

Article

Deciphering Spatial Patterns in Groundwater Quality Across Nouvelle-Aquitaine, France: A Multivariate Analysis of Two Decades of Monitoring Data

Mouna El Jirari ¹, Abdoul Azize Barry ², Abderrahim Bousousi ¹, Zouhair Zeiki ¹, Meryem Ayach ³, Mohamed Sadiki ¹, Abdelhak Bouabdli ¹, Meryem Touzani ⁴, Muriel Guiraud ⁵, Vincent Valles ^{6,7} and Laurent Barbiero ^{8,*}

- ¹ Laboratoire de Géosciences, Faculté des Sciences, Université Ibn Tofaïl, Kenitra 14000, Morocco; mouna.eljirari@uit.ac.ma (M.E.J.); abderrahim.bousousi@uit.ac.ma (A.B.); zouhair.zeiki@uit.ac.ma (Z.Z.); mohamed.sadiki@uit.ac.ma (M.S.); bouabdli@uit.ac.ma (A.B.)
- ² Geoscience and Environment Laboratory (LaGE), Department of Earth Sciences, Joseph Ki-Zerbo University, Ouagadougou 7021, Burkina Faso; azize.barry@ujkz.bf
- ³ Geosciences, Water and Environment Laboratory, Faculty of Sciences Rabat, Mohammed V University, Rabat 10000, Morocco; meryem_ayach@um5.ac.ma
- ⁴ National Institute of Agronomic Research, Rabat 10060, Morocco; meryem.touzani@inra.ma
- ⁵ Agence Regionale de Santé Occitanie, 10 Chemin du Raisin, CEDEX 9, 31050 Toulouse, France; muriel.guiraud@ars.sante.fr
- ⁶ Mixed Research Unit EMMAH (Environnement Méditerranéen et Modélisation des Agro-Hydrosystèmes), Hydrogeology Laboratory, Avignon University, 84916 Avignon, France; vincent.valles@outlook.fr
- ⁷ Soil Science Department (FSTBM), Faculté des Sciences et Techniques, P.O. Box 523, Beni Mellal 23000, Morocco
- ⁸ 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, a vital resource for drinking water supply, must be managed sustainably to ensure its availability and quality. In France, the SISE-Eaux database on water intended for human consumption, archived by the Regional Health Agencies (ARS) since 1990, constitutes a rich source of information. This study focused on the groundwater of the Nouvelle-Aquitaine region, the largest administrative region in metropolitan France, covering 84,061 km² with 6 million inhabitants. It is based on a 22-year data extraction, resulting in a matrix of 121,649 observations and 51 physico-chemical and bacteriological parameters. Following logarithmic transformation of the data and fitting of variograms using the mean value of each parameter for each sampling point, the spatial distribution of numerous parameters across the region is presented. From this initial sparse matrix, a dense matrix of 23,319 samples (rows) and 15 key parameters (columns) was selected for a multivariate approach. A Principal Component Analysis (PCA) was used to condense the information and create summary maps capturing over 68% of the information contained in the dense matrix. The combined results of the multivariate analysis (dense matrix) and the distribution of individual parameters (sparse matrix) highlight the diversity of sources contributing to the spatial variability of groundwater, such as the role of lithology, the origin and pathways of fecal contamination, and the influence of redox processes. Neither the large size of the study area nor the high number of parameters proved to be an obstacle to the analysis. The understanding of ongoing processes and the factorial axis distribution maps, which enable the spatial representation of these mechanisms, can be used to facilitate groundwater monitoring and protection.



Academic Editor: Hailong Yin

Received: 18 January 2026

Revised: 9 February 2026

Accepted: 13 February 2026

Published: 14 February 2026

Copyright: © 2026 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: groundwater database; bacteriological composition; chemical composition; principal component analysis; clustering

1. Introduction

Groundwater represents a major global freshwater reserve, significantly contributing to drinking water supplies, agriculture, industry, and the maintenance of ecosystems [1–5]. However, these natural reservoirs, which are primarily recharged by the infiltration of precipitation—a spatially variable process—are rendered particularly vulnerable by anthropogenic pressures and climate change. These pressures alter the spatial and temporal distribution of recharge, making the understanding of their dynamics and their interactions with surface water essential for sustainable management and the preservation of their quality [6,7].

Furthermore, groundwater quality is a complex concept encompassing numerous parameters. These include natural factors, such as geology and major ions, but also anthropogenic factors such as heavy metals, microbiological, organic, and inorganic parameters, as well as xenobiotic compounds originating from human activities, notably pesticides and their metabolites. This complexity, which stems from the multitude of diverse mechanisms influencing these parameters, makes its study particularly challenging [8].

Concurrently, improvements in analytical techniques and reductions in their costs have enabled the monitoring of this resource to generate a much larger quantity of samples and parameters than in the past. While this evolution opens up new perspectives, it also poses significant methodological challenges. Multivariate statistical methods are now frequently used as tools to analyze environmental databases in order to determine the origin, types of contamination, and spatial distribution [9–13]. Analyses based on multivariate statistics (Principal Components Analysis, Hierarchical Clustering, Analysis of Variance) are often combined with complementary geostatistical processing [14–16], which helps both in interpreting the origin of groundwater diversity and in better understanding its spatial distribution. It therefore becomes necessary to determine how to effectively utilize these data on a regional scale, assess the relevance of producing a map for each of the hundreds of available parameters, and identify approaches for integrating all this information in a coherent manner.

In France, this challenge can be addressed using tools such as the SISE-Eaux database (Health-Environment Water Information System) [17,18]. Managed by the General Directorate of Health (DGS) and the Regional Health Agencies (Agences Régionales de Santé—ARS), this database has compiled, since the 1990s, all data acquired through the sanitary control of drinking water supply points. It aggregates data on both groundwater and surface water, as well as the results of analyses performed on raw water or within the distribution network, making it an essential tool for monitoring and management. Moreover, due to the multitude of parameters it contains—including chemical elements, bacteriological data, and hundreds of pesticides and metabolites—it offers the possibility of exploring water quality at a regional scale and addressing the challenges posed by managing large datasets [19].

The present work builds upon previous results obtained by our research group, whose objectives are to improve the management and monitoring of groundwater quality using the parameters available in the SISE-Eaux database. Previous work has progressively refined the methodology for analyzing the information contained in the database and has identified new constraints for its utilization. The advancements can be summarized as follows. Initially, maps of individual parameter values were produced to visualize ground-

water quality spatially, providing detailed information. However, this approach becomes less effective when the number of parameters is high, due to redundancy and consequent information overload. The development of synthetic macro-parameters through factorial axes from Principal Component Analysis, which significantly reduces the dimensionality of the data space with minimal loss of information, allows for the identification of the most significant parameters and highlights the relationships between variables [20]. The logarithmic transformation of the data helped to limit the impact of extreme values (low or high) that could, during analysis, mask certain processes responsible for water diversity [21,22]. The subsequent approach involved developing a parameter typology based on common behaviors and selecting a representative parameter for each parameter class. The subsequent approach involved developing a parameter typology and selecting a representative parameter for each parameter class [22–25], with the aim of reducing the costs of water quality monitoring and surveillance. Finally, the separate analysis of spatial and temporal variances within the datasets enabled the identification of seasonal characteristics or long-term trends of ongoing processes, as well as their spatial extent [26].

This study aims to better understand the ongoing processes responsible for the diversity in groundwater quality in the Nouvelle-Aquitaine region. It is based on a regional extraction from the SISE-Eaux database. Compared to previous advancements, whose objective was the development and refinement of a relevant method for analyzing the information contained in such a database, three aspects require strengthening. The first concerns a potential limitation regarding the size of the region considered. In these prior studies, the size of the investigated regions ranged from 8722 km² to 72,724 km², representing a wide range of surface areas. However, the study of the vast Occitanie region, spanning two contrasting slopes in terms of climate and land use, demonstrated the necessity of subdividing the study area to, once again, avoid masking certain processes [27], which, during the analysis, may obscure certain processes responsible for the physico-chemical and bacteriological variability of groundwater. This finding, therefore, raises the question of the relevance of these approaches for large regions, which will be tested on the Nouvelle-Aquitaine region, the largest administrative region in France. The second point concerns the number of parameters. In these multi-parameter studies based on extractions from databases such as SISE-Eaux, the number of parameters has been reduced, typically to between 12 and 20 (major ions, a few metals or nitrogen species, as well as fecal contamination parameters). A larger number of parameters (51) will be considered, thereby increasing the diversity of parameter types by applying the analysis to major ions, a broad panel of microbiological parameters, trace elements, qualitative criteria, and mineral or organic suspended matter. Finally, all these studies have highlighted that fecal contamination of groundwater is generally associated with multiple factorial axes (PCs), underscoring the complexity of contamination conditions and pathways. The study will therefore also focus on this aspect.

2. Materials and Methods

2.1. Nouvelle-Aquitaine Region

The Nouvelle-Aquitaine region covers an area of 84,036 km² in the far south-west of metropolitan France (Figure 1a). It comprises 12 administrative departments: Charente, Charente-Maritime, Gironde, Landes, Pyrénées-Atlantiques, Deux-Sèvres, Vienne, Haute-Vienne, Creuse, Corrèze, Dordogne, and Lot-et-Garonne. The region is bordered to the south by the Pyrenees mountain range and to the west by the Atlantic Ocean. Its northern and eastern boundaries do not correspond to distinct geographical features but are the result of an administrative reform for territorial consolidation in 2015. The altitude ranges from 0 m on the Atlantic coast to 2974 m at Pic Palas near the Spanish border.

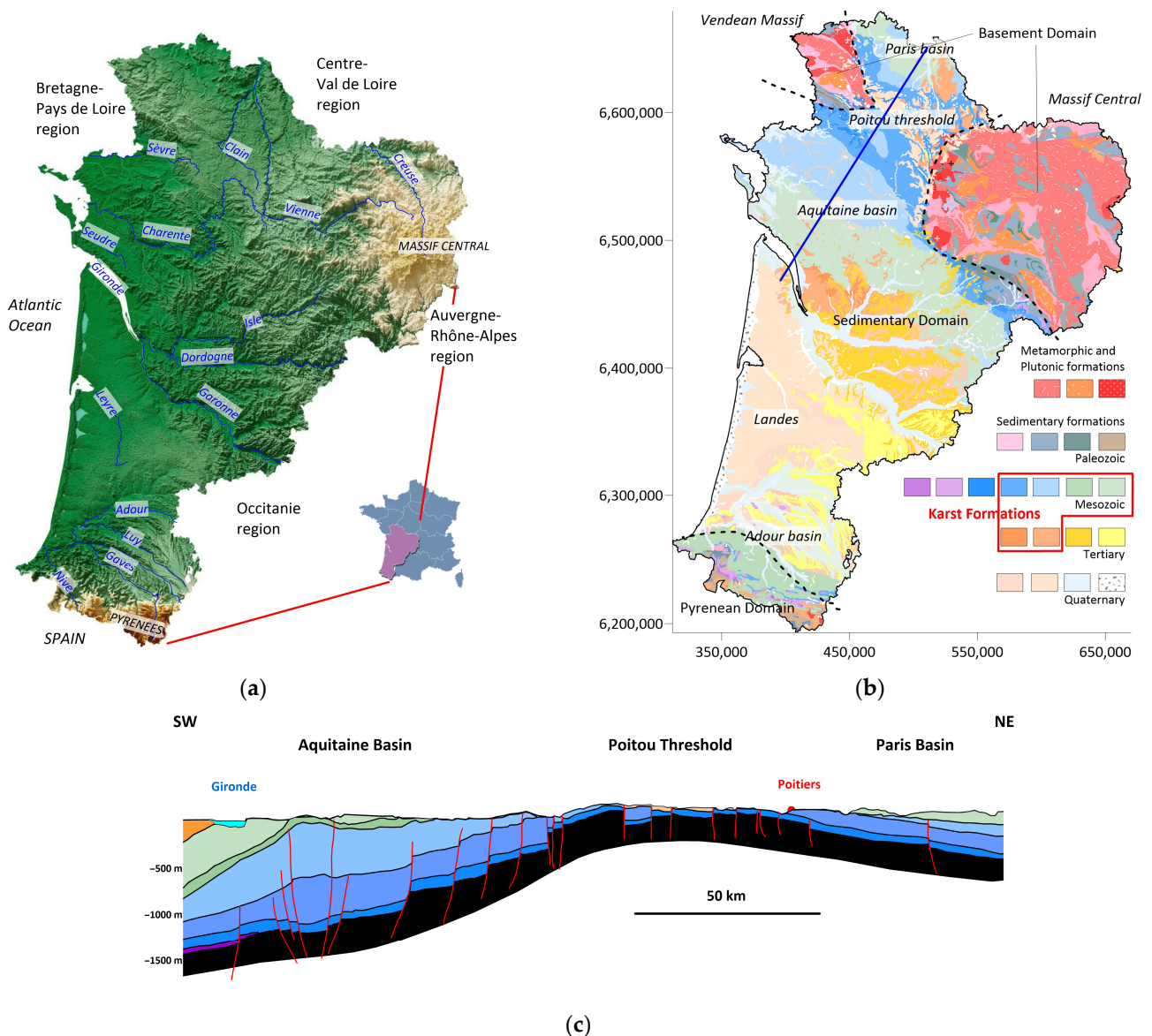


Figure 1. (a) Location and main physical characteristics of the Nouvelle Aquitaine region; (b) simplified geological map of the Nouvelle Aquitaine region (UTM coordinates in m); (c) Simplified geological cross-section (blue line on (b)) from the Paris Basin to the Gironde (adapted from BRGM, <https://www.brgm.fr/fr/implantation-regionale/nouvelle-aquitaine>, accessed on 1 January 2025, and AGSO-AGBP, https://www.agso.net/sites/agso.net/IMG/pdf/livret_guide_seuil_poitou.pdf, accessed on 11 November 2025).

The region is predominantly characterized by an oceanic climate in the west, with several variations (Aquitaine Oceanic climate, North-West Oceanic climate, Limousin Oceanic climate, Basque Oceanic climate), locally influenced by a continental climate in the north-east, and a Pyrenean Mountain climate in the extreme south. Two major hydrographic basins drain the region: the Adour-Garonne basin in the south (covering 71% of the region) and the Loire-Bretagne basin in the north (29% of the region), beyond the Seuil du Poitou (Poitou threshold [28]), which marks the boundary separating the Paris Basin from the Aquitaine Basin, and where Hercynian crystalline bedrock and Meso-Cenozoic sedimentary formations outcrop.

The region's geology (Figure 1b) is divided into three major domains: the sedimentary domain (Paris Basin and Aquitaine Basin), the basement domain of the Massif Central and the Vendean Massif (in the far north-west of the region), and the Pyrenean domain. This

results in a great diversity of numerous aquifers, which can be unconfined or confined, and of sedimentary, alluvial, basement, or karstic type.

The basement formations are naturally impermeable, but weathering and/or fracturing render them permeable, promoting preferential water flow and providing storage capacity. Weathering produces arenaceous (weathered rock) materials, which, depending on the initial mineralogy and texture of the rock, can vary from sandy to clayey. Sandy and coarse-grained arenites exhibit the highest permeabilities and storage capacities. Groundwater resources in these areas are generally modest and very unevenly distributed. Arenite is generally thinner and coarser on steep slopes, while it is more developed and finer-textured in low-gradient areas. The thickness of these weathered materials usually does not exceed 15 m. In unaltered rock, water circulates only through open fissures, mainly found between 10 and 100 m deep, which exhibit variable permeability depending on their degree of clogging. On the Poitou threshold, the aquifers consist of relatively thin sedimentary formations overlying the basement (Figure 1c). On either side of the threshold, groundwater generally flows toward the Aquitaine Basin to the southwest and the Paris Basin to the northeast. These sedimentary basins contain complex aquifer systems, characterized by an alternation of unconfined aquifers near the surface and along the basin margins, with confined aquifers at greater depths. The Tertiary detrital deposits (calcareous molasse) form heterogeneous, multi-layered systems. Jurassic and Cretaceous karst formations exhibit dual porosity (fissures and conduits), which facilitates rapid water flow and strong responsiveness to rainfall, while also making them highly vulnerable to surface contamination. However, flow dynamics are also influenced by elevation; it is faster toward the Poitou threshold, where elevations range between 100 and 300 m, and slower toward the two sedimentary basins, where elevations drop to between 5 and 50 m. Anoxic conditions may develop in groundwater in these areas. Soils developed over limestone and molasse rocks are generally brown soils with finer textures and, therefore, higher filtering capacity than the arenaceous materials developed over crystalline basement. Coastal aquifers, located along the Atlantic coast, occur in sandy and calcareous formations. They benefit from high precipitation and rapid infiltration but remain exposed to saltwater intrusion and climate variability. The limestone and molassic formations are covered by Quaternary aeolian sands, a few meters thick, forming the Landes triangle of approximately 14,000 km².

Agriculture occupies a significant portion of the territory, with large-scale crop farming and viticulture concentrated in the plains and foothills (Figure 2a). The basement areas support pastures for cattle farming, which is also secondarily present in marshland areas (Marais Poitevin, Marais de Rochefort-Saintes). The region is also the leading region in France for the production of goat cheese and milk, primarily limited to the Poitou threshold, and for ewe's milk cheese in the far south (Pays Basque, Béarn). Major urban centers such as Bordeaux, Poitiers, Limoges, and Bayonne structure the regional landscape and exert pressure on agricultural land and water resources.

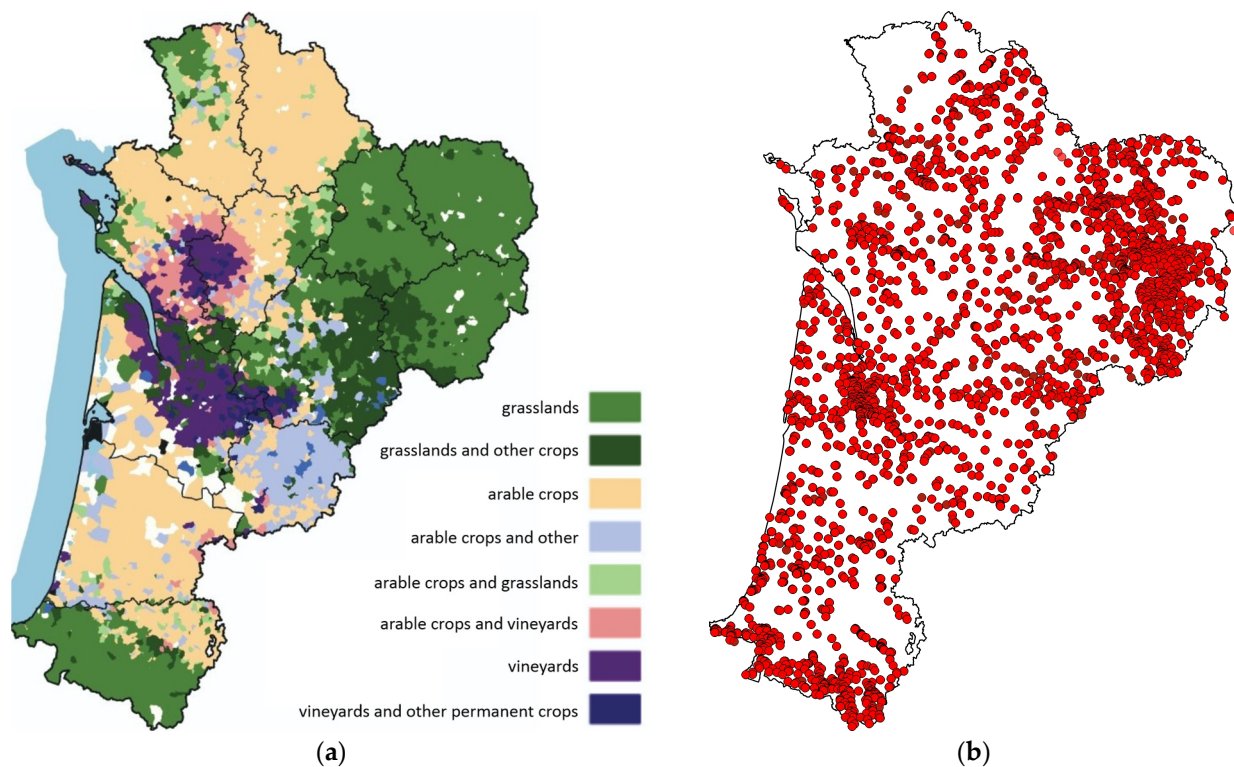


Figure 2. (a) Land use in Nouvelle Aquitaine (year 2020, adapted from [29]); (b) Distribution of the 2946 sampling points for the dense matrix.

2.2. Database

SISE-Eaux is the primary French database containing information on the quality of groundwater intended for human consumption. It is managed by the French Ministry of Health but is administered at the regional and departmental level by the Regional Health Agencies (ARS, Agences Régionales de Santé) and has been populated with data since the 1990s [30,31]. This study is based on data extraction from this database, covering the entire Nouvelle-Aquitaine administrative region over a period of 22 years, from 9 January 2002 to 9 April 2024. Given that the data are sensitive as they relate to public water supply, the geographical coordinates of the abstraction points are not provided, and their location on the maps has been adjusted to the nearest kilometer, which does not affect the results obtained. All analyses were performed by laboratories approved by the ARS, holding international accreditation and analytical quality certification. The extraction concerns only groundwater, and only raw water samples—collected at abstraction points and not treated for potability—were included in this study.

The number of parameters analyzed for each water sample varies depending on the type of analysis, which itself is determined by the monitoring history. During the initial connection of a catchment to a water supply network, a comprehensive analysis is typically conducted, encompassing several hundred parameters. These parameters may have evolved since the database's inception due to analytical advancements and the emergence of new contaminants. Generally, a full analysis includes physical characteristics, chemical characteristics (major, minor, and trace ions (metals, metalloids . . .)), a broad bacteriological panel, a wide range of pesticides and related products, as well as antibiotic substances. If public health issues are detected, subsequent analyses are usually basic (physicochemistry, major ions, nitrogen species, and fecal bacteriology) supplemented with parameters targeting the identified problems. The frequency of sanitary controls is not regular and depends mainly on two factors, primarily the size of the population served by the water catchments. Sampling is frequent in areas with large populations, but controls

are much less frequent for catchments supplying only a few houses. The frequency also depends on any history of contamination or health issues. Catchments or sectors that regularly show non-compliance are monitored more frequently. Despite this, knowledge of the zoning of non-compliance risks remains approximate, which also justifies this study.

The data extraction therefore resulted in a matrix containing numerous empty cells—a sparse matrix of 121,649 rows (water samples) and 51 columns (parameters). Among these 51 parameters, we find conventional parameters such as major ions, key physicochemical characteristics, nitrogen species including Kjeldahl nitrogen (typically an indicator of recent pollution or direct pathways to the aquifer), several trace elements and/or redox-sensitive metals (Fe, Mn, As, Se, Pb, Cu, Zn, Hg, Cr, Br, Ni, Cd, F), and a diverse range of bacteriological parameters. These include bacteria of fecal origin (*Enterococci* and *Escherichia coli*, denoted as Enter. and E.coli), total coliforms (Colif.) and thermotolerant coliforms (Therm. Col.), sulfite-reducing bacteria and their spores (SO₃ red., sp. SO₃ red.), bacteria cultivable at 22 °C and 37 °C, *Pseudomonas aeruginosa* (Pseudom.), *Legionella pneumophila* (Legion.), cysts of *Giardia intestinalis* (Giardia), and oocysts of *Cryptosporidium* (Ooc. Crypto.). The comprehensive analysis of this sparse matrix will be the subject of a separate publication.

Following the removal of empty cells, a dense matrix of 23,319 observations and 15 parameters was selected for this study. The selection of parameters is based on a compromise between the number of available data points (rows) for each parameter (columns) and the number of parameters for a multifactorial approach. The 15 retained parameters are: electrical conductivity at 25 °C (EC), indicators of fecal bacteriology (Enter. and E.coli), major ions (Na⁺, K⁺, Ca²⁺, Mg²⁺, Cl[−], SO₄^{2−}, HCO₃[−]), nitrogen species (NO₃[−], NO₂[−], NH₄⁺), turbidity (Turb.), and pH. For clarity in the subsequent text, a distinction is made between a parameter (e.g., SO₄) and its corresponding ion (e.g., SO₄^{2−}). The spatial distribution of the study's sampling points is presented in Figure 2b. The dense matrix includes data from 2946 distinct sampling points, with between 1 and 15 samples collected per point, representing an average of 7.92 samples per sampling point over the extraction period.

2.3. Data Conditioning

Previous studies based on extractions from the Sise-Eaux database revealed that extreme values are likely to obscure the detection of certain key mechanisms in the acquisition of the physicochemical and bacteriological characteristics of water. A logarithmic transformation of the data overcomes this problem [21], which has been tested through the construction of quantile-quantile plots and Kolmogorov–Smirnov normality test, adapted to high-dimensional statistical distributions. Consequently, the data were transformed according to the formula $y = \log_{10}(x + DL)$, where x and DL represent the measurement of the physicochemical or bacteriological parameter X and its detection limit, respectively. For each analyzed parameter, the detection limits were provided by the laboratories. Only pH, which already corresponds to the logarithmic transformation of the chemical activity of H₃O⁺, was retained without conditioning. This data processing aims to bring the distribution of each parameter closer to a normal distribution and, most importantly, to limit the influence of extreme values.

2.4. Principal Component Analysis

To reduce the dimensionality of the data hyperspace, a Principal Component Analysis (PCA) was performed on the log-transformed data using the dense matrix. This treatment aims to identify and rank the sources of variability within the dataset [9]. The PCA was based on the correlation matrix, thus considering standardized variables, which allows for the integration of parameters of very different natures and units, or even unitless parameters like pH. Furthermore, it was performed by diagonalizing the correlation matrix.

Under this condition, the factorial axes are orthogonal to each other and are therefore theoretically associated with independent processes responsible for the variability of water quality. Factorial axes accounting for approximately 80% of the information were retained. The remaining factorial axes, which explain a low percentage of the variance and are generally considered statistical noise [32], were eliminated.

2.5. Parameter Clustering

An unsupervised ascending hierarchical clustering (AHC) [33,34] was performed on the bacteriological and physicochemical parameters based on the first factorial axes from the PCA. This was performed to understand the diversity and similarities in the distribution of parameters at the regional scale, leading to a parameter typology. Dissimilarity was measured using Euclidean distance.

2.6. Parameter Mapping

The selection of parameters when transitioning from the sparse matrix to the dense matrix is likely to introduce a bias in the information contained in the database extraction, as all samples for which at least one parameter was not reported were excluded. Therefore, the spatial analysis and mapping of individual parameters were conducted using the sparse matrix (121,649 rows). Spatial distribution maps of the parameters were generated by kriging based on experimental variograms [35], which were studied and adjusted to a spherical model with a nugget effect using the least squares method [36]. This approach allowed for the selection of the variogram models best suited to each parameter, thereby improving the accuracy of the interpolations and the reliability of the produced maps. Variograms measure the evolution of the semi-variance between pairs of points as a function of the distance separating them [37]. However, in the extracted dataset, water samples were collected at different points and on different dates, thus incorporating both spatial and temporal variability. In this context, variograms were calculated using the mean values at each sampling point [38]. Consequently, the variograms reflect a semi-variance less influenced by temporal variability, although variograms calculated in this manner inevitably still include a proportion of temporal variability, albeit reduced, as the sampling spanned 22 years and the collections were not synchronous across the different points. Raw and directional variograms with a rotation step of 15° were calculated to detect any anisotropy in the distribution of parameter values. For reference, the number of analyzed water samples used to calculate the variograms and maps for major elements, Electrical Conductivity (EC), *E. coli*, and *Enterococcus* ranges between 2946 and 4027 data points per parameter. This number is slightly lower for nitrates, metals (Fe, Mn), and coliform, cultivable, and sulfite-reducing bacteria (1452 to 1589). The quality of the interpolation was evaluated by comparing the calculated value to the average of the measured values at each sampling point.

2.7. Seasonality

To analyze the seasonal variance of fecal contamination over the 22-year dataset, the deviations from the mean for the *E.coli* parameter were sorted by day of the year and presented graphically. This approach reflects the local temporal variability for a given sampling station. An analysis of variance was performed to quantify the impact of seasonality (winter, spring, summer, late summer + autumn) on the explained variance for fecal contamination.

3. Results

3.1. Distribution of Parameters

The mapping results for the various parameters were classified by their degree of similarity. No significant anisotropy in the distribution of values was detected for any of the parameters. The first group was related to water mineralisation, comprising Electrical Conductivity (EC), Dry Residue, Calcium (Ca), and Bicarbonate (HCO_3) (Figure 3). These parameters exhibited low values over the crystalline bedrock of the Massif Central and the axial zone of the Pyrenees. The areas with the highest mineralisation were located in Charente/Seudre/Sèvre basins and, to a slightly lesser extent, over the Tertiary molasse deposits of the Garonne Basin. The Quaternary aeolian sands of the Landes region showed low mineralisation. It is noteworthy that a minor discrepancy existed between EC and the (Ca + HCO_3) pair in the lower Garonne valley. This area exhibited a peak in overall mineralisation, yet it did not correspond to the maxima for Calcium and Bicarbonate. This indicated a significant contribution from other major ions to the water mineralisation in this sector. The distribution of $\log(\text{Na})$ and $\log(\text{SO}_4)$ values, and secondarily of $\log(\text{Cl})$, confirmed this observation (Figure 3). Their spatial patterns were quite similar to those of the major ions previously described, but with a distinct specificity for the lower Garonne valley, where concentrations were at a maximum. The Mg/Ca ratio, which typically reflects the duration of water-rock interaction in carbonate environments [39,40], highlighted the distinct nature of the lower Garonne valley, showing elevated Mg/Ca ratios. The regression between the averages of the measured values and the calculated values (shown for the $\log(\text{EC})$ parameter, Figure 4a) indicates that the generated maps are reliable for the major ion chemistry.

Therefore, in general, the major ions exhibited an upstream/downstream distribution pattern that aligned with the hydrological, hydrogeological, and topographic gradients, and which had a strong lithological control.

The structure of the microbiological parameters was fundamentally different from that of the major ions and, more broadly, from all physicochemical parameters (Figure 5). This spatial structure indeed revealed a strong local component and high temporal variability. Nevertheless, regional structural trends were discernible. The maximum values were located on the foothills of the Massif Central (i.e., the northern bank of the lower Garonne valley) and the Pyrenees, as well as on the Poitou threshold. It should be recalled that for fecal contamination, the drinking water standard is the total absence of bacteria, i.e., a value of zero according to the formula presented in Section 2.3. The correlation between the measured and calculated values is significantly weaker for fecal contamination parameters (shown for $\log(\text{E.coli})$, Figure 4b) than for major ions and electrical conductivity. It should also be noted, however, that the lower number of data points for the parameters *Pseudomonas aeruginosa*, *Legionella pneumophila*, *Giardia intestinalis* cysts, and *Cryptosporidium* oocysts (Figure 5g–j) resulted in less precise maps compared to the other parameters. In the most extreme case, this lack of data for Cyanobacteria, with only 5 sampling points across the entire region, made it impossible to produce a meaningful map.

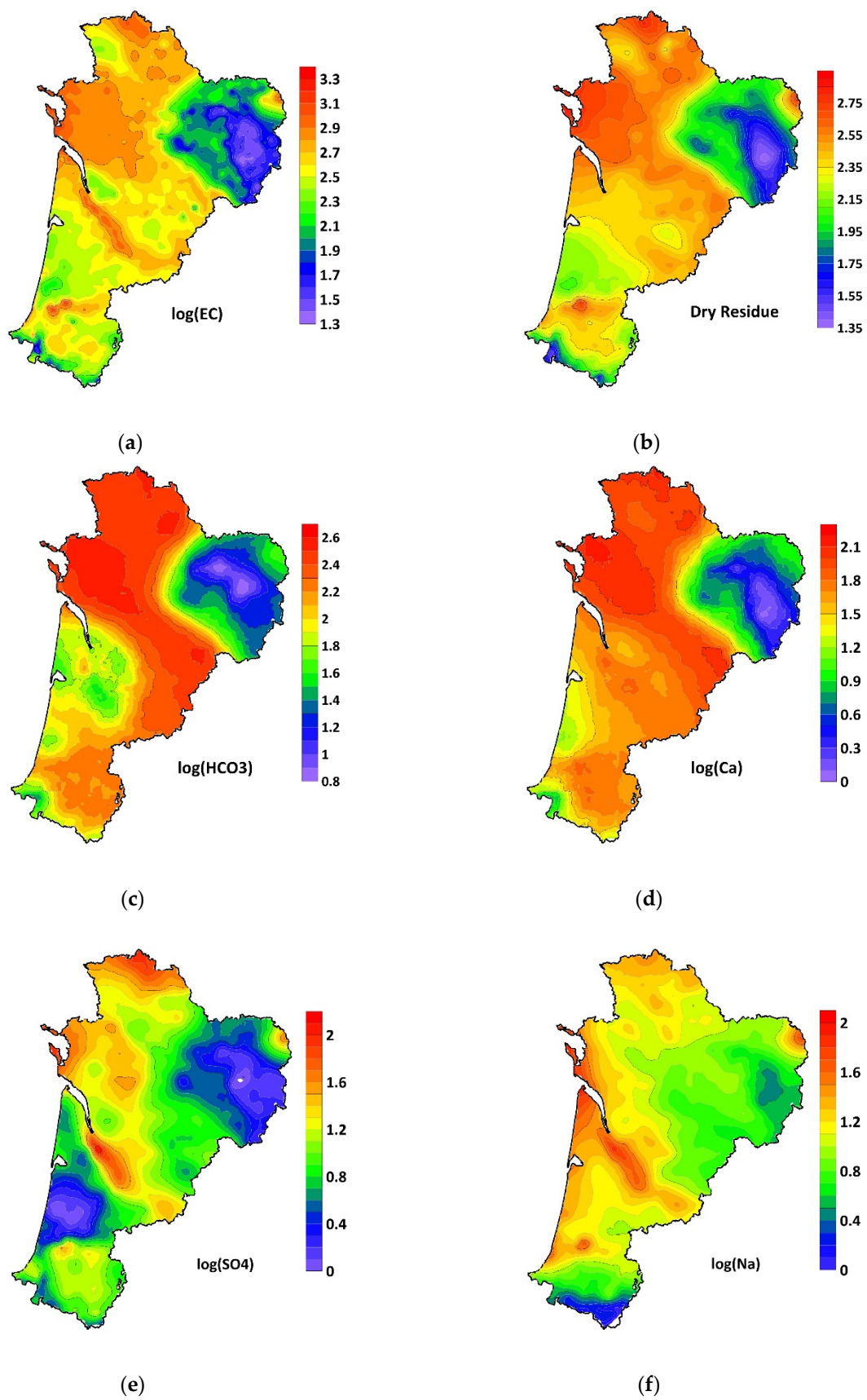


Figure 3. Cont.

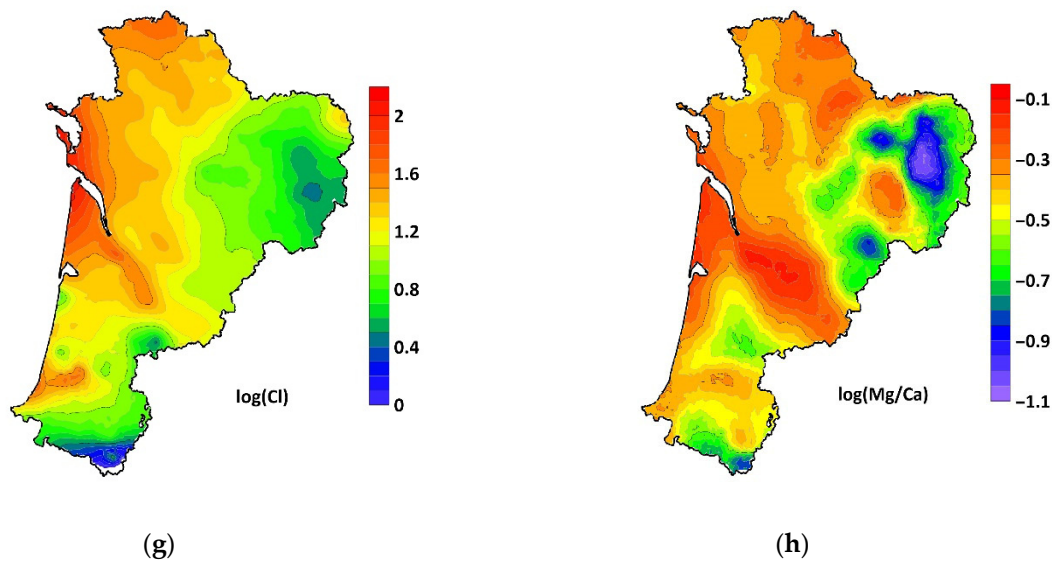


Figure 3. Log values distribution for: (a) EC; (b) Dry residue (TDS); (c) HCO_3^- ; (d) Ca; (e) SO_4 ; (f) Na; (g) Cl; (h) Mg/Ca ratio.

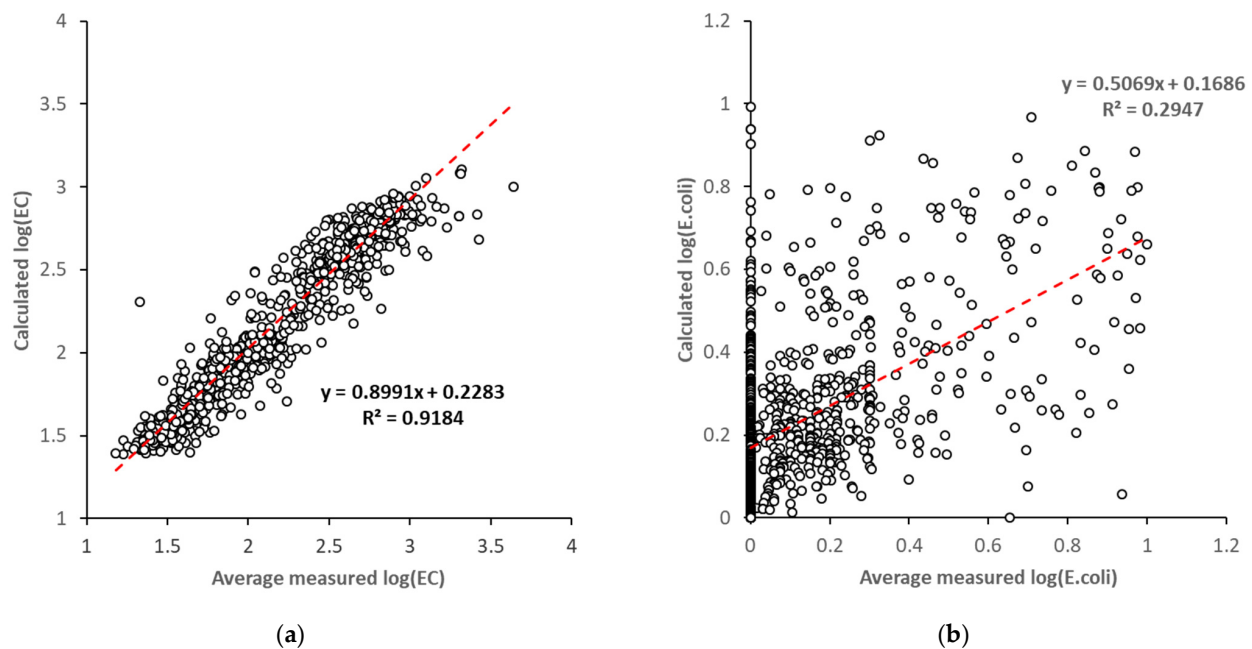


Figure 4. For each catchment point, regression between the averages of the measured values and the calculated values for (a) log(EC) and (b) log(*E.coli*).

Regarding the different nitrogen species and the metals Mn and Fe, a markedly different distribution pattern emerged (Figure 6). The nitrate (NO_3^-) and ammonium (NH_4^+) maps were almost inverse of one another. Nitrate concentrations were minimal in the “Landes triangle,” while maximum values were found on the Poitou threshold and in the Dordogne and Adour valleys. Nitrites (NO_2^-) were present in the crystalline sector of the north-western part of the region, and to a lesser extent in the Massif Central and the Pyrenees. The distribution of iron and manganese resembles that of ammonium. In contrast, the analysis of “Kjeldahl N”—which represents the total quantity of organic nitrogen and ammonium in a water sample—exhibited a distribution pattern distinct from the inorganic nitrogen species. The lowest values were located in the far north of the region within the Loire basin, and secondarily on the Pyrenean foothills.

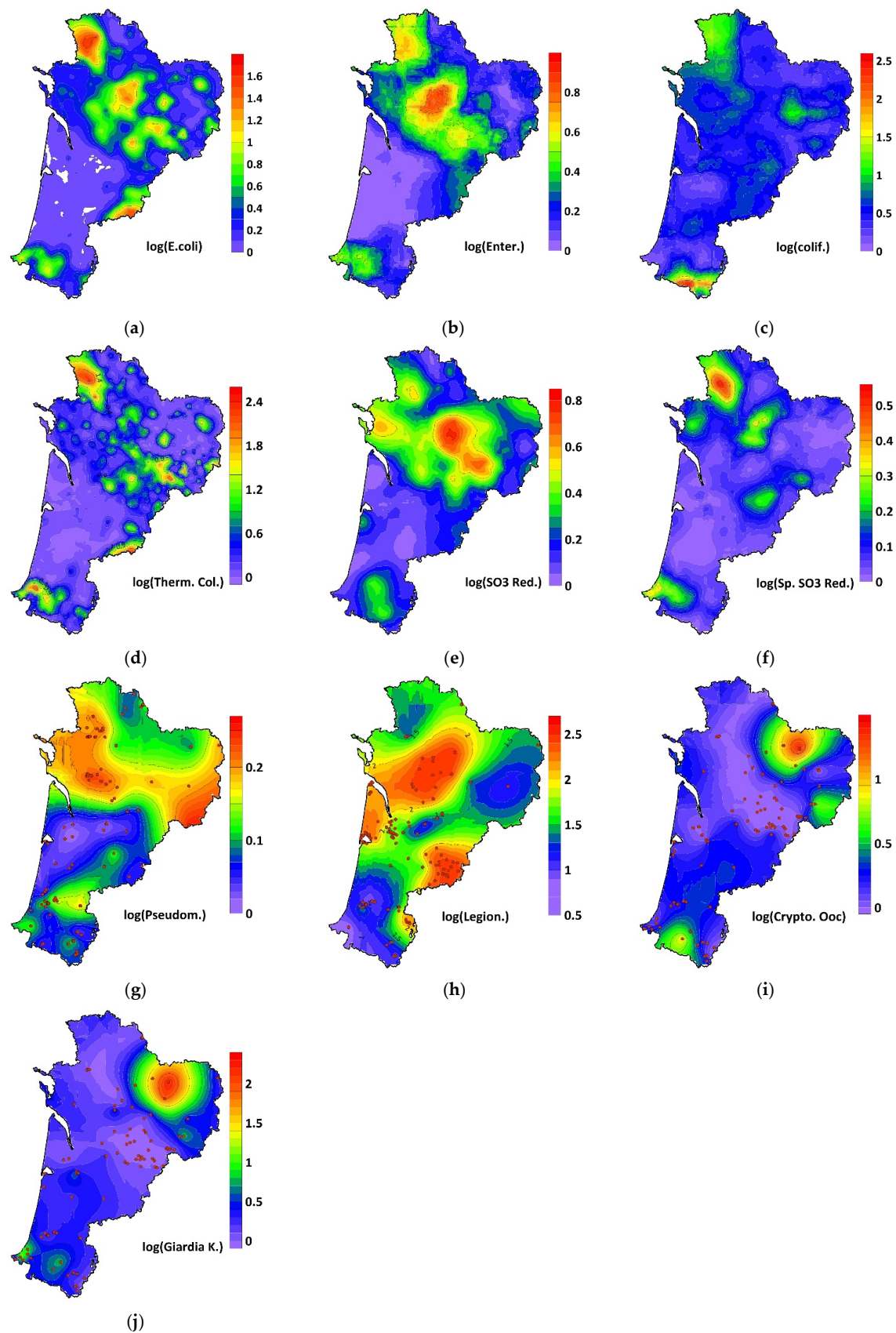


Figure 5. Log values distribution for: (a) *E.coli*; (b) *Enter.*; (c) Total coliforms; (d) Thermotolerant coliforms; (e) Sulfite-reducing bacteria; (f) Spores of sulfite-reducing bacteria; (g) *Pseudomonas aeruginosa*; (h) *Legionella pneumophila*; (i) Oocysts of *Cryptosporidium*; (j) Cysts of *Giardia intestinalis*. Red circles denote sampling points for (g–j).

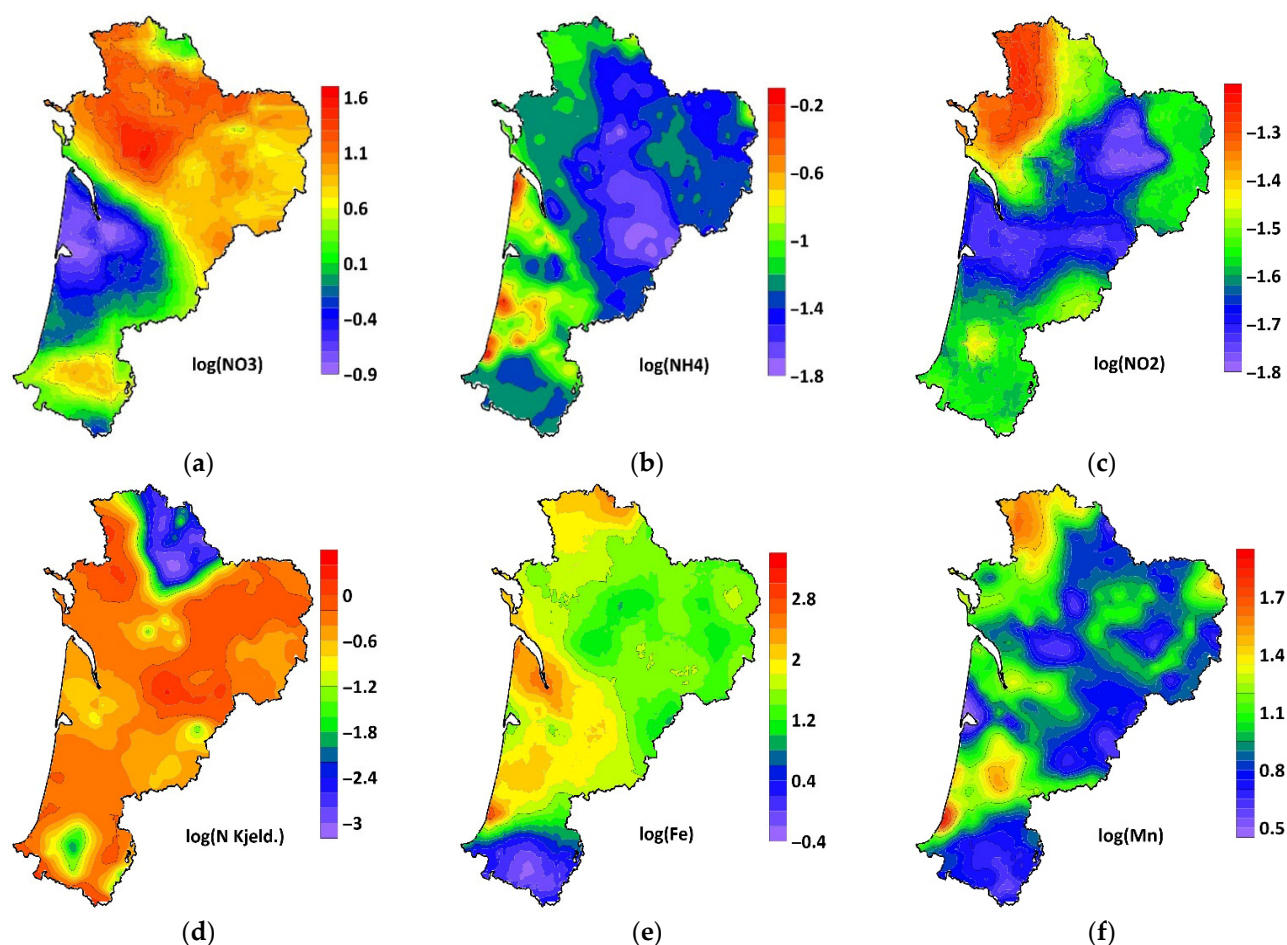


Figure 6. Log values distribution for: (a) NO_3 ; (b) NH_4 ; (c) NO_2 ; (d) N Kjeldahl; (e) Fe; (f) Mn.

3.2. Analysis Based on the Dense Matrix

Table 1 reports the values of the mean and standard deviation for both matrices (sparse matrix and dense matrix) for the parameters $\log(\text{EC})$, which is well correlated with major ions, and $\log(E.coli)$, representing fecal contamination. The data show very similar means and standard deviations between the matrices. Furthermore, it is observed that the difference between the means for the two matrices is small compared to the standard deviation, indicating that the “dense matrix” subset is well representative of the overall “sparse matrix.”

Table 1. Comparison of mean values and standard deviations for the parameters $\log(E.coli)$ and $\log(\text{EC})$ between the sparse matrix and the dense matrix.

	Sparse Matrix	Dense Matrix		Sparse Matrix	Dense Matrix
$\log(E.coli)$			$\log(\text{EC})$		
Numb. of values	41,096	23,319	Numb. of values	62,678	23,319
Mean	0.54	0.71	Mean	2.40	2.42
St. deviation	0.90	1.02	St. deviation	0.51	0.43

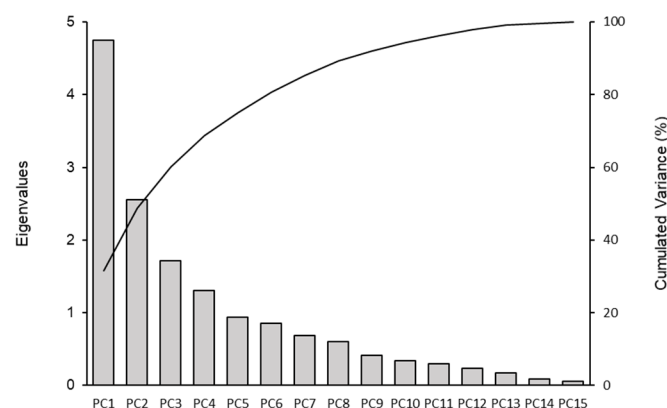
Descriptive statistics for the 15 parameters are presented in Table 2. The most significant variations were observed for parameters NO_2 , NH_4 , HCO_3 and EC.

Table 2. Descriptive statistics for the 15 parameters of the dense matrix (23,319 observations, log-transformed values).

Parameter	Minimum	Maximum	Mean	St. Deviation	Var. Coef.
NO ₃	−2.000	2.000	0.513	0.737	1.44
Enter.	0.000	5.540	0.578	0.801	1.38
Turb.	−2.000	2.978	−0.068	0.674	9.92
EC	0.000	3.628	2.283	0.691	0.30
NH ₄	−2.000	0.684	−1.228	0.355	0.29
pH	−11.250	0.000	−5.847	2.934	0.50
NO ₂	−2.000	0.286	−1.545	0.207	0.13
SO ₄	−1.000	3.417	0.932	0.567	0.61
Cl	−0.699	2.778	1.124	0.411	0.37
E.coli	0.000	5.540	0.712	1.019	1.43
Ca	−1.000	2.720	1.445	0.616	0.43
Mg	−1.000	2.493	0.667	0.446	0.67
Na	−0.699	2.623	0.994	0.397	0.40
K	−1.658	1.455	0.199	0.346	1.74
HCO ₃	−0.886	3.076	2.019	0.549	0.27

3.2.1. Principal Component Analysis (PCA)

Principal component analysis was performed on the dense matrix, comprising 23,319 samples from 2946 monitoring points (Figure 2b). It is important to note that this analysis integrated both spatial and temporal variability. Distinguishing between these two sources of variability will be the focus of subsequent work. The inertia of the factorial axes is presented in Figure 7.

**Figure 7.** Inertia of the factorial axes (PC₁ to PC₁₅) from the Principal Component Analysis performed on log-transformed data.

The first four factorial (Figure 8) axes accounted for 68.8% of the total variance, which was substantial given the number of parameters included (15). These axes had eigenvalues greater than unity, meaning they carried more information than any single original parameter (Table 3). This result demonstrated the effective dimensionality reduction of the data hyperspace. Furthermore, the first six factorial axes were required to account for 80% of the information within the dataset, underscoring the complexity of groundwater quality variations at this spatial and temporal scale. Table 3 details the contributions of the various parameters to the first four factorial axes (which explain between 31.6% and 8.7% of the variance), which will form the basis for our interpretation. The contributions of the parameters are also reported on the correlation circles (Figure 8). Subsequent factorial axes contribute a more limited share of the variance, are more challenging to interpret, and will therefore be considered part of the statistical background noise of the analysis.

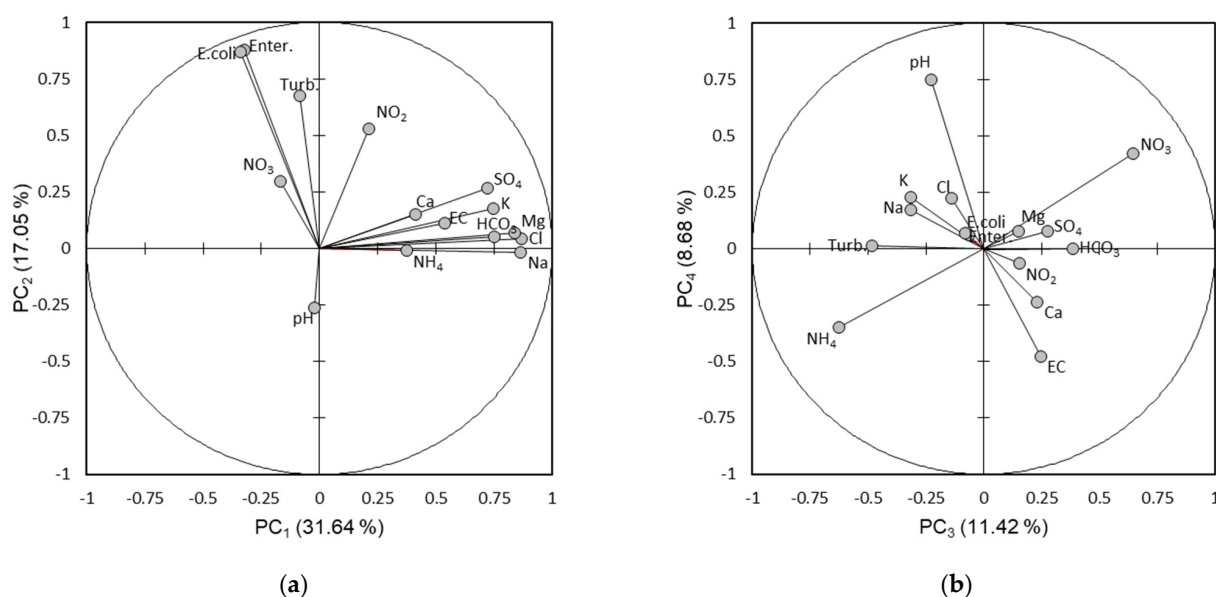


Figure 8. Distribution of parameters in the factorial planes: (a) PC₁–PC₂ and (b) PC₃–PC₄.

Table 3. Contribution of parameters to the first four factorial axes.

	PC ₁	PC ₂	PC ₃	PC ₄
Eigenvalue	4.746	2.558	1.712	1.302
Parameter				
NO ₃	−0.167	0.294	0.648	0.420
Enter.	−0.324	0.876	−0.062	0.060
Turb.	−0.083	0.674	−0.483	0.012
EC	0.537	0.113	0.246	−0.478
NH ₄	0.375	−0.010	−0.626	−0.351
pH	−0.021	−0.261	−0.228	0.750
NO ₂	0.214	0.529	0.154	−0.066
SO ₄	0.723	0.267	0.277	0.077
Cl	0.868	0.041	−0.137	0.223
E.coli	−0.339	0.868	−0.079	0.070
Ca	0.411	0.152	0.232	−0.237
Mg	0.838	0.071	0.152	0.075
Na	0.864	−0.016	−0.315	0.170
K	0.746	0.178	−0.317	0.227
HCO ₃	0.752	0.050	0.386	0.001

The first principal component (PC₁) was strongly defined by major ions and conductivity. This represented a mineralization gradient, which contrasted mineralized waters—particularly those exhibiting a sodium sulfated-chlorinated facies—with diluted waters that were incidentally contaminated with *E. coli* and *Enterococcus*. High values are observed in the metamorphic and crystalline zone of the Massif Central, the high-altitude areas of the Pyrenees, the lower Adour valley, and the lower Garonne valley (Figure 9). Low values extend across broad areas, including the Landes region and the Pyrenean foothills, the Dordogne basin, and the zone of secondary and tertiary rocks in the Aquitaine Basin adjacent to the Massif Central. The second factorial axis (PC₂) highlighted the specific association between fecal contamination and nitrites, or more generally, the combination of nitrite and nitrate. Turbidity also contributed strongly to this axis. High PC₂ values are associated with the crystalline massifs in the northern part of the region and the Poitou Threshold, the upper Dordogne basin, the Landes region, the upper part of the Garonne basin, and, to a lesser extent, the Pyrenees. Low values are located in the Garonne Val-

ley (both right and left banks), the northern part of the Aquitaine Basin, and the Paris Basin. The third principal component (PC_3) reflected the opposition between nitrate and ammonium—that is, the two forms of nitrogen, the former is stable in oxidizing conditions and the latter is stable in reducing environments. PC_3 exhibits maximum values primarily along the coastline. The lowest values cover the remainder of the region, specifically areas of medium and high altitude, irrespective of lithology. The fourth principal component (PC_4) was primarily defined by pH and nitrate with positive loadings, and by electrical conductivity and ammonium with negative loadings.

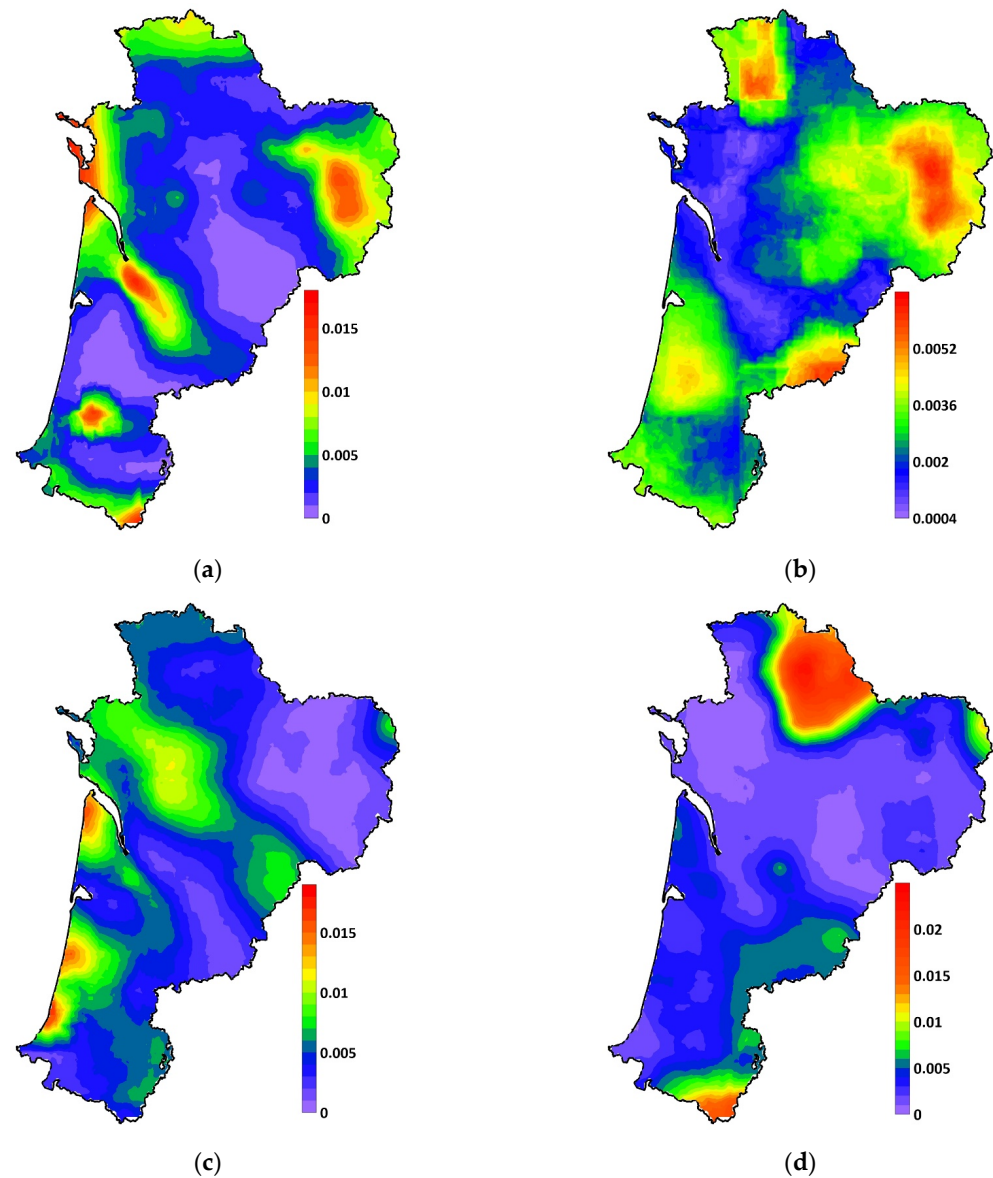


Figure 9. Distribution of the sampling point coordinates on PC_1 to PC_4 (a–d).

3.2.2. Parameters Clustering

The hierarchical classification was based on the PCA results, meaning it considered essential information after eliminating statistical noise. The dendrogram (Figure 10) makes it possible to distinguish three main sets of parameters. The first consisted of fecal bacteriology, associated with water turbidity, and to a lesser extent, with the presence of nitrites. The second was pH, which appeared as an isolated parameter. Finally, the third concerned major ions and electrical conductivity, secondarily associated with other forms of nitrogen (NO_3 and NH_4).

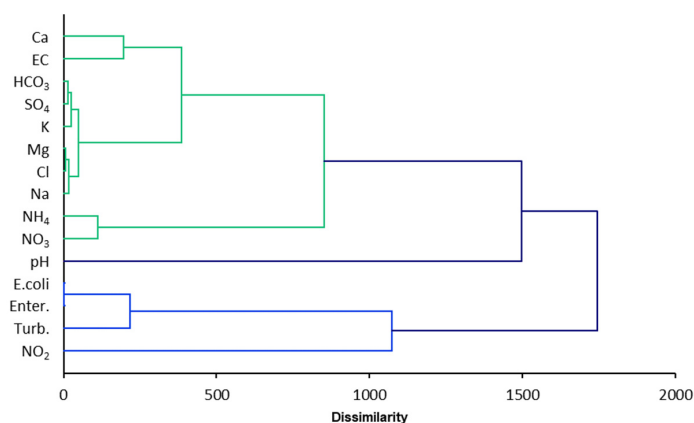


Figure 10. Typology of the parameters based on the first six factorial axes (80% of the variance).

3.3. Seasonality for Fecal Contamination Parameters

Figure 11 shows a compilation of 22 years of temporal variance for the E.coli parameter at a sampling station affected by fecal contamination, plotted against the day of the year. It reveals low variance from January to June and significantly higher variance from July to the end of November.

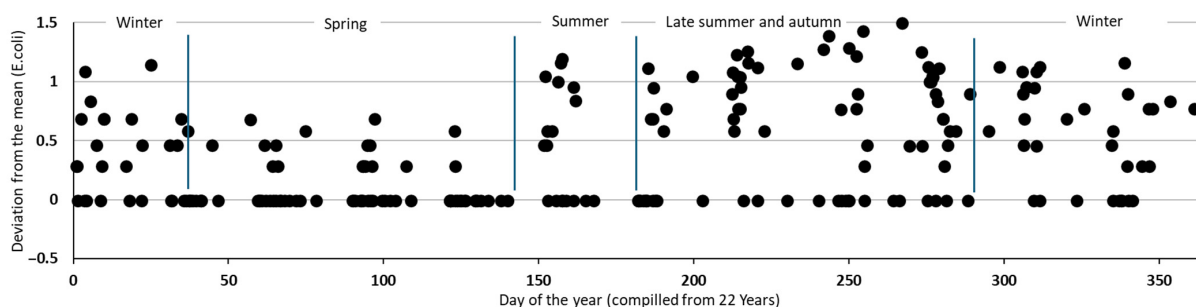


Figure 11. Seasonal trend by day of the year for the bacteriological parameter E.coli over 22 years of the sampling period.

The analysis of variance conducted using the compartmentalization into four periods of the year (Figure 11: spring from day 37 to 140, summer from 141 to 181, late summer and autumn from 182 to 289, and winter from 290 to 36) reveals an R^2 coefficient of 0.282 (RMSE = 0.415).

4. Discussion

The analysis conducted on the groundwater of the Nouvelle-Aquitaine region highlights several aspects, namely the role of lithology, land use, and redox processes in water diversity.

4.1. Role of Lithology in Water Diversity

The chemical diversity of the groundwater in Nouvelle-Aquitaine is strongly structured by lithology and landscape position [41]. The regional scale of the geological structures explains the high correlation coefficient obtained during cross-validation for major ions and electrical conductivity (Figure 4a). Major ions, particularly Ca, and electrical conductivity, which appear separately from other major ions in the parameter classification, reflect this lithological influence. In the crystalline zones of the Massif Central and the axial zone of the Pyrenees, composed of sparingly soluble granites and gneisses, the waters exhibit low values of Ca, HCO_3 and reduced Electrical Conductivity. The low mineralization is less pronounced in the crystalline sector of the far north-west of the region, not because the phenomenon is lesser, but due to the low number of sampling

points (Figure 2b), introducing a bias in the presented maps (Figure 3a). The low solubility of these deep-origin crystalline rocks, combined with a colder and wetter climate due to higher altitude, limits dissolution, leading to poorly mineralized waters. Similarly, the Quaternary eolian sands of the Landes [42], due to their siliceous and permeable nature, also result in poorly mineralized waters.

In contrast, the sedimentary limestone areas of the Paris and Aquitaine basins, Poitou threshold, and the alluvium of the Garonne Valley, as well as the calcareous molasse sectors of the Adour basin, present more highly mineralized waters, reflecting the higher solubility of the rocks. The deep aquifers there ensure regular circulation and homogeneous mineralization. This mineralization is favored by the duration of water-rock interaction, which results in an upstream-downstream minerality gradient, observed, for example, in the Garonne Valley, notably with a higher Mg/Ca ratio. The first factorial axis of the Principal Component Analysis, accounting for 31.6% of the variance, highlights this regional geological structuring, but not exclusively. It also reflects a vulnerability to bacterial contamination of the poorly mineralized groundwater, as well as the distinction between old groundwater and recent infiltration surface water, as will be discussed later.

4.2. Importance of Agricultural Activities and Land Use

4.2.1. Livestock Farming and Bacteriology

The first factorial axis (PC₁) exhibits a complex interplay between rock type, water-rock interaction time, and vulnerability to fecal contamination in the context of low-mineralization waters. Although this contamination is primarily associated with the second factorial axis (PC₂), its presence on PC₁ underscores the importance of fecal contamination as a criterion and demonstrates the complex determinism involving at least two independent processes. Similar findings have been reported in other French administrative regions, notably Provence-Alpes-Côte d'Azur, Auvergne-Rhône-Alpes, Bourgogne-Franche-Comté, Occitanie, and Corsica [20,24,25,27,38]. All these studies, as well as the present one, highlight both the localised nature of faecal contamination and its high temporal variability, all of which depend on numerous factors we will detail below. This very localised structure and the strong component of temporal variability explain the weak relationship obtained during cross-validation (Figure 4b) and, consequently, the lower reliability of the generated maps (Figure 5).

On factorial axis 1 (PC₁), these dilute waters are recent, and their position is largely independent of their chemical signature, with the exception of a minor contribution from nitrates. It is noteworthy that water turbidity does not contribute to this factorial axis, suggesting at least partial soil filtration and indicating infiltration from rainfall and runoff through likely gently sloping soils with low stone or gravel content. PC₁ thus reflects a general bacteriological pollution pressure, independent of the level of protection of the water abstraction points. On the PC₁ distribution map (Figure 9), mineralized waters correspond to low values, while dilute waters correspond to the highest values.

The second component of fecal contamination, represented by the second axis (PC₂), is the primary source of variability. It is associated with turbidity and indicates the contribution of surface waters and/or preferential flow along fractures [43–45]. The PC₂ distribution map reflects the pollution pressure associated with the vulnerability of the water abstraction points.

The seasonality of log(*E.coli*) variance at a station affected by fecal contamination highlights that this variance increases from the month of June (Day 140), coinciding with the first rains following the manure spreading period in livestock farming areas. The contamination persists regularly until the end of November (Day 290). This observation, as well as the prevalence of fecal contamination in dilute waters, reflects:

1. The impact of late summer and autumn storms, which generate significant runoff of poorly mineralized water. This aspect was initially proposed as a hypothesis and was later confirmed by distinguishing spatial and temporal variances on a more comprehensive time series dataset [26]. Seasonal analysis highlights more pronounced contamination during the second half of August, a period known for its severe thunderstorms. The relatively low R^2 coefficient from the analysis of variance (0.282) can be explained by the occurrence of short-duration, isolated, and high-intensity events. These events generate runoff but have a limited temporal impact and do not affect all sampling points, only the most vulnerable ones;
2. The vulnerability of certain sectors, which are poor in divalent ions and have low mineralization, where particle flocculation is limited, thereby promoting the transport of bacteria to the aquifers.

Thus, groundwater fecal contamination results from the interaction between livestock-related inputs and local environmental characteristics. The most impacted groundwater is located in:

1. The foothills of the Massif Central, coinciding with areas of mixed (dairy and beef) cattle farming and multi-species herbivore rearing;
2. The Pyrenees, where frequent contamination is associated with sheep and goat farming on the humid Atlantic slopes, where precipitation facilitates the transport of bacteria to shallow aquifers.

In the impacted sectors, the concurrent elevation of both nitrates and Kjeldahl nitrogen indicates continuous pollution [46], comprising both recent inputs (Kjeldahl N) and older degradation products (Nitrates).

The two fecal contamination parameters (*E. coli* and *Enterococci*) appear almost indistinguishable in the hierarchical parameter classification, with near-zero dissimilarity, confirming a common origin and pathway [47]. Conversely, their strong association with turbidity underscores that suspended solids act as a vector for transporting fecal bacteria to the aquifer [48–50], probably in a context of permeable, shallow soils, or rapid preferential flow pathways. Furthermore, the correlation between bacteriological parameters and nitrogen compounds reflects organic inputs from livestock farming.

Therefore, aquifer fecal contamination depends on both surface inputs from livestock farming [51] and the environmental characteristics that control transport. Areas where these factors combine are the most vulnerable.

4.2.2. Fertilization and Nitrogen Species

The spatial distribution of nitrogen species reflects not only inputs from livestock farming but also the direct influence of agricultural practices and fertilization types. Elevated nitrate concentrations are found in the Mesozoic sedimentary areas of the Aquitaine Basin and the Poitou threshold, which are predominantly occupied by large-scale cereal and oilseed crops. These areas are subject to regular applications of mineral fertilizers, particularly those supplying both nitrate and ammonium simultaneously. The highly mobile nitrate readily infiltrates permeable and well-drained soils, whereas ammonium is more retained in clay and organic horizons. Water-saturated environments, characterized by a shallow water table, limit nitrogen oxidation and promote the persistence of ammonium. Clustering confirms this pattern: the relationship between nitrate, Ca, and HCO_3 indicates that high nitrate levels are associated with well-aerated carbonate environments conducive to nitrification, as well as intensive agricultural zones. In contrast, the sandy, wet areas within the Quaternary sands of the Landes region [42], which are lightly cultivated, show low nitrate levels but higher ammonium concentrations.

4.3. Redox Processes in the Landscapes of Nouvelle-Aquitaine

The third factorial axis (PC₃) reflects redox conditions, illustrating an opposition between nitrate and ammonium—the two stable forms of nitrogen in oxidizing and reducing environments, respectively. In other words, nitrates are primarily present in oxidizing, high-permeability, and sloped areas, such as sandy and gravelly soils, where rapid water flow allows oxygenation and stabilizes nitrate in solution, as observed in the Pyrenean foothills. Conversely, in low-altitude, low-slope areas like the Quaternary sands of the Landes, where the water table is shallow and water circulation is slow, conditions are reduced. Nitrification is limited, nitrates are scarce, while ammonium, Fe, and Mn (soluble under reducing conditions) exhibit higher concentrations. In this regard, it should be noted that Fe, Mn, and NH₄ show a fairly similar distribution at the regional scale (Figure 6b,e,f).

A specific observation from the clustering concerns nitrites, which show an unexpected correlation with turbidity. This association appears to indicate that, in some well-oxygenated but organic matter-poor areas, the transport of fine particles promotes the presence of low nitrite concentrations. Turbidity, linked to erosion, topography, and agricultural activities, reflects the suspension of these particles, which are often associated with organic carbon. The observed nitrites may thus originate from the partial transformation of nitrate in these oxidizing environments with low biological activity.

The distribution of fecal bacteriology differs from that of other microorganisms, such as *Pseudomonas*, which are more influenced by environmental characteristics than by livestock inputs. This distinction demonstrates that some bacteria are genuine indicators of fecal contamination, while others reflect local environmental conditions. Sulfite-reducing bacteria also exhibit a specific distribution. Their occurrence is linked to anoxic or low-oxygen zones, often in shallow or stagnant groundwater. These bacteria are involved in denitrification processes, and their occurrence can be compared with that of fecal bacteriology to identify areas where organic inputs are being mobilized and transformed, particularly in the bedrock of the far north-west of the region, on the edge of the Massif Central, and on the few crystalline outcrops of the Poitou threshold.

5. Conclusions

This work follows several studies aimed at extracting and analyzing the information contained in the Sise-Eaux database for several French administrative regions. Nouvelle-Aquitaine is the largest region (84,036 km²), with an area nearly three times that of the Provence-Alpes-Côte d'Azur region, which was the subject of a comparable analysis. This analysis consisted of a logarithmic transformation of the data, a principal component analysis, a mapping of macro-parameters, and a clustering of parameters. The results are consistent, indicating that the analytical method is robust and applicable over large areas.

In practice, the data collected by Regional Health Agencies (ARS) is used in two types of summaries. On the one hand, each year, to detect the presence of non-compliances in water intended for distribution to the drinking water network and to monitor the evolution of the number of non-compliances over time. On the other hand, to produce maps of individual parameter values, sometimes several hundred are considered if xenobiotic parameters are considered. The issue with parameter maps is linked to strong redundancies, typically for electrical conductivity (EC) and major ions, and for fecal contamination.

The factorial axis approach makes it possible to synthesize the information, but also to separate independent sources (in the mathematical sense) of variability and the processes responsible for this variability. Thus, fecal contamination, which is the primary quality concern for raw water intended for human consumption (as it is the main cause of non-compliance), exhibits two independent components reflecting two sources of variability. One appears to be more related to lithology and the spatial distribution of agricultural

activities, while the other is more associated with environmental vulnerability during rainfall events, with a strong link to turbidity. A single *E. coli* count does not indicate the underlying mechanism leading to this contamination, which multivariate analysis can reveal by distinguishing factorial axes, each associated with a series of factors that may influence contamination. Knowing this, quality monitoring can be optimized by focusing on the most vulnerable lithological/land-use sectors or on wells vulnerable during specific periods—namely after rainfall events likely to generate runoff leading to groundwater contamination. Quantifying and mapping these factors separately provides valuable support when making decisions about which protective measures should be implemented.

The automatic classification of parameters, based on the majority of the information, reveals distinct parameter families and redundancies within these families. For instance, *E. coli* and *Enterococcus* provide entirely redundant information. The same is true, to a lesser extent, for major ions. However, caution must be exercised regarding fecal contamination parameters, which may show redundancy within a given dataset, but not universally. They nonetheless exhibit different survival rates and transport properties, and regulations often require their concurrent monitoring.

Therefore, this study confirms that for a regional-scale assessment of multiparameter water quality, several simplification options are available. One can either work with a few principal components (PCs), each synthesizing the information from several parameters and isolating the different mechanisms involved, or by selecting a single representative parameter from a group of parameters that are highly correlated with it.

This approach thus streamlines the regional-scale study of multiparameter water quality data for water intended for human consumption.

Author Contributions: Conceptualization, M.S. and V.V.; methodology, V.V., L.B., and M.E.J.; software, M.E.J., M.T. and A.B. (Abderrahim Bousouis); validation, M.E.J., M.A., Z.Z., M.G. and A.A.B.; formal analysis, M.E.J.; investigation, M.E.J.; resources, L.B. and M.G.; data curation, V.V., M.G. and L.B.; writing—original draft preparation, M.E.J.; writing—review and editing, L.B. and V.V.; visualization, M.E.J.; supervision, A.B. (Abdelhak Bouabdli), M.S. and V.V.; project administration, A.B. (Abdelhak Bouabdli) and M.S.; 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 datasets presented in this article are not easily accessible for reasons of sensitivity to possible malicious acts. Requests for access to the datasets should be addressed to the Health Agency ARS of the Nouvelle-Aquitaine region.

Acknowledgments: We are deeply grateful to the Regional Health Agency of the Nouvelle-Aquitaine region for their assistance during the data extraction process.

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

References

1. 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](#)]
2. Jakeman, A.J.; Barreteau, O.; Hunt, R.J.; Rinaudo, J.-D.; Ross, A.; Arshad, M.; Hamilton, S. *Integrated Groundwater Management: Concepts, Approaches and Challenges*, 1st ed.; Jakeman, A.J., Barreteau, O., Hunt, R.J., Rinaudo, J.-D., Ross, A., Eds.; Springer: Heidelberg, Germany, 2016; ISBN 978-3-319-23576-9.
3. Priyan, K. Issues and Challenges of Groundwater and Surface Water Management in Semi-Arid Regions. In *Groundwater Resources Development and Planning in the Semi-Arid Region*; Pande, C.B., Moharir, K.N., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 1–17, ISBN 978-3-030-68124-1.
4. Syafiuddin, A.; Boopathy, R.; Hadibarata, T. Challenges and Solutions for Sustainable Groundwater Usage: Pollution Control and Integrated Management. *Curr. Pollut. Rep.* **2020**, *6*, 310–327. [[CrossRef](#)]

5. Closas, A.; Villholth, K.G. Groundwater Governance: Addressing Core Concepts and Challenges. *WIREs Water* **2020**, *7*, e1392. [CrossRef]
6. 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]
7. Li, S.; Wang, Z.; Lai, C.; Lin, G. Quantitative Assessment of the Relative Impacts of Climate Change and Human Activity on Flood Susceptibility Based on a Cloud Model. *J. Hydrol.* **2020**, *588*, 125051. [CrossRef]
8. Kemper, K.E. Groundwater—From Development to Management. *Hydrogeol. J.* **2004**, *12*, 3–5. [CrossRef]
9. 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]
10. Halwani, D.A.; Jurdi, M.; Abu Salem, F.K.; Jaffa, M.A.; Amacha, N.; Habib, R.R.; Dhaini, H.R. Cadmium Health Risk Assessment and Anthropogenic Sources of Pollution in Mount-Lebanon Springs. *Expo. Health* **2020**, *12*, 163–178. [CrossRef]
11. Boente, C.; Matanzas, N.; García-González, N.; Rodríguez-Valdés, E.; Gallego, J.R. Trace Elements of Concern Affecting Urban Agriculture in Industrialized Areas: A Multivariate Approach. *Chemosphere* **2017**, *183*, 546–556. [CrossRef]
12. Wu, J.; Li, P.; Wang, D.; Ren, X.; Wei, M. Statistical and Multivariate Statistical Techniques to Trace the Sources and Affecting Factors of Groundwater Pollution in a Rapidly Growing City on the Chinese Loess Plateau. *Hum. Ecol. Risk Assess. An. Int. J.* **2020**, *26*, 1603–1621. [CrossRef]
13. Orecchia, C.; Giambastiani, B.M.S.; Greggio, N.; Campo, B.; Dinelli, E. Geochemical Characterization of Groundwater in the Confined and Unconfined Aquifers of the Northern Italy. *Appl. Sci.* **2022**, *12*, 7944. [CrossRef]
14. Chilès, J.; Delfiner, P. *Geostatistics: Modeling Spatial Uncertainty*, 1st ed.; Wiley: Toronto, ON, Canada, 1999.
15. Schiavo, M.; Giambastiani, B.M.S.; Greggio, N.; Colombani, N.; Mastrocicco, M. Geostatistical Assessment of Groundwater Arsenic Contamination in the Padana Plain. *Sci. Total Environ.* **2024**, *931*, 172998. [CrossRef]
16. Luque-Espinar, J.A.; Chica-Olmo, M. Impacts of Anthropogenic Activities on Groundwater Quality in a Detritic Aquifer in SE Spain. *Expo. Health* **2020**, *12*, 681–698. [CrossRef]
17. Gran-Aymeric, L. Un Portail National Sur La Qualite Des Eaux Destinees a La Consommation Humaine. *Tech. Sci. Méthodes* **2010**, *12*, 45–48. [CrossRef]
18. Chery, L.; Laurent, A.; Vincent, B.; Tracol, R. *Echanges SISE-Eaux/ADES: Identification Des Protocoles Compatibles Avec Les Scénarios d'échange SANDRE*; SANDRE: Orléans, France, 2011.
19. 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]
20. 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]
21. 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]
22. Mohsine, I.; Kacimi, I.; Valles, V.; Leblanc, M.; El Mahradi, 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]
23. Ayach, M.; Lazar, H.; Lamat, C.; Bousouis, A.; Touzani, M.; El Jarjini, Y.; Kacimi, I.; Valles, V.; Barbiero, L.; Morarech, M. Groundwaters in the Auvergne-Rhône-Alpes Region, France: Grouping Homogeneous Groundwater Bodies for Optimized Monitoring and Protection. *Water* **2024**, *16*, 869. [CrossRef]
24. 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. [CrossRef]
25. 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]
26. Bousouis, A.; Bouabdli, A.; Ayach, M.; Lazar, H.; Ravung, L.; Valles, V.; Barbiero, L. Discrimination of Spatial and Temporal Variabilities in the Analysis of Groundwater Databases: Application to the Bourgogne-Franche-Comté Region, France. *Water* **2025**, *17*, 384. [CrossRef]
27. 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]
28. Gaillard, T.; Moreau, M.; Mari, J.-L. Seismic and Stratigraphic Characterization of Karstogenic Horizons in a Sequence of Carbonate Deposits: Example of the Dogger Limestones of the Poitou Threshold. *E3S Web Conf.* **2024**, *504*, 05005. [CrossRef]
29. DRAAF L'occupation Des Sols Agricoles En Nouvelle- Aquitaine Entre 1970 et 2020. Available online: https://draaf.nouvelle-aquitaine.agriculture.gouv.fr/IMG/pdf/agrestena_etudes_48_avril2024_assolements_v7.pdf (accessed on 7 August 2025).

30. Pouey, J.; Galey, C.; Chesneau, J.; Jones, G.; Franques, N.; Beaudeau, P.; Groupe des Référents Régionaux EpiGEH; Mouly, D. Implementation of a National Waterborne Disease Outbreak Surveillance System: Overview and Preliminary Results, France, 2010 to 2019. *Eurosurveillance* **2021**, *26*, 2001466. [[CrossRef](#)]
31. Beaudeau, P.; Pascal, M.; Mouly, D.; Galey, C.; Thomas, O. Health Risks Associated with Drinking Water in a Context of Climate Change in France: A Review of Surveillance Requirements. *J. Water Clim. Change* **2011**, *2*, 230–246. [[CrossRef](#)]
32. 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](#)]
33. Madhulatha, T.S. An Overview on Clustering Methods. *IOSR J. Eng.* **2012**, *2*, 719–725. [[CrossRef](#)]
34. Bouguettaya, A.; Yu, Q.; Liu, X.; Zhou, X.; Song, A. Efficient Agglomerative Hierarchical Clustering. *Expert. Syst. Appl.* **2015**, *42*, 2785–2797. [[CrossRef](#)]
35. Webster, R.; Olliver, M.A. *Geostatistics for Environmental Scientists*, 2nd ed.; John Wiley & Sons: Chichester, UK, 2007; p. W11416, ISBN 9780470028582.
36. Bárdossy, A. Copula-Based Geostatistical Models for Groundwater Quality Parameters. *Water Resour. Res.* **2006**, *42*, W11416. [[CrossRef](#)]
37. Cressie, N. The Origins of Kriging. *Math. Geol.* **1990**, *22*, 239–252. [[CrossRef](#)]
38. Ayach, M.; Lazar, H.; Bousouis, A.; Touiouine, A.; Kacimi, I.; Valles, V.; Barbiero, L. Multi-Parameter Analysis of Groundwater Resources Quality in the Auvergne-Rhône-Alpes Region (France) Using a Large Database. *Resources* **2023**, *12*, 143. [[CrossRef](#)]
39. Batiot, C.; Emblanch, C.; Blavoux, B. Carbone Organique Total (COT) et Magnésium (Mg²⁺): Deux Traceurs Complémentaires Du Temps de Séjour Dans l’aquifère Karstique. *Comptes Rendus Geosci.* **2003**, *335*, 205–214. [[CrossRef](#)]
40. Garry, B.; Blondel, T.; Emblanch, C.; Sudre, C.; Bilgot, S.; Cavaillou, A.; Boyer, D.; Auguste, M. Contribution of Artificial Galleries to the Knowledge of Karstic System Behaviour in Addition to Natural Cavern Data. *Int. J. Speleol.* **2008**, *37*, 75–82. [[CrossRef](#)]
41. Pacheco, F.; van der Weijden, C.H. Contributions of Water-Rock Interactions to the Composition of Groundwater in Areas with a Sizeable Anthropogenic Input: A Case Study of the Waters of the Fundão Area, Central Portugal. *Water Resour. Res.* **1996**, *32*, 3553–3570. [[CrossRef](#)]
42. Bertran, P.; Allenet, G.; Gé, T.; Naughton, F.; Poirier, P.; Goñi, M.F.S. Coversand and Pleistocene Palaeosols in the Landes Region, Southwestern France. *J. Quat. Sci.* **2009**, *24*, 259–269. [[CrossRef](#)]
43. 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](#)]
44. Abbas, A.; Baek, S.; Silvera, N.; Soulileuth, B.; Pachepsky, Y.; Ribolzi, O.; Boithias, L.; Cho, K.H. In-Stream Escherichia Coli Modeling Using High-Temporal-Resolution Data with Deep Learning and Process-Based Models. *Hydrol. Earth Syst. Sci.* **2021**, *25*, 6185–6202. [[CrossRef](#)]
45. Boithias, L.; Choisy, M.; Souliyaseng, N.; Jourden, M.; Quet, F.; Buisson, Y.; Thammahacksa, C.; Silvera, N.; Latsachack, K.; Sengtaheuanghoung, O.; et al. Hydrological Regime and Water Shortage as Drivers of the Seasonal Incidence of Diarrheal Diseases in a Tropical Montane Environment. *PLoS Negl. Trop. Dis.* **2016**, *10*, e0005195. [[CrossRef](#)]
46. Ogawa, Y.; Nishikawa, M.; Nakasugi, O.; Ii, H.; Hirata, T. Determination of the Abundance of ΔN in Nitrate Ion in Contaminated Groundwater Samples Using an Elemental Analyzer Coupled to a Mass Spectrometer. *Analyst* **2001**, *126*, 1051–1054. [[CrossRef](#)]
47. Haack, S.K.; Fogarty, L.R.; Wright, C. Escherichia Coli and Enterococci at Beaches in the Grand Traverse Bay, Lake Michigan: Sources, Characteristics, and Environmental Pathways. *Environ. Sci. Technol.* **2003**, *37*, 3275–3282. [[CrossRef](#)] [[PubMed](#)]
48. Soupier, M.L.; Mostaghimi, S.; Dillaha, T. Attachment of Escherichia Coli and Enterococci to Particles in Runoff. *J. Environ. Qual.* **2010**, *39*, 1019–1027. [[CrossRef](#)]
49. Pandey, P.K.; Kass, P.H.; Soupier, M.L.; Biswas, S.; Singh, V.P. Contamination of Water Resources by Pathogenic Bacteria. *AMB Express* **2014**, *4*, 51. [[CrossRef](#)]
50. Farrell, C.; Hassard, F.; Jefferson, B.; Leziart, T.; Nocker, A.; Jarvis, P. Turbidity Composition and the Relationship with Microbial Attachment and UV Inactivation Efficacy. *Sci. Total Environ.* **2018**, *624*, 638–647. [[CrossRef](#)] [[PubMed](#)]
51. Peyraud, J.-L.; Cellier, P.; Aarts, F.; Béline, F.; Bockstaller, C.; Bourblanc, M.; Delaby, L.; Dourmad, J.Y.; Dupraz, P.; Durand, P.; et al. Nitrogen Flows and Livestock Farming: Lessons and Perspectives. *Adv. Anim. Biosci.* **2014**, *5*, 68–71. [[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.