A self-organizing map approach to characterize hydrogeology of the fractured Serra-Geral transboundary aquifer

Fabio Iwashita, Michael J. Friedel and Francisco J. F. Ferreira

ABSTRACT

The aim of this work is to understand the exchange of water between the Serra Geral aquifer system (SGAS) and Guarani aquifer system (GAS). The objectives are two-fold. First, introduce the capability of the modified self-organizing maps (MSOM) as an unbiased nonlinear approach to estimate missing values of hydrochemistry and hydraulic transmissivity associated with the SGAS, a transboundary groundwater system spanning parts of four South American countries. Second, identify areas with potential connectivity of the SGAS with the GAS based on analysis of the spatial variability of key elements and comparison with current conceptual models of hydraulic connectivity. The MSOM is employed to calculate correlations (trends) between 27 variables from 1,132 wells. Hydraulic transmissivity is calculated from specific capacity values from well-pump tests in 157 locations. Hydrochemical facies estimates appear unbiased and consistent with current conceptual-connectivity models indicating that vertical fluxes from GAS are influenced by geological structure. The MSOM provides additional spatial estimates revealing new areas with likely connections between the two aquifer systems.

Key words | fractured aquifer, Guarani aquifer, hydrochemistry, machine learning, Serra Geral

Fabio Iwashita (corresponding author)
Earth Sciences Department,
University of Florence,
Florence,
Italy
and
Geosciences Department,
University Los Andes,
Bogota,
Colombia
E-mail: f.iwashita1@uniandes.edu.co

Michael J. Friedel
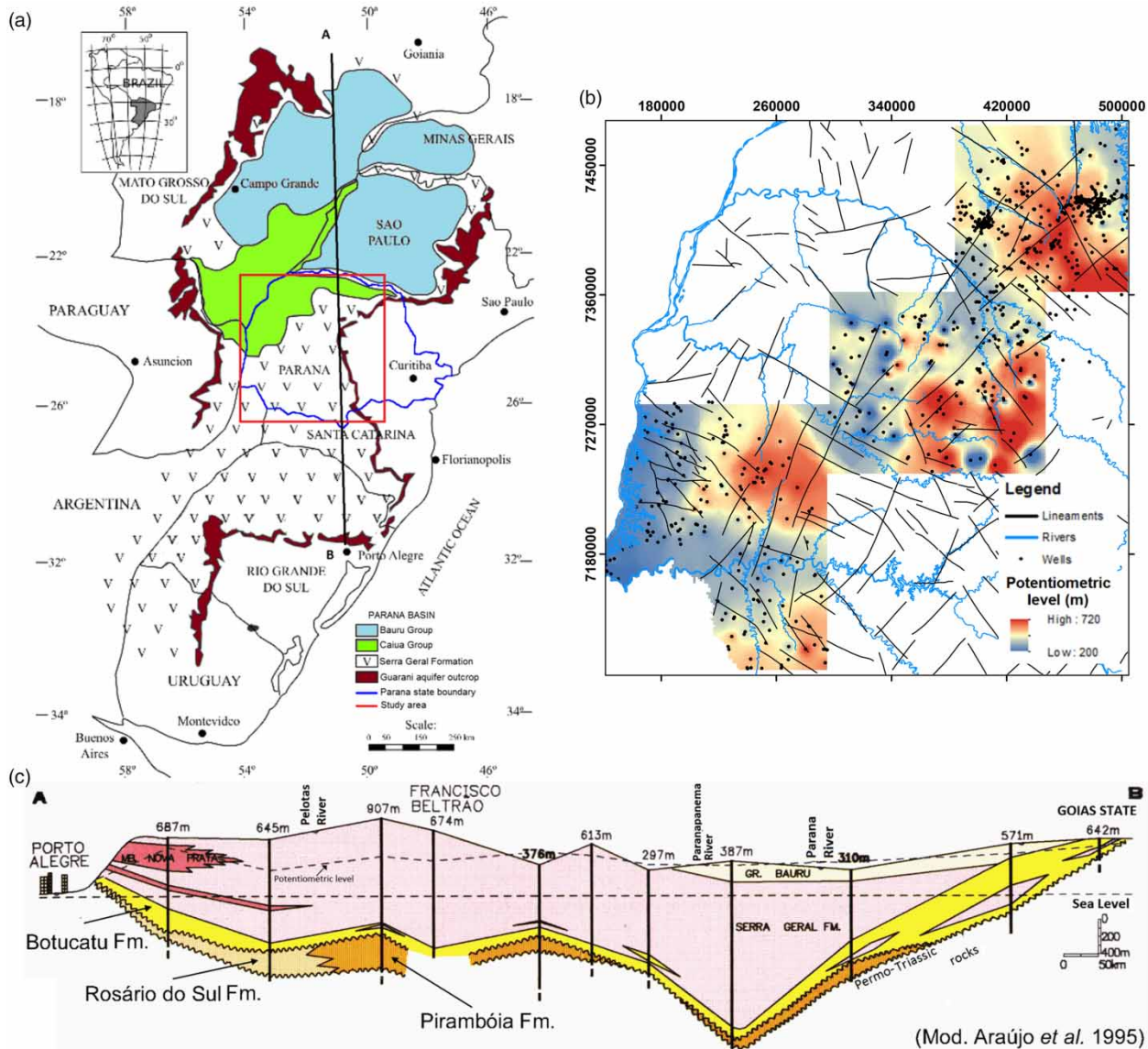GNS Science Hydrogeology,
Lower Hutt,
New Zealand

Francisco J. F. Ferreira
Geology Department,
Parana Federal University,
Curitiba,
Brazil

INTRODUCTION

The Serra Geral aquifer system (SGAS) is one of the largest and most important in Brazil. It is an unconfined and fractured transboundary aquifer formed within the sequence of lower Cretaceous Parana flood basalts (Leinz 1949). The total outcropping of this aquifer reaches 1.2 million km². The SGAS covers four Brazilian states and parts of three countries: Argentina, Uruguay and Paraguay (Figure 1). The basalt overlies the Guarani Aquifer System (GAS) in the middle sections of the Parana basin is an equivalent porous aquifer composed primarily by sandstone from the Botucatu and Pirambóia Formations (Sracek & Hirata 2002; Wendland et al. 2007). Despite the apparent confinement of the GAS, many studies have identified hydraulic communication between the two systems based on hydrochemistry (Fraga 1986). In this scenario, the movement of

stored water from the GAS to the SGAS is possible when there are geological discontinuities associated with favorable hydraulic conditions, such as when the GAS potentiometric level is higher than the actual depth of the aquifer (Figure 2).

Mapping the hydraulic connections and spatial distribution of hydrochemical facies is an essential step in the implementation of numerical models. Furthermore, the estimates forming a consistent regional dataset can be used to construct variograms of aquifer properties to assist in the calibration of a groundwater model (Friedel & Iwashita 2013; Friedel et al. 2016). A common approach used to identify areas of hydraulic connectivity between adjacent aquifers is to analyze the relative difference in concentrations and spatial distribution among the GAS and SGAS waters
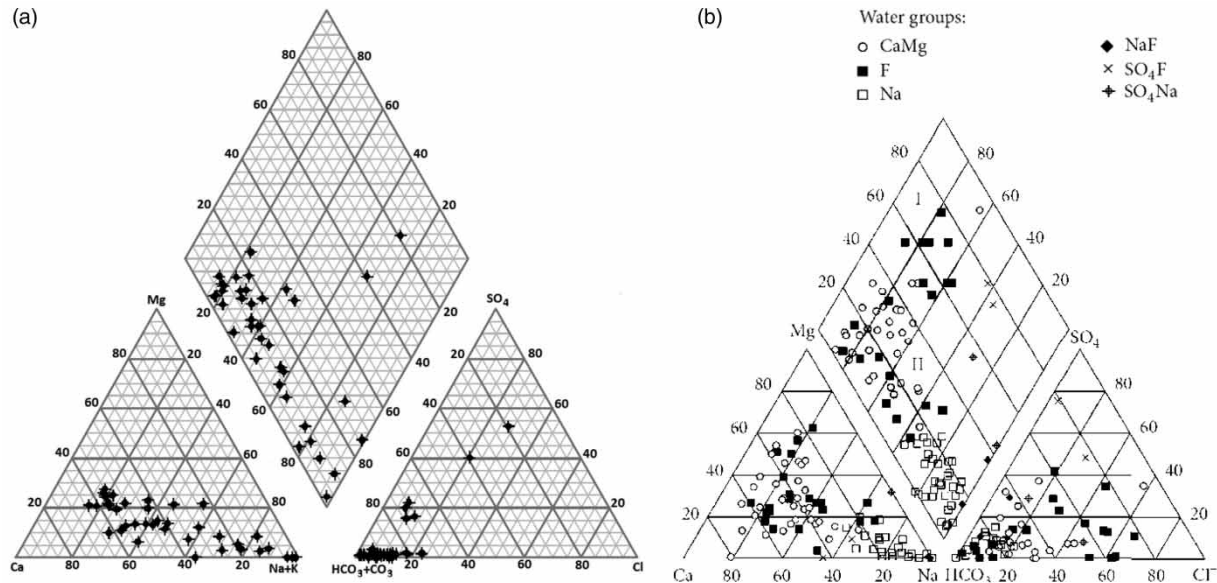
Uncorrected Proof

2   F. Iwashita *et al.* │ Using self-organizing maps to characterize hydrogeology of Serra-Geral aquifer   **Hydrology Research** │ in press │ 2017

Figure 1 │ Geological setting. (a) The Serra Geral Formation limits (note that it is comprised in four different South American countries (modified from Nanni *et al.* 2009); (b) Elevation map with lineaments (Soares *et al.* 1982; Zalán *et al.* 1987; Artur 1998) and rivers draped over; and (c) profile of the geological setting.

(Bittencourt *et al.* 2003; Portela Filho 2003; Ferreira *et al.* 2005; Rosa Filho *et al.* 2006; Silva 2007; Mocellin 2009; Nanni *et al.* 2009; Bongiolo *et al.* 2014).

The integrated analysis of geological structures and variation of chemical concentration for conservative elements is a preferred methodology to identify hydraulic connections between the Serra Geral and the Guarani aquifers (Ferreira *et al.* 2005). Remote sensing and geophysical data are used to support the study of aquifers when characterizing geological structures (Nanni *et al.* 2009), or as a predictive variable in the absence of monitoring wells (Souza Filho *et al.* 2010). Additionally, there is a possible influence of morphological features on water chemistry, particularly where a thick layer of soil combined with high clay content may prevent recharge to the SGAS section (Nanni *et al.* 2009). When these conditions are associated with vertical faults, the relative concentration of Guarani trace elements increase and the connection between the two systems becomes more evident.

Uncorrected Proof

3   F. Iwashita *et al.* | Using self-organizing maps to characterize hydrogeology of Serra-Geral aquifer   **Hydrology Research** | **in press** | **2017**

Figure 2 | Piper diagram from water well samples of the SGAS: (a) in Parana State (Athayde *et al.* 2007); (b) across Rio Grande do Sul State (Nanni *et al.* 2013).

The availability of hydrochemical measurements characterizing the SGAS includes pH and major ions; however, these data sets are incomplete (sparse). Most parametric statistical methods, such as analysis of variance (Winter *et al.* 2006), requires a complete data set. Other multivariate statistical methods, such as cluster analysis (Suk & Lee 1999), principal component analysis (Astel *et al.* 2007) and factor analysis, require the computation of eigenvalues and eigenvectors (Neter *et al.* 1996), based on a complete data set. One alternative to deal with data sparseness are imputation methods (Malek *et al.* 2008). These imputations methods comprise statistical and mathematical approaches to estimate missing values in datasets based on a combination of the available data (Dickson & Giblin 2007); however, these methods are linear approaches. The self-organizing map (SOM) is a nonlinear vector-quantization technique (Kohonen 1984) that is sometimes used to estimate missing values (Wang 2003; Malek *et al.* 2008), such as in precipitation and run-off processes (Kalteh & Hjorth 2009). Some other related studies include the characterization and survey of groundwater chemistry (Hong & Rosen 2001; Lu & Lo 2002; Sánchez-Martos *et al.* 2002; Nakagawa *et al.* 2016), groundwater levels (Han *et al.* 2016), soil hydraulic properties (Iwashita *et al.* 2012), air quality datasets (Junninen *et al.* 2004), detection of errors in

large datasets (Fessant & Midenet 2002), and address spatial continuity problems for groundwater model calibration (Friedel & Iwashita 2013). The SOM can characterize high-dimensional datasets, representing them in two or three dimensions projected onto maps composed by code vectors (ASCE 2000a, 2000b). Each code vector has the same dimension as the input data array. Through an iterative process, the SOM is trained to fit input data, with each sample associated to n-dimensional weight vector (Kohonen 2001). The SOM's attribute of learning vector quantization preserves topological relations among samples making it an inherently robust estimation method (Dickson & Giblin 2007).

The aim of this study is to identify the spatial exchange of groundwater between the Serra Geral and Guarani aquifers. The objective is to compare model-estimated values of hydrochemistry and hydraulic conductivity with published conceptual models of the Serra Geral fractured aquifer in the State of Parana, Brazil. To achieve this aim and objective, the following tasks are undertaken: (1) use the cross-component planes of SOM weights to determine Pearson's correlation index between hydrochemical elements, relief morphometry, and aeromagnetic data; (2) use modified self-organizing maps (MSOM) to estimate missing values in the groundwater database; (4) apply the k-means method to find relevant clusters (Davies & Bouldin 1977)

Uncorrected Proof

4   F. Iwashita *et al.* | Using self-organizing maps to characterize hydrogeology of Serra-Geral aquifer   **Hydrology Research** | in press | 2017

of hydrochemical facies; and (5) identify possible flux connections between the Serra Geral and the Guarani aquifers based on anomalous spatial hydrochemistry behavior.

STUDY AREA

The SGAS covers approximately 1.2 million km^2 in the State of Parana (Figure 1). The SGAS is the primary source of water for the state (Athayde *et al.* 2007; Manasses 2009) and is characterized as anisotropic and fractured crystalline rocks. In this system, water flows through fractures characterized as cracks and gaps opened by tectonic displacements and weathering. The relative storage capacity of fissured aquifers like the SGAS depends on the fracture density, the width of fractures (gaps), and the communication between these structures (Fraga 1986). For these reasons, the groundwater yield from wells drilled into the SGAS generally depends on the number and density of fractures. The SGAS is hosted mainly by volcanic rocks as tholeiitic

basalts, andesites, rhyolites and rhyodacites (Harris & Milner 1997). The thickness of the Serra Geral Formation volcanic rocks increases from east to west, reaching 1,500 meters at the center of the Parana Basin (Peate *et al.* 1988). The outcropping rocks generally present aphanitic and other micro-crystalline textures with massive or vesicular-amygdalous structure. Dikes and sills of tholeiitic and rhyodacitic composition are widespread (Turner *et al.* 1999).

The main processes affecting water chemistry of the Serra Geral aquifer are weathering of the basaltic rocks and associated equilibrium with secondary minerals. The geochemical interaction between percolating water to the aquifer rock along the recharge and discharge zones are important in defining the hydrochemical characteristics (Bittencourt *et al.* 2003). Given the lithologic characteristics of Serra Geral Formation, the SGAS water is classified as calcium and magnesium bicarbonate-rich (Fraga 1986) (Figure 2). However, the type of water can also be affected by mixing of water with the underlying aquifer (Figure 3). The time interval water stays in contact with soluble aquifer materials and is positively correlated with the total dissolved
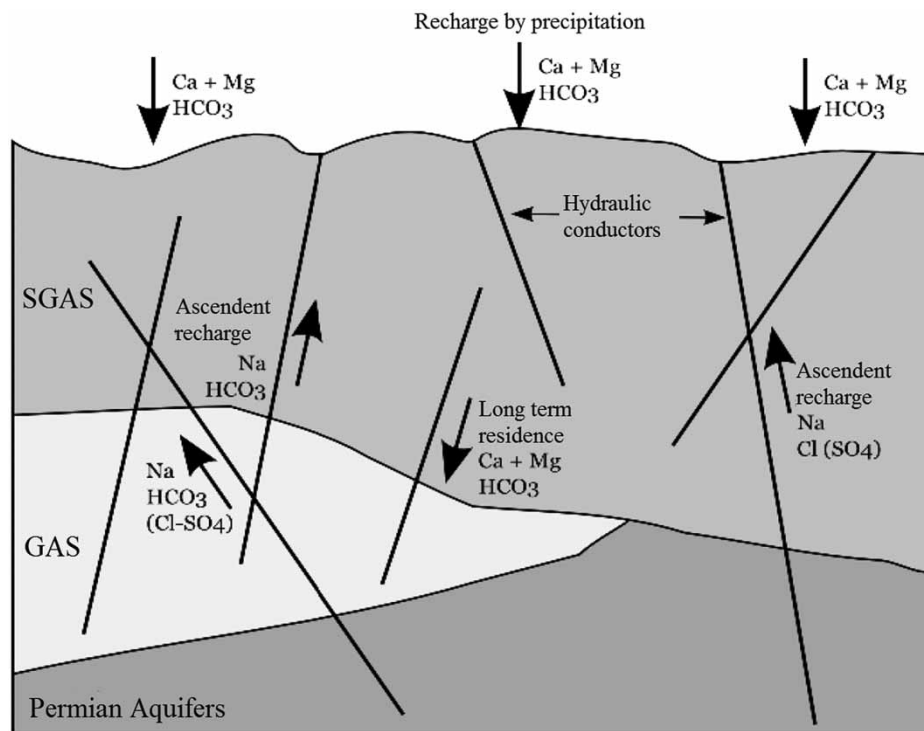

Figure 3 | Conceptual model of the connectivity between the Guarani (GAS) and the Serra Geral (SGAS) aquifer system, representing the influence of geological structures and hydrochemistry characteristics (Nanni *et al.* 2009).

solids (TDS) content. The hydrochemistry of the GAS is highly variable, especially in confined areas, where the water composition appears as facies variations, or by mixtures associated with sandstone fractures (Gastmans *et al.* 2010). In deep confined areas, the GAS waters are not suitable for public supply given the high content of TDS and concentration of sulfates and fluorides above the recommended limits for human consumption (Nanni *et al.* 2009). The waters in the GAS outcropping region is characterized as calcium bicarbonate-rich, changing to sodium bicarbonate with increasing concentrations of chloride and sulfate towards deeper confined areas (Rabelo & Wedland 2009). The variation in water composition is attributed to calcium–sodium exchange and carbonate dissolution resulting in a sodium bicarbonate groundwater. At least part of the sodium, chloride, and sulfate are attributed to dissolution of evaporates associated with the Pirambóia Formation (Gastmans *et al.* 2010).

METHODS

In phreatic aquifers, such as the SGAS, the surface hydrology and groundwater chemistry need to be investigated as coupled processes. Groundwater chemistry may be related to surface and near-surface hydrologic phenomena with variables such as relief, soil texture and magnetic properties useful as predictor variables. These variables can potentially describe the connection between near-surface phenomena such as soil chemical weathering (Iwashita *et al.* 2011) and transport processes, making it possible to predict groundwater quality (Souza Filho *et al.* 2010; Friedel *et al.* 2012) and groundwater recharge (James *et al.* 2010). The dataset used in this study is comprised of 27 variables, including relief features (5), variables derived from digital elevation model geoprocessing (2), airborne magnetic (1) and hydrochemical and well parameters (19) across 1,132 samples.

To model the spatial distribution of aquifer chemistry and hydraulic properties, several steps are adapted from in Iwashita *et al.* (2012) and presented in Figure 4. First, the well parameters acquired from pump tests were used to calculate hydraulic transmissivity using McLin's (2005) method. Second, standardization was used to ensure that no single
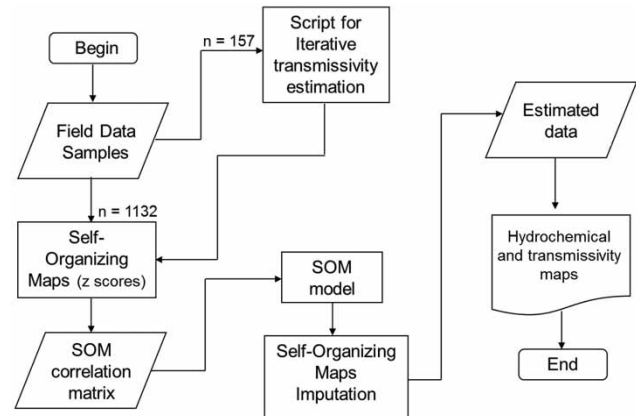
Figure 4 | Flowchart of the proposed method to imputate missing values for hydrochemistry and transmissivity data.

variable would dominate the vector-quantization process (Kalteth *et al.* 2008). The standardization is based on application of the z-score transformation given by:

$$z_i = \frac{x_i - \bar{x}_i}{s_i}$$

where z is the standardized value; x is the raw score; $\bar{x}$ is the sample average, and s is the sample standard deviation, i is an index for each variable.

Third, the cross component planes of the SOM weights (Kohonen 2001) are used to quantify nonlinear relations among all of the variables. The correlation matrix is calculated from a complete dataset after the missing values are estimated through the imputation process. The SOM data mining and component planes analysis is carried out using the SiroSOM (CSIRO Exploration & Mining) graphical user interface (GUI). This GUI provides an interface between data sets and available functions in the freely available SOM Toolbox (Vesanto *et al.* 2000). Fourth, the k-means clustering technique is used to classify the SOM into statistically relevant groups based on the Davies–Bouldin index (Davies & Bouldin 1977). Fifth, the hydrochemical elements are projected into a continuous surface to evaluate their spatial distribution. In the final step, the spatial pattern of chemical elements is compared to previous conceptual models established in the literature to identify regions in the study area where potential connections may exist between the aquifer systems.

Uncorrected Proof

6   F. Iwashita *et al.* | Using self-organizing maps to characterize hydrogeology of Serra-Geral aquifer   Hydrology Research | in press | 2017

Self-organizing maps

The estimation of environmental variables across the model domain is undertaken using the SOM technique (Kohonen 1984). The SOM can be considered a type of clustering technique based on vector quantization. The term self-organizing is based on the unsupervised nature of the algorithm having the ability to organize information without any prior knowledge of an output pattern. The basic process involves training, diversity, and estimation. The SOM training process provides a means of representing multidimensional data in lower dimensional space than the original data set (Kohonen 2001). The process of reducing the dimensionality is essentially a data compression technique known as vector quantization (Hastie *et al.* 2002). The output consists of neurons organized on a two-dimensional rectangular grid (map). Each neuron in the map is represented by a multi-dimensional weight vector M_{ij}, $i = 1,..., k_x$, $j = 1,...,k_y$, where in a rectangular SOM, k_x is the number of rows, and k_y, is the number of columns, and the dimension n of the neuron is the same as the number of input variables, $M_{ij} = \{m_{ij}^1, m_{ij}^2, \ldots, m_{ij}^n\}$. Each neuron is connected to the adjacent neuron through a functional neighborhood relation (Vesanto & Alhoniemi 2000). Individual data samples are associated to a vector with properties that reflect its contributions relative to the other variables. From this cloud of data vectors, a best matching unit (BMU) is iteratively determined by minimizing the Euclidean distance measure for each variable (Vesanto & Alhoniemi 2000; Kohonen 2001).

To quantify the success of this topology-preservation step, the network performance is analyzed by computing the quantization error:

$$E(G, X) = \frac{1}{N} \sum_{i \in Q} \sum_{j=1}^{M} h_{i,l} ||x_j - w_i||^2$$

where w_i are weight vectors assigned to a fixed number of N neurons in the map grid G, x_j are the M input data vectors, $h_{i,I}$ is a neighborhood function, $||x_j - w_i||$ is the Euclidian norm, and I is the best matching unit (BMU) vector:

$$||x - m_c|| = \min_i ||x - m_i||$$

where $||\circ||$ is the Euclidian distance, x is the input vector, m is the weight vector and c is the neuron whose vector is nearest to the input vector x.

The topographic error, ET, indicates how well the trained network keeps the topography of the data analyzed, it is a measure (percentage) of the number of node vectors that are adjacent in n-dimensional space, but are not adjacent on the resulting self-organized map. Computationally, the topographic error is given by:

$$E_{TE} = \frac{1}{m} \sum_{p=1}^{m} u(X_p)$$

If the neuron-winner of vector X_p is closest to the neuron, that is, the distance from X_p to it is the smallest one, then $u(X_p) = 0$, otherwise $u(X_p) = 1$.

The resulting maps are organized in such a way that similar data are mapped to the same or nearby nodes, and dissimilar data are mapped to nodes with greater separation distances. The estimates of variables are taken directly from the BMU vectors (Fessant & Midenet 2002; Wang 2003; Friedel 2016). This study uses an alternative estimation scheme where the associated best-matching unit vectors are supplied as initial values. The final values are arrived at iteratively based on minimization of an objective function comprising the topographical error vector and quantization error vector. The estimation of missing values (often referred to as imputation) is done simultaneously for all variables across the hypersurface (Kalteh & Hjorth 2009). For more details about SOM training and estimation refer to Kohonen (2001) and Vesanto & Alhoniemi (2000).

According to Kalteth *et al.* (2008) the map size, i.e. the number of nodes that will be projected into the map, plays an important role on the training process, as they determine the number of clusters where the samples will be assigned to. Vesanto *et al.* (2000) proposed a heuristic method to calculate the number of nodes based on a formula and the ratio between the two largest eigenvalues from the covariance matrix. However, this approach would not be practicable on a database with missing values or categorical variables. An alternative approach would be to find a suitable map dimension with small topographical error.

Uncorrected Proof

7   F. Iwashita *et al.* | Using self-organizing maps to characterize hydrogeology of Serra-Geral aquifer   **Hydrology Research** | in press | 2017

The map is characterized by 22 by 16 nodes (352 neurons), and during the initial step, the rough training was performed using a Gaussian neighborhood with initial and final radius of 28 and seven units, respectively, for 30 iterations. The fine training step had 600 iterations and used Gaussian neighborhood with initial and final radius of seven and one units, respectively.

Hydrochemical data

The hydrochemical data are provided by government institutions responsible for water monitoring and analysis in the State of Parana; these agencies include the Waters Institute (Aguas Parana), in charge of providing authorizations for groundwater extraction, and the Sanitation Company (SANEPAR), that collects, treats and provides water for human consumption. The dataset collected and analyzed by these two institutions from 1997 to 2008 is comprised of 27 variables sampled from 1,132 wells (Table 1), including major cations and anions and well-field parameters. Specific dates of collection can be found in Portela Filho (2003), Barros (2007) and Mocelin (2009). The entire dataset is public and available online, however, the documentation detailing analytical methods was not disclosed. Additionally, given the constant advancements in chemical analytical instruments and development of new protocols, such databases often differ in which set of elements are analyzed, and in the accuracy and limit of detection.

Hydraulic transmissivity data

The Serra Geral aquifer transmissivity is estimated using specific capacity data calculated from pumping tests conducted in 157 wells by the Águas do Parana Institute. The

Table 1 | Acronyms for employed variables and their respective unit

Category	Description	Acronym	Unit	Samples	Estimated
Morphometric	Aspect	cosAspect	–	1,132	0
	Elevation	ELEV	m	1,132	0
	Horizontal curvature	HOR-CURV	^{o}/m	1,132	0
	Slope	SLOPE	degrees	1,132	0
	Vertical curvature	VERT-CURV	^{o}/m	1,132	0
GIS	Flow accumulation	FLOWACC	integer	1,132	0
	Distance from lineaments	LINEADIST	m	761	371
Geophysical	Aeromagnetic	AEROMAG	hT	1,126	6
Well information	Potentiometric level	POTENC	m	1,025	107
	Depth	DEPTH	m	1,074	58
	Drawdown	DRAWDOWN	m	1,008	124
	Yeld	YELD	m^3/h	1,029	103
	Specific capacity	CAPAC	m^3/h·m	1,007	125
	Transmissivity	Transm	l/m	157	975
Hydrochemical	Calcium	Ca	mg/L	332	800
	Magnesium	Mg	mg/L	318	814
	Sodium	Na	mg/L	259	873
	Potassium	K	mg/L	255	877
	Chloride	Cl	mg/L	255	877
	Sulfate	SO4	mg/L	197	935
	Carbonate	HCO3	mg/L	54	1,078
	Bicarbonate	CO3	mg/L	287	845
	pH	pH	mg/L	198	934
	Total dissolved solids	TDS	mg/L	205	927
	Free CO2	CO2	mg/L	75	1,057
	Nitrate	NO3	mg/L	125	1,007
	Carbonate-bicarbonate	CO3-HCO3	mg/L	163	969
	Fluoride	F	mg/L	92	1,040

Uncorrected Proof

8   F. Iwashita *et al.* | Using self-organizing maps to characterize hydrogeology of Serra-Geral aquifer   Hydrology Research | in press | 2017

calculations are based on a modified version of the Bradbury–Rothschild iterative solution technique (Bradbury & Rothschild 1985) and adapted using MATLAB (McLin 2005):

$$T = \frac{Q}{4\pi(s_t - s_w)}\left[\ln\left(\frac{2.25Tt}{r_w^2 s}\right) + 2s_p\right]$$

where T = aquifer transmissivity (m^2/d), Q = well discharge (m^3/min), s_t = is total drawdown observed in a production well, s_w = drawdown due to well loss S = aquifer storage coefficient (dimensionless), t = time since pumping began (min), r_w = effective wellbore radius (cm), and s_p = a partial penetration factor (dimensionless).

An iterative solution is required to calculate T (Bradbury & Rothschild 1985). An initial guess is proposed for T on the right-hand side of the equation, and an updated solution for T is obtained from the left-hand side. This updated solution is then used on the right-hand side of the equation, and a new T is computed. This iterative process continues until a predefined tolerance criterion is reached (McLin 2005). These calculated well transmissivities are then incorporated into the used training set for the SOM. The computed model estimates missing transmissivity values at places where the pumping tests information is incomplete.

Aeromagnetic data

The aeromagnetic data are provided by PETROBRAS and pre-processed by the Applied Geophysics Laboratory of Research (LPGA) at the Parana Federal University. The raw data comprise a series of aerial surveys conducted in the Parana Basin during the 1980s. This survey is conducted with NS-trending lines spaced at 2 km, a flight height of 500 m and sampling intervals of approximately 100 m. Control lines spaced at 20 km and perpendicular to the acquisition lines also are acquired during the survey. The pre-processing comprised the generation of regular grids (500 × 500 m) by the minimum curvature method (Briggs 1974). The residual magnetic field is then calculated, and artifacts (noise) along the flight lines eliminated using the micro-leveling technique (Minty 1991). Thus, the final magnetic data set represent the micro-leveled anomalous magnetic field (nT).

Topographical dataset

Characterization of the topographic relief is possible using elevation data provided by the Shuttle Radar Topographic Mission (http://edcsns17.cr.usgs.gov/NewEarthExplorer/). The digital elevation model associated with these data are provided by the United States Geological Survey on a lattice with 90-m spatial resolution (Farr & Kobrick 2000). The Topodata project, conducted by the Brazilian National Institute for Space Research-INPE (Valeriano *et al.* 2009), created derived geomorphometric metrics with a 30-m resolution. The geomorphometric features provide a way to extract morphometric features, such as slope, aspect (hillslope orientation), vertical and horizontal curvature, and accumulated hydrological flux (Jenson & Domingue 1988).

Model evaluation

The model framework is evaluated using a Bootstrap approach (Kohavi 1995). The basis of the Bootstrap cross-validation is the leave-one-out strategy. This strategy requires leaving one data value out of the training set while creating a new SOM which is then used to estimate the missing value. Because a new SOM is created up to 30 times for each value under scrutiny, it forms the basis for the stochastic framework from which residuals are used to evaluate error statistics and model bias (Figure 5). The bootstrap is carried out according to the following steps: (a) The SOM vector framework is calculated using the entire dataset; (b) The first sample of the chosen dependent variable to be evaluated is extracted from the dataset; (c) The SOM framework is calculated again and the missing value is estimated; (d) The previous step is repeated for each sample of the set (number of samples vary for each analyzed element); (e) Steps c and d are repeated 30 times for each variable, i.e. the model has to run approximately 50,000 times; (f) The residues of predicted values are analyzed and the average of predictions of 30 runs is compared to the observed values.

RESULTS AND DISCUSSION

Limited data availability and high spatial variability of the data promote increasing amounts of uncertainty in model
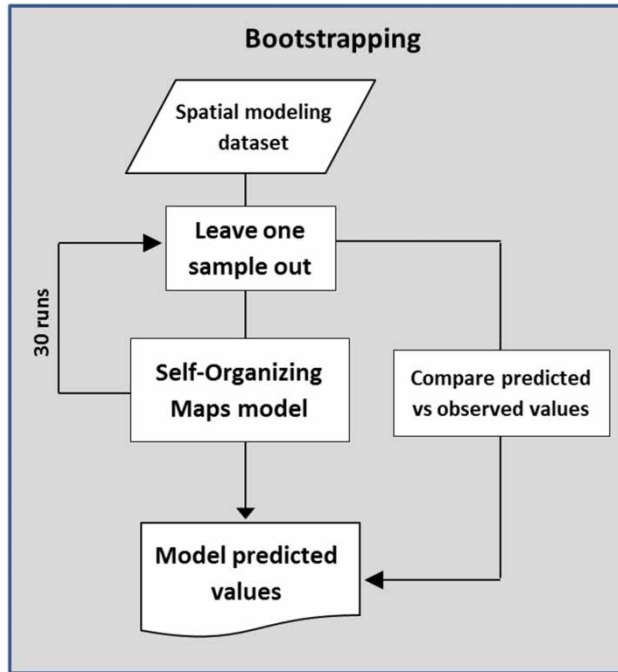
Uncorrected Proof

9   F. Iwashita *et al.* | Using self-organizing maps to characterize hydrogeology of Serra-Geral aquifer   Hydrology Research | in press | 2017

Figure 5 | Flowchart for the Bootstrap model evaluation.

predictions (Hornberger 1998). Scarce data sets can result in biased predictions (Dickson & Giblin 2007) requiring a modified scheme based on bootstrapping (Breiman 1996). The SOM algorithm is objective, but there is subjectivity when choosing the set of data variables as potential predictors, and the samples are spatially limited with varying levels of uncertainty in their measurements and observations. For these reasons, the reliability of the SOM as a model to predict soil geochemical variables is evaluated using cross-validation.

In this study, the SOM appears to be an unbiased estimator as indicated by the one-to-one correspondence and constant variance for calcium, bicarbonate, sodium, sulfate, TDS, fluoride, chloride, nitrate and hydraulic transmissivity (Figure 6). A comparison of median values, as most elements present a lognormal distribution, between observed and estimated values is presented in Table 2. Additionally, it is possible to analyze the consistency of the relation between the estimated elements using a piper diagram (Figure 7). Modeled results are in correspondence with the piper diagrams
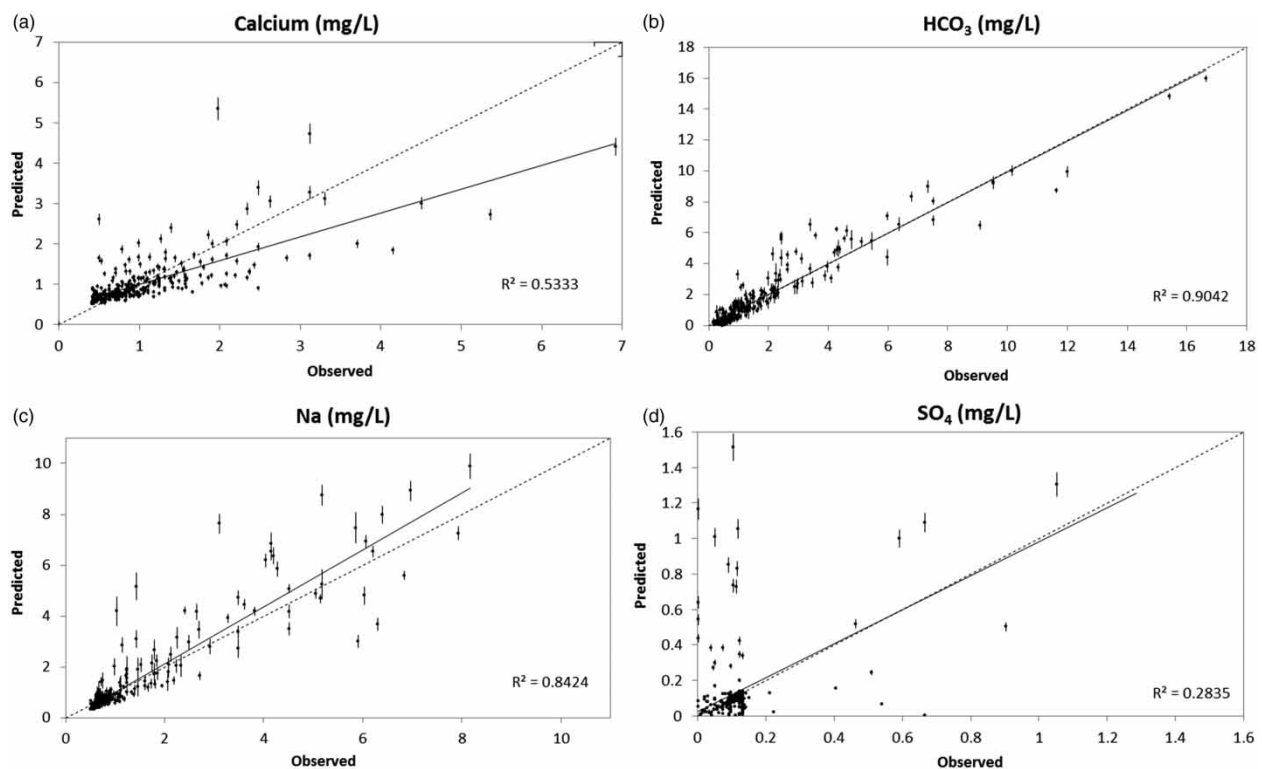

Figure 6 | Model evaluation plots using a Bootstrap approach, one to one correspondence for reference purpose (dashed line) and average predicted values from 30 runs for each sample versus observed values for (a) total dissolved solids, (b) fluoride, (c) chloride, (d) nitrate, (e) calcium, (f) bicarbonate, (g) sodium, and (h) sulfate. *(continued.)*
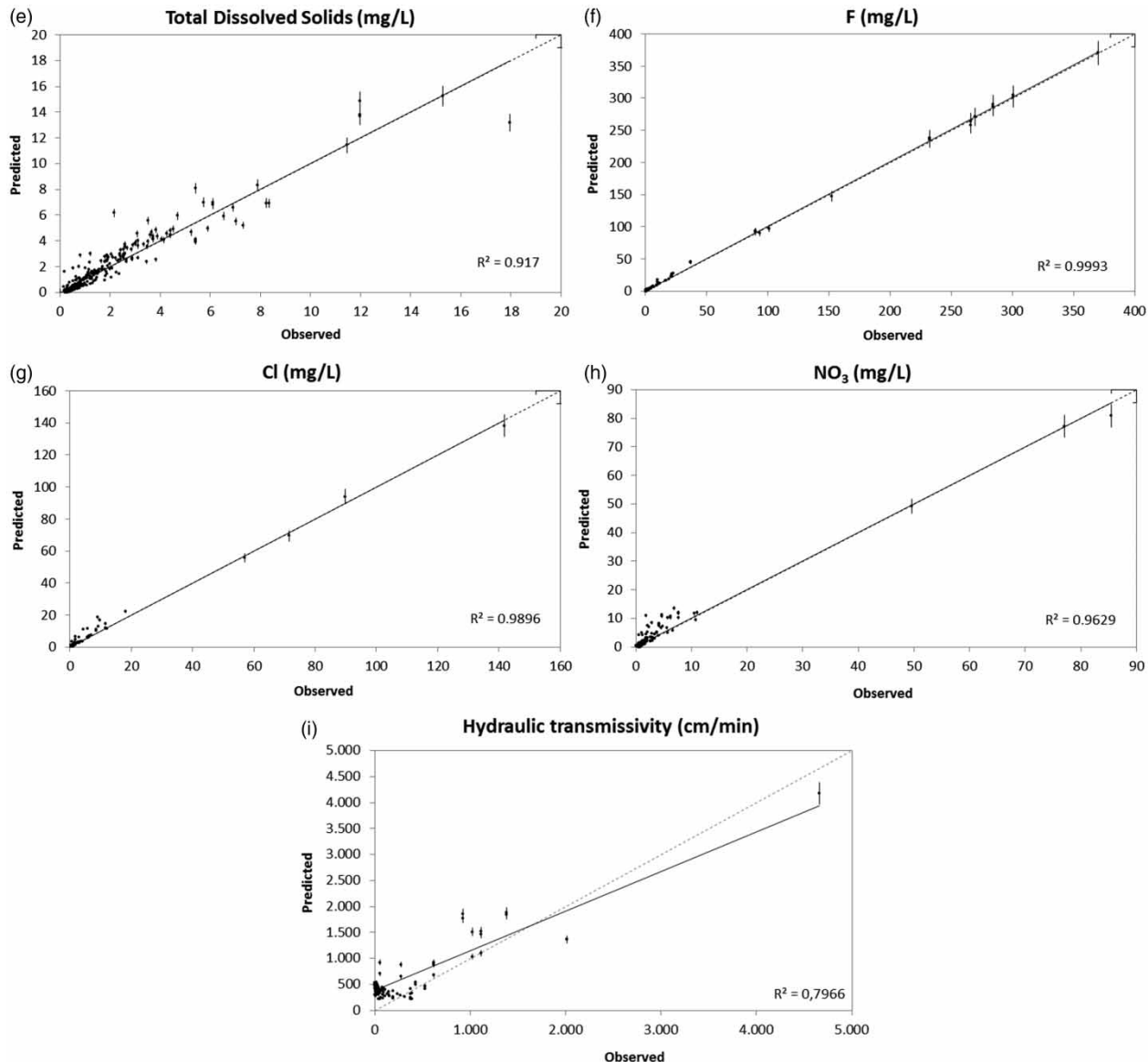
Uncorrected Proof

10   F. Iwashita *et al.* | Using self-organizing maps to characterize hydrogeology of Serra-Geral aquifer   **Hydrology Research** | in press | 2017

Figure 6 | continued

obtained by Athayde *et al.* (2007) and Nanni *et al.* (2013) indicating that relationships among estimated elements are preserved. In other words, the estimated values maintained their statistical relations, for example, high concentrations of Ca are associated with high concentrations of Mg. This is an important indication of robustness associated with the proposed modeling framework despite the high number of sparse explanatory variables.

To identify potential areas of hydraulic connectivity between SGAS and GAS, the spatial distribution of

hydrochemical variables are compared to information described in the literature. Under favorable potentiometric conditions (Figure 1), the waters from the GAS ascend through geological structures (open fault planes) to the SGAS modifying the typical hydrochemical signature of the aquifer (Nanni *et al.* 2009).

The component planes (Figure 8) reveal interesting aspects of the training data that include correlation, dissimilarity, and grouping. Similarity in the color patterns, such as Ca and Mg, indicate a strong positive correlation. This is

Uncorrected Proof

11   F. Iwashita *et al.* | Using self-organizing maps to characterize hydrogeology of Serra-Geral aquifer   Hydrology Research | in press | 2017

Table 2 | Comparison of median values between observed versus estimated values

	Median values	
	Observed	Estimated
Ca mg/L	13.00	11.21
Mg mg/L	3.38	2.73
Na mg/L	9.90	8.64
K mg/L	0.60	0.61
Cl mg/L	1.85	1.88
SO4 mg/L	1.00	1.18
HCO3 mg/L	71.29	64.97
NO3 mg/L	0.26	0.51
F mg/L	0.10	0.11
TDS mg/L	121.00	121.52
Transm cm/min	118.98	167.07

an interesting asset for exploratory analysis, especially when supported by the correlation matrix (Table 3) calculated after the topological evaluation. For example, calcium and magnesium have a correlation of 0.97. Sodium, chloride and sulfate are all positively correlated according to the component plots. Sulfate correlates 0.85 and 0.52 with sodium and chloride, respectively. The high correlation between calcium and magnesium is justified by the fact that both are products from dissolution of basaltic rocks forming minerals. Hydrochemical facies from SGAS, with longer residence times, are calcic-bicarbonate or calcic-magnesium bicarbonate (Fraga 1986). Sodium bicarbonate-rich waters differ in the composition to solutions formed due to leaching of the Serra Geral Formation basaltic rocks. The sodium content is attributed to several sources, for example, the alteration of albite, input by diffusion loading of halite and mirabilite weathering from GAS (Sracek & Hirata 2002). Therefore, anomalous concentrations of sodium bicarbonate on the SGAS waters may be related to GAS, indicating a hydraulic connection between the two systems. The bicarbonate anion is the most abundant in the SGAS and GAS. This anion is usually originated from the dissolution of carbon dioxide mutually present in the atmosphere and soil, reacting with percolating waters or from basalt silicates hydrolysis. Thus, the low bicarbonate concentrations are linked to either recent recharged waters or with waters with a long residence time (from silicate weathering) (Bittencourt 1978).

Component plane plots of pH, carbonate and bicarbonate show partial inter-correlation (Figure 8). The carbonate-bicarbonate content is related to pH of the solution. In neutral and weakly alkaline conditions, the presence of bicarbonate is higher than the carbonate. From $pH = 8.30$, the concentration of carbonate increases gradually until it replaces the presence of bicarbonate (Mocellin 2009). For the SGAS, pH alkaline values are attributed to the influence of groundwater fluxes from the GAS aquifer. This is because with increasing alkalinity and pH values there is a carbonate imbalance leading to calcium depletion, causing an increase in sodium concentration (Silva 2007). There is a noteworthy negative correlation between fluoride (−0.19), TDS (−0.44) and chloride (−0.26) concentrations and distance from the nearest lineament. Areas near lineaments are associated with higher concentrations of certain elements, strengthening the hypothesis of the role of structural conditions in control of waters from GAS. The same table shows a positive correlation among fluoride, sodium, chloride and sulfate; all considered typical elements of the GAS.

The purpose of generating continuous surfaces by a simple interpolation method is to analyze the spatial behavior of the imputation method results. Therefore, the interpolated surface should not be interpreted as estimation. In addition, the choice of a simple interpolation method like inverse-distance weighting allows a visualization of the influence from each sample on the continuous surface, especially of those anomalous values that may cause the bull's eye effect. Another assistance for a critical spatial analysis are different representations for samples with measured points and points where values are estimated, illustrating the representatives of training points.

For the purpose of discussion of the results, from hereby the study area will be divided into three subareas as shown in Figure 9(a). The calcium map (Figure 9(a)) shows low concentrations in most part of subarea II and in the northeast and southeast sectors of subarea III. Regions with low levels of calcium may be related to a possible connection with GAS, due to the proportional decline of this element. A low content of calcium can also be an indication of areas with meteoric waters, especially if regions with low calcium content coincide with low levels of bicarbonate (Figure 9(b)), which according to Fraga (1986), are associated with recent recharge areas. In contrast, high
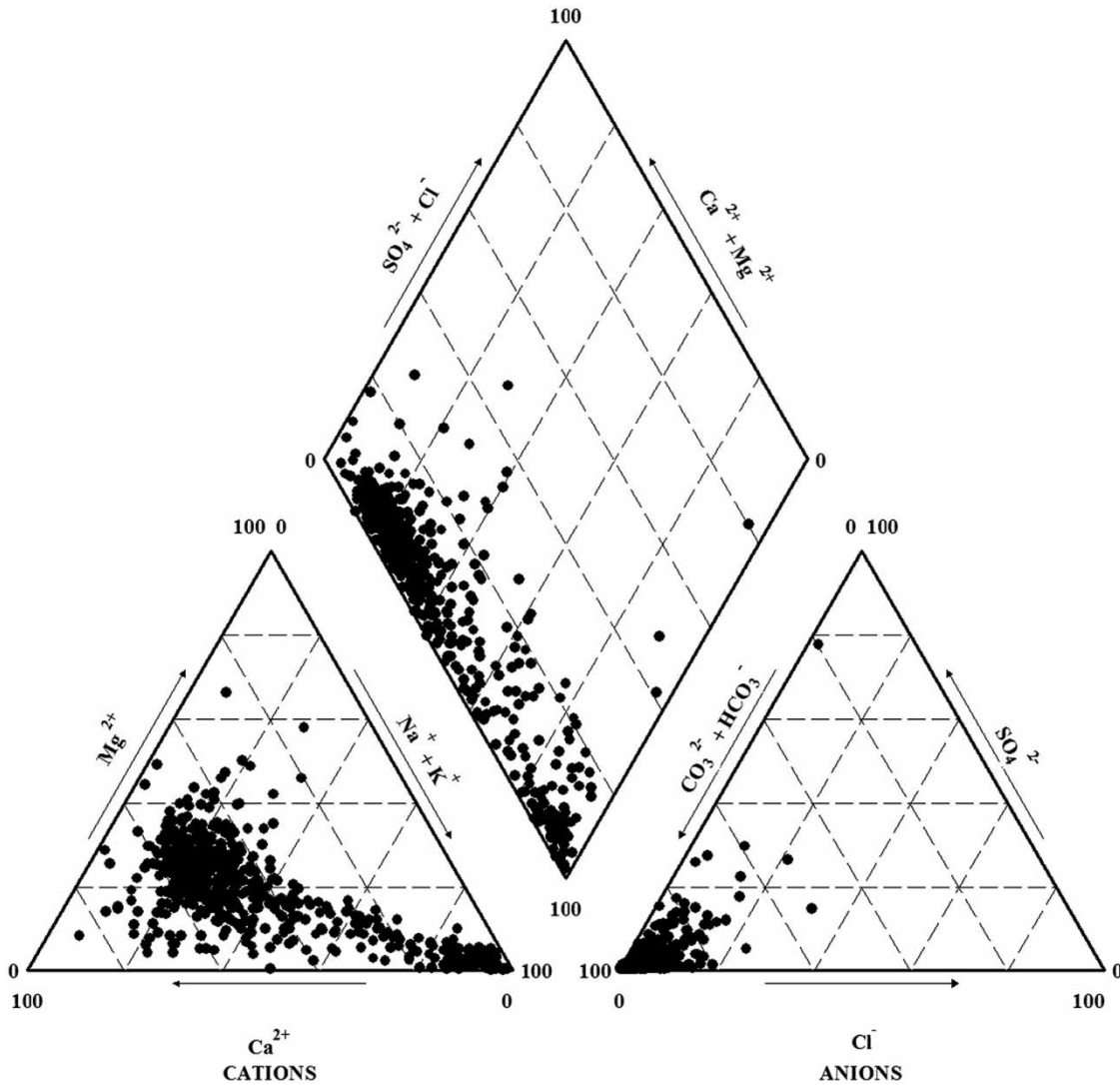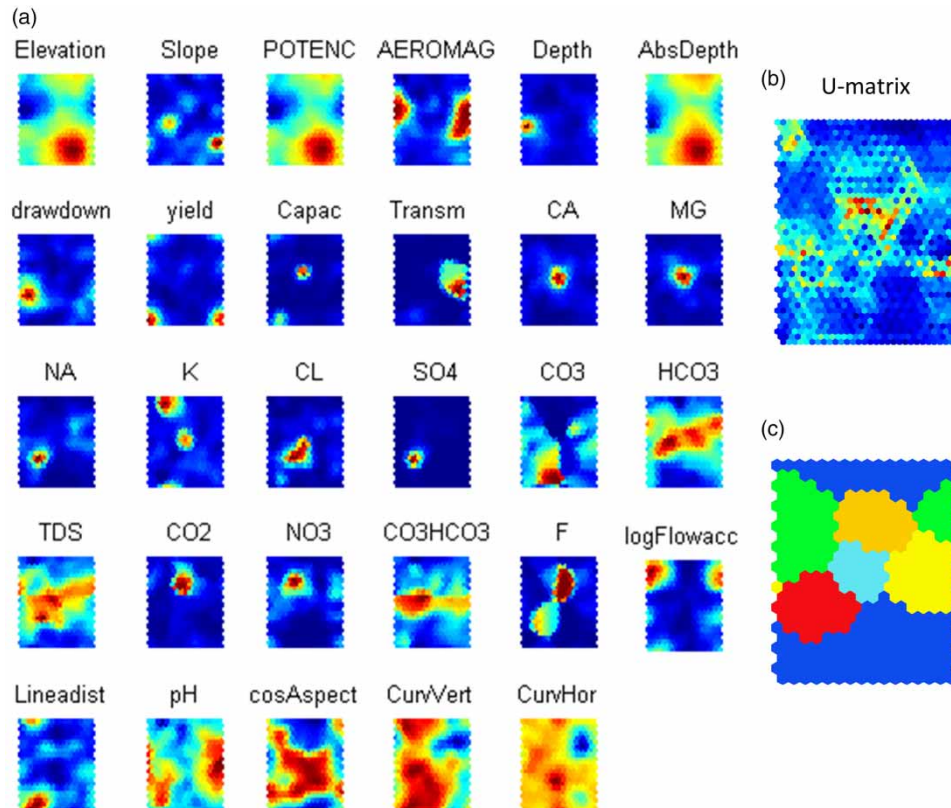
Uncorrected Proof

12   F. Iwashita et al. | Using self-organizing maps to characterize hydrogeology of Serra-Geral aquifer   **Hydrology Research | in press | 2017**

Figure 7 | Piper diagram with estimated values by SOM.

concentrations of calcium and TDS is commonly an indication of waters with long residence. Likewise, high levels of bicarbonate could indicate a connection with the GAS or a confined section of the SGAS; when these elements associated with long residence are considered together, such as calcium and bicarbonate, it is possible to identify areas across the northeast of subarea II where they are positively correlated, with high content values, pointing out to water confinement. The southern part of subarea III shows low values for calcium and high values for bicarbonate, indicating an upward flow originated from the GAS to SGAS. To the southeast of subarea III, evidences for a hydraulic

connection between the two aquifer systems are strengthened by the observation of high sodium and sulfate content (Figure 9(g) and 9(h)), taken as characteristic of GAS waters. The effusive rocks of the Parana basin are sulfide-poor, as in other forms of sulfur. Higher sulfur contents are attributed to contamination from underlying aquifers or mineralized, pyrite-rich intrusions (Bittencourt et al. 2003).

The TDS comprise the sum of all the present mineral constituents in solution, having a direct relationship with the mineralogical rock composition and the time of groundwater percolation/residence within the system, thus reflecting the chemical weathering of rock forming minerals.

Uncorrected Proof

13   F. Iwashita *et al.* │ Using self-organizing maps to characterize hydrogeology of Serra-Geral aquifer   **Hydrology Research** │ **in press** │ **2017**

Figure 8 │ Component planes from SOM used to visualize nonlinear correlation. All variables were standardized using z-score, (b) U-Matrix. (c) U-Matrix classified using k-means technique.

A typical feature of confined waters is high content of TDS, caused by a long period of residence, whereas in SGAS waters, high TDS is another sign of connection between the two systems (Fraga 1986). These indications are reinforced when high TDS is followed with high concentrations of other elements such as fluoride and favorable hydraulic conditions, detectable through differences in potentiometric levels, leading to an upward flow from GAS.

The northern section of subarea II and majority of subarea III in its northeastern portion (Figure 9(e)) display high levels of TDS, potentially indicating either the aquifer's connectivity or confinement. The low TDS content plus low levels of calcium and bicarbonate observed at the central sector of subarea II reinforces the hypothesis of areas with recent recharge. These same areas have lower concentrations of fluoride, which has a positive correlation with sodium, sulfate and chloride, characteristic elements found in the GAS.
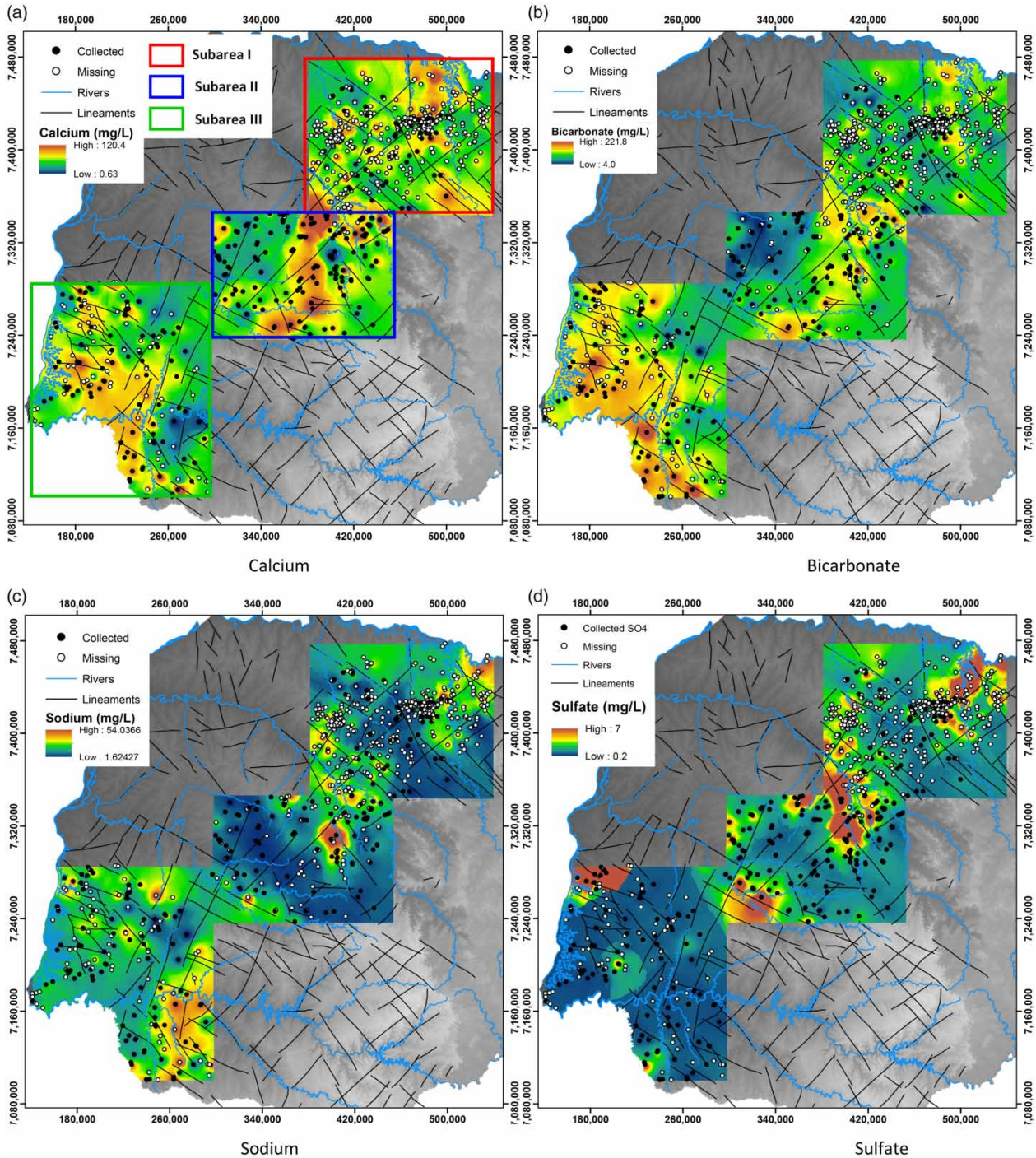
For Fraga (1986), the presence of fluoride in the SGAS is associated with upward flow of alkaline waters from the GAS. By contrast, Nanni *et al.* (2009) argued that the origin of fluorine needs further investigation because elevated concentrations could be the result of the SGAS secondary mineral weathering. The high concentrations of fluoride in the northern portion of subarea I is spatially coincident with anomalously high values found in surface waters (Licht 2001), while the fluoride content in groundwater is attributed to deep geological structures, features that can be captured by aeromagnetic survey. Fluoride transported by upward flow from the GAS is connected to high fracture density, thereby facilitating the flow from one aquifer to the other. High fluoride values could be enhanced by surficial processes, such as the presence of a thick layer of soil and a high proportion of clay preventing the recharge of the SGAS (Nanni *et al.* 2009).

The land use and physical features also influence the chemical characteristics of recharging waters. Nitrate (Figure 9(g)) usually has low concentrations in natural waters, but high concentrations of nitrates in wells may

Table 3 | Correlation matrix calculated from SOM after the imputation process

	Log flow acc	Curv-Hor	Curv-Vert	cosAspect	pH	Lineadist	F	CO3HCO3	NO3	CO2	TDS	HCO3	CO3	SO4	Cl	K	Na	Mg	Ca	Transmiss	Capac	yield	drawdown	AbsDepth	Depth	AEROMAG	POTENC	Slope
logFlowacc	1																											
CurvHor	−0.26	1																										
CurvVert	−0.19	0.74	1																									
cosAspect	−0.09	0.34	0.27	1																								
pH	0.04	−0.33	−0.34	0.2	1																							
Lineadist	0.24	0.02	0.26	−0.02	0.06	1																						
F	−0.34	−0.05	−0.12	−0.11	−0.12	−0.19	1																					
CO3HCO3	−0.26	0.05	−0.03	0.41	0.37	−0.31	0.16	1																				
NO3	0	0.02	−0.16	−0.1	−0.06	−0.12	0.21	0.39	1																			
CO2	−0.17	−0.03	−0.19	−0.36	−0.44	−0.18	0.62	−0.01	0.61	1																		
TDS	−0.23	0.01	−0.09	0.52	0.39	−0.44	0.34	0.78	0.32	0.02	1																	
HCO3	−0.1	−0.19	−0.28	0.41	0.4	−0.41	0.23	0.84	0.46	0.1	0.9	1																
CO3	−0.12	0.22	0.35	0.31	0.32	0.42	0.02	−0.01	−0.32	−0.41	0.12	−0.1	1															
SO4	−0.17	0.18	0.3	0.08	0.24	−0.06	0.37	0.15	−0.04	−0.11	0.34	0.09	0.34	1														
Cl	−0.31	0.37	0.3	0.45	0.08	−0.26	0.38	0.32	−0.12	−0.11	0.61	0.3	0.38	0.52	1													
K	0.26	0.21	0.38	−0.01	−0.52	0.14	0.02	−0.1	0.04	0.14	−0.11	−0.09	−0.29	−0.13	0.07	1												
Na	−0.1	−0.11	0.02	0.17	0.61	−0.09	0.34	0.43	0.02	−0.19	0.61	0.44	0.37	0.85	0.47	−0.29	1											
Mg	−0.28	0.13	−0.01	0.43	−0.19	−0.34	0.38	0.58	0.34	0.27	0.65	0.61	−0.13	−0.03	0.51	0.3	0.01	1										
Ca	−0.25	0.15	0	0.47	−0.14	−0.37	0.35	0.57	0.24	0.18	0.66	0.6	−0.12	−0.01	0.6	0.33	0.05	0.97	1									
Transmiss	−0.08	−0.25	−0.22	0.17	0.54	0.09	−0.07	0.33	−0.03	−0.11	0.24	0.36	0.11	−0.13	−0.12	−0.19	0.22	−0.03	−0.03	1								
Capac	0.09	0	0.06	0.1	−0.17	0.3	0.18	0.16	0.16	0.15	0.1	0.1	−0.08	−0.1	0.09	0.37	−0.12	0.53	0.48	−0.07	1							
yield	0.29	−0.03	−0.11	0	0	0.42	−0.24	−0.15	0.02	−0.13	−0.26	−0.16	−0.04	−0.18	−0.15	0.02	−0.21	−0.12	−0.09	−0.11	0.36	1						
drawdown	−0.02	0.07	0.03	0.08	0.51	−0.03	0.05	0.14	−0.09	−0.32	0.2	0.05	0.26	0.53	0.21	−0.3	0.6	−0.24	−0.18	−0.11	−0.24	−0.07	1					
AbsDepth	−0.34	0.05	0.09	0.05	−0.27	−0.05	−0.12	−0.41	−0.57	−0.11	−0.41	−0.47	−0.04	−0.13	0.04	−0.03	−0.35	−0.08	−0.06	0.03	−0.12	−0.11	−0.22	1				
Depth	−0.02	0.07	0.1	0	0.46	0.09	0.04	0.39	0.01	−0.25	0.25	0.2	0.26	0.34	0.13	−0.2	0.5	−0.12	−0.08	0.02	−0.05	−0.01	0.66	−0.49	1			
AEROMAG	0.14	−0.17	−0.35	0	0.67	−0.08	−0.29	0.35	0.17	−0.2	0.27	0.34	0.12	−0.11	−0.16	−0.34	0.2	−0.15	−0.14	0.49	−0.15	0	0.17	−0.38	0.32	1		
POTENC	−0.31	0.05	0.1	0.06	−0.13	0.02	−0.16	−0.36	−0.63	−0.24	−0.42	−0.48	0.04	−0.05	0.03	−0.11	−0.24	−0.19	−0.14	0	−0.15	−0.05	0.02	0.94	−0.18	−0.32	1	
Slope	−0.21	0.12	0.02	0.2	0.21	0.04	0.02	0.2	−0.06	−0.09	0.07	0.12	0.1	0.13	0.13	−0.12	0.08	0.03	0.05	−0.08	−0.05	0.03	0.2	0.18	0.17	0.05	0.28	1
Elevation	−0.38	0.07	0.13	0.05	−0.11	−0.02	−0.12	−0.31	−0.64	−0.24	−0.38	−0.45	0.06	−0.01	0.08	−0.12	−0.19	−0.16	−0.11	0.03	−0.16	−0.12	0.03	0.93	−0.13	−0.29	0.99	0.27

Uncorrected Proof

15   F. Iwashita *et al.* │ Using self-organizing maps to characterize hydrogeology of Serra-Geral aquifer   **Hydrology Research │ in press │ 2017**

Figure 9 │ Imputed values (white dots) and sampled wells (black dots) with respective continuous surfaces calculated by the inverse distance weighting (IDW) method. (a) Calcium (mg/L), (b) Bicarbonate (mg/L), (c) Sodium (mg/L) and (d) Sulfate (mg/L), (e) Total dissolved solids (mg/L), (f) Fluoride (mg/L), (g) Nitrate (mg/L), (h) Chloride (mg/L), (i) Specific capacity (m³/h/m) and (j) classified clusters. *(continued.)*

result from direct infiltration of surface water or from polluted water percolating through soil into the phreatic aquifer. Nitrate has a high spatial variability. In many

groundwater systems nitrate is unlikely to have a relation with geological formations. Natural waters may contain large quantities of nitrate without causing serious health
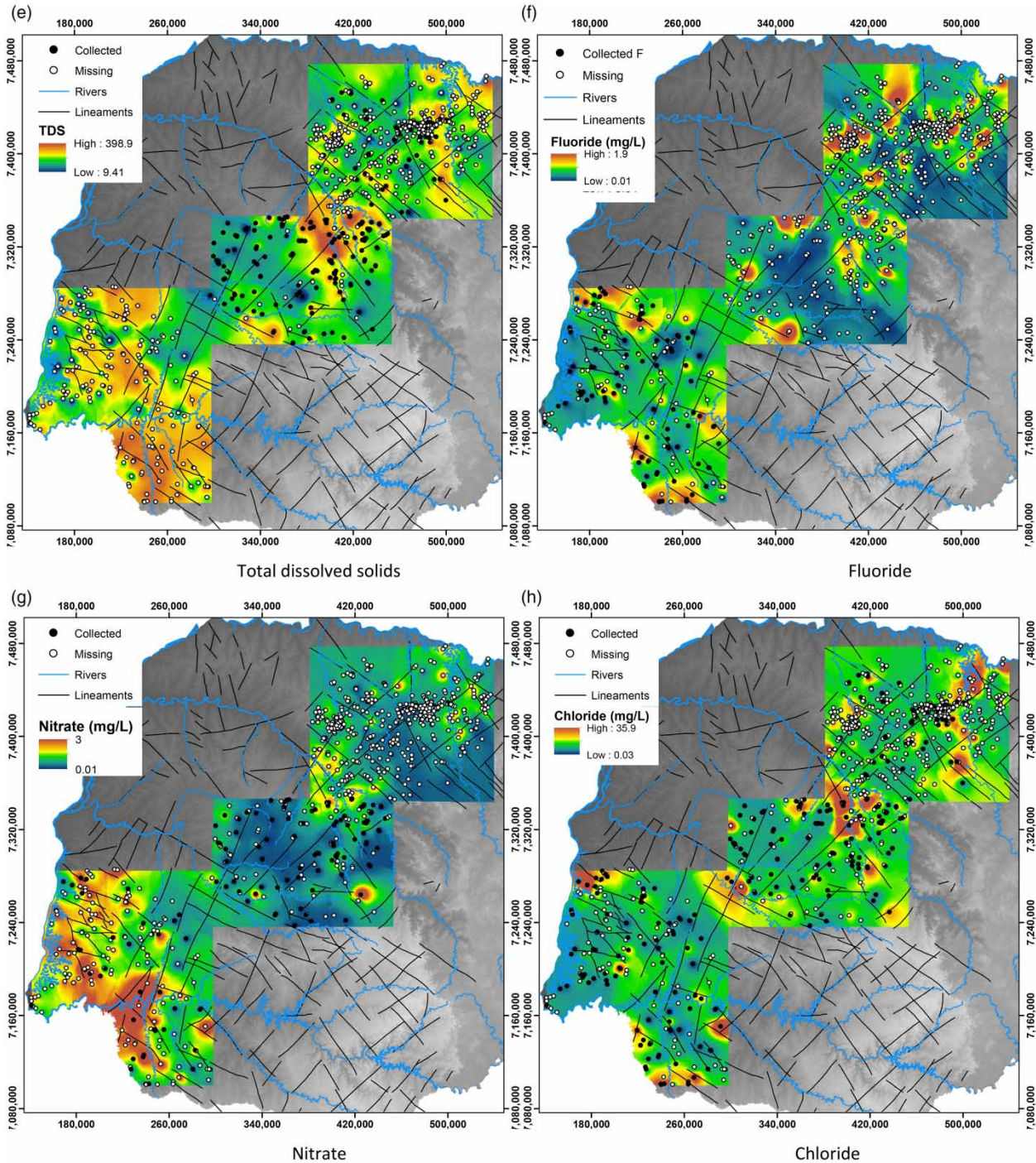
Uncorrected Proof

16   F. Iwashita *et al.* | Using self-organizing maps to characterize hydrogeology of Serra-Geral aquifer   **Hydrology Research** | in press | 2017

Figure 9 | Continued.

problems, but levels exceeding 5 mg/L represent an indicator of possible contamination by animal wastes or fertilizers (Rebouças & Fraga 1988). The high nitrate content

observed in the northeast and southwest portions of subarea I is attributed to surface contamination, given the proximity to two urban centers (Londrina and Maringá). Furthermore

Uncorrected Proof

17   F. Iwashita *et al.* | Using self-organizing maps to characterize hydrogeology of Serra-Geral aquifer   **Hydrology Research** | in press | 2017

transmissivity

Clusters
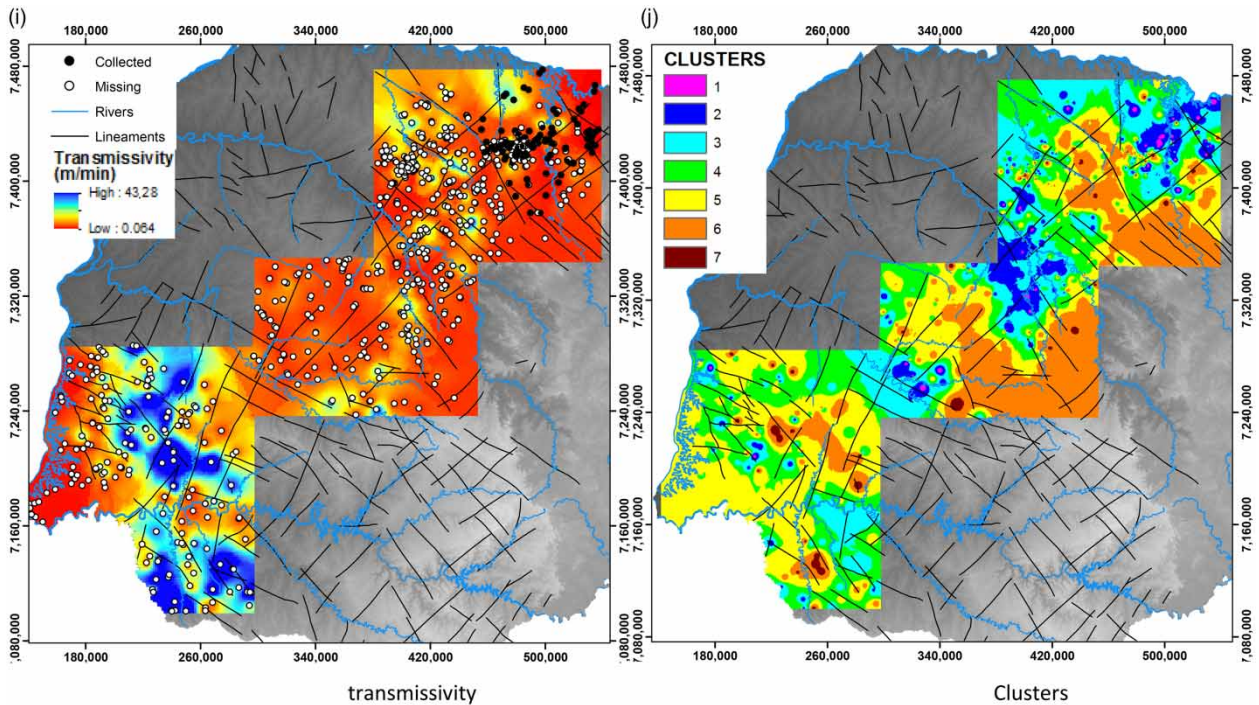
Figure 9 | Continued.

the area is intensely cultivated with cotton, coffee and soy (Licht 2001).

An alternative to trace-surface contamination is the chloride content (Figure 9(h)). Chloride is considered a highly mobile ion through most aquifer systems. Its source can be either anthropogenic or natural. In the GAS, the chloride is likely associated with evaporitic rocks and from weathering of micas present in the Pirambóia and Botucatu Formations (Gastmans *et al.* 2010), while in the SGAS, the chloride reflects surface intakes, weathering of basalt secondary minerals, such as chlorite and potential upward flow from GAS. The non-reactive characteristics of chloride make hydraulic properties as support variables useful to analyze its spatial content variation. In Table 3, chloride has a positive correlation with fluoride (0.38) and TDS (0.61) suggesting long-term residence waters. Additionally, chloride has negative a correlation with distance to lineaments (−0.26) associated with vertical faults reinforcing the likelihood for hydraulic connections with the GAS.

The specific capacity refers to how much the water level decreases as a function of a given yield rate, i.e. it describes the aquifer's capacity for water supply and storage. In a fractured aquifer, the specific capacity is related to the density of structural discontinuities. The hydraulic transmissivity, presented in Figure 9(i), in the northern portion of the Serra Geral aquifer (Fraga 1986) exhibits low transmissivity, whereas the southern Serra Geral aquifer exhibits higher values in areas close to central parts of the Parana basin likely to experience preferential flow under high potentiometric gradients. The potentiometric level can be used to describe the hydrogeological flow direction in isotropic aquifers, like the GAS, which is a porous aquifer hosted by the Botucatu and Pirambóia Formations. However, considering the anisotropic feature of the SGAS at a regional scale, the low potentiometric level could indicate a preferred path taken by the groundwater.

The variability of transmissivity depends on several factors, including potentiometric level and negative pressure. The horizontal movement of water is conditioned by the presence of discontinuities caused by the horizontal heterogeneity originated from a series of overlapping outflows (Rebouças & Fraga 1988). The transmissivity map (Figure 9 (i)) represents mostly estimated values. The training dataset is located in the northeastern portion of subarea I, where

Uncorrected Proof

18   F. Iwashita *et al.* | Using self-organizing maps to characterize hydrogeology of Serra-Geral aquifer   Hydrology Research | in press | 2017

data variability is high and suitable for SOM training to prevent overfitting (Rallo *et al.* 2002). The calculated hydraulic transmissivity supports hydrochemical analysis and is a parameter for numerical modeling of groundwater flow and solute transport.

The cluster map (Figure 9(j)) summarizes U-matrix values. The U-matrix is a bi-dimensional representation for dissimilarities of n-dimensional code vectors (Ultsch 2003). K-means clustering was performed over the U-matrix, while the Davies–Bouldin Index pointed out that seven clusters would present a good balance between diversity and model simplicity. Clusters one, two and three represent areas with potential connectivity between the SGAS and the GAS. The three clusters occur in areas with high TDS, sodium, chloride and sulfate, reinforcing the suggestion of connection. Specifically, cluster one is associated with outliers of sulfate and chloride. Cluster four corresponds to transitions between waters with typical Serra Geral hydrochemical facies and areas indicating connectivity with the Guarani aquifer. Cluster five has characteristics of long residency (high TDS) but without elements often associated with Guarani waters. Cluster six (Figure 9(j)) is quite predominant throughout Subareas I and II, it relates to a signature of a recent recharge and less weathered rock, expressed by its hydraulic and hydrochemical properties (low TDS). In addition, cluster six occurs across SGAS recharge zones, water divides with high elevation. The spatial location of this cluster and the values of the elements (low calcium, bicarbonate and TDS) reinforces the characterization of these waters as SGAS hydrochemical facies. Subarea I is greatly influenced by the density and magnitude of vertical structures, likely responsible for the diversity of clusters in this subarea. Cluster seven is influenced by high contents of fluoride within areas of cluster six and according to Nanni *et al.* (2013) it is possible that high concentrations of fluoride are associated with deep fault zones connecting SGAS, GAS and older aquifers below it.

CONCLUSIONS

A spatial hydrochemical model of the Serra Geral fractured aquifer is developed based on the spatial variation of chemical elements and compared with models previously established in the literature. Using the MSOM, missing values of hydrochemistry and hydraulic transmissivity could be estimated.

The following remarks are given with respect to the objectives of this study:

(1) the correlation matrix, calculated from the SOM estimated dataset, provides parametrical relations between hydrochemical elements and explanatory variables. MSOM estimation provided unbiased hydrochemical correlations supported by the literature;

(2) the k-means clustering technique classified variables based on their topological similarity with groups reflecting: (i) hydrochemical facies of recent recharge areas; (ii) potential connectivity between GAS e SGAS; (iii) regions featuring transition; (iv) water with confinement and log residence traits; and (v) typical hydrochemical signature of SGAS;

(3) the proposed method used to estimate hydraulic transmissivity is adequate to cope with incomplete information in the well database, proving to be an important advantage when constructing numerical groundwater models;

(4) the analysis of spatial distribution of chemical elements and cluster maps revealed regions with potential hydraulic connections between the Serra Geral and GASs;

(5) the proposed method is suitable to survey hydrochemistry and groundwater physical properties, revealing and quantifying relationships in a large set of variables, which would not be possible to observe using parametric, multivariate statistical approaches.

Research institutes and regulatory agencies are continually collecting water samples and monitoring wells, but the analysis equipment precision and scope of each dataset differs as methods and sampling equipment change. The proposed method provides a way to integrate data from different sources creating a larger and more comprehensive base for groundwater-quality analysis and modelling.

Finally, the proposed modeling method is anticipated to be useful as an alternative for studies analyzing hydrochemistry and producing input parameters for numerical modeling of the Serra Geral and Guarani aquifers. Furthermore, this approach can help to overcome problems related to scarcity of physical and hydrochemical datasets for

Uncorrected Proof

19   F. Iwashita *et al.* | Using self-organizing maps to characterize hydrogeology of Serra-Geral aquifer   **Hydrology Research | in press | 2017**

groundwater characterization, particularly for numerical modeling related problems that rely on spatially consistent sets of input parameters. Estimating these missing variables without trends or bias, while preserving spatial autocorrelation, is an important asset for establishing initial conditions and defining prior information for inverse modeling.

ACKNOWLEDGEMENTS

REFERENCES

Araújo, L., França, A. & Potter, P. 1995 Arcabouço hidrogeológico do aqüífero Gigante do Mercosul (Brasil, Argentina, Uruguai e Paraguai): Fm Botucatu, Pirambóia, Rosário do Sul, Buena Vista, Misiones e Tacuarembó (Hydrogeologic framework of the Giante Aquifer of Mercosul-Brazil, Argentina, Uruguay and Paraguay). *Anais do 9° Encontro de perfurados de poços* 110–120.

Artur, P. C. 1998 *Paleolineamentos na Bacia do Paraná: favorabilidade para a acumulação de hidrocarbonetos (Parana Basin Paleolineaments: Favorability for Hydrocarbon Acumulation)*. Master's Thesis. Geology Department, Parana Federal University. Curitiba. 202 p.

ASCE Task Committee on application of Artificial Neural Networks in Hydrology 2000a Artificial neural networks in hydrology I: preliminary concepts. *J. Hydrol. Eng.* **5**, 115–123.

ASCE Task Committee on application of Artificial Neural Networks in Hydrology 2000b Artificial neural networks in hydrology II: hydrologic applications. *J. Hydrol. Eng.* **5**, 124–137.

Astel, A., Tsakoviski, S., Barbieri, P. & Simeinov, V. 2007 Comparison of self- organizing maps classification approach with cluster and principal components analysis for large environmental datasets. *Water Res.* **41**, 4566–4578.

Athayde, G. B., Müller, C. V., Rosa Filho, E. F. & Hindi, E. C. 2007 Estudo sobre os tipos das águas do aquífero Serra Geral no município de Marechal Cândido Rondon – PR (Study of water characteristics from Serra Geral aquifer in Marechal Candido Rondon city – PR). *Águas Subterrân.* **21**, 111–122.

Benavides, A., Everett, M. E. & Pierce Jr., C. 2009 Unexploded ordnance discrimination using time-domain electromagnetic induction and self-organizing maps. *Stoch. Environ. Res. Risk Assess.* **23**, 169–179.

Bittencourt, A. V. L., Rosa-Filho, E. F., Hindi, E. C. & Buchman-Filho, A. C. 2003 A influência dos basaltos e de misturas com águas de aqüíferos sotopostos nas águas subterrâneas do Sistema Aqüífero Serra-Geral na bacia do rio Piquiri Paraná – BR (The influence of basalts and mixture from overlaid aquifers in the Serra Geral Aquifer Systems Waters across Piriqui river basin). *Rev. Águas Subterrân.* **17** (1), 67–75.

Bongiolo, A. B. S., Ferreira, J. F. F., Bittencourt, A. V. L. & Salamuni, A. 2014 Connectivity and magnetic-structural compartmentalization of the Serra Geral and Guarani aquifer systems in central state of Paraná (Paraná basin Brazil). *Rev. Brasil. Geofís.* **32** (1), 141–160.

Bradbury, K. R. & Rothschild, E. R. 1985 A computerized technique for estimating the hydraulic conductivity of aquifers from specific capacity data. *Ground Water* **23**, 240–246.

Breiman, L. 1996 Bagging predictors. *Mach. Learn.* **24** (2), 123–140.

Briggs, I. C. 1974 Machine contouring using minimum curvature. *Geophysics* **39** (1), 39–48.

Davies, D. L. & Bouldin, D. W. 1977 A cluster separation measure IEEE transactions on pattern. *Anal. Mach. Intell.* **1**, 224–227.

Dickson, B. L. & Giblin, A. 2007 An evaluation of methods for imputation of missing trace element data in groundwaters. *Geochemistry: Explor. Environ. Anal.* **7**, 173–178.

Efron, B. & Tibshirani, R. J. 1993 An introduction to the bootstrap. In: *Monographs on Statistics and Applied Probability*, Vol. 57. Chapman & Hall, London, 436 pp.

Farr, T. G. & Kobrick, M. 2000 Shuttle radar topography mission produces a wealth of data. *Eos Trans. Am. Geophys. Union* **81** (48), 583–585.

Ferreira, F. J. F., Portela Filho, C. V., Rosa Filho, E. F. & Rostirolla, S. P. 2005 Conectividade e compartimentação dos sistemas aqüíferos Serra Geral e Guarani na região central do arco Ponta Grossa (Bacia do Paraná Brasil) (Connectivity and compartmentation of Serra Geral and Guarani aquifers across central region of Ponta Grossa arch). *Rev. Latino Am. Hidrogeol.* **5**, 61–74.

Fessant, F. & Midenet, S. 2002 Self-organizing map for data imputation and correction in –surveys. *Neur. Comput. Appl.* **10**, 300–310.

Fraga, C. G. 1986 *Introdução ao Zoneamento do Sistema Aqüífero Serra Geral no Estado do Paraná (Introduction to the Zoning of Serra Geral Aquifer System in Paraná State)*. Master's Thesis, Geosciences Institute, Universidade de São Paulo (USP), São Paulo, Brazil.

Friedel, M. J. 2016 Estimation and scaling of hydrostratigraphic units: application of unsupervised machine learning and multivariate statistical techniques to hydrogeophysical data. *Hydrogeol. J.* **24** (8), 2103–2122.

Friedel, M. J. & Iwashita, F. 2013 Spatial hybrid modeling for application to environmental inverse problems. *Environ. Modell. Softw.* **43**, 60–79.

Uncorrected Proof

20   F. Iwashita et al. | Using self-organizing maps to characterize hydrogeology of Serra-Geral aquifer   Hydrology Research | in press | 2017

Friedel, M. J., Souza Filho, O., Iwashita, F., Moreira Silva, A. & Yoshinaga, S. P. 2012 Data-driven modeling for groundwater exploration in fractured crystalline terrain northeast Brazil. Hydrogeol. J. 20, 1061–1080.

Friedel, M. J., Esfahani, A. & Iwashita, F. 2016 Toward real-time three-dimensional mapping of surficial aquifers using a hybrid modeling approach. Hydrogeol. J. 24 (1), 211–229.

Gastmans, D., Chang, H. K. & Hutcheon, I. 2010 Groundwater geochemical evolution in the northern portion of the Guarani Aquifer System (Brazil) and its relationship to diagenetic features. Appl. Geochem. 25, 16–33.

Han, J. C., Huang, Y., Li, Z., Zhao, C., Cheng, G. & Huang, P. 2016 Groundwater level prediction using a SOM-aided stepwise cluster inference model. J. Environ. Manage. 182, 308–321.

Harris, C. & Milner, S. 1997 Crustal origin for the Paraná rhyolites: discussion of 'Description and petrogenesis of the Paraná Rhyolites southern Brazil' by Garland et al. 1995. J. Petrol. 38, 299–302.

Hastie, T., Tibshirani, R. & Friedman, J. 2002 The Elements of Statistical Learning. Springer-Verlag, Berlin, 533 pp.

Hong, Y. & Rosen, M. R. 2001 Intelligent characterization and diagnosis of the groundwater quality in a urban fractured-rock aquifer using an artificial neural network. Urban Water 3, 193–204.

Hornberger, G. M. (ed.). 1998 Elements of Physical Hydrology. JHU Press, Baltimore.

Iwashita, F., Friedel, M. J., Souza Filho, C. R. & Fraser, S. J. 2011 Hillslope chemical weathering across Paraná state, Brazil: a data mining-GIS hybrid approach. Geomorphology 132, 167–175.

Iwashita, F., Friedel, M. J., Ribeiro, G. F. & Fraser, S. J. 2012 Intelligent estimation of spatially distributed soil physical properties. Geoderma 170, 1–10.

Jenson, S. K. & Domingue, J. O. 1988 Extracting topographic structure from digital elevation data for geographic information system analysis. Photogram. Eng. Remote Sens. 54, 1593–1600.

Junninen, H., Niska, H., Tuppurainen, K., Ruuskanen, J. & Kolehmainen, M. 2004 Methods for imputation of missing data in air quality data sets. Atmos. Environ. 38, 2895–2907.

Kalteh, A. M. & Hjorth, P. 2009 Imputation of missing values in precipitation-runoff process database. Hydrol. Res. 40, 420–432.

Kohavi, R. 1995 A study of cross-validation and bootstrap for accuracy estimation and model selection. Int. Joint Conf. Artif. Intell. 14 (2), 1137–1145.

Kohonen, T. 1984 Self-organization and Associative Memory. Springer, Berlin.

Kohonen, T. 2001 Self-organizing Maps, 3rd edn. Springer-Verlag, Berlin.

Leinz, V. 1949 Contribuição à geologia dos derrames basálticos do sul do Brasil (A contribution to the geology of basalts leakages in South Brazil). USP-Faculdade de Filosofia. Ciencias e Letras-Departamento de Geologia e Palentologia.

Licht, O. A. B. 2001 A geoquímica multielementar na gestão ambiental (The Multi-Element Geochemistry for Environmental Management). PhD Thesis, Faculdade de Geologia Universidade Federal do Paraná Brazil.

Lu, R. & Lo, S. 2002 Diagnosing reservoir water quality using self-organizing maps and fuzzy theory. Water Res. 36, 2265–2274.

Malek, M. A., Harun, S., Shamsuddin, S. M. & Mohamad, I. 2008 Imputation of time series data via Kohonen self-organizing maps in the presence of missing data. Eng. Technol. 41, 501–506.

Manasses, F. 2009 Caracterização hidroquímica da água subterrânea da formação Serra Geral na região sudoeste do estado do Paraná (Hydrochemical Characterization of Groundwater in Serra Geral Formation Across Southeast Parana State). Master's Thesis, Universidade Federal do Paraná Curitiba Brazil, 136 p.

McLin, S. G. 2005 Estimating aquifer transmissivity from specific capacity using MATLAB. Ground Water 43, 611–614.

Minty, B. R. S. 1991 Enhancement and presentation of airborne geophysical data. AGSO J. 17, 63–75.

Mocellin, R. C. 2009 Conectividade e compartimentação magnética-estrutural dos sistemas aquíferos serra geral e guarani na região sudoeste do Estado do Parana (Bacia do Paraná Brasil) (Magnetic-Structural Connectivity and Compartmentation of Serra Geral and Guarani Aquifer Systems in Southeast Area in Parana State). Master's Thesis, Universidade Federal do Paraná Curitiba Brazil, 231 p.

Nakagawa, K., Amano, H., Kawamura, A. & Berndtsson, R. 2016 Classification of groundwater chemistry in Shimabara, using self-organizing maps. Hydrol. Res. 48 (3), 840–850.

Nanni, A., Roisenberg, A., Marimom, M. P. C. & Viero, A. P. 2009 The hydrochemical facies and anomalous fluoride content in Serra Geral aquifer system southern Brazil: a GIS approach with tectonic and principal component analysis. Environ. Geol. 58, 1247–1255.

Nanni, A., Roisenberg, A., de Hollanda, M. H. B. M., Marimon, M. P. C., Viero, A. P. & Scheibe, L. F. 2013 Fluoride in the Serra Geral aquifer system: source evaluation using stable isotopes and principal component analysis. J. Geol. Res. Article ID 309638, 9 pages. dx.doi.org/10.1155/2013/309638.

Neter, J., Kutner, M. N., Nachtssheim, C. J. & Wasserman, W. 1996 Applied Linear Statistical Models, 4th edn. WCB/McGraw-Hill, Boston.

Peate, D. W., Mantovani, M. S. M. & Hawkesworth, C. J. 1988 Geochemical stratigraphy of Paraná continental flood basalts: borehole evidence. Rev. Brasil. Geoci. 18, 212–221.

Penn, B. S. 2005 Using self-organizing maps to visualize high-dimensional data. Comput. Geosci. 31, 531–544.

Portela Filho, C. V. 2003 Condicionamento estrutural-magnético do sistema aqüífero Serra Geral na região central do arco de Ponta Grossa (Bacia do Paraná) e sua conectividade com o sistema aqüífero Guarani (Structural-Magnetic Conditioning of Serra Geral Aquifer in Ponta Grossa Arch Central Region). Master's Thesis. Universidade Federal do Paraná Curitiba Brazil, 163 p.

Rabelo, J. L. & Wedland, E. 2009 Assessment of groundwater recharge and water fluxes of the Guarani aquifer systems Brazil. Hydrogeol. J. 17, 1733–1748.

Uncorrected Proof

21   F. Iwashita *et al.* | Using self-organizing maps to characterize hydrogeology of Serra-Geral aquifer   **Hydrology Research** | in press | 2017

Rallo, R., Ferre-Gine, J., Arenas, A. & Giralt, F. 2002 Neural virtual sensor for the inferential prediction of product quality form process variables. *Comput. Chem. Eng.* **26**, 1735–1754.

Rebouças, A. C. & Fraga, C. G. 1988 Hidrogeologia das rochas vulcânicas do Brasil. *Rev. Águas Subterrân.* **12**, 29–55.

Ritter, H. & Schulten, K. 1986 On the stationary state of Kohonen's self-organizing sensory mapping. *Biol. Cybernet.* **54** (2), 99–106.

Rosa Filho, E. F., Bittencourt, A. V. L., Hindi, E. C. & Bittencourt, A. 2006 Groundwater types and structural conditioning study of the Guarani aquifer system in the western of Paraná state (Brazil). *Águas Subterrân.* **20**, 39–48.

Sánchez-Martos, F., Aguilera, P. A., Garrido-Frenich, A., Torres, J. A. & Pulido-Bosch, A. 2002 Assessment of groundwater quality by means of self-organizing maps: application in a semi-arid area. *Environ. Manage.* **30**, 716–726.

Silva, A. B. 2007 *Conectividade e compartimentação magnética-estrutural dos sistemas aqüíferos Serra Geral e Guarani na região central do estado do Paraná (Magnetic-Structural Connectivity and Compartmentation of Serra Geral and Guarani Aquifers in Parana State Central Region).* Master's Thesis, Universidade Federal do Paraná Curitiba Brazil, 182 p.

Soares, P. C., Barcellos, P. E., Csordas, S. M., Mattos, J. T., Baliero, M. G. & Meneses, P. R. 1982 Lineamentos em imagens Landsat e Radar e suas implicações no conhecimento tectônico da Bacia do Paraná (Lineaments in Landsat and Radar images and its implication on Parana Basin tectonic knowledge). *II Brazilian Remote Sensing Symposium*, CNPq-INPE, Brasília.

Souza Filho, O. A., Silva, A. M., Remacre, A. Z., Sancevero, S. S., McCafferty, A. E. & Perrotta, M. M. 2010 Using helicopter electromagnetic data to predict groundwater quality in fractured crystalline bedrock in a semi-arid region Northeast Brazil. *Hydrogeol. J.* **18**, 905–916.

Sracek, O. & Hirata, R. 2002 Geochemical and stable isotopic evolution of the Guarani aquifer system in the state of São Paulo Brazil. *Hydrogeol. J.* **10**, 643–655.

Suk, H. & Lee, K. K. 1999 Characterization of a groundwater hydrochemical system through multivariate analysis: clustering into groundwater zones. *Ground Water* **37**, 358–366.

Thiede, D. S. & Vasconcelos, P. M. 2010 Parana flood basalts: rapid extrusion hypothesis confirmed by new 40ar/39ar results. *Geology* **38**, 747–750.

Turner, S. P., Peate, D. W., Hawkesworth, C. J. & Mantovani, M. S. M. 1999 Chemical stratigraphy of the Paraná basalt succession in western Uruguay: further evidence for the diachronous nature of the Paraná magma types. *J. Geodynam.* **28**, 459–469.

Ultsch, A. 2003 *U*-matrix: a Tool to Visualize Clusters in High Dimensional Data.* Fachbereich Mathematik und Informatik, Marburg.

Valeriano, M. M., Rosetti, D. F. & Albuquerque, P. C. G. 2009 TOPODATA: desenvolvimento da primeira versão do banco de dados geomorfométricos locais em cobertura nacional (TOPODATA: development of the first version of a geomorphometric database with national coverage). In: *XIV Simpósio Brasileiro de Sensoriamento Remoto.* Natal, Brasil, pp. 5499–5506.

Vesanto, J. & Alhoniemi, E. 2000 Clustering of the self-organizing map. *IEEE Trans. Neural Netw.* **11**, 586–600.

Vesanto, J., Himberg, J., Alhoniemi, E. & Parhankangas, J. 2000 *SOM Toolbox for Matlab 5.* Helsinki University of Technology, Finland.

Wang, S. 2003 Application of self-organizing maps for data mining with incomplete data sets. *Neural Comput. Appl.* **12**, 42–48.

Wendland, E., Barreto, C. & Gomes, L. H. 2007 Water balance in the Guarani aquifer outcrop zone based on hydrogeologic monitoring. *J. Hydrol.* **342**, 261–269.

Winter, C. L., Guadagnini, A., Nychka, D. & Tartakovsky, D. M. 2006 Multivariate sensitivity analysis of saturated flow through simulated highly heterogeneous groundwater aquifers. *J. Comput. Phys.* **217**, 166–175.

Zalán, P. V., Wolff, S., Conceição, J. C. J., Astolf, M. A. M., Vieira, I. S., Appi, V. T. & Zanotto, O. A. 1987 Tectônica e Sedimentação da Bacia do Paraná (Parana Basin Tectonic and Sedimentation). In: *III South Brazilian Geology Symposium.* Brazilian Geological Society 1, Curitiba, pp. 441–473.