Groundwater

Inspecting the Reliability of Geochemical Facies Identified for the Waterworks' Capture Zone in Germany

by Majid Taie Semiromi[1], Steven Böttcher[2], and Christoph Merz[2,3]

Abstract

Little research attention has been given to validating clusters obtained from the groundwater geochemistry of the waterworks' capture zone with a prevailing lake-groundwater exchange. To address this knowledge gap, we proposed a new scheme whereby Gaussian finite mixture modeling (GFMM) and Spike-and-Slab Bayesian (SSB) algorithms were utilized to cluster the groundwater geochemistry while quantifying the probability of the resulting cluster membership against each other. We applied GFMM and SSB to 13 geochemical parameters collected during different sampling periods at 13 observation points across the Barnim Highlands plateau located in the northeast of Berlin, Germany; this included 10 observation wells, two lakes, and a gallery of drinking production wells. The cluster analysis of GFMM yielded nine clusters, either with a probability $\geq$0.8, while the SSB produced three hierarchical clusters with a probability of cluster membership varying from $<$0.2 to $>$0.8. The findings demonstrated that the clustering results of GFMM were in good agreement with the classification as per the principal component analysis and Piper diagram. By superimposing the parameter clustering onto the observation clustering, we could identify discrepancies that exist among the parameters of a certain cluster. This enables the identification of different factors that may control the geochemistry of a certain cluster, although parameters of that cluster share a strong similarity. The GFMM results have shown that from 2002, there has been active groundwater inflow from the lakes towards the capture zone. This means that it is necessary to adopt appropriate measures to reverse the inflow towards the lakes.

Introduction

Establishing a rigorous advanced warning system for a sustainable groundwater management strategy is mandatory, which means it is important to enhance the current understanding of spatiotemporal groundwater geochemical evolution and its controlling factors (Liu et al. 2015). In particular, this requires directing attention to the groundwater geochemistry dataset associated with the capture zone of waterworks where drinking water for domestic consumption is produced.

The evolution of groundwater geochemical facies and their spatiotemporal variability have mainly been characterized by natural processes (e.g., lithology, water-rock interactions, and hydrogeological conditions), and anthropogenic activities (e.g., mining, agriculture, industry, construction of dams, and groundwater overdrafts (Hadj et al. 2014; Carucci et al. 2012; Güler et al. 2012; Sbarbati et al. 2015; He et al. 2012; Wang et al. 2013; Golian et al. 2019; Ostad-Ali-Askari et al. 2019); Taie Semiromi and Koch 2020). The geochemical facies of groundwater also tend to vary over time and space (Yang et al. 2020).

Groundwater geochemical datasets are intrinsically multivariate; this means that each groundwater parameter represents the physical, chemical, and biological properties that have been imprinted on the parameter (Liu et al. 2021). Therefore, delineating the spatiotemporal evolution of groundwater geochemistry of a dataset requires the implementation of cluster analysis, in which the dataset is split into several clusters. Indeed, observations/samples of a certain cluster illustrate the

[1]Corresponding author: Leibniz Centre for Agricultural Landscape Research (ZALF), Working Group "Lowland Hydrology and Water Management", Eberswalder Straße 84, 15374 Müncheberg, Germany; majid.taie@zalf.de

[2]Leibniz Centre for Agricultural Landscape Research (ZALF), Working Group "Lowland Hydrology and Water Management", Eberswalder Straße 84, 15374, Müncheberg, Germany

[3]Hydrogeology Group, Institute of Geological Sciences, Freie Universität Berlin, Malteserstr. 74-100, 12249, Berlin, Germany

Article impact statement: The probability of cluster membership quantified using an algorithm should be validated against another probabilistic-based classifier.

Received May 2021, accepted December 2021.

doi: 10.1111/gwat.13168

greatest similarity, while the similarity among clusters is minimal; this demonstrates the dominant controlling processes shaping each cluster (Güler and Thyne 2004a, 2004b; Pacheco Castro et al. 2018; Pant et al. 2018; Templ et al. 2008).

Clustering methods may be broadly classified into two groups: (1) hierarchical clustering methods, which consists of agglomerative and divisive methods; and (2) partitioning clustering methods, which consists of distance-based, model-based, and density-based approaches (Fraley and Raftery 1998). A review of clustering algorithms and their development was discussed by Saxena et al. (2017). These approaches have gained increasing popularity to identify hydrogeochemical facies and detect their spatiotemporal evolution over time (Kim et al. 2003; Simeonov et al. 2003; Shrestha and Kazama 2007; Cloutier et al. 2008; Nguyen et al. 2015; Wang et al. 2015).

Clustering is considered a propitious approach that has been widely used in several multivariate analyses to delineate spatiotemporal patterns embedded in a groundwater geochemistry dataset (e.g., Cloutier et al. 2008; Güler and Thyne 2004a, b; Yang et al. 2020; Papaioannou et al. 2010; Woocay and Walton 2008). However, the validity of procured geochemical clusters has largely remained a formidable challenge because clustering is a subjective process (Dougherty and Brun 2004).

To address this issue, clustering methods with probabilistic outputs may be very useful when an observation is assigned into a cluster while its membership coefficient is quantified in terms of probability (Aguilera et al. 2013). Regardless, relying on one specific probabilistic approach to determine geochemical facies and its controlling factors for hydrologically complex catchments may not be reliable. This is because the estimated probability of membership for a specific cluster may have considerable uncertainty. This is particularly the case for the capture zone of waterworks where the monthly and seasonal water level fluctuations of lakes, groundwater, and streamflow within the capture zone may provoke a movement in the hydrological boundary conditions. This movement influences the groundwater dynamics including the reversal of previously established hydraulic gradients and subsequent alterations in the geochemical environment (Ferone and Devito 2004; Vepraskas et al. 2020).

The paucity of stratigraphic knowledge for most aquifers across the world and the anthropogenic impacts from activities such as water extraction for drinking water can cause unexpected effects on surface water and groundwater quality and quantity. These effects add to the complexity of the dataset of interest, increasing the uncertainty associated with cluster identification.

Thus, this serves as the point of departure for current research through which we have aimed to appraise a probabilistic classifier to cluster the geochemical dataset of the waterworks' capture zone. It also seeks to concurrently validate the obtained clusters against that of another probabilistic classifier; such a comparison has been poorly documented in the literature. To address this knowledge gap, two algorithms previously not used for clustering a groundwater geochemical dataset were selected; Gaussian finite mixture modeling (GFMM) (Scrucca et al. 2016) and spike-and-slab Bayesian model (SSB) (Partovi Nia and Davison 2012). Successful applications of GFMM (e.g., Ellefsen et al. 2014; Ellefsen and Smith 2016; Scrucca 2016; Marbac et al. 2017; Popp et al. 2019; Saranya et al. 2020; Zhou and Wang 2020), and SSB (e.g., Tadesse et al. 2005; Partovi Nia 2009; Partovi Nia and Davison 2012; Anderson and Vehtari 2017; Canale et al. 2017; Cao et al. 2019; Bai et al. 2021), have been widely documented for datasets other than those relating to groundwater geochemistry.

Consequently, clusters representing the highest probability of occurrence may be considered for future water quality management of waterworks while downweighing clusters with the lowest probability of occurrence. It is expected that the risk of failure for precautionary and protection measures in line with appropriate water quality management is reduced (Aguilera et al. 2013) when using highly reliable groundwater geochemical clusters.

This study will also examine the complementary implications of these two probabilistic clustering approaches, as one is fed with replicated observations (i.e., SSB) and the other is input with all individual observations (i.e., GFMM) when used together for a specific hydrogeochemical dataset. The suitability of this comparison will be assessed for hydrogeochemical parameters associated with 13 groundwater observations, two lakes, and a gallery of drinking water production wells. The residents in the northeast of Berlin are highly dependent on the capture zone of the waterworks; as such, subtle evidence signifying an early warning to water quality endangerment may be decoded by taking advantage of the possible complementary outcomes of these algorithms.

Study Area

The study area is located approximately 20 km northeast of Berlin, Germany (Figure 1a) on the Barnim Highlands plateau. The topography is characterized by a hummocky landscape of gently rolling hills intersected by a glacial valley that is slightly southward sloping. The valley contains a small stream, the Fredersdorfer Mühlenfließ, that passes through two lakes (i.e., Fängersee [Fae] and Bötzsee [Boe]), in the southern part of the area. At the outlet of the downstream lake (Boe, Figure 2), the annual mean discharge has declined by 50% over the last three decades, from 0.3 m^3/s in 1980 to 0.15 m^3/s in 2009. In addition to climatic and land-use changes, the local waterworks have potentially influenced the hydrology of the region in recent years (Atlas 2018). The annual precipitation in the region for 1951 to 2012 varies between 345 and 794 mm/year. Annual atmospheric water balances of the region show a mean annual deficit of approximately 82 mm/year for the 1981 to 2006 period (Germer, et al. 2011). The dominant land-use type in the catchment is a mixed forest where approximately 69% is dominated by pines. The remaining area almost

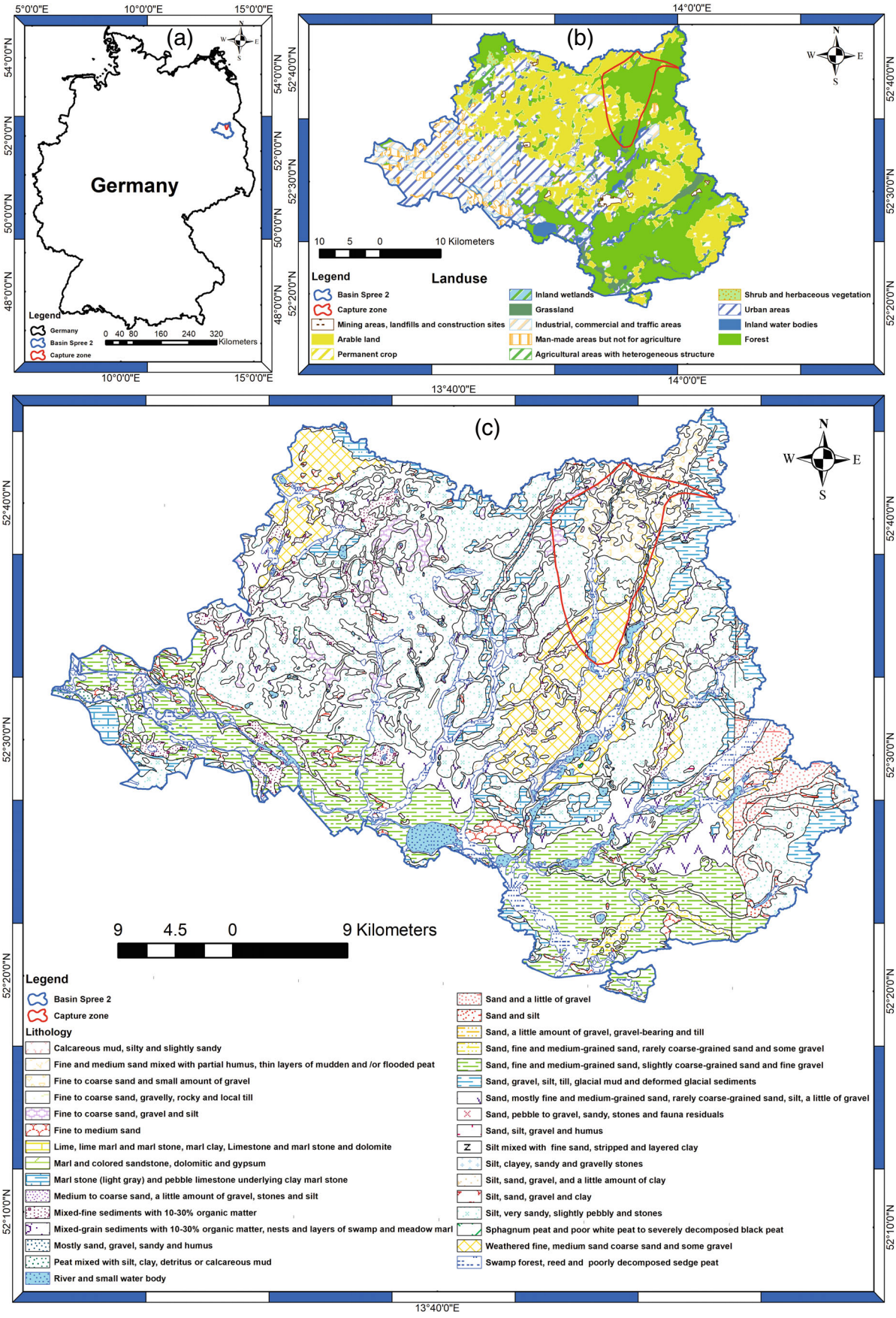

Figure 1. The study area: (a) map of Germany and position of the application area; (b) land use/land cover of the Spree 2 basin and the capture zone of waterworks; and (c) the lithology of the Spree 2 basin and the capture zone.

exclusively consists of agriculture, while other patches of residential land-use constitute less than 1% of the study area (Figure 1b).

Waterworks located in the catchment have been operational since 1977. Groundwater is extracted from a 1.5 km gallery in length, and is comprised of 12 production wells (herein referred as WW); the gallery extends from the Northwest to Southeast (Figure 2). The mean daily production between 2011 and 2013 was 4475 m^3/d, with a range from 1622 to 10,751 m^3/d. Due to elevated levels of urbanization in the peri-urban zone of Berlin, the production capacity of waterworks has been increased, including the expansion of the gallery in 2014. According to the official hydrogeological map of Brandenburg (State Office for Mining Geology and Raw Material of Brandenburg 2012), the capture zone spans a 93 km^2 area. Traditionally, fixed-radius and analytical approaches have been used to delimit a capture zone (Frind and Molson 2018). However, advanced techniques have come to the fore; this includes techniques such as the particle tracking method in which a groundwater model is built to configure the flow system and track particles along flow lines. The capture zone of these waterworks was delineated using long-term groundwater head data by the State Office for Mining Geology and Raw Material of Brandenburg (2012). Although the capture zone states the overall extent of the study area, this study was particularly focused on the vicinity of waterworks (see Figure 2). Lake Bötzsee (Boe) was located approximately 200 m in the west of the gallery with a length of 2.8 km; it extends in the north-south direction, with a maximum width and depth of 450 and 13.6 m, respectively. Lake Fängersee (Fae) is located approximately 400 m upstream of WW, and is approximately half the size of Boe, with a maximum known depth of 5.5 m (Figure 2).

The geology in the study area is formed by a series of layered Pleistocene and Tertiary sediments that are approximately 150 to 200 m thick, with a lower confining bed of Oligocene marine Rupel clay. The series consists of a complex interplay of glacial deposits from the Pleistocene and permeable marine and limnic sediments of the Upper Oligocene and Miocene. The series may be divided into an upper unconfined system of shallow Weichselian and late Saalian sediments. Underneath these sediments is a thick confined aquifer system of the early Saalian and Elster layers, and Upper Oligocene and Miocene sediments.

In general, a shallow (i.e., 5 to 10 m) unconfined aquifer is separated from the thick (140 to 190 m) lower confined aquifer by a 15 to 20 m thick layer of Saalian sediments. The stratification is known to be convoluted and disturbed towards the northern part of the capture zone (State Office for Mining Geology and Raw Material of Brandenburg 2012). The confined and unconfined aquifers consist of multiple permeable sediment layers partially disconnected by layers of till, which are thus hydraulically connected.

In the northeast of Germany, groundwater levels and landscape runoff have largely been in decline for over three decades (Suckow et al. 2002; Lahmer 2003; Germer et al. 2011; Merz and Pekdeger 2011); regional climate studies suggest further decreases over the next decades (Gerstengarbe et al. 2003, 2013; Held et al. 2013). Thus, water resource management for this region requires a thorough assessment of possible adaptions and measures to counteract or mitigate severe consequences, such as decreasing groundwater heads and surface water levels and declining groundwater and surface water quality.

Methods

Sampling and Analytical Procedures

There were 25 monitoring wells located in the vicinity of the production gallery; of these, 10 wells were selected based on the distance and direction from the waterworks and depths (Table S1, Supporting Information). All observation wells were assumed to be hydraulically affected by water extraction (Böttcher et al. 2014) (Figure 2).

Sampling took place from September 2011 to August 2013 at bimonthly intervals. A total of 131 samples were analyzed from 10 groundwater observation wells and two lakes. The temporal resolution of sampling for all observation points is provided in Table S2. Groundwater sampling was carried out after a minimum pumping duration of 45 min until geochemical parameters remained stable; a Grundfoss BMI/MP1-230 V immersion pump was used to sample groundwater.

The pH, redox potential, dissolved oxygen, electrical conductivity (EC), and temperature were measured in the field during sampling. Samples were filtered using 0.22 μm membrane filters to exclude suspended solids such as precipitated iron (Fe^{-}) and manganese (Mn^{-}) (hydr)oxides and colloids. Samples taken for cation analysis were preserved in concentrated nitric acid (HNO_3^-).

Data Preprocessing

The 13 hydrochemical parameters used as inputs for the clustering algorithms were: pH, EC, dissolved organic carbon (DOC), Na^+, K^+, Mg^{2+}, Mn^{2+}, Fe^{2+}, Ca^{2+}, Cl^-, HCO_3^-, PO_4^{3-}, and SO_4^{2-}.

All samples were checked for ionic balance and excluded when the charge balance error exceeded 5%; a 131 × 13 data matrix remained viable for data analysis. The ranges of all dataset variables after preprocessing are provided in Table S1 for each site. Samples were distributed in a relatively similar manner among all sites with at least 10 samples per site, with the exception of G23; this site was included at a later stage of the monitoring program and contributed only five samples.

To have the equal effect of variables on the clustering approach, the 131 samples were standardized to mean zero and variance one by subtracting each sample variable from

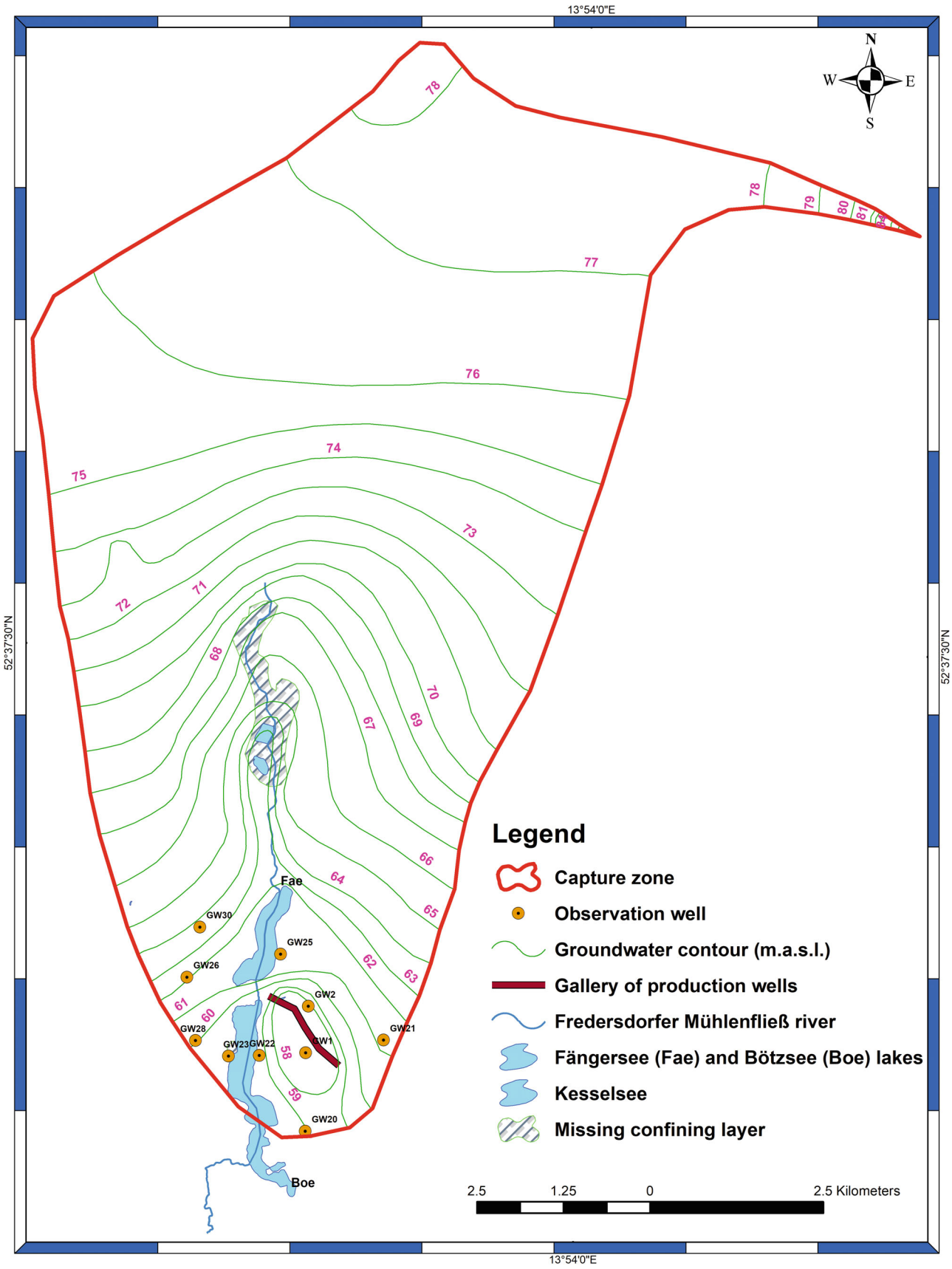

Figure 2. Map of the capture zone of waterworks in the southern part of the study area along with the Fredersdorfer Mühlenfließ river and the two lakes (Fängersee [Fae] and Bötzsee [Boe]) (Ministry of Environment Health and Consumer Protection of Brandenburg 2009; Böttcher et al. 2014).

its mean and dividing by its standard deviation (Iwamori et al. 2017).

Clustering Approaches

GFMM Clustering Algorithm

Finite normal mixture models as a model-based clustering method with probabilistic outputs are the most common for the clustering of a wide range of datasets (e.g., Marbac et al. 2017; Saranya et al. 2020; Zhou and Wang 2020).

Assuming an aquifer system with two regions that have different geochemical characteristics, the geochemical parameters of each region are characterized by a specific probability density function (PDF) (Ellefsen et al. 2014). The PDF of the geochemical properties of region 1 is: $f(Z\,|\,\theta_1)$; θ_1 shows the parameters of that specific PDF, and Z is the value or concentration of each geochemical element. Notably, geochemical elements have already been through the transformation procedure, being standardized to mean zero and variance one. The parameters of the PDF for the geochemical properties of region 2 are represented by θ_2. Therefore, when a certain sample denoted by i, falls within region 1, the resultant $f(Z_i\,|\,\theta_1)$ will normally be large-valued, while $f(Z_i\,|\,\theta_2)$ will be small-valued. Conversely, when sample, i, falls within region 2, this relationship is reversed (Ellefsen et al. 2014).

Thus, the PDF for the entire aquifer $p(Z)$ was a weighted summation of the two PDFs for the two regions, calculated as: $p(Z) = \lambda_1 f(Z\,|\,\theta_1) + \lambda_2 f(Z\,|\,\theta_2)$. The coefficients, λ_1 and λ_2, are the areas of regions 1 and 2, respectively, divided by the area of the entire aquifer. Each weight/coefficient is the relative contribution of the corresponding PDF to $p(Z)$.

In practice, as there are several regions/sampling areas with associated geochemical properties, the finite mixture model is generalized to J regions. Therefore, we have:

$$p(Z) = \sum_{j=1}^{J} \lambda_j f\left(Z\,|\,\theta_j\right) \tag{1}$$

The PDF $p(Z)$ should be considered a mixture of J PDFs, each reflecting the geochemical properties of a specific region within the entire study area (e.g., aquifer). The weights fall within the range: $0 \le \lambda_j \le 1$ and $\sum_{j=1}^{J} \lambda_j = 1$ (McLachlan and Peel 2000). As a result, a finite mixture model is considered to be able to appropriately characterize the mathematical representation of the geochemical properties of a specific application area.

The probability that sample, i, is attributed to PDF, j, is represented by the conditional probability:

$$g_{ij} = \frac{\lambda_j f\left(Z_i\,|\,\theta_j\right)}{p\left(Z_i\right)} \tag{2}$$

where $i = 1, \ldots, n$ where n is the number of samples (Fraley and Raftery 2002). Thus, as g_{ij} increase, sample i becomes increasingly resembled by PDF j. A "cluster" constitutes samples in which $g_{ij} \ge 0.5$ for the PDF j. Accordingly, the number of clusters is identical to the number of PDFs in the finite mixture model; this is the premise for this approach being named "mixture-model clustering" (Ellefsen et al. 2014).

The parameters of the mixture model, θ_j, were unidentified; therefore, they were estimated using the log-likelihood function: $l(\theta; 1, \ldots, n) = \sum_{i=1}^{n} \log\left(f\left(n_i; \theta\right)\right)$. As it is difficult to conduct the direct maximization of the log-likelihood, the Expectation-Maximization (EM) algorithm was utilized to procure the maximum likelihood estimator (MLE) (Dempster et al. 1977; McLachlan and Peel 2000).

One of the most important challenges in clustering is identifying the optimal number of clusters/components. Additionally, the PDF base on which clustering is carried out should be assigned in GFMM. The 14 models were tested in the current study (Table 1), as available in the Mclust package in R (https://cran.r-project.org/web/packages/mclust/mclust.pdf); each model was associated with a PDF that represented a structure type, volume, shape, and orientation. We used Bayesian information criterion (BIC) to specify the optimal number of clusters and the model appropriate for clustering. In GFMM, the Gaussian mixture model dimensional reduction (GMMDR) developed by Luca Scrucca (2010) was used to project observations onto a reduced subspace; in turn, summary plots were able to assist with visualizing the clustering typology. This method was premised on the eigen decomposition of an appropriate kernel matrix with unknown parameters obtained from the number of PDFs in the Gaussian mixture model; this was fitted well to the dataset of interest. Similar to principal component analysis (PCA), transformed observations were referred to as "directions," comparable to principal components; details of classification using GFMM were described by Scrucca et al. (2016).

Spike-and-Slab Bayesian (SSB) Clustering Algorithm

Assigning observations into different clusters in Bayesian clustering is considered a statistical parameter. Thus, we postulated a Bayesian model for the dataset that was conditional on the grouping configuration. To that end, a prior distribution was appointed for the clusters, in which a search algorithm was applied to determine the maximum posteriori grouping. The search algorithm used in this study, as implemented in the bclust package, was an agglomerative search approach because it illustrates a dendrogram which provides a visual assist to other possible groupings.

Based on a dataset with observations assigned to C clusters of sizes, $T_1, \ldots, T_C$, with complete $T = \sum_{c=1}^{C} T_c$ clustering individuals; then, a multinomial-Dirichlet distribution (Heard et al. 2006) as the allocation prior is assumed as follows:

$$f(C) \propto \frac{(C-1)!T_1! \ldots T_c!}{T(T+C-1)!} \tag{3}$$

Table 1
Parameterizations of Multidimensional Data Available in the Mclust Package, and Associated Geometric Properties, as Per Scrucca et al. (2016)

Model	Distribution	Volume	Shape	Orientation
EII	Spherical	Equal	Equal	—
VII	Spherical	Variable	Equal	—
EEI	Diagonal	Equal	Equal	Coordinate axes
VEI	Diagonal	Variable	Equal	Coordinate axes
EVI	Diagonal	Equal	Variable	Coordinate axes
VVI	Diagonal	Variable	Variable	Coordinate axes
EEE	Ellipsoidal	Equal	Equal	Equal
EVE	Ellipsoidal	Equal	Variable	Equal
VEE	Ellipsoidal	Variable	Equal	Equal
VVE	Ellipsoidal	Variable	Variable	Equal
EEV	Spherical	Equal	Equal	Variable
VEV	Spherical	Variable	Equal	Variable
EVV	Diagonal	Equal	Variable	Variable
VVV	Diagonal	Variable	Variable	Variable

$$f(C \mid y) = k^{-1} f(y \mid \mathrm{C}) f(C) \tag{4}$$

where $f(C \mid y)$ is the marginal density of the dataset for known allocation clusters (c) obtained from Equation 5 (Partovi Nia 2009), and $k > 0$ is a fixed value for a given dataset. This value may be disregarded in numerical calculations as it has no effect on agglomerative clustering.

$$f(y) = \prod_{v=1}^{V} f(y_v) \tag{5}$$

where the univariate random variable, y_v, represents the clustering sample dataset T ($T = 1, \ldots, T_c$) in cluster C ($c = 1, \ldots, C$), recorded on the continuous variable, v ($v = 1, \ldots, V$). The marginal density for each variable is a convex combination of the spike-and-slab densities. The spike-and-slab distribution and its parameters are discussed by Partovi Nia and Davison (2012).

To apply SSB in practice, in the beginning, each sample was considered an individual cluster. Thus, the number of samples was equal to that of clusters, meaning that $C = T$ and, as such, the number of samples of cluster, c, was $T_c = 1$ for all, $c = 1, \ldots, C$; then, pairwise merges were applied. To do so, the clustering posterior (Equation 4) was computed and the merge that maximized (Equation 4) was applied. The log posterior, $g_c = logf(C \mid y)$, was the most optimal merge with c clusters; this was adopted to represent the dendrogram height (Partovi Nia and Davison 2012). Based on Equation 4, clusters c_1 and c_2 were to join to build a new cluster c; as a result, $T_c = T_{c1} + T_{c2}$. In doing so, the algorithm continues examining all pairwise merges again and it is proceeded until all clusters are combined; therefore, all samples are within one cluster. The most suitable grouping, determined via the posterior as an objective function in an agglomerative procedure, maximizes g_c over $c = 1, \ldots, C$. The groupings pertaining to g_c were reserved in agglomerative order as clusters (c) is increased; hence, a dendrogram illustration was possible. Although a monotone height function is a requisite to create a dendrogram, g_c is not necessarily monotone. For this reason, a transformation, which is $g_{max} = \max(g_c)$, is applied that assumes that $c_{max} = argmax(g_c)$ is the number of clusters maximizing g_c. When $c > c_{max}$, the dendrogram height becomes negative as per: $h_c = g_{max} - g_c$, and therefore, the formula is revised as: $h_c = g_c - g_{max}$, which is positive. If g_c is unimodal, h_c is monotone by definition; as such, splitting the dendrogram at zero provides the grouping that maximizes g_c. As plotting a dendrogram in R requires having positive heights, h_c is revised as: $h_c - \min(hc)$ (Partovi Nia and Davison 2012).

In this Bayesian clustering approach, the importance of variables for clustering was represented in the Bernoulli random variable, δ_v. Note that the log posterior was utilized as an objective function to identify the optimal number of clusters of SSB.

The details of the SSB model are described by Partovi Nia and Davison (2012). To apply SSB, the "bclust" R package (https://cran.r-project.org/src/contrib/Archive/bclust/) was utilized, as it was developed on basis of the presented approach.

Evaluation of Optimal Clusters Identified by GFMM and SSB

To appraise the optimal cluster numbers identified by BIC and the log-posterior for GFMM and SSB, respectively, we applied 30 indices provided by the NbClust R package (https://cran.r-project.org/web/packages/NbClust/NbClust.pdf). These indices were applied to ascertain the most appropriate number of clusters that may exist in the geochemical dataset. To calculate the 30 indices, we selected "euclidean" to compute the dissimilarity index. Then, according to the majority rule, the cluster number with the highest frequency (i.e., based on the results of these 30 indices), was selected as the optimal clustering number.

Assessment of Clusters Via Standard Piper Diagram and PCA Clusters

To examine whether the clustering results were plausible, we applied the Piper diagram (Piper 1944; Russoniello and Lautz 2020) to clusters of the classification approach (GFMM or SSB) in which the number of clusters was more similar to that identified by the NbClust R package. We deployed Piper to reveal the major cations and anions, thereby establishing the predominant water type. To create Piper, we used the publicly available GW_Chart program (Winston 2000).

The PCA was deployed to further investigate the suitability of the resultant clusters and ascertain the dominant factors underlying the geochemistry of observation points (i.e., groundwater observation wells, lakes, drinking production wells). Similar to the Piper diagram, we used PCA to assess the clusters from the classification

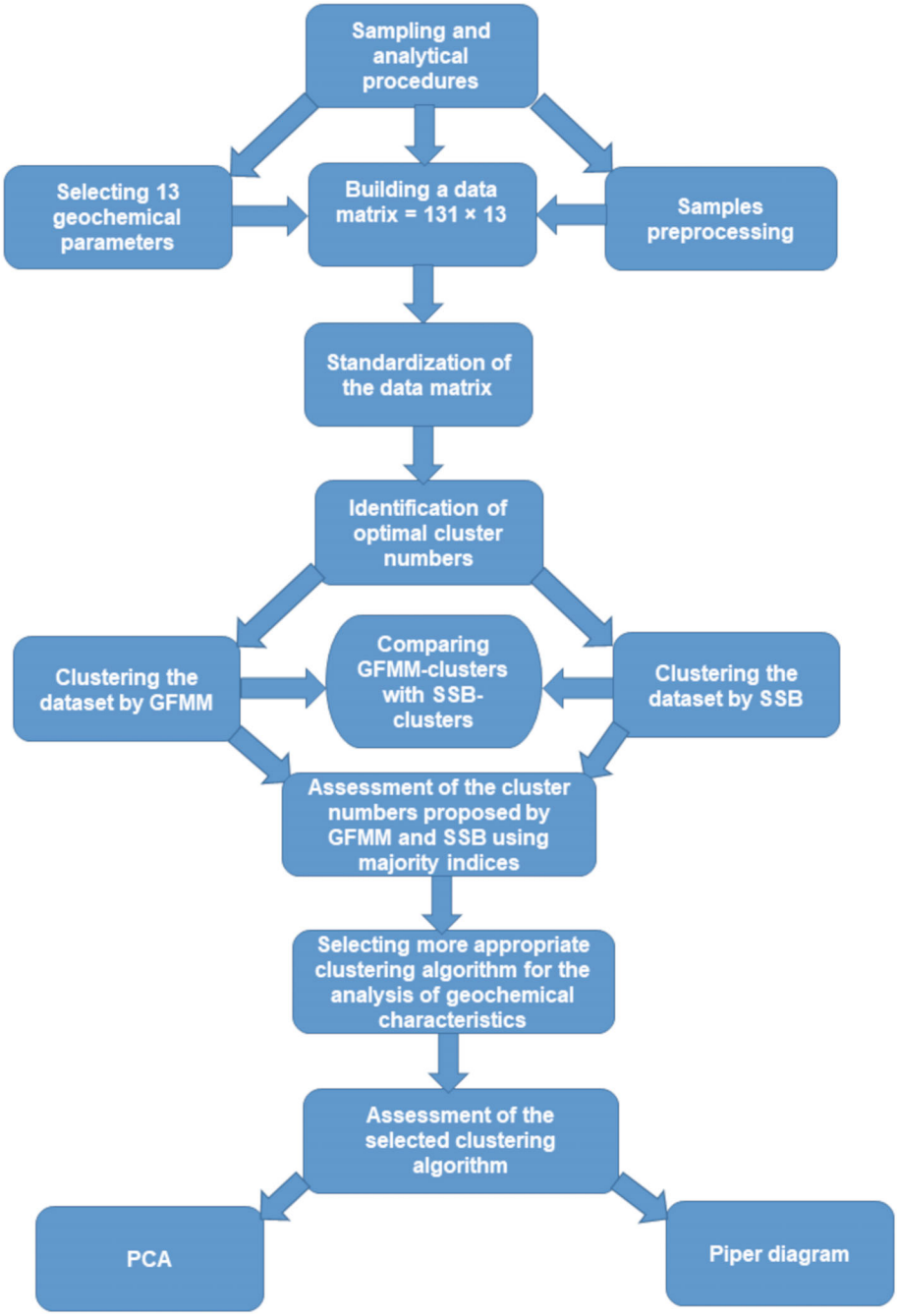

Figure 3. Methodological workflow that commences from sampling to assessing the selected classifier.

approach (GFMM or SSB) in which the number of clusters were more close to that specified by several indices of the NbClust R package.

The methodological steps adopted in the current study were represented using a flowchart (Figure 3).

Results

Ascertaining the Optimal Number of Clusters

The resulting optimal number of clusters are illustrated in Figures 4a-4c; the optimal number of clusters identified via BIC, for GFMM was nine (Figure 4a). According to the highest BIC value, the most suitable model was the "ellipsoidal, equal volume, shape, and orientation" (EEE) (Table 1), as shown by Figure 4a.

The optimal number of clusters identified via maximum of the log posterior for the SSB model was three (Figure 4b).

As the clusters from SSB were presented using a dendrogram, the number of clusters was more subjective (Partovi Nia and Davison 2012) (Figure 7).

Based on the majority indices as implemented in the NbClust R package, five objective functions/indices representing the highest frequency suggest that two and nine are the optimal number of clusters. As samples were

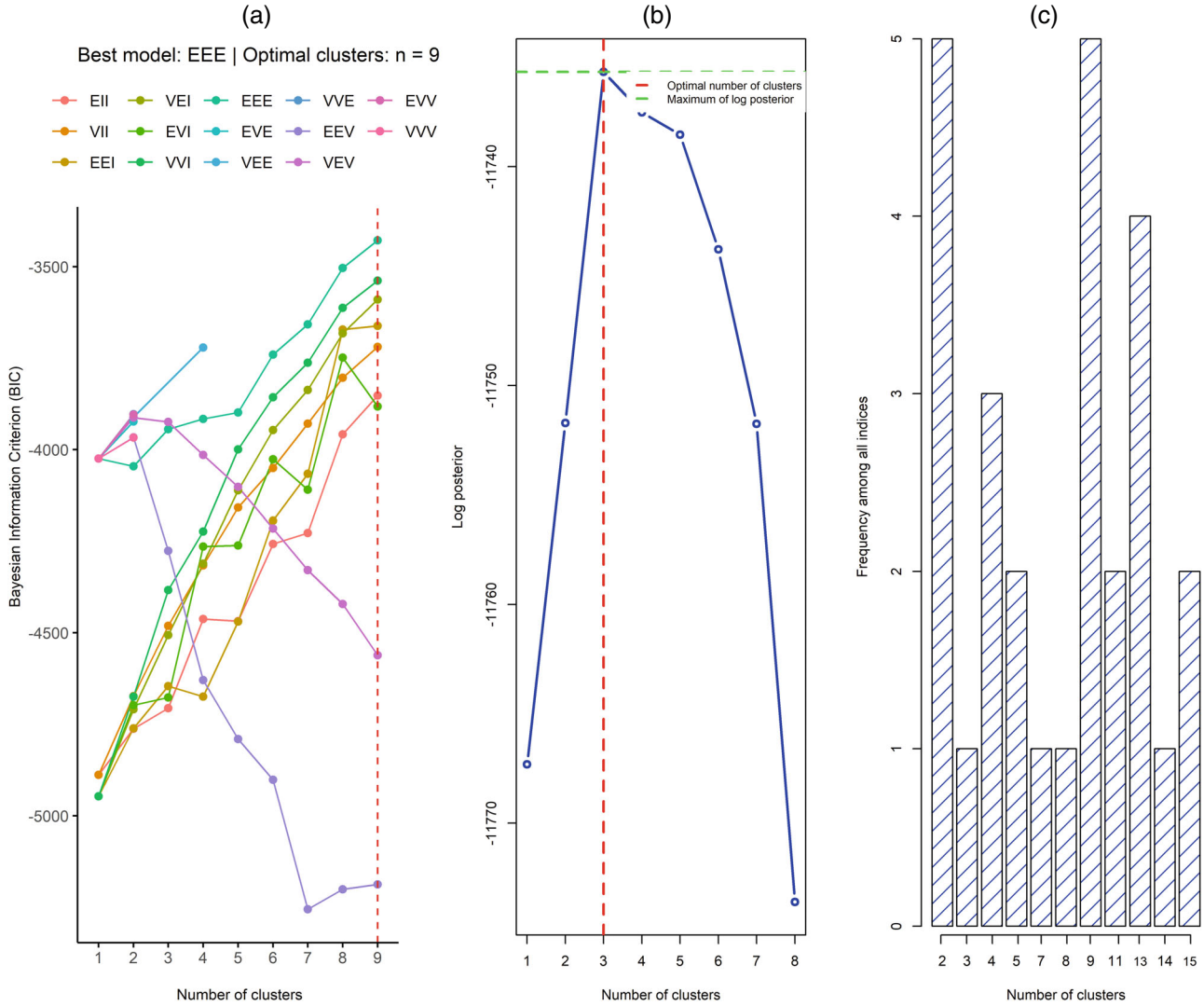

Figure 4. Identification of optimal clusters: (a) the optimal number of clusters and the suitable model (Table 1), identified for the clustering approach using GFMM (note that the best model was identified based on the greatest BIC); (b) the optimal number of clusters for SSB (similarly, note that the optimal number of clusters was based on the greatest log posterior); and (c) the most suitable cluster number specified using the majority approach following Charrad et al. (2014).

collected from different sources (i.e., lake, observation wells, and gallery of production wells), and distributed across the region, we regarded nine as the number of clusters identified using the frequency of majority indices (Figure 4c).

Classification of Observations and Quantification of Uncertainty Using GFMM

Figure 4 shows that the EEE model (Table 1) of the GFMM fitted by the Expectation-Maximization (EM) algorithm yielded nine clusters. As stated earlier, the EEE model may plausibly fit to the mixture components of geochemical parameters (Figure 4a) where −1187.50 and −3428.04 were the log-likelihood and BIC, respectively. Given the 131 observations in the dataset, 13, 10, 6, 42, 9, 17, 12, 11, and 11 of the observations/samples (Table S1), were classified in the clusters of 1 to 9, respectively (Figure 5); in other words, approximately 10%, 7%, 4%, 32%, 6%, 13%, 9%, 8%, and 8% of observations were categorized from C1 to C9, respectively. Therefore, C4 and C3 with 42 and six members comprise the largest and smallest clusters, respectively.

One of the biggest advantages of the GFMM is that it is able to demonstrate the probability assignment of each sample to its most suitable cluster. Interestingly, nearly each observation point corresponded to one cluster and was represented by a strong probability, $P \geq 0.80$ (Figure 5). All observations of GW1, sampled at different time periods (Table S1), made C1; the same was true for observations of GW2, GW26, GW28, GW30, Boe, and Fae that structured the clusters C2, C7, C8, C9, and C5, respectively. However, some observations did not fall into one cluster exclusively. For instance, GW2_11 was structured into C3, while the remainder of observations pertaining to GW2 were categorized as C2. Likewise, GW28_2 fell under C3, while the remaining

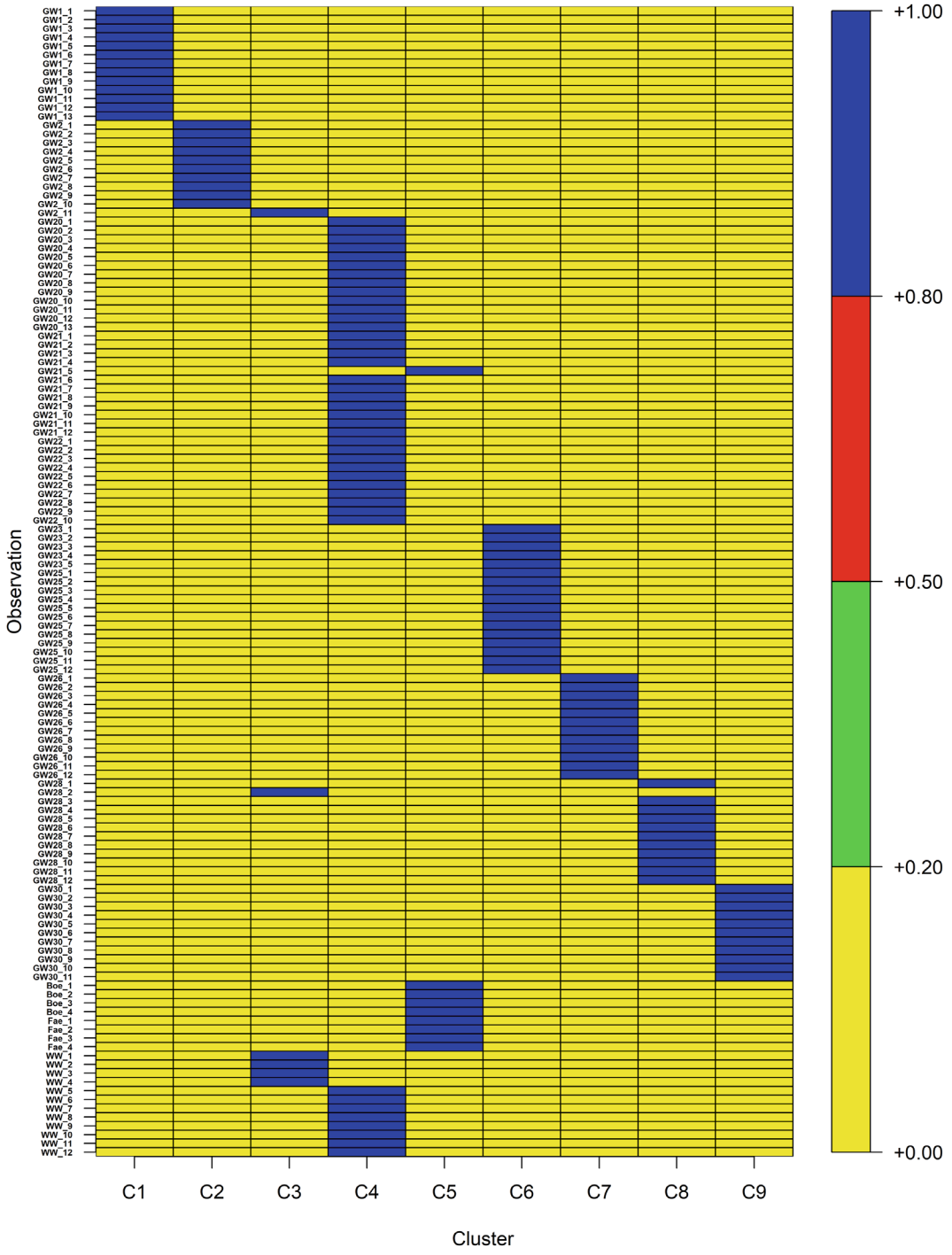

Figure 5. Assignment of sampling points to its highly probable cluster (i.e., clusters 1 to 9). Note that no observation fell into the probability category of 0.20 to 0.50 and 0.50 to 0.80.

observations belonging to GW28 were represented by C8. Moreover, observations for some of the sampling points showed considerable similarity, thus structuring into one cluster. In this regard, C3 was formed by observations from GW2, GW8, and WW_1 to WW_4; C4 was produced by observations from GW20, GW21 (except for GW21_5), GW22, and WW_5 to WW_12, while C6 included all observations from GW23 and GW25.

Figure 6 illustrates a two-dimensional (2D) data plot projected onto the first two directions specified with uncertainty boundaries and data points, that were marked according to the corresponding mixture component. Uncertainty was indicated through a grayscale, whereby darker regions reflected higher uncertainty. In the context of uncertainty boundaries among clusters, it was evident that most of the nine clusters showed a narrow uncertainty boundary where they were satisfactorily distinguished from each other, albeit with a marginal overlap between

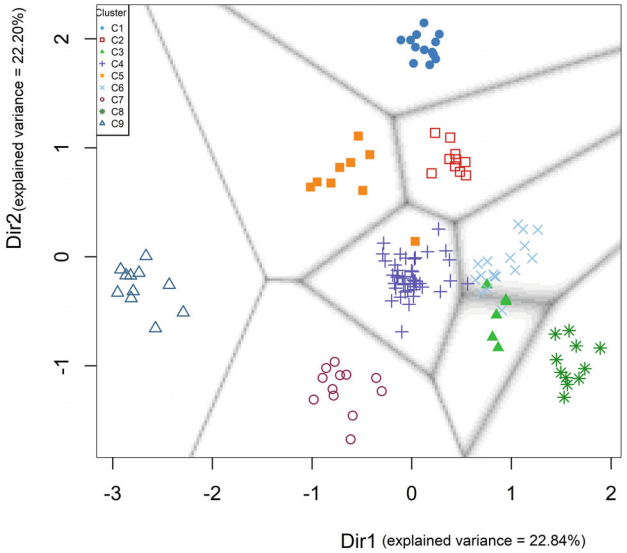

Figure 6. The nine identified clusters of GFMM with uncertainty areas projected onto the first two GMMDR directions (Dir1 and Dir2).

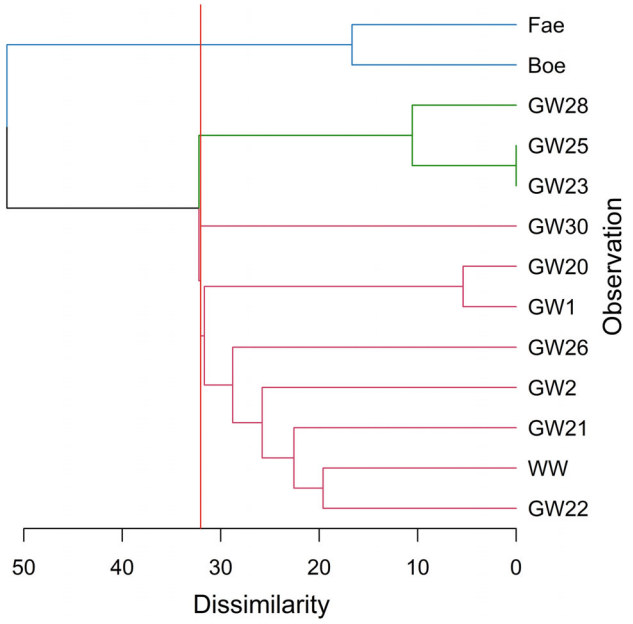

Figure 7. Hierarchical clusters detected using the SSB model.

C3, C4, and C6 (Figure 6). Interestingly, one of the samples of C5 (i.e., GW21_5) was assigned to C4, given the delineated uncertainty boundaries among the clusters; Figure 5 shows that C5 was fully structured with the two lake samples and a sample of an observation well (i.e., GW21_5) (see Table S1).

Discerning Geochemical Clusters Using the SSB Model

According to the dendrogram in Figures 7 and 8 and the phenon line drawn at the dissimilarity index of 32, three clusters (G1 to G3) were identified for the 13 observation points consisting of 131 samples via SSB. Figure 7 shows that observations at the two lakes, Fae and Boe, formed G1 with an almost strong similarity or a low dissimilarity index. The observations of sampling points GW28, GW25, and GW23 structured G2, in which the former did not share a striking similarity with the latter two; GW25 and GW23 exhibited a close similarity as quantified by the dissimilarity index at almost "zero" (Figure 7). The largest cluster, G3, consisted of eight observation points, in which a wide range of similarity/dissimilarity was observed. Notably, this cluster itself was comprised of three sub-clusters. With the exception of GW30 which represents a stand-alone component within the entire cluster, the remaining sampling points were grouped together, albeit with different patterns and numbers. In this regard, GW20 and GW1 were clustered together with a remarkable similarity, while GW26, GW2, GW21, WW, and GW22, presenting with an order of increasing similarity or decreasing dissimilarity, were patterned on one sub-cluster.

In addition to the hierarchical clustering illustrated in Figure 7, an image plot was superimposed onto the dendrogram (Figure 8), in which observation points and variables were clustered in rows and columns, respectively. The rainbow color scheme used in this plot was used to illustrate minimum and maximum values, which were shown in red and magenta, respectively. Therefore, the corresponding colors of intermediate values, depending on their proximity to the minimum and maximum, fell within this rainbow-colored range. Figure 8 shows that Fae and Boe created G1 with an almost strong similarity shared by three variables, including Na^+, PO_4^{3-}, and Fe^{2+}; these were found to be almost identical, thus also forming one cluster. Due to the similarity for SO_4^{2-} (orange, Figure 8), for observation points GW28, GW25, and GW23 (G2), it was structured as one cluster itself. The observation points, GW23 and GW25, created a sub-cluster with striking similarity (Figure 7); they benefited from three variables, including Fe^{2+}, SO_4^{2-}, and pH that fell within the same cluster-variable. The same held true for GW28 and GW25, where SO_4^{2-} shared similarities between them; likewise, DOC, SO_4^{2-}, and SO_4^{2-} provided other cluster-variables between GW28 and GW23. The stand-alone sub-cluster GW30 (Figure 8) demonstrated meaningful similarity with GW20 and GW26 due to the shared cluster-variable, DOC and Mg^{2+}; the latter showed only similarity between GW30 and GW26. Likewise, GW30 displayed a certain similarity with GW22, WW, GW21, and GW1 in accordance with the shared cluster-variables, K^+, Na^+, PO_4^{3-}, and pH, respectively. Furthermore, there was no cluster-variable within the largest cluster (G3) that consisted of GW30, GW20, GW1, GW26, GW2, GW21, WW, and GW22. However, there were cluster-variables within the sub-clusters of the largest cluster. In this regard, DOC, SO_4^{2-}, and Cl^- were cluster-variables of GW20 and GW1, while GW26, GW2, GW21, WW, and GW22 had only one cluster-variable; Ca^{2+}. The sampling points, GW26, GW2, and GW21, shared a close similarity by having four cluster-variables consisting of

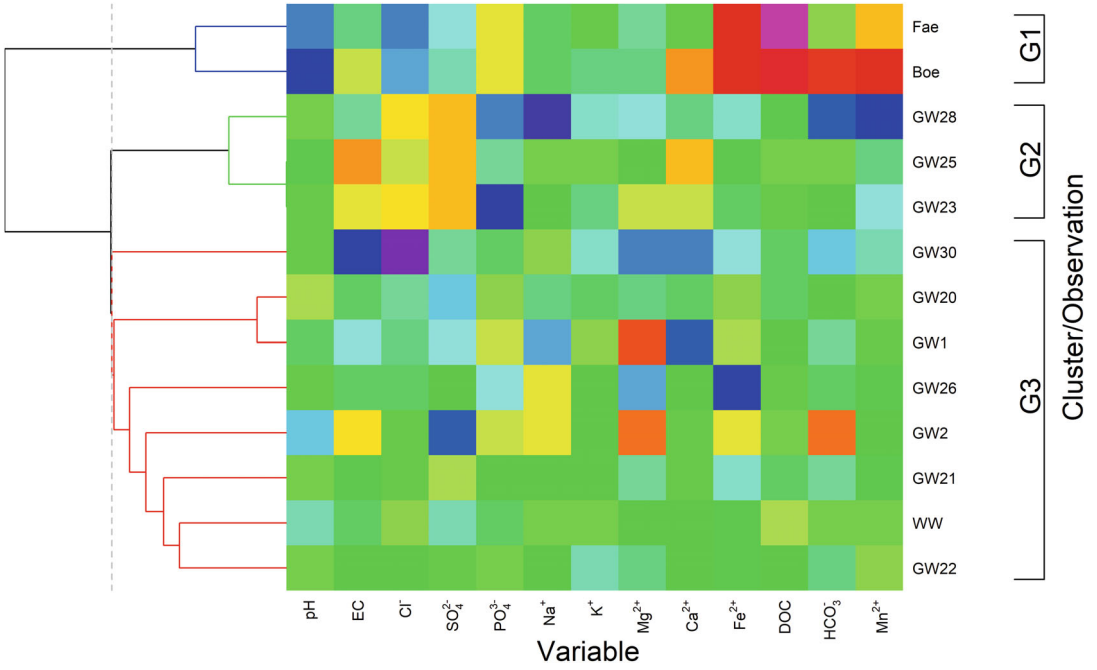

Figure 8. Superimposing an image plot onto a dendrogram tree obtained from applying the SSB algorithm (Figure 7), in which the clustering of observation points and variables/parameters were incorporated.

Mn^{2+}, Ca^{2+}, K^{+}, and Cl^{-}. The same was true for WW and GW22, in which four cluster-variables existed; Mn^{2+}, Fe^{2+}, Ca^{2+}, and EC.

Quantifying the Probability of Observation Assignments to each Cluster Via SSB

As each of the clusters identified using GFMM were broadly representative of one observation point (see Figure 5), we computed the probability assignment of each observation point to its most probable cluster (Figures 7 and 8). This enabled a comparison of the membership probabilities estimated by GFMM and SSB for each observation point/cluster.

The probability to which each observation point was clustered using SSB is illustrated by Figure 9a; these probabilities are given in ascending order, where the highest and lowest probabilities pertained to GW28 and GW21 with a probability >0.9 and 0.15, respectively. In comparison with GFMM, the probabilities calculated using SSB showed only GW28, Boe, GW30, and GW2 fell within the probability category of 0.8 to 1. By contrast, all observations pertaining to each cluster (C1 to C9) were classified as the narrow probability range of 0.8 to 1, as approximated using GFMM. Counter-intuitively, some observation points structuring one sub-cluster such as GW23 and G25 and sharing a striking similarity (Figure 7), suggested a middle probability of membership (approximately 0.5) in this sub-cluster; the same held true for GW20 and GW1.

Given the spike-and-slab models used in this approach, we were able to determine the parameters that most influenced the clustering typology. As these parameters were arranged based on variable importance

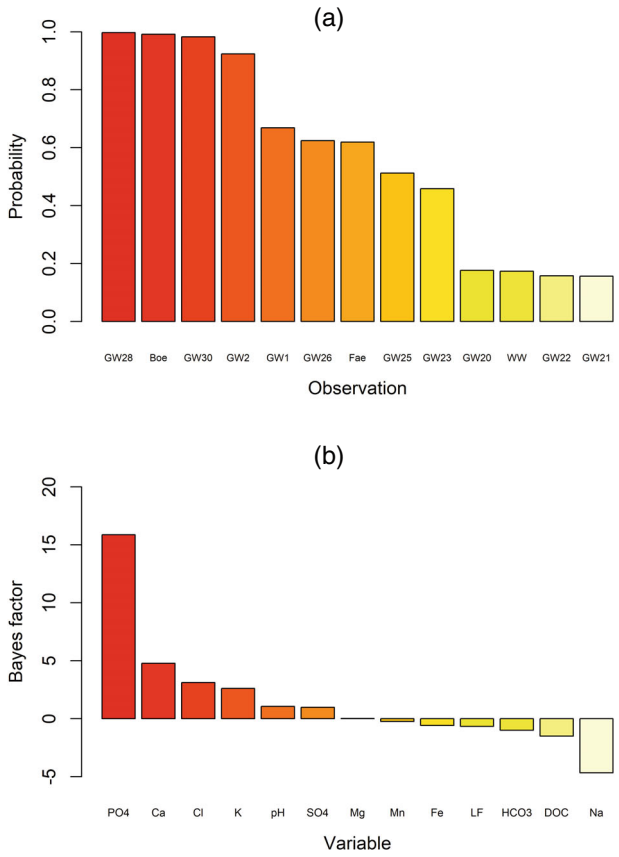

Figure 9. (a) The probability that relates each observation point to clusters determined using the SSB model (Figures 7 and 8). Note that the observation points are given in ascending order of probability; and (b) the most important variables (in order of importance), influencing the clustering of SSB as quantified using the log Bayes factor.

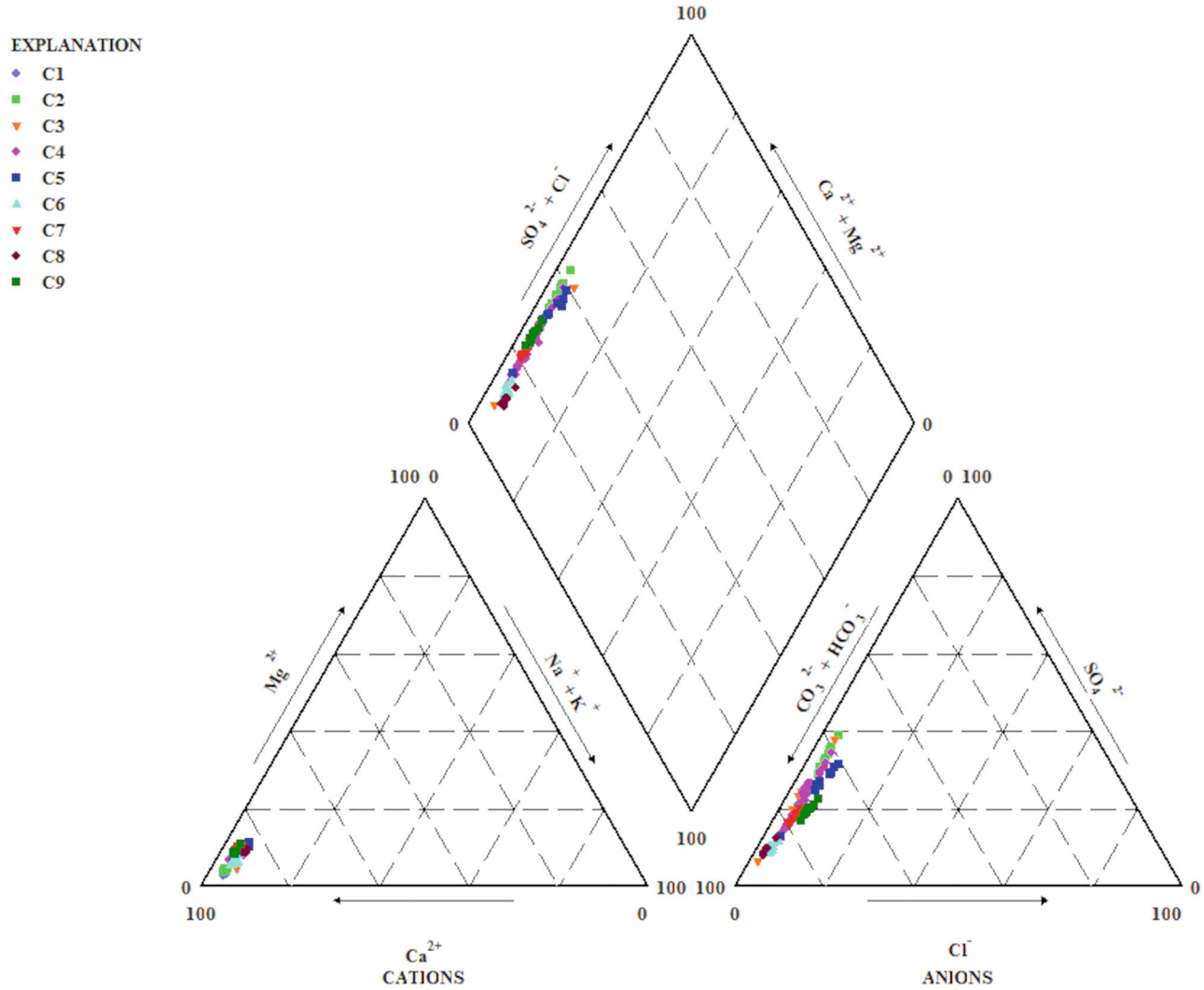

Figure 10. Hydrogeochemical facies discerned using the Piper diagram for the nine clusters of GFMM.

(Figure 9b), it was observed that only three variables had an impact on clustering (i.e., Na^+, SO_4^{2-}, and Cl^-), while the remaining variables had no effect on clustering.

Evaluation of GFMM-Clusters with the Piper Diagram and PCA

As the number of clusters identified by GFMM was exactly the same as the number of clusters obtained from the majority index (Figures 4a and 4c) (i.e., nine), the plausibility of clustering results from GFMM was examined using the Piper diagram and PCA (Figures 10 and 11). As such, the C1 to C9 clusters were illustrated on the Piper diagram and a 2D plot in which the *x* and *y* axis represented the first and second principal components, respectively. Figure 10 shows that clusters were almost well distinguished from each other while being grouped all together as a Ca^{2+}-HCO_3^- water type, in which Ca^{2+} and HCO_3^- were the predominant cation and anion for all clusters, respectively. As all clusters (C1 to C9) fell into one water type category, C1 and C6 had the highest and lowest Ca^{2+} concentrations, respectively; these are well illustrated by the boxplot (Figure S1). Likewise, the highest and lowest HCO_3^- concentrations were associated with clusters C8 and C2, respectively. Despite the striking similarity among clusters that fell into one water type category (Figure 10), these clusters did not completely overlap; this supports the distinction created with the nine clusters from GFMM.

The PCA of the dataset provided 13 eigenvalues (corresponding to the variance of the 13 geochemical parameters), and 13 eigenvectors (corresponding to 13 components that were linear combinations of 13 geochemical parameters). Table 2 lists the first four principal components in which the eigenvalues were >1, thus explaining approximately 75% of the variance. To evaluate and interpret principal components with respect to possible underlying processes, the Pearson correlation coefficients of the 13 original standardized hydrochemical parameters and the principal components (i.e., the Loadings) were calculated; this described how much each variable/parameter contributes to a particular PC. Due to the high variability attributable to these two principal components, further analysis was conducted for these particular principal components.

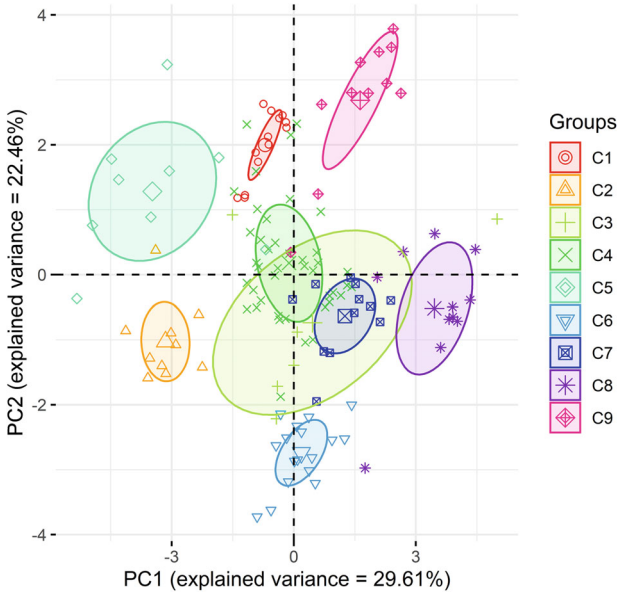

Figure 11. Principal component scores for the first two components. Samples belonging to one cluster are illustrated with the same color. Note that the nine ellipses correspond to the nine identified clusters (C1 to C9) from GFMM.

Table 2
Principal Component Loadings, Eigenvalues, and the Explained Variances for the First Four Components

Variable	PC1	PC2	PC3	PC4
pH	−0.55*	0.00	0.03	0.28
EC	0.40	0.79**	−0.02	−0.15
Cl^-	−0.16	0.78**	0.43	−0.26
SO_4^{2-}	−0.60*	0.60*	−0.15	−0.26
PO_4^{3-}	0.65*	−0.45	0.16	0.04
Na^+	0.39	0.32	−0.40	0.69**
K^+	0.31	0.19	0.32	0.32
Mg^{2+}	0.55*	0.14	0.72**	0.00
Ca^{2+}	0.35	0.80*	−0.38	−0.11
Fe^{2+}	0.76**	−0.19	0.20	−0.44
DOC	−0.33	0.32	0.53*	0.60*
HCO_3^-	0.83**	0.39	−0.15	0.13
Mn^{2+}	0.73**	−0.20	−0.23	0.12
Eigenvalue	3.82	2.89	1.54	1.40
Explained variance (%)	29.61	22.46	11.95	10.89

Note that the absolute values that were approximately ≥0.5 (moderate loadings), and ≥0.7 (strong loadings) for each component are denoted by "*" and "**", respectively.

To further inspect the clustering results from GFMM, according to the nine clusters (Figure 5), nine ellipses corresponding to C1 to C9 were established on a plot of principal component scores relating to the first two components (PC1 and PC2) (Figure 11). Clusters were mainly largely separated from each other, although C4 and C8 showed a large and small degree of overlap with C9, respectively; C7 formed as a sub-cluster within C3.

Discussion

Analysis of Clusters Delineated by GFMM and SSB

The nine delineated clusters closely corresponded to the 13 observation points in the dataset (Figure 5); each cluster resembled one or two observation points. This particularly holds true for C1, C2, C5, C6, C7, C8, and C9, which represented GW1, GW2, Boe/Fae, GW23/GW25, GW26, GW28, and GW30, respectively. Subtle differences observed between clusters suggest the possibility of a mixture condition in which geochemical characteristics may be ascribed to the geochemistry of the lakes, groundwater observations wells, and the gallery of drinking water production wells.

The nine clusters of GFMM were distinguished on the Piper diagram with some marginal overlaps (Figure 10). Yang et al. (2020) showed that seven clusters resulting from hierarchical cluster analysis were reasonably reflected by the Stiff and Piper diagrams. This indicates that standard techniques such as the Piper diagram are still a useful tool to assess an advanced clustering approach. Marginal overlaps among clusters may be associated with similarly high concentrations of $Ca^{2+}$and HCO_3^-, and similarly low concentrations of Cl^- and Mg^{2+} (see Figure S1). The viability of the GFMM algorithm was more evident when it could differentiate nine geochemical clusters for what had been characterized as one water type (i.e., Ca^{2+}-HCO_3^-), by the Piper diagram; this signifies that GFMM may successfully detect even subtle differences in the dataset.

To examine the clustering of GFMM with PCA, nine obtained clusters were projected onto a plot created using the first two principal components (Figure 11). The size of the ellipses demonstrates the degree of similarity between samples captured by an ellipse; a higher density of samples within one ellipse indicates stronger similarities between samples, implying a smaller level of uncertainty reflected in a certain cluster (Scrucca et al. 2016). For instance, C3 represented the largest ellipse while being recognized as the smallest cluster with only six samples (members). This suggests a low degree of similarity between its members and indicates considerable uncertainty where two samples were across the uncertainty boundary and one sample overstepped the boundary (Figure 6). By contrast, C4 was the largest cluster with 42 members; it showed a higher similarity between its samples, yielding a smaller ellipse. In this regard, C1, C2, and C6 with a high density of samples within their ellipses, explained a remarkable similarity found between members of each cluster. Given the distinct geochemical characteristics of these two lakes (Figure S1), the GW21_5 should have already been assigned to C4, which was also assigned to other samples of observation well GW21. Rectifying the misleading assigned probability of GW21_5 to C5 was suggested by elliptical clusters in Figure 11; this was further advocated by the uncertainty boundaries in Figure 6. Although the successful application using PCA to assess a clustering method was demonstrated (e.g., Iwamori et al. 2017; Yang

et al. 2020), Liu et al. (2021) drew a comparison between t-distributed Stochastic Neighbor Embedding (t-SNE) and PCA. They found that t-SNE showed superiority over PCA, in which members of each cluster plotted over a 2D t-SNE were relatively well agglomerated together rather than those of PCA.

By superimposing an image plot onto a dendrogram tree derived from the SSB algorithm (Figure 8), we identified the discrepancy among parameters of a certain cluster. This means that disparities among individual samples of geochemical facies will not be hidden; this is of paramount importance to future water quality management. For instance, GW28, GW25, and GW23 were grouped as G2 despite the distinct colors (considerable differences) for Mn^{2+} and Ca^{2+}. The same held true for G1, in which distinct colors, particularly for Mn^{2+}, HCO_3^-, and DOC, were easily recognized. Thus, different factors may control the geochemistry of a certain cluster, although the samples of one cluster share a strong similarity.

By comparison, the probabilities of all GFMM clusters had a probability membership $\geq$0.8, while the cluster membership probability of observation points of SSB (nearly correspondent to the nine clusters from GFMM), ranged from $<$0.2 to $>$0.8. Therefore, the probability assignment of only four observation points (GW28, Boe, GW30, and GW2) to specific clusters was nearly identical; that is $\geq$0.8 for both methods. The probability assignment of three observation points varied from 0.6 to 0.8 (GW1, GW26, and Fae). Likewise, the probability of membership of observation points, GW25 and GW23, varied from 0.4 to 0.6. Only the probability of membership of two clusters fell below 0.2; this should be treated with caution in the context of water quality management.

Due to the highly distinguishable geochemistry of the two lakes, C5 (Figure 5; GFMM) and G1 (Figure 8; SSB) were majorly formed based on all samples pertaining to the two lakes, and to only one sample from a groundwater observation point (i.e., GW21_5; see Figure 5).

Note that the most effective parameters on the clustering using SSB (i.e., Na^+, SO_4^{2-}, and Cl^-) were consistent with the strong and moderate loadings of PC4, PC1/PC2, and PC2, respectively.

Strengths and Weaknesses of GFMM and SSB for Clustering

To further assess the effectiveness of GFMM and SSB for classification purposes, we applied them to another dataset with the known number of clusters; this was the Oslo dataset (see Figures S3 and S4). Theoretically, we expected nine clusters corresponding to nine different plants from the Oslo dataset (Liu et al. 2021). The findings demonstrate that GFMM failed to distinguish the measurement of the nine plants into nine clusters; rather, the result was three clusters (Figure S3). By contrast, SSB could successfully differentiate the dataset into nine groups, albeit this data had a wide range of similarity among the groups (Figure S4).

Although GFMM was unsuccessful in differentiating the nine different materials of plants into nine clusters, it could satisfactorily classify the geochemical parameters of the capture zone of the waterworks in this study as nine clusters. Each cluster corresponded to an observation point; the exception to this was for GW20, GW21, GW22, and WW, where their parameters constituted one cluster (i.e., C4) (Figure 5). This is likely because GW20, GW21, and GW22 were evenly distributed around the waterworks (i.e., WW), with an average distance ranging from 700 to 1000 m; as a result, a similar geochemical pattern for these points was detected by GFMM. The clusters of GFMM were congruent with the classification carried out by the PCA and Piper diagram. Therefore, the Oslo dataset or similar datasets such as the Taiyuan karst water (Ma et al. 2011) and Jianghan Plain groundwater (Yang et al. 2020) may only be used as auxiliary indices to assess the performance of a clustering method. However, making a conscious decision on the efficacy of a clustering method also requires the soft knowledge of the practitioner, as is the case for the geochemical dataset of this study and the results obtained from applying GFMM.

The strength and weakness of SSB and GFMM should be considered when applying them to a dataset, specifically to a geochemical dataset obtained from an aquifer or waterworks. As SSB can operate based on replications, it can yield a clear-cut hierarchal group, in which entire observations belonging to a certain group may easily be visualized, and the similarity/dissimilarity among groups may easily be detected (see Figure 7). In addition SSB makes it possible to discern the extent of similarity among observations points and compare the degree of similarity among parameters/variables of obtained groups. Thus, the most important parameters that create distinctions between groups may be assigned. GFMM yields clusters over the time series and therefore, the geochemical evolution may be detected. This means that any turning point or break in the clusters, normally due to a significant change within the area of interest, may effectively be identified. This was the case for shifting the geochemical pattern of WW from C3 to C4 from 2002. It seems that GFMM was sensitive to the nature of the dataset of interest, thus a preprocessing of the dataset should be carefully conducted prior to its use. Due to its high sensitivity, preprocessing should be carried out using different transformation methods. In this regard, one of the datasets with known clusters (e.g., the Oslo dataset) may be utilized to ascertain the proper transformation technique. Generally, it is suggested that GFMM and SSB are applied to the dataset of interest to take advantage of the complementary outputs of these approaches alongside the quantified uncertainty of each obtained cluster. This is of huge importance when using these datasets for water quality management of waterworks or an aquifer.

Conclusion

It is hugely challenging to delineate the geochemical facies of the capture zone of waterworks and identify controlling factors influencing geochemistry. This is particularly relevant for waterworks that are dynamically

influenced by lake-aquifer flux exchange. To identify the spatiotemporal geochemical evolution and its driving factors, a number of methods were developed; these methods ranged from classic diagrams such as Stiff and Piper to advanced clustering/classification algorithms such as multivariate statistics and machine learning algorithms. Despite the significant advances in recently developed methods for this purpose, there is considerable uncertainty ascribed to the ascertained clusters; this imposes severe constraints on the suitability of obtained clusters for groundwater quality management of the capture zone of waterworks.

To address this issue, we proposed a scheme where two classifiers with cluster memberships given in terms of probability were appraised against each other. For this purpose, two sound clustering algorithms were deployed to evaluate the validity of clusters obtained from either method; these algorithms were GFMM and SSB. The algorithms were applied to 13 hydrochemical parameters collected during different sampling periods at 13 observation points (including 10 observation wells, two lakes, and a gallery of drinking production wells). These observation points were located across the Barnim Highlands plateau to the northeast of Berlin, Germany. We also drew comparisons between the most important parameters controlling the geochemistry of clusters obtained from GFMM and those obtained from SSB. The results demonstrate that GFMM produced nine clusters with a probability of membership that was $\geq$0.8 for either cluster. SSB yielded three major hierarchal clusters in which the probability of observations ranged from $<$0.2 to $>$0.8. Given the comparison between clusters delineated by GFMM and SSB, the probability assignment of only four observations to specific clusters was nearly identical; this means that the probability of observations was $\geq$0.8 for both methods. The results indicated that three parameters (i.e., Na^+, SO_4^{2-}, and Cl^-), given in order of importance, could influence the classification carried out by SSB. Furthermore, the clustering result from GFMM was in good agreement with the clustering obtained from PCA and the Piper diagram.

Our findings also show that the geochemical cluster (identified by GFMM) of the capture zone of the waterworks resembled that of the lakes from 2002. This means that the geochemical properties of the waterworks have switched from C3 to C4. Therefore, we postulate that since this time, a noticeable inflow from the lakes into the capture zone of the waterworks has been activated. This is a result of the groundwater head drawdown of the capture zone due to the overutilization of groundwater.

In conclusion, the complementary implications of GFMM and SSB that enhance the current understanding of underlying geochemical processes relating to the capture zone of the waterworks are highly useful to inform future groundwater quality management and planning.

Acknowledgments

The authors wish to thank the German Federal Ministry of Education and Research for funding the research project INKA BB (INKAA BB, 2010). We would also like to thank Bernd Schwien and Dorith Henning for collecting the field data. Open Access funding enabled and organized by Projekt DEAL.

Data Availability Statement

Data archiving is currently underway. Therefore, the dataset used in this research will be stored in the Leibniz Centre for Agricultural Landscape Research (ZALF) Data Repository and will be made publicly available through scientific publications. The ZALF Open Research Data may be accessed through: https://open-research-data.zalf.de/default.aspx.

Authors' Note

The authors do not have any conflicts of interest or financial disclosures to report.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article. Supporting Information is generally *not* peer reviewed.

Appendix S1. The Oslo dataset.
Appendix S2. General description of Spike-and-slab Bayesian (SSB).
Appendix S3. Geochemical characteristics of the nine procured clusters of GFMM.
Appendix S4. Geochemical characterization of the principal components.
Appendix S5. First principal component.
Appendix S6. Second principal component.
Appendix S7. Third principal component.
Appendix S8. Fourth principal component.
Appendix S9. References.
Table S1. Filter screen depth intervals and average values and standard deviation for all field parameters and variables used. Concentrations are given in mmol/L.
Table S2. The sampling periods for each of the observation points
Figure S1. Box plots of the 13 hydrochemical parameters for the nine identified clusters of GFMM.
Figure S2. Pairwise scatterplots for the 13 hydrochemical parameters with respect to the nine captured clusters of GFMM.
Figure S3. Clusters obtained from applying GFMM to the Oslo dataset. Note that the subscript numbers denote the 40 sites over the 120-km transect.
Figure S4. Clusters obtained from applying SSB to the Oslo dataset.

References

Aguilera, P.A., A. Fernández, R.F. Ropero, and L. Molina. 2013. Groundwater quality assessment using data clustering based on hybrid Bayesian networks. *Stochastic Environmental*

Research and Risk Assessment 27, no. 2: 435–447. https://doi.org/10.1007/s00477-012-0676-8

Anderson, M.R., and A. Vehtari. 2017. Bayesian inference for spatio-temporal spike-and-slab priors. *Journal of Machine Learning Research* 18, no. 1: 1–58.

Atlas, Berlin Environmental. 2018. Groundwater Levels of the Main Aquifer and Panke Valley Aquifer (Edition 2018). Edited by Senate Department for Urban Development and Housing.

Bai, R., V. Rockova, and E. George. 2021. Spike-and-slab meets LASSO: A review of the spike-and-slab LASSO. In *Handbook of Bayesian Variable Selection*, ed. Mahlet Tadesse, and Marina Vannucci, 528. New York: CRC Press.

Böttcher, S., C. Merz, G. Lischeid, and R. Dannowski. 2014. Using Isomap to differentiate between anthropogenic and natural effects on groundwater dynamics in a complex geological setting. *Journal of Hydrology* 519, no. Part B: 1634–1641. https://doi.org/10.1016/j.jhydrol.2014.09.048

Canale, A., A. Lijoi, B. Nipoti, and I. Prünster. 2017. On the Pitman–Yor process with spike and slab base measure. *Biometrika* 104, no. 3: 681–697. https://doi.org/10.1093/biomet/asx041

Cao, T., X. Zeng, J. Wu, D. Wang, Y. Sun, X. Zhu, J. Lin, and Y. Long. 2019. Groundwater contaminant source identification via Bayesian model selection and uncertainty quantification. *Hydrogeology Journal* 27, no. 8: 2907–2918. https://doi.org/10.1007/s10040-019-02055-3

Carucci, V., M. Petitta, and R. Aravena. 2012. Interaction between shallow and deep aquifers in the Tivoli Plain (Central Italy) enhanced by groundwater extraction: A multi-isotope approach and geochemical modeling. *Applied Geochemistry* 27, no. 1: 266–280. https://doi.org/10.1016/j.apgeochem.2011.11.007

Charrad, M., N. Ghazzali, V. Boiteau, and A. Niknafs. 2014. NbClust: An R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software* 61, no. 6: 1–36. https://doi.org/10.18637/jss.v061.i06

Cloutier, V., R. Lefebvre, R. Therrien, and M.M. Savard. 2008. Multivariate statistical analysis of geochemical data as indicative of the hydrogeochemical evolution of groundwater in a sedimentary rock aquifer system. *Journal of Hydrology* 353, no. 3: 294–313. https://doi.org/10.1016/j.jhydrol.2008.02.015

Dempster, A.P., N.M. Laird, and D.B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)* 39, no. 1: 1–22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

Dougherty, E.R., and M. Brun. 2004. A probabilistic theory of clustering. *Pattern Recognition* 37, no. 5: 917–925.

Ellefsen, K.J., and D.B. Smith. 2016. Manual hierarchical clustering of regional geochemical data using a Bayesian finite mixture model. *Applied Geochemistry* 75: 200–210. https://doi.org/10.1016/j.apgeochem.2016.05.016

Ellefsen, K.J., D.B. Smith, and J.D. Horton. 2014. A modified procedure for mixture-model clustering of regional geochemical data. *Applied Geochemistry* 51: 315–326.

Ferone, J.M., and K.J. Devito. 2004. Shallow groundwater-surface water interactions in pond-peatland complexes along a Boreal Plains topographic gradient. *Journal of Hydrology* 292, no. 1–4: 75–95. https://doi.org/10.1016/j.jhydrol.2003.12.032

Fraley, C., and A.E. Raftery. 2002. Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association* 97, no. 458: 611–631.

Fraley, C., and A.E. Raftery. 1998. How many clusters? which clustering method? answers via model-based cluster analysis, Department of Statistics University of Washington, Technical Report no. 329.

Frind, E.O., and J.W. Molson. 2018. Issues and options in the delineation of well capture zones under uncertainty. *Groundwater* 56, no. 3: 366–376.

Germer, S., K. Kaiser, O. Bens, and R.F. Huettl. 2011. Water balance changes and responses of ecosystems and society in the Berlin-Brandenburg region—A review. *Die Erde* 142, no. 1–2: 65–95.

Gerstengarbe, F.W., P.C. Werner, H. Osterle, and O. Burghoff. 2013. Winter storm- and summer thunderstorm-related loss events with regard to climate change in Germany. *Theoretical and Applied Climatology* 114, no. 3–4: 715–724. https://doi.org/10.1007/s00704-013-0843-y

Gerstengarbe, F.W., F. Badeck, F. Hattermann, V. Krysanova, W. Lahmer, P. Lasch, M. Stock, F. Suckow, F. Wechsung, and P.C. Werner. 2003. Studie zur klimatischen Entwicklung im Land Brandenburg bis 2055 und deren Auswirkugnen auf den Wasserhausalt, die Forst- und Landwirtschaft sowie die Ableitung erster Perspektiven. *PIK Report* 83: 1–96.

Golian, M., H. Katibeh, V.P. Singh, K. Ostad-Ali-Askari, and H. Tavasoli Rostami. 2019. Prediction of tunnelling impact on flow rates of adjacent extraction water wells. *Quarterly Journal of Engineering Geology and Hydrogeology* 53: 236–251. https://doi.org/10.1144/qjegh2019-055

Güler, C., and G.D. Thyne. 2004a. Delineation of hydrochemical facies distribution in a regional groundwater system by means of fuzzy c-means clustering. *Water Resources Research* 40, no. 12: 1–12. https://doi.org/10.1029/2004wr003299

Güler, C., and G.D. Thyne. 2004b. Hydrologic and geologic factors controlling surface and groundwater chemistry in Indian Wells-Owens Valley area, southeastern California. USA. *Journal of Hydrology* 285, no. 1–4: 177–198.

Güler, C., M.A. Kurt, M. Alpaslan, and C. Akbulut. 2012. Assessment of the impact of anthropogenic activities on the groundwater hydrology and chemistry in Tarsus coastal plain (Mersin, SE Turkey) using fuzzy clustering, multivariate statistics and GIS techniques. *Journal of Hydrology* 414-415: 435–451. https://doi.org/10.1016/j.jhydrol.2011.11.021

Hadj, A., N. Friha, K. Chkir, B. Zouari, P.D. Hamelin, and A. Aigoun. 2014. Hydro-geochemical processes in the complexe terminal aquifer of southern Tunisia: An integrated investigation based on geochemical and multivariate statistical methods. *Journal of African Earth Sciences* 100: 81–95. https://doi.org/10.1016/j.jafrearsci.2014.06.015

He, J., J. Ma, P. Zhang, L. Tian, G. Zhu, W.M. Edmunds, and Q. Zhang. 2012. Groundwater recharge environments and hydrogeochemical evolution in the Jiuquan Basin, Northwest China. *Applied Geochemistry* 27, no. 4: 866–878. https://doi.org/10.1016/j.apgeochem.2012.01.014

Heard, N.A., C.C. Holmes, and D.A. Stephens. 2006. A quantitative study of gene regulation involved in the immune response of Anopheline mosquitoes: An application of Bayesian hierarchical clustering of curves. *Journal of the American Statistical Association* 101, no. 473: 18–29.

Held, H., F.W. Gerstengarbe, T. Pardowitz, J.G. Pinto, U. Ulbrich, K. Born, M.G. Donat, M.K. Karremann, G.C. Leckebusch, P. Ludwig, K.M. Nissen, H. Osterle, B.F. Prahl, P.C. Werner, D.J. Befort, and O. Burghoff. 2013. Projections of global warming-induced impacts on winter storm losses in the German private household sector. *Climatic Change* 121, no. 2: 195–207. https://doi.org/10.1007/s10584-013-0872-7

Iwamori, H., K. Yoshida, H. Nakamura, T. Kuwatani, M. Hamada, S. Haraguchi, and K. Ueki. 2017. Classification of geochemical data based on multivariate statistical analyses: Complementary roles of cluster, principal component, and independent component analyses. *Geochemistry Geophysics Geosystems* 18, no. 3: 994–1012. https://doi.org/10.1002/2016gc006663

Kim, J.H., B.W. Yum, R.H. Kim, D.C. Koh, T.J. Cheong, J. Lee, and H.W. Chang. 2003. Application of cluster analysis for the hydrogeochemical factors of saline groundwater in Kimje, Korea. *Geosciences Journal* 7, no. 4: 313–322.

Lahmer, W. 2003. Trend analyses of percolation in the state of Brandenburg and possible impacts of climate change. *Journal of Hydrology and Hydromechanics* 51, no. 3: 196–209.

Liu, H., J. Yang, M. Ye, S.C. James, Z. Tang, J. Dong, and T. Xing. 2021. Using *t*-distributed stochastic neighbor embedding (t-SNE) for cluster analysis and spatial zone delineation of groundwater geochemistry data. *Journal of Hydrology* 597: 1–13.

Liu, F., X.F. Song, L.H. Yang, D.M. Han, Y.H. Zhang, Y. Ma, and H.M. Bu. 2015. The role of anthropogenic and natural factors in shaping the geochemical evolution of groundwater in the Subei Lake basin, Ordos energy base, Northwestern China. *Science of the Total Environment* 538: 327–340. https://doi.org/10.1016/j.scitotenv.2015.08.057

Ma, R., Y. Wang, Z. Sun, C. Zheng, T. Ma, and H. Prommer. 2011. Geochemical evolution of groundwater in carbonate aquifers in Taiyuan, northern China. *Applied Geochemistry* 26, no. 5: 884–897.

Marbac, M., C. Biernacki, and V. Vandewalle. 2017. Model-based clustering of Gaussian copulas for mixed data. *Communications in Statistics—Theory and Methods* 46, no. 23: 11635–11656. https://doi.org/10.1080/03610926.2016.1277753

McLachlan, G., and D. Peel. 2000. *Finite Mixture Models. Wiley Series in Probability and Statistics*. New York: John Wiley & Sons Inc.

Merz, C., and A. Pekdeger. 2011. Anthropogenic changes in the landscape hydrology of the Berlin-Brandenburg region. *Die Erde* 142, no. 1–2: 21–39.

Ministry of Environment Health and Consumer Protection of Brandenburg. 2009. Environmental Data Brandenburg 2008/2009. http://www.mugv.brandenburg.de/cms/detail.php/bb1.c.302494.de

Nguyen, T.T., A. Kawamura, T.N. Tong, N. Nakagawa, H. Amaguchi, and R. Gilbuena. 2015. Clustering spatio-seasonal hydrogeochemical data using self-organizing maps for groundwater quality assessment in the Red River Delta, Vietnam. *Journal of Hydrology* 522: 661–673. https://doi.org/10.1016/j.jhydrol.2015.01.023

Ostad-Ali-Askari, K., H. Ghorbanizadeh Kharazi, M. Shayannejad, and M.J. Zareian. 2019. Effect of management strategies on reducing negative impacts of climate change on water resources of the Isfahan–Borkhar aquifer using MODFLOW. *River Research and Applications* 35, no. 6: 611–631. https://doi.org/10.1002/rra.3463

Pacheco Castro, R., J. Pacheco Avila, M. Ye, and A. Cabrera Sansores. 2018. Groundwater quality: Analysis of its temporal and spatial variability in a karst aquifer. *Groundwater* 56, no. 1: 62–72.

Pant, R.R., F. Zhang, F.U. Rehman, G. Wang, M. Ye, C. Zeng, and H. Tang. 2018. Spatiotemporal variations of hydrogeochemistry and its controlling factors in the Gandaki River basin, Central Himalaya Nepal. *Science of the Total Environment* 622: 770–782.

Papaioannou, A., A. Mavridou, C. Hadjichristodoulou, P. Papastergiou, O. Pappa, E. Dovriki, and I. Rigas. 2010. Application of multivariate statistical methods for groundwater physicochemical and biological quality assessment in the context of public health. *Environmental Monitoring and Assessment* 170, no. 1: 87–97. https://doi.org/10.1007/s10661-009-1217-x

Partovi Nia, V., and A.C. Davison. 2012. High-dimensional Bayesian clustering with variable selection: The R Package bclus. *Journal of Statistical Software* 47, no. 5: 1–22.

Partovi Nia, V. 2009. Fast high-dimensional Bayesian classification and clustering. Ph.D. dissertation, Ecole Polytechnique Federale de Lasuanne. http://library.epfl.ch/en/theses/?nr=4482

Piper, A.M. 1944. A graphic procedure in the geochemical interpretation of water-analyses. *Transactions, American Geophysical Union* 25, no. 6: 914. http://dx.doi.org/10.1029/tr025i006p00914

Popp, A.L., A. Scheidegger, C. Moeck, M.S. Brennwald, and R. Kipfer. 2019. Integrating Bayesian groundwater mixing modeling with on-site helium analysis to identify unknown water sources. *Water Resources Research* 55, no. 12: 10602–10615. https://doi.org/10.1029/2019wr025677

Russoniello, C.J. and L.K. Lautz. 2020. Pay the PIED Piper: Guidelines to Visualize Large Geochemical Datasets on Piper Diagrams. *Groundwater* 58, no. 3: 464–469. http://dx.doi.org/10.1111/gwat.12953

Saranya, S., S. Poonguzhali, and S. Karunakaran. 2020. Gaussian mixture model based clustering of manual muscle testing grades using surface electromyogram signals. *Physical and Engineering Sciences in Medicine* 43: 837–847. https://doi.org/10.1007/s13246-020-00880-5

Saxena, A., M. Prasad, A. Gupta, N. Bharill, O.P. Patel, A. Tiwari, M.J. Er, D. Weiping, and C.T. Lin. 2017. A review of clustering techniques and developments. *Neurocomputing* 267: 664–681.

Sbarbati, C., N. Colombani, M. Mastrocicco, R. Aravena, and M. Petitta. 2015. Performance of different assessment methods to evaluate contaminant sources and fate in a coastal aquifer. *Environmental Science and Pollution Research* 22, no. 20: 15536–15548. https://doi.org/10.1007/s11356-015-4731-0

Scrucca, L. 2016. Identifying connected components in Gaussian finite mixture models for clustering. *Computational Statistics & Data Analysis* 93, no. C: 5–17. https://doi.org/10.1016/j.csda.2015.01.006

Scrucca, L., M. Fop, T.B. Murphy, and A.E. Raftery. 2016. mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *R Journal* 8, no. 1: 289–317.

Scrucca, L. 2010. Dimension reduction for model-based clustering. *Statistics and Computing* 20, no. 4: 471–484. https://doi.org/10.1007/s11222-009-9138-7

Shrestha, S., and F. Kazama. 2007. Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. *Environmental Modelling & Software* 22, no. 4: 464–475. https://doi.org/10.1016/j.envsoft.2006.02.001

Simeonov, V., J.A. Stratis, C. Samara, G. Zachariadis, D. Voutsa, A. Anthemidis, M. Sofoniou, and T. Kouimtzis. 2003. Assessment of the surface water quality in Northern Greece. *Water Research* 37, no. 17: 4119–4124. https://doi.org/10.1016/S0043-1354(03)00398-1

State Office for Mining Geology and Raw Material of Brandenburg. 2012. Hydrogeologische Karten Brandenburg. Cottbus.

Suckow, F., P. Lasch, and F. Badeck. 2002. Auswirkungen von Kimaveränderungen auf die Grundwasserneubildung. *Eberswalder Forstliche Schriftreihe* 15: 36–44.

Tadesse, M.G., N. Sha, and M. Vannucci. 2005. Bayesian variable selection in clustering high-dimensional data. *Journal of the American Statistical Association* 100, no. 470: 602–617. https://doi.org/10.1198/016214504000001565

Taie Semiromi, M., and M. Koch. 2020. How do gaining and losing streams react to the combined effects of climate change and pumping in the Gharehsoo river basin, Iran? *Water Resources Research* 56, no. 7: 1–35. https://doi.org/10.1029/2019wr025388

Templ, M., P. Filzmoser, and C. Reimann. 2008. Cluster analysis applied to regional geochemical data: Problems and possibilities. *Applied Geochemistry* 23, no. 8: 2198–2213. https://doi.org/10.1016/j.apgeochem.2008.03.004

Vepraskas, M.J., R.W. Skaggs, and P.V. Caldwell. 2020. Method to assess climate change impacts on hydrologic boundaries of individual wetlands. *Wetlands* 40, no. 2: 365–376. https://doi.org/10.1007/s13157-019-01183-6

Wang, H., X. Jiang, L. Wan, G. Han, and H. Guo. 2015. Hydrogeochemical characterization of groundwater flow systems in the discharge area of a river basin. *Journal of Hydrology* 527: 433–441. https://doi.org/10.1016/j.jhydrol.2015.04.063

Wang, P., J. Yu, Y. Zhang, and C. Liu. 2013. Groundwater recharge and hydrogeochemical evolution in the Ejina Basin, Northwest China. *Journal of Hydrology* 476: 72–86. https://doi.org/10.1016/j.jhydrol.2012.10.049

Winston, R. B. 2000. Graphical User Interface for MODFLOW, Version 4: U.S. Geological Survey Open-File Report 00-315, 27 p., https://doi.org/10.3133/ofr00315.

Woocay, A., and J. Walton. 2008. Multivariate analyses of water chemistry: Surface and ground water interactions. *Groundwater* 46, no. 3: 437–449. https://doi.org/10.1111/j.1745-6584.2007.00404.x

Yang, J., M. Ye, Z. Tang, T. Jiao, X. Song, Y. Pei, and H. Liu. 2020. Using cluster analysis for understanding spatial and temporal patterns and controlling factors of groundwater geochemistry in a regional aquifer. *Journal of Hydrology* 583: 1–15. https://doi.org/10.1016/j.jhydrol.2020.124594

Zhou, R.G., and W. Wang. 2020. Online nonparametric Bayesian analysis of parsimonious Gaussian mixture models and scenes clustering. *ETRI Journal* 43, no. 1: 74–81. https://doi.org/10.4218/etrij.2019-0336