CrossRef]
13. Fitch, P.; Brodaric, B.; Stenson, M.; Booth, N. Integrated Groundwater Data Management BT. In *Integrated Groundwater Management: Concepts, Approaches and Challenges*; Jakeman, A.J., Barreteau, O., Hunt, R.J., Rinaudo, J.-D., Ross, A., Eds.; Springer International Publishing: Cham, Germany, 2016; pp. 667–692. ISBN 978-3-319-23576-9.
14. Condon, L.E.; Kollet, S.; Bierkens, M.F.P.; Fogg, G.E.; Maxwell, R.M.; Hill, M.C.; Fransen, H.-J.H.; Verhoef, A.; Van Loon, A.F.; Sulis, M.; et al. Global Groundwater Modeling and Monitoring: Opportunities and Challenges. *Water Resour. Res.* **2021**, *57*, e2020WR029500. [CrossRef]
15. Lall, U.; Josset, L.; Russo, T. A Snapshot of the World's Groundwater Challenges. *Annu. Rev. Environ. Resour.* **2020**, *45*, 171–194. [CrossRef]
16. Chery, L.; Laurent, A.; Vincent, B.; Tracol, R. *Echanges SISE-Eaux/ADES: Identification des Protocoles Compatibles Avec les Scénarios D'échange SANDRE*; Vincennes/Orléans, France, 2011. Available online: https://infoterre.brgm.fr/rapports/RP-59211-FR.pdf (accessed on 6 November 2024).
17. Loftis, J.C. Trends in groundwater quality. *Hydrol. Process.* **1996**, *10*, 335–355. [CrossRef]
18. Gleeson, T.; Cuthbert, M.; Ferguson, G.; Perrone, D. Global Groundwater Sustainability, Resources, and Systems in the Anthropocene. *Annu. Rev. Earth Planet. Sci.* **2020**, *48*, 431–463. [CrossRef]
19. Elshall, A.S.; Arik, A.D.; El-Kadi, A.I.; Pierce, S.; Ye, M.; Burnett, K.M.; Wada, C.A.; Bremer, L.L.; Chun, G. Groundwater sustainability: A review of the interactions between science and policy. *Environ. Res. Lett.* **2020**, *15*, 93004. [CrossRef]
20. Mays, L.W. Groundwater Resources Sustainability: Past, Present, and Future. *Water Resour. Manag.* **2013**, *27*, 4409–4424. [CrossRef]
21. De Graciansky, P.-C.; Roberts, D.G.; Tricart, P. Chapter Nine—The Tethyan Margin in Corsica. In *The Western Alps, from Rift to Passive Margin to Orogenic Belt*; Graciansky, P.-C.D., Roberts, D.G., Eds.; Elsevier: Amsterdam, The Netherlands, 2011; Volume 14, pp. 183–188. ISBN 09282025.
22. Gilli, E.; Nicod, J. Karsts et grottes de Corse. *Karstologia* **2010**, *55*, 260–261.
23. Nguyen-Thé, D.; Palvadeau, E.; Sinzelle, B. Atlas Cartographique des Aquifères Littoraux de Corse. Available online: https://infoterre.brgm.fr/rapports/RP-58166-FR.pdf (accessed on 6 November 2024).
24. ODDC. Observatoire du Développement Durable de Corse. Available online: http://www.oddc.fr/modules.php?name=SimpleProfil&op=showonedoc&id=23&mmg=9,513 (accessed on 6 November 2024).
25. Lazar, H.; Ayach, M.; Barry, A.; Mohsine, I.; Touiouine, A.; Huneau, F.; Mori, C.; Garel, E.; Kacimi, I.; Valles, V.; et al. Groundwater bodies in Corsica: A critical approach to GWBs subdivision based on multivariate water quality criteria. *Hydrology* **2023**, *10*, 213. [CrossRef]
26. Steinskog, D.J.; Tjøstheim, D.B.; Kvamstø, N.G. A Cautionary Note on the Use of the Kolmogorov–Smirnov Test for Normality. *Mon. Weather Rev.* **2007**, *135*, 1151–1157. [CrossRef]
27. Marden, J.I. Positions and QQ Plots. *Stat. Sci.* **2004**, *19*, 606–614. [CrossRef]
28. Cousineau, D.; Chartier, S. Outliers detection and treatment: A review. *Int. J. Psychol. Res.* **2010**, *3*, 58–67. [CrossRef]
29. Mohsine, I.; Kacimi, I.; Abraham, S.; Valles, V.; Barbiero, L.; Dassonville, F.; Bahaj, T.; Kassou, N.; Touiouine, A.; Jabrane, M.; et al. Exploring Multiscale Variability in Groundwater Quality: A Comparative Analysis of Spatial and Temporal Patterns via Clustering. *Water* **2023**, *15*, 1603. [CrossRef]
30. Jabrane, M.; Touiouine, A.; Bouabdli, A.; Chakiri, S.; Mohsine, I.; Valles, V.; Barbiero, L. Data Conditioning Modes for the Study of Groundwater Resource Quality Using a Large Physico-Chemical and Bacteriological Database, Occitanie Region, France. *Water* **2023**, *15*, 84. [CrossRef]
31. Pearson, K.L., III. On lines and planes of closest fit to systems of points in space. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1901**, *2*, 559–572. [CrossRef]
32. Helena, B.; Pardo, R.; Vega, M.; Barrado, E.; Fernandez, J.M.; Fernandez, L. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water Res.* **2000**, *34*, 807–816. [CrossRef]
33. Rezende-Filho, A.T.; Valles, V.; Furian, S.; Oliveira, C.M.S.C.; Ouardi, J.; Barbiero, L. Impacts of lithological and anthropogenic factors affecting water chemistry in the upper Paraguay River Basin. *J. Environ. Qual.* **2015**, *44*, 1832–1842. [CrossRef]
34. Gleser, L.J. A Note on the Sphericity Test. *Ann. Math. Stat.* **1966**, *37*, 464–467. [CrossRef]
35. Bouguettaya, A.; Yu, Q.; Liu, X.; Zhou, X.; Song, A. Efficient agglomerative hierarchical clustering. *Expert Syst. Appl.* **2015**, *42*, 2785–2797. [CrossRef]
36. Day, W.H.E.; Edelsbrunner, H. Efficient algorithms for agglomerative hierarchical clustering methods. *J. Classif.* **1984**, *1*, 7–24. [CrossRef]
37. Bousouis, A.; Bouabdli, A.; Ayach, M.; Ravung, L.; Valles, V.; Barbiero, L. The Multi-Parameter Mapping of Groundwater Quality in the Bourgogne-Franche-Comté Region (France) for Spatially Based Monitoring Management. *Sustainability* **2024**, *16*, 8503. [CrossRef]
38. Tiouiouine, A.; Yameogo, S.; Valles, V.; Barbiero, L.; Dassonville, F.; Moulin, M.; Bouramtane, T.; Bahaj, T.; Morarech, M.; Kacimi, I. Dimension reduction and analysis of a 10-year physicochemical and biological water database applied to water resources intended for human consumption in the provence-alpes-cote d'azur region, France. *Water* **2020**, *12*, 525. [CrossRef]

39. Mohsine, I.; Kacimi, I.; Valles, V.; Leblanc, M.; El Mahrad, B.; Dassonville, F.; Kassou, N.; Bouramtane, T.; Abraham, S.; Touiouine, A.; et al. Differentiation of multi-parametric groups of groundwater bodies through Discriminant Analysis and Machine Learning. *Hydrology* **2023**, *10*, 230. [CrossRef]
40. Jabrane, M.; Touiouine, A.; Valles, V.; Bouabdli, A.; Chakiri, S.; Mohsine, I.; El Jarjini, Y.; Morarech, M.; Duran, Y.; Barbiero, L. Search for a Relevant Scale to Optimize the Quality Monitoring of Groundwater Bodies in the Occitanie Region (France). *Hydrology* **2023**, *10*, 89. [CrossRef]
41. Ayach, M.; Lazar, H.; Bousouis, A.; Touiouine, A.; Kacimi, I.; Valles, V.; Barbiero, L. Multi-Parameter Analysis of Groundwater Resources Quality in the Auvergne-Rhô ne-Alpes Region (France) Using a Large Database. *Resources* **2023**, *12*, 869.
42. Lambert, D.; Mallet, M.; Ducrocq, V.; Dulac, F.; Gheusi, F.; Kalthoff, N. CORSiCA: A Mediterranean atmospheric and oceanographic observatory in Corsica within the framework of HyMeX and ChArMEx. *Adv. Geosci.* **2010**, *26*, 125–131. [CrossRef]
43. Adler, B.; Kalthoff, N.; Kohler, M.; Handwerker, J.; Wieser, A.; Corsmeier, U.; Kottmeier, C.; Lambert, D.; Bock, O. The variability of water vapour and pre-convective conditions over the mountainous island of Corsica. *Q. J. R. Meteorol. Soc.* **2016**, *142*, 335–346. [CrossRef]
44. Pandey, P.K.; Kass, P.H.; Soupir, M.L.; Biswas, S.; Singh, V.P. Contamination of water resources by pathogenic bacteria. *AMB Express* **2014**, *4*, 51. [CrossRef]
45. Li, P.; Karunanidhi, D.; Subramani, T.; Srinivasamoorthy, K. Sources and Consequences of Groundwater Contamination. *Arch. Environ. Contam. Toxicol.* **2021**, *80*, 1–10. [CrossRef]
46. Gallay, A.; De Valk, H.; Cournot, M.; Ladeuil, B.; Hemery, C.; Castor, C.; Bon, F.; Mégraud, F.; Le Cann, P.; Desenclos, J.C. A large multi-pathogen waterborne community outbreak linked to faecal contamination of a groundwater system, France, 2000. *Clin. Microbiol. Infect.* **2006**, *12*, 561–570. [CrossRef]
47. Fitts, C.R. 11—Groundwater Contamination. In *Groundwater Science (Second Edition)*; Fitts, C.R., Ed.; Academic Press: Boston, MA, USA, 2013; pp. 499–585. ISBN 978-0-12-384705-8.
48. Pachepsky, Y.A.; Shelton, D.R. Escherichia Coli and Fecal Coliforms in Freshwater and Estuarine Sediments. *Crit. Rev. Environ. Sci. Technol.* **2011**, *41*, 1067–1110. [CrossRef]
49. European Parliament; Council of the European Union. Directive (EU) 2020/2184 of the European Parliament and of the Council of 16 December 2020 on the Quality of Water Intended for Human Consumption (Recast) (Text with EEA Relevance). Available online: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32020L2184 (accessed on 6 November 2024).