Discovering geochemical patterns using self-organizing neural networks: a new perspective for sedimentary provenance analysis

Juan Pablo Lacassie[a,*], Barry Roser[b,1], Javier Ruiz Del Solar[c,2], Francisco Hervé[a,3]

[a] *Departamento de Geología, Universidad de Chile, Casilla 13518, Correo 21, Santiago, Chile*
[b] *Department of Geoscience, Shimane University, Matsue 690-8504, Japan*
[c] *Departamento de Ingeniería Eléctrica, Universidad de Chile, Casilla 412-3, 6513027, Santiago, Chile*

Abstract

An unsupervised neural network technique, Growing Cell Structures (GCS), was used to visualize geochemical differences between sandstones of four different sedimentary provenance groups: P1 (mafic), P2 (intermediate), P3 (felsic), and P4 (recycled). Multidimensional data of four sandstone data sets comprising major elements, log-normalized major elements, trace elements, and high field strength elements (HFSE) were projected into colored two-dimensional maps that can be visually and quantitatively interpreted. The cluster structure and variable distributions produced show that each sedimentary provenance group can be distinguished in the neural maps according to a unique combination of major or trace element concentrations. In these terms, the distinguishing features of each provenance group are: P1—high Fe_2O_3t, TiO_2, MgO, MnO, CaO, P_2O_5, Sc, V, Cr, and Cu; P2—intermediate Fe_2O_3t, TiO_2, MgO, MnO, CaO, Sc, V, and Cu; P3—intermediate to high K_2O, intermediate SiO_2 and Al_2O_3, low Fe_2O_3t and TiO_2, and intermediate to low Nb, Rb, and Th; P4—high SiO_2, Y, Nb, Rb, Th, Ba, and Zr, coupled with low Al_2O_3, CaO, Na_2O, Fe_2O_3t, MgO, MnO, and TiO_2. The elemental associations in P1, P2, and P3 reflect petrogenetic evolution of first-cycle sources, whereas the associations in P4 are compatible with the combined effects of recycling, weathering, and heavy mineral concentration.

Keywords: Neural networks; Visual analysis; Unsupervised learning; Sedimentary provenance; Geochemistry; Sandstone

1. Introduction

1.1. Geochemical study of terrigenous sedimentary rocks

The geochemical composition of terrigenous sedimentary rocks is a function of a complex interplay of variables. Processes such as weathering, transport, diagenesis, sorting, and heavy mineral concentration (Johnsson, 1993; McLennan et al., 1993) may modify compositions inherited from source. If the influence of

* Corresponding author. Fax: +56-2-6963050.
E-mail addresses: jlacassi@cec.uchile.cl (J.P. Lacassie), roser@riko.shimane-u.ac.jp (B. Roser), jruizd@cec.uchile.cl (J. Ruiz Del Solar), fherve@cec.uchile.cl (F. Hervé).
[1] Fax: +81-852-326469.
[2] Fax: +56-2-6720162.
[3] Fax: +56-2-6963050.

these processes is minor, and if transport distances are short, the compositions of first-cycle sandstones predominantly reflect the nature and proportion of their detrital components and hence their provenance (e.g., Bhatia, 1983; Roser and Korsch, 1988). If intense weathering or extensive recycling is involved, the composition of derived sediments may be severely modified from that of their original source terranes. However, in these cases, distinctive geochemical signatures of the processes involved may in turn be imparted to the sediments (Nesbitt and Young, 1984; McLennan et al., 1993).

Major and trace element geochemistry provides clues to both provenance type and tectonic setting, and much effort has been directed at determining which elements are the more discriminating. However, the large number of variables involved can hamper clear visualization and interpretation of the geochemical information as a whole. Most of the pioneering work that deals with provenance analysis of multidimensional major and trace element geochemical data sets was carried out using multiple discriminant analysis (MDA; Bhatia, 1983; Roser and Korsch, 1988) or principal component analysis (PCA; Bhatia and Crook, 1986). Both are multivariable statistical techniques that are used to visualize first-order differences between groups of samples. MDA is a "supervised" technique that will not distinguish "natural" groups within sets of data, as it relies upon prior knowledge of the groupings (Le Maitre, 1982). In contrast, PCA can distinguish between natural groups without prior assumptions. However, being a type of linear fitting, non-linear relations between samples or variables are rejected.

In this study, we use Growing Cell Structures or GCS (Fritzke, 1996), a specific type of artificial neural network, to visualize geochemical differences between four first-order sedimentary provenance categories of terrigenous sedimentary rocks: P1 (mafic), P2 (intermediate), P3 (felsic), and P4 (recycled), as defined by Roser and Korsch (1988).

1.2. Artificial neural networks, self-organizing maps (SOM), and GCS

The basic task of any artificial neural network is to learn how to map real observations into classes in a decision space, using a given decision or neural rule. This mapping has different names in different contexts. Those most often used include pattern recognition, pattern classification, or simply classification. The adaptive process of "teaching" a given mapping to a neural network is known as training, and the set of data used for training purposes is known as a training set. In contrast, those data used for testing the performance of the neural network after training is completed is known as a test set. If such training is carried out using as a training set in which the classification of the observations (samples or features) is known, the training process is known as "supervised" learning. During supervised learning, the parameters of the neural network (synaptic weights or connections) are adaptively adjusted in response to successive presentations of the training set. The learning process proceeds in this way: (i) a sample is introduced into the neural network under training; (ii) the neural network produces an answer or decision, i.e., assigns a class to the sample; (iii) an error signal is generated by comparing the decision of the network to the actual class of the sample; (iv) this error signal is used for adapting the parameters of the neural network; (v) this process is repeated using many examples until stable values for the network parameters are attained (i.e., the neural network has learnt). In contrast, "unsupervised" learning is carried out without using any a priori classification of the samples. In this case, the parameters of the neural network are adjusted taking into account measures among the samples, such as distances. When unsupervised learning is employed, terms such as clustering, self-organizing, self-discovering or self-learning are used, rather than classification. The samples or features to be self-organized are normally multidimensional observations. For this reason, they are considered as vectors (input or feature vectors) and vector notation is employed (e.g., distance, cluster, and input space).

Kohonen's SOM (Kohonen, 1995) constitute a very well-known and widely used neural network model which employs unsupervised learning. SOM adapts to the training data (input vectors) in a way such that a high-dimensional input space is projected onto a two-dimensional rectangular array of interconnected units. This two-dimensional array is known as a feature map, and the arrangement of its units (connected by edges) is known as map topology. The feature map is a topology-preserving map (Kohonen, 1995), i.e., simi-

Table 1
Provenance groups, sample suites, and sources

Group	Nme	Nte	Sample suite	Source
P1*	35	31	Maitai Terrane (New Zealand)	Roser and Korsch (1988); Roser (unpublished data)†
P1	33		Solomons Islands–Woodlark Basin seafloor sediments	Crook et al. (1984)
P1*	23	24	Izu–Bonin island arc (Japan)	Gill et al. (1994)†
P1*	20	19	Brook St. Terrane (N.Z.)	Roser (unpublished data)†
P1	9		Baldwin Formation–Tamworth Group (Australia)	Chappell (1968); Bhatia (1985)
P1	6		Oceanic Island Arc sandstone averages	Bhatia (1983)
P1*	6	6	Kays Creek Formation, Caples Terrane (N.Z.)	Roser et al. (1993)†
Sub-total	132	80		
%	15.2	12.8		
P2*	46	46	Tuapeka Group (N.Z.)	Roser et al. (1993)†
P2*	35	34	North Range Group, Murihiku Terrane (N.Z.)	Roser et al. (2000a, 2002) †
P2*	31	26	Waipapa Terrane (New Zealand)	Roser and Korsch (1988); Palmer et al. (1995)†
P2	22		Uyak Complex (Kodiak Island, Alaska)	Connelly (1978)
P2*	24	24	Upper Peak and Bold Peak Fmtns, Caples Terrane (N.Z.)	Roser et al. (1993)†
P2	9		Hill End Suite (Australia)	Bhatia (1985)
P2	6		Continental Island Arc sandstone averages	Bhatia (1983)
P2	4		Baldwin Formation (Australia)	Chappell (1968)
Sub-total	177	130		
%	20.4	20.7		
P3*	256	241	Torlesse Terrane (Rakaia and Pahau terranes, N.Z.)	Roser et al. (1995a)†
P3*	30	29	Kaihikuan–Warepan Murihiku Terrane (N.Z.)	Roser et al. (2000a, 2002) †
P3	22		St. Ynez Mts. (California, USA)	Van de Kamp et al. (1976)
P3	17		Shimanto Belt (Shikoku, Japan)	Shinjoe (1995)†
P3	14		Franciscan Complex (California, USA)	Bailey et al. (1964)
P3	9		Holocene Sands (Salton Basin, California, USA)	Van de Kamp et al. (1976)
P3	7		Hodgkinson Basin (Australia)	Bhatia (1985)
Sub-total	355	270		
%	40.9	43.1		
P4*	82	77	Buller Terrane (N. Z.)	Roser et al. (1995b, 1996) †
P4*	43	43	Tanabe Group (Honshu, Japan)	Roser et al. (2000b)†
P4	20		Horton Group (Nova Scotia, Canada)	Murphy (2000)†
P4*	14	13	Snowy Mountains (Australia)	Wyborn and Chappell (1983)
P4*	14	14	Takaka Terrane (N. Z.)	Roser et al. (1995b, 1996) †
P4	13		Libby Creek Group (Wyoming, USA)	Crichton and Condie (1993)†
P4	10		Bendigo–Cookman Suite (Australia)	Bhatia (1983, 1985)
P4	8		Pongola Supergroup (South Africa)	Wronkiewicz and Condie (1987)†

(*continued on next page*)

Table 1 (*continued*)

Group	Nme	Nte	Sample suite	Source
Sub-total	204	147		
%	23.5	23.4		
Total	868	627		

Number of analyses (Nme) refers to the major element data set ($\Sigma n = 868$). Number of analyses (Nte) refers to the trace element data set ($\Sigma n = 627$) drawn from the astered sample suites, excluding a small number of samples with highly anomalous abundances or missing values for single elements.

* Trace element data sources.

† Analyses drawn from literature and unpublished data, not used in Roser and Korsch (1988).

lar input vectors correspond to units of the feature map that are close to each other. Hence, the underlying structure of the high-dimensional input space is maintained and can be discovered by exploring the two-dimensional feature map. As similar input vectors group into subsets called clusters, the cluster structure of the input space also appears in the two-dimensional feature map. One limitation of SOM is that its map structure must be specified in advance. This is generally not possible in an optimal way, since a necessary piece of information, the probability distribution of the input vectors, is usually not available.

GCS is an extension and development of SOM, whose map can be described as a two-dimensional array of interconnected units organized in the form of triangles. In contrast to SOM feature maps, GCS map topology is not fixed, but is generated gradually during an incremental learning process which assumes no prior knowledge of group affiliation. Starting from an initial topology of three units organized as one two-dimensional triangle, GCS adapts the overall map structure by inserting new units into those regions that represent large portions of the input data (Fritzke, 1994). The topology is thus determined by the problem being examined.

In this study, we used GCS to project large multi-dimensional geochemical data sets into two-dimensional representations or feature maps where the information most relevant to the clustering task is discovered through visual inspection.

2. Methods

We have used literature-derived data sets, mostly of ancient sedimentary successions (Table 1) to examine geochemical patterns in sandstones of different sedimentary provenance types. These include data from different geographic regions and of differing age. The data sets thus incorporate both spatial and temporal diversity.

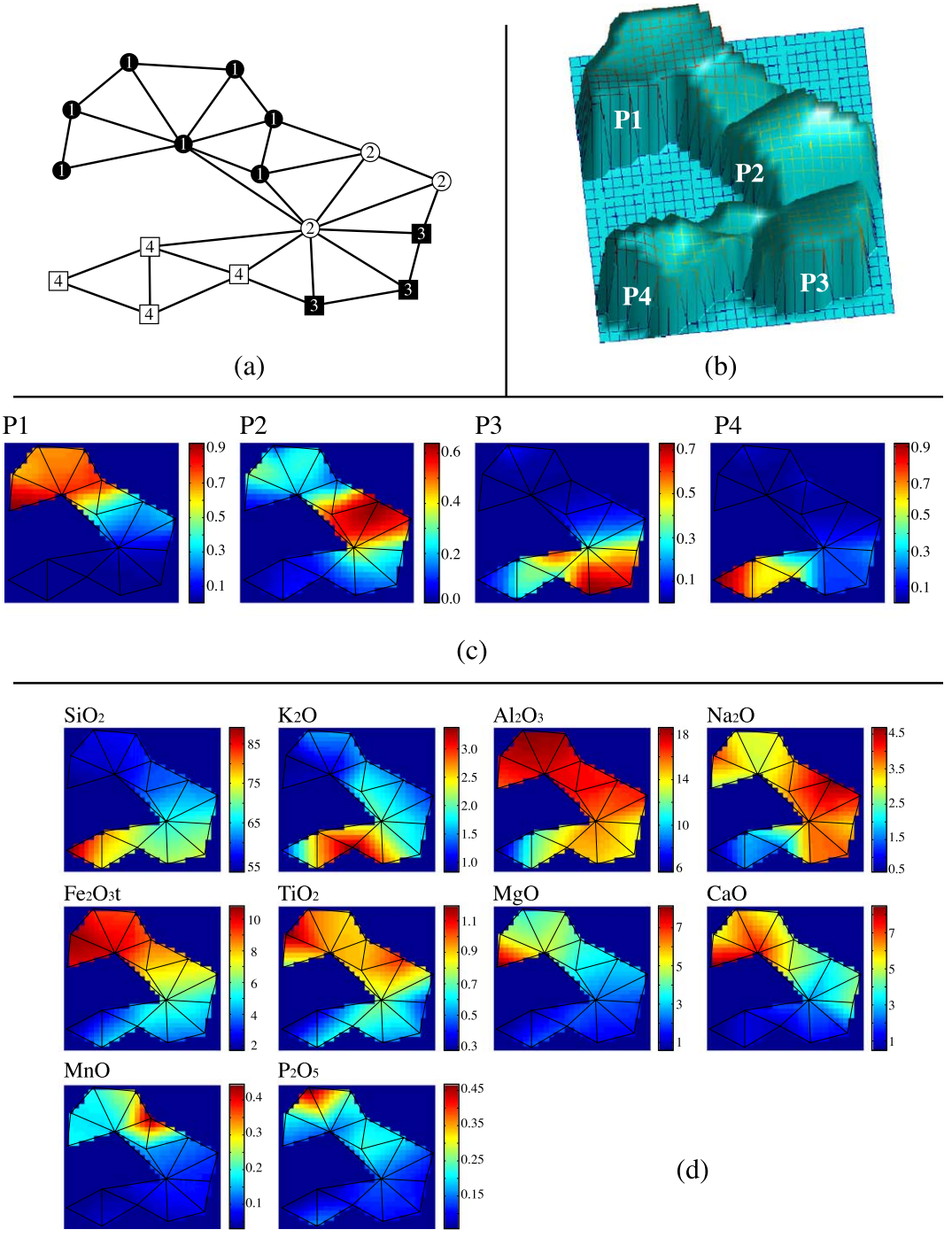
Four data sets were constructed using selected sandstone analyses. The sandstones suites chosen have well-established provenance type, not determined solely by geochemical means, and metamorphic grade lower than upper greenschist to lower amphibolite facies.

The data sets are:

(1) Major elements: comprising 868 sandstone analyses with concentration data (wt.%) for the 10 major and minor oxides (SiO_2, TiO_2, Al_2O_3, K_2O, Na_2O, CaO, Fe_2O_3t, MgO, MnO and P_2O_5) conventionally analyzed.
(2) Log-normalized major elements: comprising log-normalized values of concentration data for 827 samples drawn from the first data set. Analyses which had zero values for any variables were dropped, and the remaining data then normalized. The aim of this was to remove closure, and thus test the effect of closure in the analysis of the first data set.

Fig. 1. Visualization of posterior probabilities and input variable distributions for a 17-unit GCS network trained on 868 cases for the major element data set. (a) Provenance-related unit distribution over the GCS map structure. (b, c) Three-dimensional (b) and two-dimensional (c) visualizations of the posterior probability distributions. Four clusters have been modeled by the network, with clear separation between each. Each cluster is associated with a specific sedimentary provenance group (P1 to P4). For the two-dimensional visualizations (c) the posterior probability values are shown by the color axes at right. (d) Two-dimensional visualizations of the distributions of the input variables. Distributions of each variable can easily be compared visually with that of the posterior probabilities, and with each other. The scales at right indicate whole rock concentrations (wt.% oxide).

(3) Trace elements: comprising 627 analyses with concentration data (ppm) for 11 trace elements (Y, Nb, Rb, Th, Ba, Zr, V, Sc, Ce, Cr and Sr) in all samples. A second analysis was also carried out on this data set adding La and Cu to test the effect of additional elements.



(4) High field strength elements (HFSE): This data set is a subset of the trace element set, restricted to HFSE. It comprises 627 analyses with concentration data (ppm) for six trace elements (Y, Nb, Th, Zr, Sc, Ce) in all samples. The aim of this data set was to test if results were improved by removal of potentially mobile elements (Rb, Ba, V, Cr, and Sr).

The sandstones in each data set were classified into one of the four first-order provenance groups defined by Roser and Korsch (1988). These are: P1 (mafic)—first-cycle basaltic and lesser andesitic detritus; P2 (intermediate)—dominantly andesitic detritus; P3 (felsic)—felsic plutonic and volcanic detritus; and P4 (recycled)—mature polycyclic quartzose detritus. Many of the sample suites employed were used by Roser and Korsch (1988) to define their classification scheme and to test their discriminant functions. They are therefore already classified according to this scheme. These data were supplemented with newer analyses drawn from the literature, along with some unpublished data (Table 1). Each of these suites were also classified into provenance groups on the basis of their discriminant function scores, and the analyses tagged with their group affiliation. To allow direct comparison between samples, all major element analyses were normalized to 100% loss-on-ignition free. The same normalization factors were applied to the trace element data.

All four data sets include sandstones of each provenance group, although the samples are not evenly distributed between groups. In the major element data set, 41% of the samples are from the P3 group, 24% from P4; and P1 and P2 comprise 15 and 20%, respectively (Table 1). The log-normalized data set has similar proportions (P3 42%; P4 21%; P1 21%; P2 16%). In the trace and HFSE data sets, 43% are from the P3 group, whereas P4, P2, and P1 comprise 23%, 21%, and 13%, respectively.

Each sandstone data set was analyzed using the GCS algorithm. The calculations were made using the GCSVIS software (Walker et al., 1999) which, for each provenance group, converts the frequency of data samples assigned to each GCS map unit (prior probability), into a prediction for a new sample associated to each GCS map unit (posterior probability). In the context of pattern recognition, the prior probability corresponds to the probability of occurrence of a given class, and thus reflects a priori knowledge. The posterior probability corresponds to the probability of a class given an observation. The posterior probabilities of the different provenance groups are then displayed on color maps, which are lain over the final GCS map structure (e.g., Fig. 1c). Alternatively, to generate a three-dimensional visualization, the posterior probabilities can be displayed as altitudes, i.e., *z*-axis values, ranging between 0 and 1 (e.g., Fig. 1b). Average values for individual input variables at each GCS map unit, can also be displayed on color maps and overlain on the final GCS structure, allowing simple visual comparison with the posterior probability color maps. (e.g., Fig. 1d). Those input variables whose color map distributions are similar to the posterior probability

Table 2
Provenance related cluster distributions of the analyzed data sets

Cluster	*N*	% P1	% P2	% P3	% P4
Major elements					
P1	7	64	12	1	1
P2	3	17	68	7	1
P3	3	2	17	85	0
P4	4	17	3	8	97
Total	17	100	100	100	100
Log-normalized major elements					
P1	3	79	21	0	0
P2	2	20	50	2	0
P3	3	1	26	95	11
P4	6	0	4	4	89
Total	14	100	100	100	100
Trace elements					
P1	4	75	8	0	0
P2	3	24	76	2	5
P3	2	1	15	79	6
P4	5	0	2	19	88
Total	14	100	100	100	100
HFSE					
P1	4	91	19	0	0
P2	3	8	70	1	1
P3	2	0	10	80	12
P4	7	1	1	19	87
Total	16	100	100	100	100

For each data set, percentage values (% P) refer to the percentage of training samples of each provenance group that were assigned to each provenance-related cluster, as defined by their respective map units (Figs. 1a, 2a, 3a, and 4a).
N = number of cluster assigned units.

Table 3
Characteristic elemental concentrations in each provenance group (P1–P4), for each of the analyzed data sets

Cluster	P1		P2		P3		P4	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Major elements (wt.%)								
SiO_2	54.86	2.84	63.03	3.30	72.54	2.47	80.44	8.05
TiO_2	0.94	0.23	0.92	0.19	0.50	0.10	0.47	0.26
Al_2O_3	17.28	1.93	16.47	1.02	14.13	1.13	10.24	4.01
Fe_2O_3t	9.93	1.55	7.22	1.30	3.50	0.85	3.15	1.91
MnO	0.20	0.12	0.12	0.04	0.05	0.02	0.04	0.04
MgO	5.20	2.17	2.81	0.71	1.25	0.42	1.31	0.93
CaO	7.01	2.78	3.16	1.14	1.58	0.71	0.74	1.02
Na_2O	3.29	0.93	4.52	0.80	3.79	0.63	1.16	0.76
K_2O	1.10	0.65	1.58	0.61	2.65	0.64	2.34	1.09
P_2O_5	0.21	0.14	0.17	0.05	0.11	0.04	0.11	0.08
N	84		121		300		198	
Log-normalized major elements (wt.%)								
SiO_2	54.64	2.72	63.19	2.80	72.40	2.95	78.88	7.76
TiO_2	0.95	0.23	0.89	0.16	0.50	0.11	0.53	0.24
Al_2O_3	17.23	1.95	16.54	1.41	14.03	1.40	10.89	3.93
Fe_2O_3t	10.08	1.41	7.18	1.01	3.55	0.90	3.57	1.87
MnO	0.20	0.12	0.12	0.05	0.05	0.03	0.05	0.08
MgO	5.37	2.07	2.75	0.63	1.27	0.45	1.54	0.92
CaO	6.98	2.81	3.28	1.12	1.81	1.30	0.79	1.09
Na_2O	3.34	0.98	4.27	0.77	3.70	0.68	1.21	0.71
K_2O	1.03	0.59	1.60	0.71	2.65	0.68	2.43	1.04
P_2O_5	0.20	0.13	0.18	0.05	0.11	0.03	0.13	0.09
N	102		87		330		154	
Trace elements (ppm)								
Ba	150.0	121.2	322.2	129.5	606.8	126.1	466.6	152.7
Cr	74.2	56.4	43.4	17.0	32.9	9.6	57.1	28.8
Nb	1.8	1.4	5.2	1.2	8.2	1.3	10.4	4.2
Rb	16.7	15.4	45.8	19.8	92.0	18.0	121.8	45.8
Sc	31.9	7.3	18.8	4.2	8.3	2.1	8.6	4.3
Sr	333.8	216.7	263.8	141.1	342.7	120.8	85.8	47.4
Th	1.6	1.1	5.7	1.8	10.9	1.6	13.3	4.2
V	315.5	87.8	182.3	48.5	73.6	20.4	61.6	29.1
Y	20.0	4.8	26.2	4.6	20.2	3.1	29.2	11.7
Zr	73.8	29.3	162.7	20.2	189.6	31.2	229.7	115.2
Ce	13.5	7.0	34.1	7.5	52.7	8.5	61.3	22.8
N	60		99		213		130	
HFSE (ppm)								
Nb	2.0	1.4	5.7	1.3	8.1	1.3	9.9	4.8
Sc	29.7	8.4	17.8	3.8	8.1	1.9	8.1	4.8
Th	1.9	1.2	6.6	1.9	10.8	1.7	12.8	5.1
Y	19.8	4.6	26.8	4.5	20.0	3.1	28.8	14.0
Zr	78.4	29.3	169.5	20.6	193.7	44.7	222.9	86.8
Ce	14.2	6.8	37.4	7.4	51.8	7.7	60.7	26.7
N	73		91		229		102	

color map distribution of a particular provenance group (e.g., Fe_2O_3t and P1 color map distributions in Fig. 1) are highly likely to be significant discriminating elements for that provenance group (Walker et al., 1999). Similarly, if two or more input variables have similar color map distributions (e.g., Th, Nb, and Ce color map distributions in Fig. 4), a close correlation exists between them.

A provenance group or class can be assigned to each unit of the final GCS map (e.g., Fig 1a). The percentage of samples assigned to each unit of the final GCS map can be computed for each provenance group (P1 to P4) using the prior probability information. The provenance group or class assigned to each map unit is that with the greatest percentage. If the difference between the greatest and the next largest percentage is small (< 10%), group assignment is determined using topological criteria, i.e., the classes assigned to the surrounding map units. This provenance group assignment allows estimation of the percentage of training samples of each provenance group assigned to each cluster, as the percentage of samples assigned to the respective map units (Table 2). It also enables, for each of the analyzed data sets, computation of the average concentrations of those samples correctly assigned to their provenance clusters (Table 3), and of the average values of selected geochemical parameters commonly used for provenance discrimination (Table 4).

3. Results

GCS analysis of the data sets resulted in four distinct network topologies or maps (Figs. 1, 2, 3, and 4). For the major element, trace element, and HFSE data sets, the GCS network modeled the problem as four clusters, which appear in the color maps as four distinct areas with high posterior probabilities (Figs. 1c, 3c, and 4c), respectively). The GCS network associated each of these areas with a specific sedimen-

Note to Table 3:
Means and standard deviations (Sd.) are expressed in weight percent (wt.%) for the major and log-normalized data sets, and parts per million (ppm) for the trace and HFSE data sets. For each provenance group, *N* refers to the number of correctly assigned training samples used for calculation of the mean and the standard deviation.

Table 4
Examples of distinctive geochemical parameters for the provenance groups (P1–P4)

Cluster	P1 (mean)	P2 (mean)	P3 (mean)	P4 (mean)
Major elements (wt.%)				
$Fe_2O_3t + MgO$	15.13	10.03	4.74	4.47
SiO_2/Al_2O_3	3.17	3.83	5.13	7.85
K_2O/Na_2O	0.33	0.35	0.70	2.03
$Al_2O_3/(CaO + Na_2O)$	1.68	2.15	2.63	5.39
N	84	121	300	198
Trace elements (ppm)				
Th/Sc	0.05	0.30	1.32	1.54
Zr/Th	44.87	28.78	17.40	17.30
K/Th (Major elements)	5524.98	2323.42	2017.92	1465.04
K/Th (Log-n Major elements)	5175.63	2344.92	2019.12	1517.52
Ti/Zr (Major elements)	26	10	6	3
Ti/Zr (Log-n Major elements)	77	33	16	14
N	60	99	213	130
HFSE (ppm)				
Th/Sc	0.06	0.37	1.33	1.57
Zr/Th	41.19	25.79	17.87	17.46
K/Th (Major elements)	4774.46	1998.81	2027.57	1523.88
K/Th (Log-n Major elements)	4472.56	2017.31	2028.78	1578.47
Ti/Zr (Major elements)	24	9	6	3
Ti/Zr (Log-n Major elements)	24	9	6	4
N	73	91	229	102

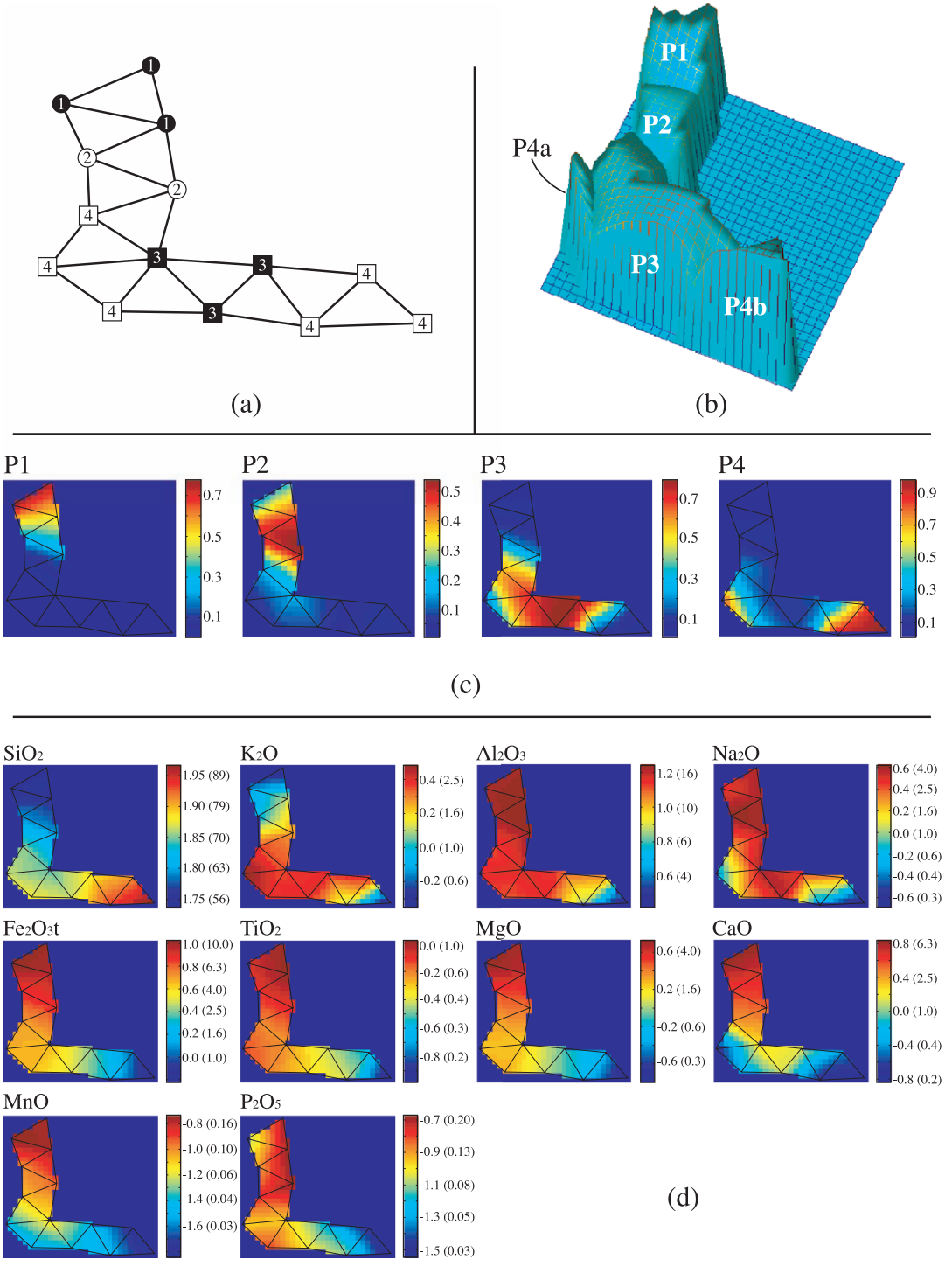
For each provenance group, *N* refers to the number of correctly assigned training samples used for calculation of the mean value.

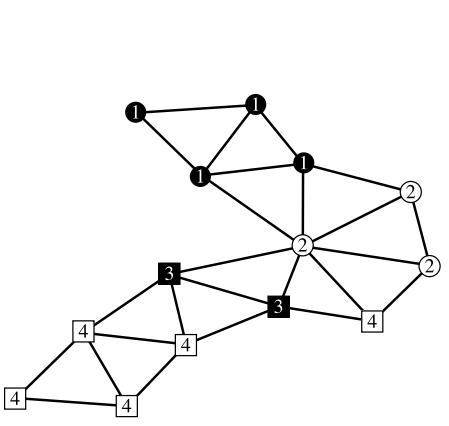
tary provenance group (P1 to P4). For the log-normalized data set, five clusters appear (Fig. 2c). Three are associated with a specific sedimentary provenance group (P1 to P3), whereas the other two are associated with P4 (P4a and P4b; Fig. 2b). For each data set, the GCS color maps show also that a more refined differentiation among sedimentary provenance groups is possible, because visual comparison between the input variable and provenance group map distributions enables identification of individual input variables (major or trace elements) with specific concentration ranges in discrete provenance subregions. For example, very high P_2O_5 concentrations appear in a subregion of P1 (Fig. 1d); very high log-normalized K_2O in the P4 subregion (Fig. 2d); high Ba in P3 (Fig. 3d); and very high Zr in P4 (Fig. 4d). There is no major overlap between neighboring clusters. Instead, the transition between them is sharp, indicating a clear decision boundary (Walker et al., 1999). Inspection reveals that all four GCS maps (Figs. 1, 2, 3, and 4) allow visualization of the role of each input variable in the overall separation of the provenance-related clusters.

3.1. Major element data set

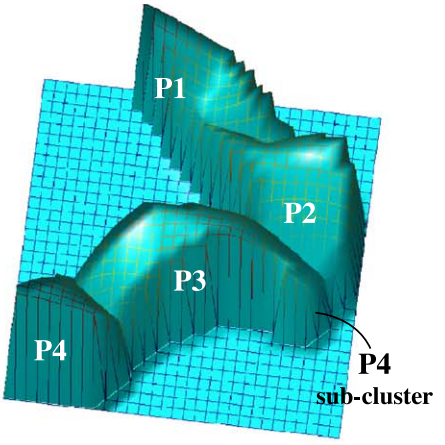
High SiO_2 values (>75 wt.%) are closely related to P4, intermediate values (70–75) to P3, intermediate to low values (60–68) to P2, and low values (<60) are typical of P1 (Fig. 1). High K_2O abundances (>2.5) are closely related to P3 and P4, whereas intermediate to

Fig. 2. Visualization of posterior probabilities and input variable distributions for a 14-unit GCS network trained on 827 cases for the log-normalized data set (log-normalized values of the concentration data of the first data set). (a) Provenance-related unit distribution over the GCS map structure. (b, c) Three-dimensional and two-dimensional visualizations of the posterior probability distributions. Five clusters have been modeled by the network, with clear separation between each. Three clusters are associated with specific sedimentary provenance groups (P1, P2, P3), and the other two with P4 (P4a and P4b). For the two-dimensional visualizations, the posterior probability values are shown by the color axes at right. (d) Two-dimensional visualizations of the input variable distributions. The scales are logarithmic (log values of whole rock concentrations). Bracketed values to the right are corresponding concentrations in weight percent.

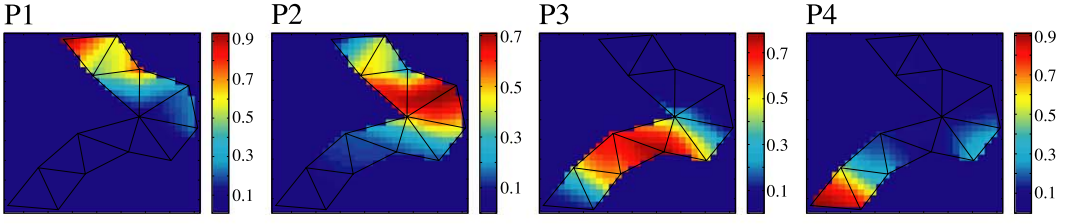

(a)
P1
P2
P4a
P3
P4b
(b)
P1
0.7
0.5
0.3
0.1
P2
0.5
0.4
0.3
0.2
0.1
P3
0.7
0.5
0.3
0.1
P4
0.9
0.7
0.5
0.3
0.1
(c)
SiO2
1.95 (89)
1.90 (79)
1.85 (70)
1.80 (63)
1.75 (56)
K2O
0.4 (2.5)
0.2 (1.6)
0.0 (1.0)
-0.2 (0.6)
Al2O3
1.2 (16)
1.0 (10)
0.8 (6)
0.6 (4)
Na2O
0.6 (4.0)
0.4 (2.5)
0.2 (1.6)
0.0 (1.0)
-0.2 (0.6)
-0.4 (0.4)
-0.6 (0.3)
Fe2O3t
1.0 (10.0)
0.8 (6.3)
0.6 (4.0)
0.4 (2.5)
0.2 (1.6)
0.0 (1.0)
TiO2
0.0 (1.0)
-0.2 (0.6)
-0.4 (0.4)
-0.6 (0.3)
-0.8 (0.2)
MgO
0.6 (4.0)
0.2 (1.6)
-0.2 (0.6)
-0.6 (0.3)
CaO
0.8 (6.3)
0.4 (2.5)
0.0 (1.0)
-0.4 (0.4)
-0.8 (0.2)
MnO
-0.8 (0.16)
-1.0 (0.10)
-1.2 (0.06)
-1.4 (0.04)
-1.6 (0.03)
P2O5
-0.7 (0.20)
-0.9 (0.13)
-1.1 (0.08)
-1.3 (0.05)
-1.5 (0.03)
(d)


1
1
1
1
2
2
2
3
3
4
4
4
4
4

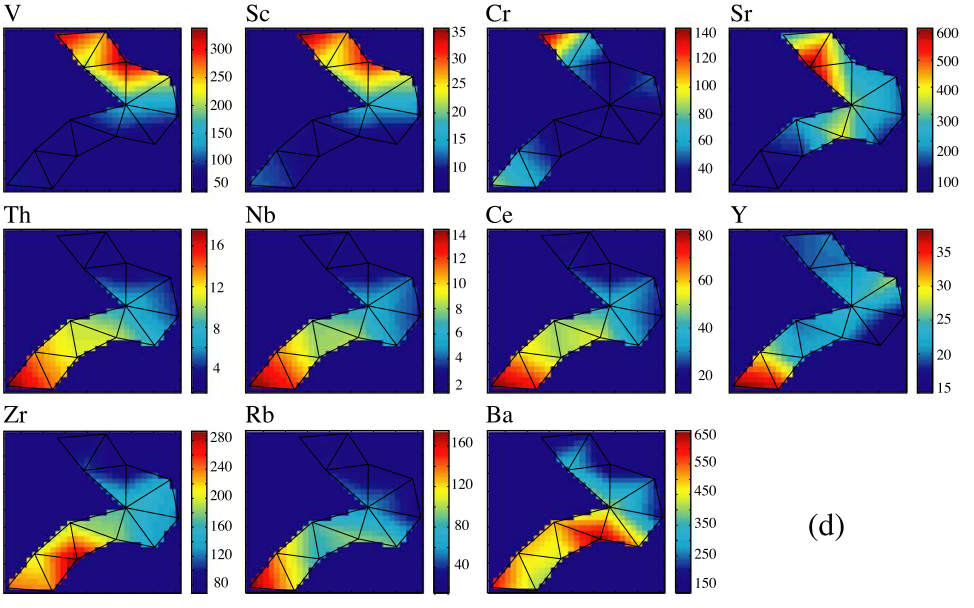
(a)


P1
P2
P3
P4
P4
sub-cluster

(b)


P1
0.9
0.7
0.5
0.3
0.1
P2
0.7
0.5
0.3
0.1
P3
0.7
0.5
0.3
0.1
P4
0.9
0.7
0.5
0.3
0.1

(c)


V
300
250
200
150
100
50
Sc
35
30
25
20
15
10
Cr
140
120
100
80
60
40
Sr
600
500
400
300
200
100
Th
16
12
8
4
Nb
14
12
10
8
6
4
2
Ce
80
60
40
20
Y
35
30
25
20
15
Zr
280
240
200
160
120
80
Rb
160
120
80
40
Ba
650
550
450
350
250
150

(d)

low values (1.3–2.2) are related to P2 and a discrete subregion of P1. High TiO_2 concentrations (>0.9) are also localized in discrete subregions of P1 and P2, whereas relatively uniform low values (close to 0.6) are associated with P3, with dispersion towards P4. High values of Al_2O_3 (>15) are evenly distributed across both P1 and P2, whereas homogeneous and moderately high values (close to 14 wt.%) are closely related to P3. Values in P4 are intermediate to low (< 12 wt.%) High values of Na_2O (>4 wt.%) are related to P2, whereas low values (< 2 wt.%) are strongly associated with P4. High Fe_2O_3t abundances (>8.5 wt.%) are closely related to P1, whereas low values (< 3.5) are related to P4. High values of CaO (>8 wt.%) and MgO (>6 wt.%) are moderately correlated and, together with high values of MnO (>0.5 wt.%) and P_2O_5 (>0.4 wt.%), are localized in discrete subregions of P1.

3.2. Log-normalized data set

After conversion of the log-normalized values, the results of the analysis of the log-*n* data set can be expressed in terms of major element concentrations. A feature of this analysis is the appearance of two clusters (4a and 4b) in the P4 group (Fig. 2). High SiO_2 concentrations (>74 wt.%) are again typical of P4 (P4b), intermediate values (68–74) are associated with P3, intermediate to low values (59–66 wt.%) are related to P2, and low values (< 59 wt.%) are characteristic of P1. Very high K_2O values (>2.7 wt.%) are typical of P4a, intermediate to high K_2O abundances (1.8–2.7 wt.%) are associated with P3, intermediate to low values (1.0–1.6 wt.%) are related to P2, and very low values (< 1.0) are related to P1 and P4b. High values of Al_2O_3 (>15 wt.%) are evenly distributed across both P1 and P2, and intermediate to moderately high Al_2O_3 abundances (11–14) extend into P3 and P4a, whereas low values (< 9 wt.%) are typical of P4b. Intermediate to high Na_2O abundances (2.0–3.6 wt.%) are distributed across P1, P2 and P3, although the greatest concentrations (>3.6 wt.%) occur in P2. Low Na_2O values (< 1.5 wt.%) are related to both P4a and 4b. TiO_2, Fe_2O_3t, and MgO are closely correlated, and high values (TiO_2 >1.0 wt.%, Fe_2O_3t> 7.9, MgO > 4.0) are related to P1, in contrast to the low values (TiO_2 < 0.5, Fe_2O_3t < 2.5, MgO < 0.8) characteristic of P4b. Intermediate to high TiO_2, Fe_2O_3t, and MgO values (0.8–1.0, 5.6–7.9, and 2.0–3.6, respectively) are associated with P2, whereas intermediate abundances (0.5–0.6, 2.5–5.0, and 0.8–2.0, respectively) are associated with P3 and P4a. CaO and MnO are also correlated. High values (CaO>5.0, MnO > 0.20) are typical of P1, whereas low concentrations (CaO < 0.6 and MnO < 0.04) are characteristic of P4. Intermediate to high CaO and MnO values (3.2–5.0 and 0.16–0.20, respectively) are associated with P2, whereas intermediate to low abundances (0.6–2.5 and 0.04–0.08, respectively) are related to P3. Very high P_2O_5 abundances (>0.20) are localized in a discrete subregion of P2, with some dispersion of uniformly high values (close to 0.16) towards P1 and the rest of P2. Low P_2O_5 values (< 0.06) are related to P4b.

3.3. Trace element data set

Trace element concentrations (ppm) also show distinct groupings (Fig. 3). Highest Y abundances (>32 ppm) are associated with P4, intermediate values (20–26) with P2 and P3, and low values (< 20) with P1. Very high Zr values (>260) are localized in discrete subregions of P3 near the P4–P3 boundary, with considerable dispersion of high values (220–260) towards P4. Intermediate Zr abundances (120–160) are typical of P2, and low values (< 120) are related to P1. High Ba values (>550) are localized in discrete subregions of P3 and P4, whereas very low values (< 250) are associated with P1. Highest abundances of Sr (>450) occur in a discrete subregion of P2 adjacent to the P1–P2 boundary, with considerable dispersion of high to intermediate values (300–450) towards both regions. Very low concentrations (< 110 ppm) are associated with P4.

Fig. 3. Visualization of posterior probabilities and input variable distributions for a 14-unit GCS network trained on 627 cases for the trace element data set. (a) Provenance-related unit distribution over the GCS map structure. (b, c) Three-dimensional and two-dimensional visualizations of the posterior probability distributions. As for Fig. 1, the network has modeled the problem as four main clusters, each associated with a specific sedimentary provenance group (P1 to P4). A weak P4 subcluster appears between P3 and P2. For the two-dimensional visualizations, the posterior probability values are shown by the color axes at right. (d) Two-dimensional visualizations of the input variable distributions. Scales for the input variables are concentrations in ppm.

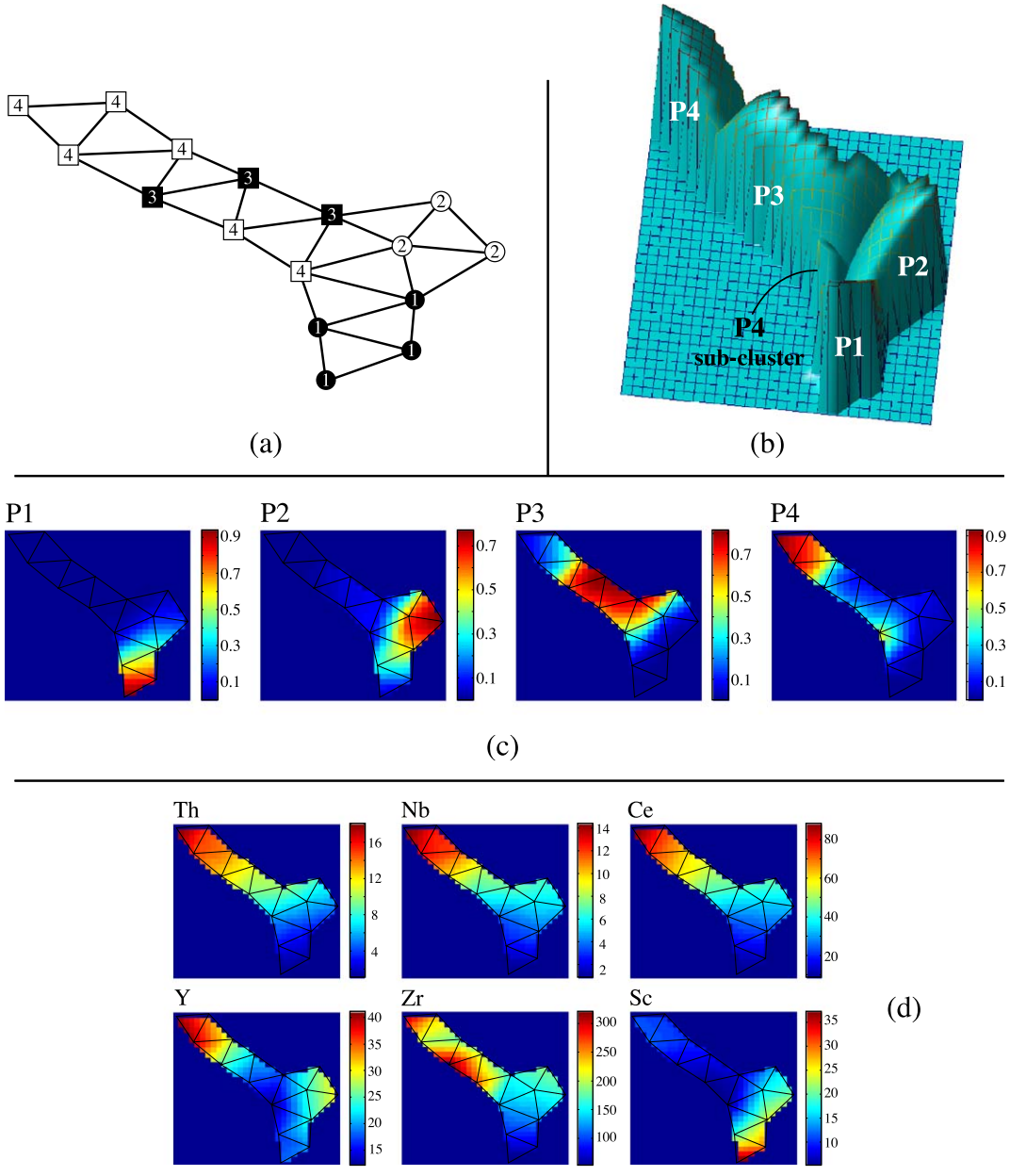


Fig. 4. Visualization of posterior probabilities and input variable distributions for a 16-unit GCS network trained on 627 cases for the HFSE data set. (a) Provenance-related unit distribution over the GCS map structure. (b, c) Three-dimensional and two-dimensional visualizations of the posterior probability distributions. The network has again modeled the problem as four clearly separated main clusters, and a weak P4 subcluster. For the two-dimensional visualizations, the posterior probability values are shown by the color axes at right. (d) Two-dimensional visualizations of the input variable distributions. Scales for the input variables are concentrations in ppm.

Nb, Rb, Th, and Ce are highly correlated. High values of all four elements (Nb > 11 ppm, Rb > 130, Th > 13, Ce > 62) are closely associated with P4, in contrast to the low values (Nb < 4, Rb < 40, Th < 4, Ce < 26) characteristic of P1. Abundances in P3 and P2 are intermediate. V and Sc are also strongly correlated,

and very high (V>260, Sc>26) and high (V>170, Sc>17) values are related to P1 and P2, respectively, in contrast to low values in P3 and P4. Highest Cr values (>90 ppm) are localized in a discrete subregion of P1. Somewhat anomalously, intermediate values (55–80) occur in both P4 and P1, whereas low values (<50) intervene in P2, P3, and in a discrete subregion of P1.

3.4. HFSE data set

The results of the analysis of the HFSE data set (Fig. 4) are in general agreement with the above results. High Y abundances (>32 ppm) are closely related to P4, intermediate to low values (25–30) to P2, and low values (<23) to P3 and P1. Very high Zr concentrations (>260) are localized in a discrete subregion of P4, and in a discrete subregion near the P4–P3 boundary, with considerable dispersion of high values (220–260) towards both P3 and P4. Intermediate to low Zr values (140–190) are associated with P2, whereas low values (<120) characterize P1. Nb, Th, and Ce are again closely correlated, with highest values (Nb>12, Th>15, Ce>70) in P4, and lowest values (Nb<3, Th<4, Ce<20) in P1. High to very high Sc abundances (Sc>26) and intermediate to high Sc values (17–22) are associated with P1 and P2, respectively, and low values (<15) with P3 and P4.

4. Discussion

Although the whole rock analyses used here were classified into provenance groups according to petrographic or geologic evidence, a priori knowledge was not used during the unsupervised training of the GCS or in determining its structure. Only after GCS determined the final structure was the class information tagged to each analysis used to distinguish the posterior probability distribution of each provenance group. Therefore, clustering or separation appears in the trained network as a natural characteristic of the samples of all four data sets analyzed.

The numerical results for the major element data set (Table 2) show that P3 and P4 are the best-defined clusters (highest percentages of correctly assigned samples), with minor dispersion of P3 towards P4 and P2. The poorer definition (lower percentages of correctly assigned samples) of the P1 and P2 clusters is mainly due to the relatively larger dispersion between these two groups, coupled with dispersion of P2 towards P3 and dispersion of P1 towards P4.

The marked differences in the color map distributions of TiO_2, MgO, MnO, CaO, and P_2O_5, between the major element and log-normalized data sets (Figs. 1 and 2) show that the analysis of the major element data set is affected by the closure of the data. Compared to the major element data set, Fe_2O_3t, TiO_2, MgO, MnO, and CaO in the log-normalized data set are more continuously distributed in their respective color maps, and display a higher inter-element correlation. This pattern is more consistent with the expected behavior of these variables as suggested by previous studies (Bhatia, 1983; Taylor and McLennan, 1985; Roser and Korsch, 1988). Therefore, removing the closure of the major element data set resulted in a general improvement in the results of the GCS analysis.

Compared to the major element data set, cluster definition in the log-normalized data set increases for P1 and P3, and decreases for P2 and P4 (Table 2). Particularly relevant is the lack of P1 dispersion towards P4, whereas poorer definition of the P2 cluster is mainly due to increased dispersion of P2 towards P1 and P3. Examination of the P4 data distribution reveals that the P4a subcluster contains a large proportion of the Buller and Snowy Mountains data (94% and 57%, respectively; Fig. 2). Conversely, the P4b subcluster contains high proportions of the Tanabe and Takaka terrane data (92% and 58%, respectively). Inspection of the P3 and P4 data shows that average TiO_2, Fe_2O_3t, and MgO concentrations in the Buller (0.65, 4.91, 2.35 wt.%, respectively) and Snowy Mountains (0.54, 3.25, 1.35 wt.%) data sets are appreciably greater than those of the Tanabe (0.25, 1.79, 0.63 wt.%) and Takaka (0.41, 2.16, 0.74 wt.%) data sets. The Buller and Snowy Mountains averages lie closer to the respective averages of the P3 Murihiku (0.61, 3.47, 1.09 wt.%), Rakaia (0.51, 3.62, 1.33 wt.%), and Pahau (0.47, 3.43, 1.12 wt.%) data sets. Hence, TiO_2, Fe_2O_3t, and MgO concentrations seem to be relevant factors responsible for the dispersion of the Buller and Snowy Mountains data towards P3 and the resulting split in the P4 cluster.

Table 2 shows that, considered as a whole, cluster definition in the trace element data set is better than that of the major element and log-normalized data sets.

However, similar to the log-normalized data set, a weak P4 subcluster appears between the P2 and P3 clusters (Fig. 3a,b,c). Inspection of the data distribution shows that the majority (93%) of the P4 data assigned to this subcluster belongs to the Tanabe Group data set. Only a small percentage (19%) of the Tanabe Group data is correctly assigned to the main P4 cluster, whereas 60% is assigned to the P4 subcluster, with moderate to high dispersion towards P3 (19%) and P1 (2%).

Two changes in the combination of variables in the trace element data set enabled analysis of the roles of La and Cu. When La was added to the original variable selection and Sr rejected, a clearer discrimination between the sedimentary provenance clusters was attained. As should be expected from their geochemical coherence, La and Ce had almost the same distribution, with high La values (>32) related to P4, intermediate values (20–32) related to P3, and low values (<20) related to P2 and P1. On the contrary, when Cu was added and Ce removed from the original variable selection, the four provenance clusters became less defined, although they were still distinguishable. Cu was moderately correlated with V and Sc, with intermediate to high Cu values (>110) associated with P1, and low values (<60) associated with P3.

Compared to the trace element data set, the HFSE analysis shows a marked improvement in definition of the P1 cluster, with a high percentage (91%) of correctly assigned samples (Table 2). Dispersion of P2 towards P1 increases and the dispersion of P4 towards P2 decreases. As for the trace element data set, the distribution of the P4-related units (Fig. 4a) suggests dispersion of P4 data towards P3. Examination of the data distribution reveals significant dispersion (32%) of the Tanabe Group data towards P3. The remainder (68%) of the Tanabe Group data are assigned to the two units that define the P4 subcluster located between the P2 and P3 clusters (Fig. 4a,b,c). Although less than for the trace element data set, a significant percentage (76%) of the P4 data assigned to these latter P4 units belongs to the Tanabe Group data set.

Examination of the individual P4 data shows that the average Sc, Cr, Th, Nb, Ce, Y, and Zr abundances of the Snowy Mountains, Buller, and Takaka data sets (Table 5) are similar to or are greater than those of the average upper continental crust (UCC) values of McLennan (2001). In contrast, except for Ce in the Takaka set, average concentrations for all six elements in the P4 Tanabe data set are significantly less than those in the above suites and in UCC (Table 5). The Tanabe averages lie close much closer to those of the P3 Rakaia, Pahau, and Murihiku data sets, in which abundances of all elements are lower than or equal to UCC composition. Therefore, for the trace and the HFS element data sets, the Sc, Cr, Th, Nb, Ce, Y, and Zr concentrations appears to control the dispersion of the Tanabe Group data towards P3, and the concentration of Tanabe samples in the P4 subcluster located between the P2 and P3 clusters. This dispersion is likely to be a product of the depositional environment and provenance of the Tanabe Group, which was deposited in a Miocene forearc basin in the Shimanto accretionary complex of Southwest Japan. Conglomerates in the lower Tanabe Group contain clasts derived from older Shimanto rocks (Tanabe Research Group, 1985, 1992, 1993), establishing its recycled nature. However, the bulk chemical composition of the Tanabe Group differs little from its Shimanto protolith (Roser et al., 2000b), suggesting that in this case, recycling has not produced the chemistry characteristic of the P4 group.

The success of the GCS technique in capturing the sedimentary provenance cluster structure and the correlations of the geochemical variables can be judged by comparison with previous numerical estimates. Being a non-linear approach, the results of the GCS analysis can only be compared indirectly with linear numerical measures. In terms of associations of elements, the observed GCS visual correlations agree with published correlation coefficients between trace

Table 5
Average concentrations (ppm) of several trace elements in P3 and P4 data sets compared to the average upper continental crust values of McLennan (2001)

Suite	Grp	Sc	Cr	Th	Nb	Ce	Y	Zr
Snowy Mountains	P4	7	55	12	10	55	27	269
Buller	P4	11	71	14	13	72	36	229
Takaka	P4	11	55	14	13	28	26	226
Tanabe	P4	4	24	9	5	35	15	174
UCC	–	13.6	83	10.7	12	64	22	190
Rakaia	P3	8	33	11	9	55	21	192
Pahau	P3	8	25	10	7	48	20	203
Murihiku	P3	11	27	11	8	53	22	204

and major elements of five eastern Australian Paleozoic graywacke suites deposited in differing tectonic settings (Bhatia and Crook, 1986). Provenances of these suites range from andesite to older metasedimentary rocks, or in the terminology of this study, from P1 to P4. The groupings of the geochemical variables which best discriminate between each provenance group as identified here are also consistent with the results of previous studies (Bhatia, 1983; Bhatia and Crook, 1986; Taylor and McLennan, 1985). The range of average P1–P3 Th/Sc ratios in the trace and HFSE data sets (Table 4) is consistent with that of Th/Sc ratios (< 0.01 to 1.8) reported for modern turbidites from active margin tectonic settings (McLennan et al., 1990). Systematic increase in Ce, Th, Nb contents and K_2O/Na_2O, SiO_2/Al_2O_3, and $Al_2O_3/(CaO+Na_2O)$ ratios from P1 to P4 and systematic decrease in V, Sc, Fe_2O_3t+MgO, K/Th, and Ti/Zr (Tables 3 and 4) are also in general agreement with the results of previous studies (Bhatia, 1983; Bhatia and Crook, 1986; Roser and Korsch, 1988).

The results also support the suggestion of Roser and Korsch (1988) that the chemistry of first-cycle sandstones of the P1–P3 groups largely reflects that of their source terranes, and that the P1–P3 succession resembles a ''differentiation'' trend between average mafic and felsic source compositions. The distinctive elemental associations found in P1 (e.g., high TiO_2, Fe_2O_3t, CaO, MgO, V, Sc, and Cu; and low Ba and Zr) are consistent with mafic parentage. Conversely, the lesser TiO_2, Fe_2O_3t, CaO, MgO, V, Sc, and Cu, and greater SiO_2, K_2O, Zr, Nb, Rb, and Th abundances identified in P3 are consistent with a felsic igneous provenance. Equally, the pattern of high SiO_2, Zr, Nb, Rb, Th, and low Al_2O_3, CaO, Na_2O, and Sr in P4 is compatible with the modifying effects of weathering and recycling, as feldspar and labile lithic fragments are lost by progressive weathering and abrasion, and quartz and resistant heavy minerals are concentrated relative to the source.

The sharpness of the GCS map boundaries between the four provenance groups (Figs. 1, 2, 3, and 4) are likely influenced by the data set used as input. Intuitively, we would expect some gradation between P1, P2, and P3 as a result of varying proportions of mafic, intermediate, and felsic detritus in individual samples and suites. Some overlap between P3 and P4 can also be expected, since many P4 sandstones are derived from felsic crystalline basement rocks. This and the effects of recycling of preexisting sediments and varying intensity of source area weathering could also produce more gradual transition between P3 and P4. The latter is supported by the dispersion of part of the P4 data towards P3, and their concentration in P4 subclusters located between the P2 and P3 clusters, for the log-*n*, trace, and HFS element data sets.

Overlap between P3 and P4 was tested by redefining the Pahau terrane samples, classified as P3 in this work, as P4. The Pahau terrane was largely derived from cannibalistic reworking of the adjacent Rakaia terrane (MacKinnon, 1983), plus a minor volcanogenic component derived from more mafic inboard terranes (Roser and Korsch, 1999). Geochemical indices reflect a shift towards the characters expected in recycled sediments, although the change in some is lessened by the volcanogenic component (Roser and Korsch, 1999). The Pahau rocks can therefore be regarded as representatives of the first stages of large-scale recycling, similar to the Tanabe suite discussed above. After GCS analysis of the redefined data set, four provenance clusters could still be distinguished. Although the relationships between P3, P2, and P1 changed a little, the distinction between P3 and P4 was less defined than in the original data set, with greater dispersion of P4 towards areas of high probability for P3. The P3 cluster was also less defined, with lower maximum probability than in the original data set, and a P4 subcluster reflecting the Pahau data appeared in the P3 cluster.

Addition of more data to this analysis could thus reduce the separation between the groups, as suggested by the gradations observed in the maps of the input variables. However, we would expect the characteristic elemental fingerprints observed in each group to be maintained. Despite the above limitation, the GCS maps provide useful information on the broad geochemical differences between sedimentary provenance types, as they highlight higher-order correlations and allow visual perception of subtle inter-relations between variables. Examination of the proportions of individual suites or subsets of the data assigned to individual map units also provides a method for identifying provenance anomalies or homogeneity within those groups. Although the clusters and elemental associations described here could also be identified on conventional variations diagrams, visual compar-

isons are facilitated by the color maps, and subsets can be readily identified, as in the case of the Tanabe data.

The application of our results to other sedimentary successions is also supported by the broad geographical distribution and varied age of the sedimentary successions analyzed, and because the GCS map topology of both data sets broadly resembles the "natural" literature-derived affinities between sedimentary provenance groups. The success of the analysis undertaken here suggests that GCS is of wide application in other branches of earth science, to identify natural clusters or anomalies and to explore the relationships between numeric variables in large data sets. After training, the neural networks of the SOM type can be used to recognize data that are similar to any of the examples shown during the training phase. Therefore, SOM neural networks trained with the analyzed sedimentary data sets can be used to classify new data of unknown sedimentary provenance type.

5. Conclusions

Our results show that significant geochemical differences between sandstones of different provenance can be visualized through GCS color maps. Their multidimensional projection of both major and trace element data resulted in "natural" separated clusters. The P1, P2, and P3 provenance groups are associated with three clearly separated clusters. The P4 provenance group is also associated a with a separate cluster, although it shows a marked tendency to be dispersed towards P3. Compared to the major element data, GCS analysis of the trace element data resulted in better-defined clusters, especially when the analysis was restricted to immobile HFSE. Both the visual and numerical results show that a unique combination of major and trace element concentrations characterize each sedimentary provenance group. In that respect, our results are similar to those of previous studies (Bhatia, 1983; Taylor and McLennan, 1985; Bhatia and Crook, 1986; Roser and Korsch, 1988). However, GCS visualization has clear interpretative and analytical advantages over the linear numerical approach of these works, because it takes into account higher-order correlations and allows visual perception of the relationships between variables. Gross aspects of the geochemical differences in the P1, P2, and P3 provenance groups appear to be strongly related to geochemical differences between their sources, whereas the elemental associations identified in P4 are compatible with the effects of weathering and recycling.

Acknowledgements

This research was supported by CONICYT under the Doctoral Fellowship Fondecyt no. 2010010 to Lacassie. Our thanks to Dr. Jacobus Le Roux, Departamento de Geología, Universidad de Chile, for his comments and discussion. We are also grateful to Keith Crook, Bryan Murphy and Mukul Bhatia for their very helpful critical reviews, which also extended the scope of this contribution. This work was conducted using the Growing Cell Structure Visualisation (GCSVIS) Toolbox, written and developed by Andrew J. Walker, Dr. Robert F. Harrison, and Dr. Simon S. Cross in the Departments of Automatic Control and Systems Engineering and Pathology at the University of Sheffield, UK.

References

Bailey, E.H., Irwin, W.P., Jones, D.L., 1964. Franciscan and related rocks and their significance in the geology of western California. Bull.-Calif., Div. Mines Geol. 183, 177.

Bhatia, M.R., 1983. Plate tectonics and geochemical composition of sandstones. J. Geol. 91, 611–627.

Bhatia, M.R., 1985. Rare earth element geochemistry of Australian Paleozoic graywackes and mudrocks: provenance and tectonic control. Sediment. Geol. 45, 97–113.

Bhatia, M.R., Crook, K.A.W., 1986. Trace element characteristics of graywackes and tectonic setting discrimination of sedimentary basins. Contrib. Mineral. Petrol. 92, 181–193.

Chappell, B.W., 1968. Volcanic greywackes from the Upper Devonian Baldwin Formation, Tamworth-Barraba district, New South Wales. J. Geol. Soc. Aust. 15, 87–102.

Connelly, W., 1978. Uyak Complex, Kodiak Island, Alaska: a cretaceous subduction complex. Geol. Soc. Amer. Bull. 89, 755–769.

Crichton, J.G., Condie, K.C., 1993. Trace elements as source indicators in cratonic sediments: a case study from the early Proterozoic Libby Creek Group, southeastern Wyoming. J. Geol. 101, 319–332.

Crook, K.A.W., Taylor, G.R., Bolton, B.R., 1984. Geochemistry of indurated and unconsolidated sediments from the seafloor of the Solomon Islands and Woodlark Basin. CCOP/SOPAC Tech. Rep. 34, 275–312.

Fritzke, B., 1994. Growing cell structures—a self-organizing net-

work for unsupervised and supervised learning. Neural Netw. 7, 1441–1460.

Fritzke, B., 1996. Growing self-organising networks—why. In: Verleysen, M. (Ed.), European Symposium on Artificial Neural Networks. D-Facto Publishers, Brussels, pp. 61–72.

Gill, B.J., Hiscott, R.N., Vidal, P., 1994. Turbidite geochemistry and evolution of the Izu–Bonin arc and continents. Lithos 33, 135–168.

Johnsson, M.J., 1993. The system controlling the composition of clastic sediments. In: Johnsson, M.J., Basu, A. (Eds.), Processes Controlling the Composition of Clastic Sediments. Spec. Pap.-Geol. Soc. Am., vol. 284. Boulder, Colorado, pp. 1–19.

Kohonen, T., 1995. Self-organising maps. In: Huang, T., Schroeder, M. (Eds.), Springer Series in Information Sciences, vol. 30. Sringer-Verlag, Berlin, Germany. 362 pp.

Le Maitre, R.W., 1982. Numerical Petrology: Statistical Interpretation of Numerical Data Elsevier, Amsterdam. 281 pp.

MacKinnon, T.C., 1983. Origin of the Torlesse terrane and coeval rocks, South Island, New Zealand. Geol. Soc. Amer. Bull. 93, 625–634.

McLennan, S.M., 2001. Relationships between the trace element composition of sedimentary rocks and upper continental crust. Geochem. Geophys. Geosyst. 2 (Paper number 2000GC000109).

McLennan, S.M., Taylor, S.R., McCulloch, M.T., Maynard, J.B., 1990. Geochemical and Nd–Sr isotopic composition of deep-sea turbidites: crustal evolution and plate tectonic associations. Geochim. Cosmochim. Acta 54, 2015–2050.

McLennan, S.M., Hemming, S., McDaniel, D.K., Hanson, G.N., 1993. Geochemical approaches to sedimentation, provenance and tectonics. In: Johnsson, M.J., Basu, A. (Eds.), Processes Controlling the Composition of Clastic Sediments. Spec. Pap.-Geol. Soc. Am., vol. 284. Boulder, Colorado, pp. 21–40.

Murphy, J.B., 2000. Tectonic influence on sedimentation along the southern flank of the late Paleozoic Magdalen basin in Canadian Appalachians: geochemical and isotopic constraints on the Horton Group in the St. Mary basin, Nova Scotia. Geol. Soc. Amer. Bull. 112, 997–1011.

Nesbitt, H.W., Young, G.M., 1984. Prediction of some weathering trends of plutonic and volcanic rocks based on thermodynamic and kinetic considerations. Geochim. Cosmochim. Acta 48, 1523–1534.

Palmer, K., Mortimer, N., Nathan, S., Isaac, M.J., Field, B.D., Sircombe, K.N., Black, P.M., Bush, S., Orr, N.W., 1995. Chemical and petrographic analyses of some New Zealand Paleozoic–Mesozoic metasedimentary and igneous rocks. Inst. Geol. Nucl. Sci. Sci. Rep. 95/16 (37 pp.).

Roser, B.P., Korsch, R.J., 1988. Provenance signatures of sandstone–mudstone suites determined using discriminant function analysis of mayor element data. Chem. Geol. 67, 119–139.

Roser, B.P., Korsch, R.J., 1999. Geochemical characterization, evolution and source of a Mesozoic accretionary wedge: the Torlesse terrane, New Zealand. Geol. Mag. 136, 493–512.

Roser, B.P., Mortimer, N., Turnbull, I.M., Landis, C.A., 1993. Geology and geochemistry of the Caples Terrane, Otago, New Zealand: compositional variations near a Permo-Triassic arc margin. In: Ballance, P. (Ed.), Sedimentary Basins of the World, 2: South Pacific Sedimentary Basins. Elsevier, Amstersdam, pp. 3–19.

Roser, B.P., Grapes, R.H., Palmer, K., 1995a. XRF analyses of sandstones and argillites from the Torlesse terrane, New Zealand. Res. Sch. Earth Sci. Vic. Univ. Wellingt., Geol. Stud., Spec. Publ. 15, 1–40.

Roser, B.P., Palmer, K., Nathan, S., Cooper, R.A., Tulloch, A.J., 1995b. Whole-rock analyses of Paleozoic sedimentary rocks from the Buller and Takaka terranes. Inst. Geol. Nucl. Sci. Sci. Rep. 95/14 (15 pp.).

Roser, B.P., Cooper, R.A., Nathan, S., Tulloch, A.J., 1996. Reconnaisance sandstone geochemistry, provenance, and tectonic setting of the lower Paleozoic terranes of the West Coast and Nelson, New Zealand. N.Z. J. Geol. Geophys. 39, 1–16.

Roser, B.P., Coombs, D.S., Korsch, R.J., Campbell, J.D., Mortimer, N., Grapes, R., 2000a. Whole-rock analyses of sandstones, siltstones, and tuffs from the Murihiku Terrane, New Zealand. Res. Rep.-Vic. Univ. Wellingt., Sch. Earth Sci. 7 (26 pp.).

Roser, B.P., Kimura, J.-I., Hisatomi, K., 2000b. Whole-rock elemental abundances in sandstones and mudrocks from the Tanabe group, Kii Peninsula, Japan. Geosci. Rep. Shimane Univ. 19, 101–112.

Roser, B.P., Coombs, D.S., Korsch, R.J., Campbell, J.D., 2002. Whole-rock geochemical variations and evolution of the Triassic–Jurassic arc-derived Murihiku Terrane, New Zealand. Geol. Mag. 139, 665–685.

Shinjoe, H., 1995. Whole rock chemistry of shale and sandstone from the Northern Shimanto belt in southwestern Shikoku. J. Sedimentol. Soc. Jpn. 41, 33–38.

Tanabe Research Group, 1985. Tanabe Group in the western coastal region (Hiki–Kamoi area) of the Kii Peninsula. Mem. Fac. Edu., Wakayama Univ., Nat. Sci. 34, 3–24 (in Japanese, English abstract).

Tanabe Research Group, H., 1992. Sedimentary facies and stratigraphy of the Asso Formation—study on the Asso Formation, Tanabe Group (Part 1). Earth Sci. (Chikyu Kagaku) 46, 369–383 (in Japanese, English abstract).

Tanabe Research Group, 1993. Sedimentary facies and sedimentary environments of the Shimomisu fan-delta—study on the Asso Formation, Tanabe Group (Part 2). Earth Sci. (Chikyu Kagaku) 47, 1–16 (in Japanese, English abstract).

Taylor, S.R., McLennan, S.M., 1985. The Continental Crust: its Composition and Evolution Blackwell, Oxford. 312 pp.

Van de Kamp, P.C., Leake, B.E., Senior, A., 1976. The petrography and geochemistry of some Californian arkoses with application to identifying gneisses and metasedimentary origin. J. Geol. 84, 195–212.

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

Wronkiewicz, D.J., Condie, K.C., 1987. Geochemistry of archean shales from the witwatersrand supergroup, South Africa: source-area weathering and provenance. Geochim. Cosmochim. Acta 51, 2401–2416.

Wyborn, L.A.I., Chappell, B.W., 1983. Chemistry of the ordovician and silurian greywackes of the Snowy Mountains, southeastern Australia: an example of chemical evolution of sediments with time. Chem. Geol. 39, 81–92.