Visualization of Volcanic Rock Geochemical Data and Classification with Artificial Neural Networks[1]

Juan Pablo Lacassie,[2,3] Javier Ruiz del Solar,[4] Barry Roser,[5] and Francisco Hervé[6]

An unsupervised neural network technique, Growing Cell Structures (GCS) was used to visualize geochemical differences between four different island arc volcanic rock types: basalts, andesites, dacites and rhyolites. The output of the method shows the cluster structure of the dataset clearly, and the relevant geochemical patterns and relationships between its variables. The data can be separated into four clusters, each associated with a specific volcanic rock type (basalt, andesite, dacite and rhyolite), according to a unique combination of major element concentrations. Following clustering, performance of the trained GCS network as a classifier of volcanic rock type was tested using two test datasets with major element concentration data for 312 and 496 island arc volcanic rock samples of known volcanic type. Preliminary classification results are promising. In the first test dataset 94% of basalts, 76% of andesites, 83% of dacites and 100% of the rhyolites were classified correctly. Successful classification rates in the second dataset were 100%, 80%, 77%, and 98% respectively. The success of the analysis suggests that neural networks analysis constitutes a useful analytical tool for identification of natural clusters and examination of the relationships between numeric variables in large datasets, and that can be used for automatic classification of new data.

KEY WORDS: neural networks; geochemistry; visual analysis; classification; volcanic rocks.

[2]Paleoproterozoic Mineralization Research Group, Department of Geology, University of Johannesburg, 2006 Auckland Park, Johannesburg, Republic of South Africa

[3]Servicio Nacional de Geología y Minería, Av. Santa María 0104, Casilla 10465, Santiago, Chile; e-mail: jlacassi@cec.uchile.cl

[4]Department of Electrical Engineering, Universidad de Chile, Casilla 412-3, Santiago, Chile; e-mail: jruizd@cec.uchile.cl

[5]Department of Geoscience, Shimane University, Matsue 690-8504, Japan; e-mail: roser@riko.shimane-u.ac.jp

[6]Department of Geology, Universidad de Chile, Casilla 13518, Correo 21, Santiago, Chile; e-mail: fherve@cec.uchile.cl

INTRODUCTION

New developments in analytical technology over the last decades have dramatically increased the volume of geochemical data collected and stored in databases. Traditional data analysis techniques have become inadequate for processing such large volumes of geochemical data and new techniques for isolation of complex correlations between variables, associations among samples, and anomaly identification are needed.

Artificial neural networks (ANN) are distributed networks of simple computing elements operating in parallel. They differ from traditional modeling approaches in that they can be trained to learn solutions rather than being programmed to model a specific problem using complex rules and mathematical routines. Connection strengths or weights between interconnected nodes are adjusted during the training or learning process, according to the input data. Consequently, ANN are capable of modeling highly complex non-linear functions, and of identifying relations in input data that are not readily apparent with common analytical techniques (e.g. Lacassie and others, 2003). Furthermore, ANN are non-parametric and thus are not constrained by the error distributions of the population, as are regression algorithms. Kohonen's Self-Organizing Map (SOM; Kohonen, 1995) is one of the best-known ANN with unsupervised learning rules, i.e. the training process is carried out without any *a priori* classification of the samples. SOM's central property is that it forms a non-linear projection or mapping of a high-dimensional (multivariate) input space on a low-dimensional output space (usually a 2-dimensional regular grid), with information on the relationships between the input data preserved as much as possible. Through the training process, neighboring input samples are mapped onto neighboring (or the same) nodes according to the metrics defined in the output space. This mapping therefore not only displays the clustering of the original data space but also preserves (at least approximately) the metric-topological relations of the data items. Furthermore, relevant geochemical information can be exhibited in a maximally concentrated form, in a way that the user does not require advanced specialist knowledge for its perception, and outlying observations can also be included in the dataset without significantly affecting the main patterns and relationships identified. Additionally, the SOM averages the input dataset in weight vectors during the training process, and thus removes noise (Vesanto and others, 1999). Although SOM has been used as an analytical tool in a variety of fields it has not been used extensively for the analysis of geochemical data.

In this paper we use Growing Cell Structures (GCS; Fritzke, 1994), an extension of the SOM algorithm, which allows simultaneous adaptation of the position of the weight vectors in the input space and the topology of the output space. Rather than using a predetermined rectangular grid of cells, as for the SOM algorithm, the GCS structure is dynamically defined during the training process. Each

GCS node has an associated activity counter, which registers the number of times that training samples are mapped onto a given node. A high counter value means high activity in this particular node. The counter information is used for deciding where new nodes should be inserted into the network. New nodes are inserted in network positions of high activity (high counter values), and which thus require higher resolution. Simultaneously, the node weight vectors are adapted: each time a training sample is mapped onto a given node, the node weight vector is moved in the direction of the sample (Kohonen, 1995). This process is repeated until stable values for the network parameters are attained, and not necessarily when the number of nodes of the network is equal to the number of input variables. After training is completed, the GCS algorithm converts the frequency of training samples assigned to each node of the final GCS map (prior probability) into a prediction for new samples associated to the nodes (posterior probability) (Fritzke, 1994; Walker, Cross, and Harrison, 1999). The prior probability corresponds to the probability of occurrence of a given class, thus it reflects *a priori* knowledge. The posterior probability corresponds to the probability of a class given an observation, and in the GCS algorithm it is determined for the activated network node after mapping a given input vector. For each class the prior probability is estimated using the node counters. The posterior probabilities estimation is the result of the network self-organization (training). The posterior probabilities of the different classes are displayed on color maps, which are lain over the final two-dimensional GCS map structure (e.g. Fig. 1(c)). Alternatively, a three-dimensional visualization can be generated, in which the posterior probabilities are displayed as altitudes (i.e. Z-axis values) ranging between 0 and 1 (e.g. Fig. 1(b)). At each node of the final GCS map the average value of each individual input variable can be computed as the corresponding average of the training samples associated to this map unit, irrespective of their individual class. This information can also be displayed on separate color maps, one for each input variable, and overlain on the final GCS structure (e.g. Fig. 1(d)).

Although GCS is primarily a clustering algorithm, after training and assignation of a class label to each network node, it can also be used for classification. The method has been tested with geochemical analyses of island arc volcanic rocks comprising four first-order categories: basalts, andesites, dacites, and rhyolites.

METHODOLOGY

Data Sources

In this study we utilize three literature-derived datasets that contain island arc volcanic rock data from different geographic regions, with concentration data

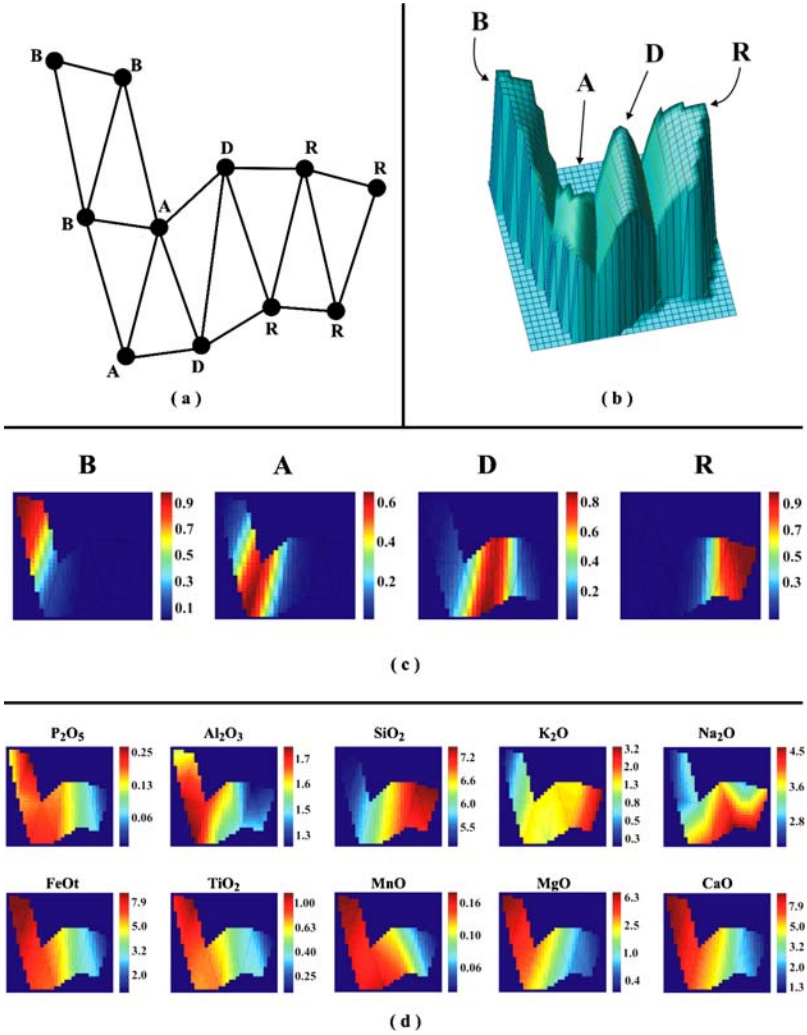

Figure 1. Visualization of posterior probabilities and input variable distributions for an 11-node GCS map trained with the 1529 samples of the Training dataset (Table 1). (a) Volcanic type-related node distribution over the GCS map structure. All nodes are associated with a specific volcanic rock type, labeled accordingly: B (basalts); A (andesites); D (dacites); R (rhyolites). (b) Three-dimensional and (c) two-dimensional visualization of the posterior probability distributions. Four clearly-separated clusters were modeled by the network. Each is associated with a specific volcanic rock type (basalts, andesites, dacites and rhyolites). For the two-dimensional visualization (c) the posterior probability values are shown by the color axes at right. (d) Two-dimensional visualization of the distributions of each input variable. The distributions of each variable can be easily compared visually with that of the posterior probabilities, and with each other. Scales at right indicate whole rock concentrations (wt.%).

(wt.%) for the ten major and minor oxides (SiO_2, TiO_2, Al_2O_3, K_2O, Na_2O, CaO, FeO_t, MgO, MnO and P_2O_5).

The first dataset (*Training dataset*) comprises 1529 whole-rock analyses of basalts (32%), andesites (24%), dacites (27%) and rhyolites (17%) (Table 1). This dataset was constructed using data from the GEOROC database, requested through the web site (http://georoc.mpch-mainz.gwdg.de/georoc/). The second dataset (*Lau-Taupo dataset*) includes data from the Lau island arc (Fiji) and the Taupo Volcanic Zone (New Zealand), comprising 312 whole-rock analyses of basalts (32%), andesites (28%), dacites (15%) and rhyolites (25%) (Table 2). The third dataset (*Honshu dataset*) is from the Honshu arc (Japan), and contains 496 volcanic whole-rock analyses of basalts (8%), andesites (63%), dacites (16%) and rhyolites (13%) (Table 2).

Data Analysis

The Training dataset was used to examine geochemical patterns of island arc volcanic rock suites. The GCS algorithm projected the samples of this dataset onto a two-dimensional feature map, and converted the frequency of training samples assigned to each node into a prediction for new samples associated to the nodes (posterior probability). The posterior probabilities of occurrence of the different volcanic rock types were then displayed over the final map structure as color maps (Fig. 1(c)). The GCS algorithm also projected average values for individual input variables at each node onto the feature map (Fig. 1(d)).

The Lau-Taupo and Honshu datasets were then used to test the performance of the GCS network trained above as a classifier of island arc volcanic rock type. The classification procedure works as follows. (i) A best-matching unit (BMU) was chosen for each sample or input vector, in both datasets as the node of the trained network whose weight vector presents the greatest similarity with the current input vector in the 10-dimensional input space. For this purpose the euclidean distance between each trained GCS network's weight vector and each input vector was calculated, and the node with the closest weight vector to the current input vector was tagged as its BMU. (ii) Then the class assigned to each sample corresponds to the class of its BMU, which in turn can be computed using the prior probability information. A more detailed explanation of the trained network class assignation method is given by Lacassie and others (2004).

The reliability of the classification model achieved by the trained GCS network was then assessed from the classification accuracy. For each of the Lau-Taupo and the Honshu datasets, the classification accuracy was evaluated separately for each volcanic rock type by the formula:

$$A_i = 100 \times (T_i / N_i)$$

Table 1. Volcanic Rock Types, Island Arcs and Sources of the Training Dataset

Type	*N*	Island Arc	Source (GEOROC reference number)
B	5	Aeolian	[3483-4371-4671-4672-6814]
	27	Aleutian	[2216-2218-2220-2233-2235-2240-2245-2501-2502-2565-2613-5183-5769-6974-6974-6992]
	1	Banda	[2703]
	21	Bismark	[2757-3473-3733-4046-4108-4590-5946-5981-6900]
	9	Great Antilles	[4429-4430-5067-5927]
	49	Izu-Bonin	[2294-2303-2305-2311-2336-2350-2798-2974-195-3292-3408-3584-4240-4568-4610-4631-4899-5068-5245-5250 -5253-5255-5256-5261-5781-6564-6679-6905]
	40	Kamchatka	[2195-2475-2481-2485-2725-3048-3272-3317-3393-4276-4648-4769-4788-4834-5033]
	7	Kermadec	[2495-3319-3370-3501-6185-6437]
	37	Kurile	[2536-2654-3042-3252-3261-3313-3316-3517–3542-3890-3919-3970-4073-4334-4799-4901-6270-6818-6973]
	29	Lesser Antilles	[1608-1701-2494-2577-2906-3286-3287-3325-3396-3472-3635-4118-4700-5712-6526-6705-6893]
	18	Luzon	[2537-2538-3520-3580-3933-3936-4096-5021-5803-6051-6053-6461-6988-7004]
	33	Mariana	[2338-2343-2371-2378-2388-2394-2400-2401-2406-2428-2430-2452-2517-2595-2954]
	58	New Hebrides	[3090-3786-3790-3875-3876-3930-3946-4032-4033-4124-4553-4624-4841-4890-5603-6302-6972]
	3	Papua New Guinea	[2974-3292-5431-6341]
	16	Ryukyu	[2974-3292-3656-4183-5299-5382-5384]
	43	Scotia	[2189-2190-2191-2193-2594-2724-2813-3991-5179-6190-6312-6650-6844]
	8	Sunda	[3313-3146-3602-4012-4583-4674-7092]
	92	Tonga	[454-2478-2495-2517-2757-2766-2771-3056-3260-3321-3324-3371-3380-3669-3678-698-3924-3927-4105–4108 -4110-4134-4242-6338-6407]
S-T	496		
A	1	Aegean	[4826]
	18	Aeolian	[3226-3483-4345-4468-4671-4672-4994]
	56	Aleutian	[2206-2214-2218-2223-2241-2245-2247-2251-2252-2503-2504-2550-2551-2565-2576-2613-3357-3503-3855-4059 -4191-5183-6582-6583-6992]
	4	Banda	[5178]
	5	Bismark	[2605-2715-3055-4954-5429]
	4	Great Antilles	[4121-4429]

Table 1. Continued

Type	*N*	Island Arc	Source (GEOROC reference number)
	25	Izu-Bonin	[2315-2801-3291-4240-4568-5245-5261-6033-6905]
	19	Kamchatka	[2195-2475-2481-3268-3272-3313-3607-4788-5738-7092]
	30	Kurile	[3042-3052-3261-3313-3316-3967-3970-4198-4334-4356-4568-6292-6973]
	63	Lesser Antilles	[1435-2625-3282-3287-3301-3579-3903-3962-4230-4255-4723-4726-5724-6046-6555]
	23	Luzon	[2488-2537-3299-3520-3522-3580-3606-3934-4096-5021-6048-6261-7004]
	1	Mariana	[2396]
	3	New Hebrides	[3090-3870]
	83	Ryukyu	[3656-3987-4183-4568-4629-5380-5383-5384-5442-6116-6117-6614]
	8	Scotia	[2193-2813-6650-6678-6844]
	7	Sunda	[2479-3602-4469-4590-4674]
	17	Tonga	[3260-4242-4245]
S-T	367		
D	22	Aegean	[4826-5012-5552-6509-6756]
	4	Aeolian	[3483-4345-4994]
	44	Aleutian	[2195-2199-2219-2220-2245-2233-2247-2252-2499-2565-2570-2574-2576-2607-2612-3855-4867-4190-5183-6578-6992]
	36	Bismark	[2480-2545-2605-2716-2717-3055-4046-4076-5622-5946-5979-6026-6612]
	4	Great Antilles	[4429]
	30	Izu-Bonin	[2271-2299-2311-2328-2342-3207-3291-4240-4568-4913-5166-5245-5250-5261-6905]
	30	Kamchatka	[2195-2475-2481-3272-3393-3607]
	1	Kermadec	[3761]
	31	Kurile	[3042-3052-3313-3316-3970-4037-4198-6292]
	17	Lesser Antilles	[2625-3282-3301-3776-3903-3904-4119-4726]
	69	Luzon	[2538-3520-3522-3580-3606-3650-4096-5002-5021-6048-6049-6051-6261-7004]
	8	Mariana	[2368-2375-2378-2435-2454-2579-3178-4590]
	23	New Hebrides	[3090-3877-4890-6302]
	3	Papua New Guinea	[4151-4995]
	16	Ryukyu	[3982-3987-4183-4568-5383-6864]
	23	Scotia	[2186-2193-2594-2724-2813-3991-6312-6678]

Table 1. Continued

Type	*N*	Island Arc	Source (GEOROC reference number)
	13	Sunda	[3113-3500-3734-4469-4590-4674-4962-6639]
	39	Tonga	[2495-2478-2601-2695-3056-3370-3371-3603-3669-3678-3698-3699-4242-6338-7092]
S-T	413		
R	15	Aeolian	[4468-4541-4721-4994-6128]
	16	Aleutian	[2246-2247-2570-2574-4867-6578-6859]
	24	Bismark	[2703-3055-4046-4076-4675-4954-5622-6026]
	45	Izu-Bonin	[2271-2294-2298-2299-2311-2328-2342-3291-3853-4240-4634-4679-4913-5239-5248-5803-6583]
	1	Kamchatka	[3393]
	1	Kermadec	[3761]
	57	Kurile	[2541-3042-3052-3252-3316-3393-3761-3967-4198-4334-5104-6269-6292]
	4	Lesser Antilles	[4726]
	5	Luzon	[2539-3522-3580-7044]
	11	Mariana	[2338-2375]
	2	New Hebrides	[4553]
	14	Papua New Guinea	[4991]
	31	Ryukyu	[2669-3408-4568-5382-5384-6115-6864]
	5	Sunda	[4469-4583-4674]
	22	Tonga	[2601-3669-3678-3698-3699-4242-4245]
S-T	253		
Total	1529		

S-T: Sub-total. *N*: Number of samples of each class.

Table 2. Volcanic Rock Types, Island Arcs and Sources of the Lau-Taupo and Honshu Datasets

Type	*N*	Island Arc	Source
Lau-Taupo dataset			
B	99	Taupo and Lau	(Cole, 1979; Cole and others, 1985)
A	87	Taupo and Lau	(Cole, 1979; Cole and others, 1985)
D	48	Taupo and Lau	(Cole, 1979; Cole and others, 1985)
R	78	Taupo and Lau	(Cole, 1979; Cole and others, 1985)
Total	312		
Honshu dataset			
B	41	Honshu	[3182-3658-3897-3976-3658-4074][a]
A	312	Honshu	[3075-3182-3187-3192-3228-3242-3271-3334-3402-3406-3510-4028-4081-4258-4261-4336-4398-4473-4697-4699-4805-4869-4942-4945-5092-5094-5097-5109-5386-5954][a]
D	80	Honshu	[2336-2804-3075-3235-3271-3334-3406-3408-3574-3595-3970-4135-4136-4141-4145-4195-4261-4338-4359-4466-4568-4734-4805-5092-5094-5097][a]
R	63	Honshu	[2336-2804-3075-3235-3271-3334-3406-3408-3574-3595-3970-4135-4136-4141-4145-4195-4261-4338-4359-4466-4568-4734-4805-5092-5094-5097][a]
Total	496		

N: Number of samples of each class.
[a] Source: GEOROC reference number.

where A_i is the classification accuracy, T_i is the number of correctly classified samples (true positives) and N_i is the total number of samples of the volcanic rock i (basalts, andesites, dacites, or rhyolites).

Finally, a second classification model was developed, based only on the information contained in SiO_2, K_2O and Na_2O.

RESULTS

The GCS analysis of the Training data set identified four clusters, which appear in the GCS color map as distinct areas with high posterior probabilities. Each cluster was associated by the GCS network to a specific rock type (basalt, andesite, dacite and rhyolite) with limited overlap between neighboring clusters, as indicated by the low percentages of samples assigned to a cluster of a different class (Fig. 1; Table 3).

The concentrations of the input variables (major elements) in the GCS map have distinct distributions which are useful for understanding the cluster structure of the data: SiO_2 and K_2O increase systematically from basalts through rhyolites, whereas the highly correlated oxides Al_2O_3, FeOt, TiO_2, MnO, MgO, CaO and P_2O_5 decrease. Only Na_2O is discontinuous, with high values associated mainly with the dacite cluster and a limited sub-region of the rhyolite cluster (Fig. 1(d)).

The classification results obtained using the Lau-Taupo and Honshu datasets as tests are shown in Table 4. Classification of the Lau-Taupo dataset was satisfactory for rhyolites (100%) and basalts (94%), but was lower for the dacites (83%) with 17% misclassified as rhyolites. Accuracy was lower still for andesites (76%) with 24% misclassified as dacites. Similarly, best classification rates in the Honshu dataset were for rhyolites (98%) and basalts (100%), with lower rates for dacites and andesites (77% and 80%, respectively). Most of the misclassified

Table 3. Distribution of the Training Dataset Samples in the Cluster Structure of the Trained Network

	Percentages of assigned samples per class			
Cluster type	B	A	D	R
B	96	0	0	0
A	4	98	1	0
D	0	2	97	1
R	0	0	2	99
Total (%)	100	100	100	100

Note. Values are the percentage of training samples of each volcanic type that were assigned to each volcanic type-related cluster.

Table 4. Confusion Matrix of the GCS Classification

	A priori class				*A priori* class[a]			
Assigned Class	B	A	D	R	B	A	D	R
Lau-Taupo dataset								
B	94	0	0	0	99	0	0	0
A	6	76	0	0	1	81	4	0
D	0	24	83	0	0	17	73	0
R	0	0	17	100	0	2	23	100
N	99	87	48	78	99	87	48	78
Honshu dataset								
B	100	9	0	0	75	10	0	0
A	0	80	9	0	25	78	11	0
D	0	11	77	2	0	11	69	3
R	0	0	14	98	0	1	20	97
N	41	312	80	63	41	312	80	63

Values are the classification rates, i.e. the percentage of samples of each volcanic type that was correctly classified.
[a]Classification results using only the information of SiO_2, K_2O and Na_2O. *N*: Number of samples of each class.

andesites and dacites of the Honshu dataset were classified as dacites (11%) and rhyolites (14%), respectively (Table 4).

DISCUSSION

The two- and the three-dimensional diagrams of Figure 1 constitute maximally concentrated projections of the relevant linear and non-linear multidimensional geochemical information of the Training dataset. These projections allow the rapid identification (visualization) of significant geochemical patterns and relationships of this multivariate dataset in a way that the user is not obliged to have a great amount of specialized knowledge for its perception. The relevant information that can be visualized includes the cluster structure of the Training dataset as described above, factors responsible for its cluster structure, and meaningful correlations between the relevant discriminating factors.

The cluster's topological relations resulting after the GCS analysis of the Training dataset (Fig. 1(b) and 1(c)) are compatible with the natural geochemical similarities between island arc volcanic rock types. These similarities are related to expected systematic changes of major element concentrations that resembles the differentiation trend from basalts to rhyolites (Fig. 1(c) and 1(d)).

The results (Table 4) suggest that andesites and dacites cannot be effectively discriminated from other volcanic rock types based on total major element information, whereas both rhyolites and basalts can be easily identified. This

indicates that the cluster boundaries between the differing lithotypes, especially between andesites and dacites, should be regarded as guides rather than rigorous separations.

Volcanic rocks are normally classified using simple chemical parameters such as the total alkali vs. silica (TAS) scheme (e.g. Cox, Bell, and Pankhurst, 1979; Rollinson, 1993), which is restricted to the linear information contained in three major oxide variables (SiO_2, K_2O and Na_2O). In sub-alkaline rocks, SiO_2 carries the greatest weight for division into the four primary lithotypes. When the GCS classification model was restricted to the TAS scheme variables, classification accuracy for the Honshu dataset decreased slightly (Table 4). This suggests that in this case relevant information was disregarded when restricting the classification to the TAS scheme variables. In contrast, classification accuracy in the Lau-Taupo dataset (Table 4) changed little.

Several opportunities exist for further research on both feature extraction and unsupervised volcanic rock classification, including exploring combined major and trace element data, in particular incorporating high field strength elements, for characterizing and distinguishing geochemical signatures of island arc volcanic rock types using ANN. If based on immobile elements, this could be useful for determining both lithotype and magma series in altered or metamorphosed volcanic rocks, where current methods using potentially mobile elements (e.g. K_2O, Na_2O) are inappropriate.

The study of Lacassie and others (2003) has shown that the GCS method offers several advantages over Principal Component Analysis (PCA), a conventionally used method for multidimensional data reduction and visualization. Although PCA is not intended for cluster analysis it corresponds to a commonly used linear-based exploratory tool which provides insightful graphical summaries of large sets of multidimensional data, and it can be used to identify structures, trends, redundancies and correlations in the data set. Both GCS and PCA were tested in the above study, using three different datasets consisting of geochemical analyses of stream sediments, sandstones of four provenance categories, and zircon grains from a wide range of igneous rock types. After being analyzed by GCS, a clustered structure was revealed in all three datasets. However, no separable features were distinguished in any of the datasets using PCA (performed on the correlation matrix, using two and three components), because this method is useful only when the data has linearly-separable features, as concluded by Lacassie and others (2003).

Similarly, the study of Lacassie and others (2004) showed that the sharpness of the GCS map boundaries between different classes is likely to be influenced by the dataset used as input. In that study, the GCS map clusters associated to the P3 (felsic provenance) and P4 (recycled) provenance groups displayed some overlap. Lacassie and others (2004) noted this was to be expected, as many P4 sandstones are derived from felsic crystalline basement rocks, and the effects of recycling of

preexisting sediments and varying intensity of source area weathering can also produce gradual transition between P3 and P4 characteristics.

CONCLUSIONS

This study shows that GCS analysis can be used as an efficient tool to improve analysis of large geochemical datasets, as it allows visualization of subtle interrelations between variables and of the cluster structure of highly dimensional data using the linear and the non-linear relationships between components.

The GCS classification scheme based on major element geochemical compositions accurately predicts volcanic rock type for basalts (94–100%) and rhyolites (98–100%) in two test sets with 312 and 496 samples, respectively. Although less successful for dacites (77–83%) and andesites (76–80%) in both datasets, these percentages are regarded as satisfactory for this preliminary application of the methodology. The results indicate that GCS could become a useful tool for classification of multidimensional island arc volcanic rock geochemical data, provided that a large and representative dataset is available for learning. Future research involving training with larger and more representative test sets and with different combination of chemical variables could improve results further.

ACKNOWLEDGMENTS

This work was supported by CONICYT under the Doctoral Fellowship Fondecyt # 2010010 to Lacassie. This manuscript was greatly improved by the very helpful critical reviews of Dr. Agnes Schumann (Free University Berlin) and Dr. Hannes Thiergärtner (Guest editor of Mathematical Geology) and Dr. Robert Howarth (University College London). Our special thanks to the members of the GEOROC Database Team. This work was conducted using the GCSVIS toolbox developed by Andrew Walker, Robert Harrison and Simon Cross at the University of Sheffield (programmed in MATLAB script; http://www.shef.ac.uk/∼path/GCSVIS/).

REFERENCES

Cole, J. W., 1979, Chemical analyses of lavas and ignimbrites of the Taupo Volcanic Zone: Victoria University of Wellington, Geol. Dep., Publ. no. 13, 31 p.

Cole, J. W., Gill, J. B., and Woodhall, B., 1985, Petrologic history of the Lau Ridge, Fiji, *in* Scholl, D.W., and Vallier, T. L., eds., Geology and Offshore resources of Pacific Island Arcs-Tonga Region. Circ.-Pac. Counc: Energy Miner. Resour., Earth Sci. Ser., Houston, Texas, Energy Miner. Resour., v. 2, p. 373–414.

Cox, K. G., Bell, J. D., and Pankhurst, R. J., 1979, The interpretation of igneous rocks: George Allen and Unwin, London, 450 p.

Fritzke, B., 1994, Growing cell structures—A self-organizing network for unsupervised and supervised learning: Neural Netw., v. 7, no. 9, p. 1441–1460.

Kohonen, T., 1995, Self-organising maps, *in* Huang, T., and Schroeder, M., eds., Springer Series in Information Sciences, v. 30: Springer-Verlag, Berlin, Germany, 362 p.

Lacassie, J. P., Ruiz-del-Solar, J., Roser, B. P., Belousova, E., Ortiz, E., Hervé, F., 2003, Discovering geochemical patterns using Self-organizing neural networks: a new perspective for analysis in the Earth Sciences: Workshop on Self-Organizing Maps, Congress Proceedings: Hibikino, p. 191–196.

Lacassie, J. P., Roser, B. P., Ruiz-del-Solar, J., Hervé, F., 2004, Visualization of geochemical datasets by using neural networks: a novel perspective for sedimentary provenance analysis: Sedimentary Geol., v. 165, no.1, p. 175–191.

Rollinson, H., 1993, Using geochemical data: evaluation, presentation, interpretation: Longman, Harlow, 352 p.

Vesanto, J., Alhoniemi, E., Himberg, J., Kiviluoto, K., and Parviainen, J., 1999, Self-organizing map for data mining in MATLAB: the SOM toolbox: Simul. News Eur., v. 25, no. 54, ARGE Simulation News, Vienna.

Walker, A. J., Cross, S. S., and Harrison, R. F., 1999, Visualisation of biomedical datasets using growing cell structure networks: a novel diagnostic classification technique: Lancet, v. 354, p. 1518–1521.