Available online at www.sciencedirect.com
SCIENCE DIRECT®
Geochimica et Cosmochimica Acta 70 (2006) 1839–1848
Geochimica
www.elsevier.com/locate/gca

# Tectonic discrimination of basalts with classification trees

Pieter Vermeesch *

*Department of Geological and Environmental Sciences, Stanford University, USA*

Received 8 April 2005; accepted in revised form 15 December 2005

**Abstract**

Traditionally, geochemical classification of basaltic rocks of unknown tectonic affinity has been performed by discrimination diagrams. Although easy to use, this method is fairly inaccurate because it only uses bi- or trivariate data. Furthermore, many popular discrimination diagrams are statistically not very rigorous because the decision boundaries are drawn by eye, and they ignore closure, thus violating the rules of compositional data analysis. Classification trees approximate the data space by a stepwise constant function, and are a more rigorous and potentially more effective way to determine tectonic affinity. Trees allow the simultaneous use of an unlimited number of geochemical features, while still permitting visualization by an easy-to-use, two-dimensional graph. Two classification trees are presented for the discrimination of basalts of mid-ocean ridge, ocean island, and island arc affinities. The first tree uses 51 major, minor, and trace elements and isotopic ratios and should be used for the classification of fresh basalt samples. A second tree only uses high field strength element analyses and isotopic ratios, and can also be used for basalts that have undergone alteration. The probability of successful classification is 89% for the first and 84% for the second tree, as determined by 10-fold cross-validation. Even though the trees presented in this paper use many geochemical features, it is not a problem if some of these are missing in the unknown sample. Classification trees solve this problem with surrogate variables, which give more or less the same decision as the primary variables. The advantages of the classification tree approach over discrimination diagrams are illustrated by a comparative test on a sample dataset of known tectonic affinities. Although arguably better than discrimination diagrams, classification trees are not perfect, and the limitations of the method are illustrated on a published dataset of basalts from the Pindos Basin (Greece).


## 1. Introduction

Igneous rocks form in a wide variety of tectonic settings, including mid-ocean ridges, ocean islands, and volcanic arcs. It is a problem of great interest to igneous petrologists to recover the original tectonic setting of mafic rocks of the past. When the geological setting alone cannot unambiguously resolve this question, the chemical composition of these rocks might contain the answer. The major, minor, and trace elemental composition of basalts shows large variations, for example as a function of formation depth (e.g., Kushiro and Kuno, 1963). Traditionally, statistical classification of geochemical data has been done with discrimination diagrams (e.g., Chayes and Velde, 1965; Pearce and Cann, 1971, 1973; Pearce, 1976; Wood, 1980; Shervais, 1982). The decision boundaries of most tectonic discrimination diagrams are drawn by eye (e.g., Pearce and Cann, 1973; Wood, 1980; Shervais, 1982). Although still widely used, these diagrams have some serious problems, including:

- Although (linear) discriminant analysis works in any number of dimensions, visual interpretation is only possible for bi- or trivariate data. By only using two or three geochemical features, a lot of information is not used, resulting in poor performance (large misclassification error).
- The geochemical discrimination diagrams of Pearce and Cann (1971, 1973) and others are often based on collections of the means of many analyses of a certain sample area. This means that, strictly speaking, such diagrams should only be used for the classification of similar data,

* Present address: Institute for Isotope Geology and Mineral Resources, ETH-Zurich, Switzerland.
*E-mail address:* pvermees@pangea.stanford.edu.

0016-7037/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.gca.2005.12.016

i.e., not individual samples but means of composites of multiple samples. Moreover, taking the mean of compositional data is a dangerous thing to do, considering the fact that elemental concentrations are subject to a constant-sum constraint, which is generally not taken into account (e.g., Chayes, 1971; Aitchison, 1986; Woronow and Love, 1990).
- Discrimination diagrams do not easily handle samples with missing data, for example if only Ti and not *V* were measured in the diagram of Shervais (1982).

As an alternative to discriminant analysis which resolves all these issues, this paper suggests classification trees, which are one of the most powerful and popular "data mining" techniques (Hastie et al., 2001). One application in which classification trees have been quite successful is email spam filtering (e.g., Hastie et al., 2001; Carreras and Marquez, 2001). Based on a large training database of predetermined genuine and spam messages, spam filters automatically generate a series of nested yes/no questions that decide which of the two categories (genuine or spam) a new message belongs to. The attributes used in a tree-based spam filter can be the frequencies of certain words or characters as a percentage of the total length of the message, the average length of uninterrupted sequences of capital letters, the total number of capital letters, etc. Spam filtering has many similarities with the problem of tectonic discrimination. In the latter case, the training data will not contain two but three classes (mid-ocean ridge, ocean island, and island arc). The attributes used for the classification will be chemical concentrations and isotopic ratios.

Section 2 will give an introduction to the construction of classification trees. As for discrimination diagrams, it is not really necessary for the end-user to know all the details of the building process, because this has to be done only once (hence this paper) after which they are very easy to use; trees in fact easier to use than discrimination diagrams. Therefore, only a brief introduction to the technique will be given, along with the necessary references for the interested reader.

In Section 3, two classification trees will be presented for the discrimination between basalts from mid-ocean ridge (MORB), ocean island (OIB), and island arc (IAB) settings, based on 756 major and trace element measurements and isotopic ratio analyses, compiled from two publicly available petrologic databases. The first tree uses all major, minor, and trace elements, and should be used for the classification of unaltered samples of basalt. The second tree only uses immobile elements and can also be used for samples that underwent some degree of weathering and/or metamorphism. Beyond this initial selection of suitable features, the construction of the trees is entirely statistical, and involves no further petrological considerations or arbitrary decision boundaries.

In Section 4, both classification trees will be tested. First, a suite of modern basalts of known tectonic affinity will be classified by trees as well as discrimination diagrams. Then, a published dataset of 20 basalts from the Pindos Basin (Greece) will be classified. This will illustrate the limitations of the tree method and serve as a cautionary note, which is valid for all statistical classification methods.

## 2. Method

The following paragraphs give a brief introduction to the theory of classification trees. They were compiled from Breiman et al. (1984), Ripley (1996), and Hastie et al. (2001). The interested reader is referred to these books for more details, while those who are merely interested in the applications can safely skip to Section 3.

### 2.1. Classification trees: introduction

Suppose we have $N$ $J$-dimensional data points $X^n = \{x_1^n, \ldots, x_j^n, \ldots, x_J^n\}$ $(1 \leqslant n \leqslant N)$ belonging to one of $K$ classes: $Y^n = c_1 | \ldots | c_k | \ldots | c_K$. For example, $\underline{X} = \{X^1, \ldots, X^n, \ldots, X^N\}$ might represent $J$ features (e.g., various major and trace element concentrations, isotopic ratios, color, weight, ...) measured in $N$ samples of basalt. $c_1, \ldots, c_K$ might then be tectonic affinity, e.g., "mid-ocean ridge," "ocean island," "island arc,"... The basic idea behind classification and regression trees (CART) is to approximate the parameter space by a piecewise constant function, in other words, to partition $\underline{X}$ into $M$ disjoint regions $\{R_1, \ldots, R_m, \ldots, R_M\}$. An example of such a partition is given in Fig. 1a. Because it is impossible to describe all possible partitions of the feature space, we restrict ourselves to a small subset of possible solutions, the recursive binary partitions (Fig. 1b). Surprising as it may seem, considering the crudeness of this approximation, trees are one of the most powerful and popular data mining techniques in existence. One of the reasons for this is that besides assuring computational feasibility, the recursive partitioning technique described next also allows the representation of the multi-dimensional decision-space as a two-dimensional tree graph (Fig. 1c). Although building a tree might seem complicated, this is not important to the user, because it only has to be done once, after which using a classification tree for data interpretation is extremely simple.

### 2.2. Building a tree

At any point in the recursive process, a partition is defined by two quantities: the split variable $j$ $(1 \leqslant j \leqslant J)$ and the split point $s(-\infty < s < \infty)$. For any given partition $R_m$ of a tree $T$, there are at most $N \times J$ possible binary subpartitions, which can be exhaustively searched. We choose the one that minimizes the "node impurity" $Q_m(T)$. Let $\hat{p}_{mk}$ be the proportion of class $k$ observations in node $m$, then

$$Q_m(T) = \sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk}). \quad (1)$$

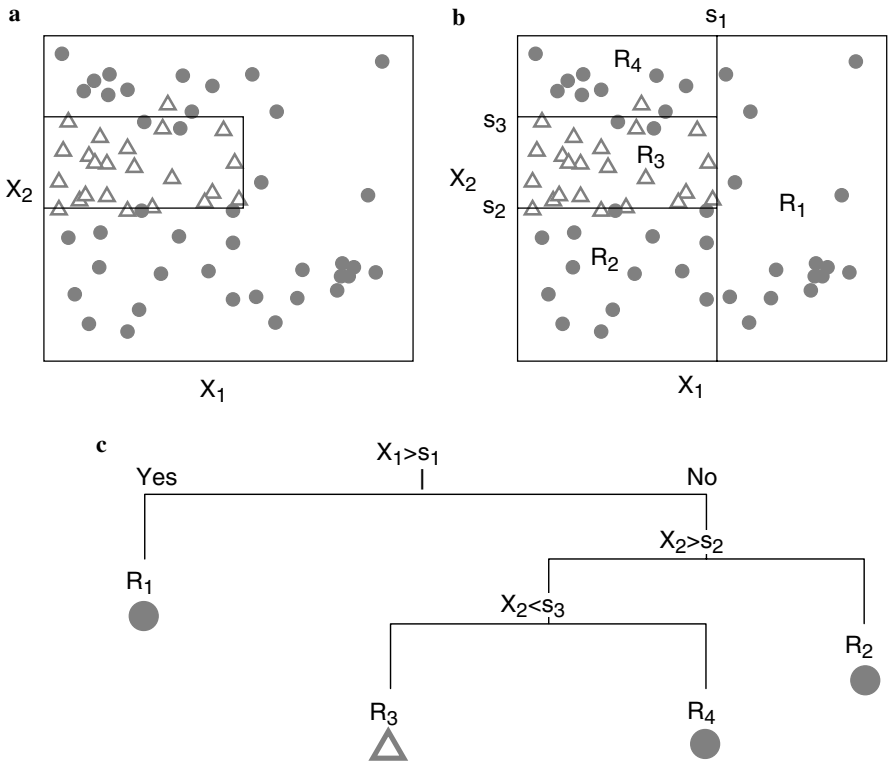


Fig. 1. An example of a bivariate ($X_1$,$X_2$) classification tree. Classification trees approximate the data space with a piecewise constant function (a). To ensure computational feasibility, a recursive binary partitioning method is used (b). Such partitions have the added advantage that the results can be visualized as a two-dimensional graph or "tree" (c). In this and subsequent trees, left branches mean "Yes" and right branches "No."

This particular form of $Q_m(T)$ is called the "Gini index of diversity," but alternatives exist. Note that if we simply used the "misclassification error" (i.e., the number of misclassified data points resulting from a candidate split) as a measure of node impurity, it would be impossible to partition the data set shown in Fig. 1 because any initial split would misclassify all the triangles, yielding the same misclassification error (=23/63 in this case). The recursive partitioning process continues until all the end-nodes are "pure," i.e., all belong to the same class. The maximum sized tree thus obtained perfectly describes the training data. In other words, it has zero bias. However, for the purpose of prediction, this tree is not optimal, because it overfits the training data, causing high variance. The tree with optimal predictive power will be smaller than the largest possible tree, and can be found by "cost-complexity pruning."

### 2.3. Pruning a tree

Define the "cost-complexity criterion" of a tree $T$ as

$$cp_\alpha(T) = \sum_{m=1}^{|T|} N_m Q_m(T) + \alpha|T| \tag{2}$$

with $|T|$ the number of terminal nodes in $T$, $N_m$ the number of observations in the $m$th terminal node, $Q_m(T)$ the "node impurity" defined by Eq. (1) and $\alpha$ a tuning parameter. For a given $\alpha \geqslant 0$, it is possible to find the subtree $T_\alpha \subset T_0$ that minimizes $cp_\alpha(T)$ over all possible subtrees of the largest possible tree $T_0$

$$T_\alpha = \underset{T \subset T_0}{\operatorname{argmin}}\, cp_\alpha(T). \tag{3}$$

Repeating this procedure for a range of values $0 \leqslant \alpha < \infty$ produces a finite nested sequence of trees $\{T_0, T_{\alpha 1}, \ldots, T_{\alpha_{\max}}\}$. Except for $T_0$, these trees will no longer have only pure end-nodes. Impure end-nodes are assigned the class that dominates in them. We then choose the value $\alpha^*$ that minimizes an estimate of future prediction error, for example by "$V$-fold cross-validation." The training data are randomly divided into $V$ (e.g., 10) fractions of equal size. We then grow $V$ overly large trees $T_0^{(v)}$, each time using all but the $v$th sample fraction. Each of these trees then goes through the pruning procedure described before, yielding $V$ nested sequences of trees $\{T_0^{(v)}, T_{\alpha 1}^{(v)}, \ldots, T_{\alpha_{\max}}^{(v)}\}, 1 \leqslant v \leqslant V$. The trees $T_\alpha^{(v)}$ were constructed without ever seeing the cases in the vth fraction. Sending the vth fraction down $T_\alpha^{(v)}$ for each $v = 1, \ldots, V$ thus yields an independent estimate of the misclassification error.

A plot of these cross-validated (CV) prediction errors versus the number of nodes in each of the nested subtrees shows a minimum at some point. As discussed before, trees with fewer nodes tend to have large bias, while the increased CV-cost of overly large trees is caused by their inflated variance. There typically exist several trees with

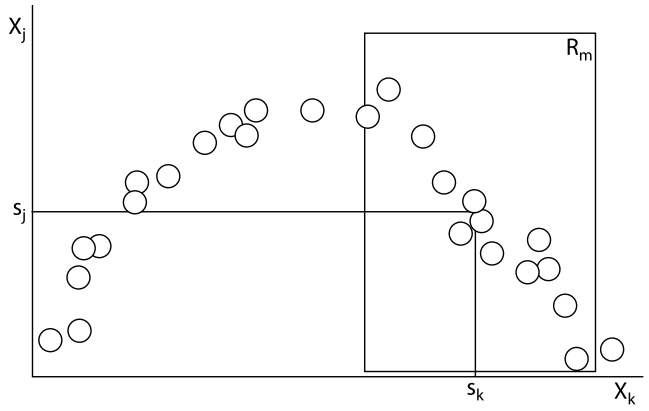


Fig. 2. Primary split variable $X_k$ and split point $s_k$ and surrogate split variable $X_j$ and split point $s_j$. Surrogates answer the question: "which other splits would classify the same objects in the same way as the primary split?"

CV-costs close to the minimum. Therefore, a "1-SE rule" is used, i.e., choosing the smallest tree whose CV misclassification cost does not exceed the minimum CV-cost plus one standard error of the CV-cost for the minimum CV-cost tree. An example of this procedure will be given in Section 3.

### 2.4. Handling missing data

One of the greatest advantages of classification trees is their ability to handle missing data. For example, in a dataset of geochemical analyses, some samples might have been analysed for major and trace elements, while others were only analysed for trace elements and stable isotopes. Yet another set of samples might have been analysed for all elements except Zr, etc. Methods like discriminant analysis cannot easily handle these situations, severely restricting their applicability and power. Both for training and prediction, trees solve the missing data problem by "surrogate splits." Having chosen the best primary predictor and split point (disregarding the missing data), the first surrogate is the predictor and corresponding split point that has the highest correlation with the primary predictor in $R_m$ (Fig. 2). The second surrogate is the predictor that shows the second highest correlation with the primary split variable and so forth.

## 3. Application to the tectonic discrimination of basalts

An extensive dataset of 756 samples was compiled from the online PETDB and GEOROC databases (http://www.petdb.org and georoc.mpch-mainz.gwdg.de). The dataset can be downloaded from electronic annex EA-1. It contains:

- 256 Island arc basalts (IAB) from the Aeolian, Izu-Bonin, Kermadec, Kurile, Lesser Antilles, Mariana, Scotia, and Tonga arcs.
- 241 Mid-ocean ridge (MORB) samples from the East Pacific Rise, Mid Atlantic Ridge, Indian Ocean, and Juan de Fuca Ridge.
- 259 Ocean-island (OIB) samples from St. Helena, the Canary, Cape Verde, Caroline, Crozet, Hawaii-Emperor, Juan Fernandez, Marquesas, Mascarene, Samoan, and Society islands.

Duplicate analyses were excluded from the database to avoid potential bias towards overrepresented samples. Fifty-one geochemical features were used:

- Major element concentrations (in weight percent): $SiO_2$, $TiO_2$, $Al_2O_3$, $Fe_2O_3$, FeO, CaO, MgO, MnO, $K_2O$, $Na_2O$, and $P_2O_5$.
- Minor and trace element concentrations (in ppm): La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu, Sc, V, Cr, Co, Ni, Cu, Zn, Ga, Rb, Sr, Y, Zr, Nb, Sn, Cs, Ba, Hf, Ta, Pb, Th, and U.
- Isotopic ratios: $^{143}Nd/^{144}Nd$, $^{87}Sr/^{86}Sr$, $^{206}Pb/^{204}Pb$, $^{207}Pb/^{204}Pb$, and $^{208}Pb/^{204}Pb$.

The fact that different measurement units are mixed and that for none of the 756 samples all 51 features were measured is not a problem for the construction of the trees, once again illustrating the robustness of the method. The classification trees discussed next were constructed using the `rpart` library of the statistical software package `R`, which can be downloaded free of charge from www.r-project.org. The actual `R`-code is provided in electronic annex EA-2.

### 3.1. A tree using major, minor, and trace elements and isotope ratios

In a first approach, all 51 features were used for the tree construction, including relatively mobile elements such as CaO and $Na_2O$. Therefore, the resulting tree should only be used on fresh samples of basalt. The largest possible tree ($T_0$) has 51 splits, and actually uses only 23 of the 51 selected features. These are: $SiO_2$, $TiO_2$, CaO, $Fe_2O_3$, MgO, $K_2O$, La, Pr, Nd, Sm, Gd, Tb, Yb, Lu, V, Ni, Rb, Sr, Y, Hf, Th, $^{87}Sr/^{86}Sr$, and $^{206}Pb/^{204}Pb$. The remaining 28 features apparently did not contain enough discriminative power. As discussed in Section 2, $T_0$ is not the best possible tree. A plot of relative cross-validation misclassification risk versus tree size shows a minimum at 18 splits (Fig. 3). Using the 1-SE rule then puts the optimal tree size at 8 splits (Fig. 3). The resulting, optimally pruned tree is shown in Fig. 4.

The classification by the optimal tree is remarkably successful. No less than 79% of all the training data correctly fall in just three terminal nodes (encircled in Fig. 4). Only 7% of the training data were misclassified, while the 10-fold cross-validation error is about 11%, corresponding to a success-rate of 89%. In other words, the probability that a sample of unknown tectonic affinity will be classified

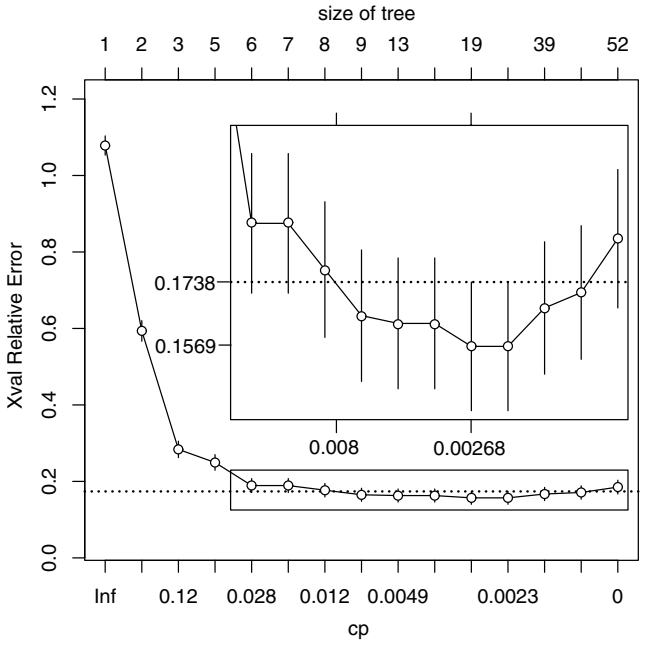


Fig. 3. Choosing the optimal tree size. The *y*-axis shows the 10-fold cross-validation error, relative to the misclassification risk of the root node (=497/756). The *x*-axis shows the size of the tree (i.e., the number of nodes) and the complexity parameter (cp). The first two splits account for 87% of the discriminative power. The inset shows a magnification of the boxed part of the curve, illustrating the 1-SE rule.

correctly is 89%. The first two splits (on $TiO_2$ and Sr) account for 87% of the discriminative power (Fig. 3). In a way, this can be seen as a justification of the use of these elements in popular discrimination diagrams such as the Ti–Zr–Sr diagram (Pearce and Cann, 1973). An analysis of $TiO_2$ and Sr alone already gives a pretty reliable classification. For example, if $TiO_2 \geqslant 2.135\%$, the tree tells us there is a 91% chance that the rock has an OIB affinity. Likewise, 87% of the training data with $TiO_2 < 2.135\%$ and $Sr < 156$ ppm are MORBs. For further discrimination, additional elements can be used, which inevitably increases the chance of missing variables. However, as discussed before, classification trees elegantly resolve this problem with surrogate split variables, which are shown in Table 1.

### 3.2. A tree of HFS elements and isotopic ratios only

Even though the classification tree built in the previous section (Fig. 4) performs very well for fresh basalts, this might not necessarily be the case for weathered or metamorphosed samples. For example, a lot of the power of this tree depends on Sr, which is considered a mobile element (e.g., Rollinson, 1993). Also MgO, Ni, and Rb are species used in Fig. 4 that are mobile to some degree. Performance of our classification, which was based on fresh samples of basalt, cannot be guaranteed if used for the tectonic discrimination of samples that have undergone alteration. Therefore, an alternative tree was built that only uses the so-called high field strength (HFS) cations, which are characterized by an ionic potential greater than two (Rollinson, 1993), as well as the isotopic ratios of Sr, Nd, and Pb, because these are considered less prone to change during alteration than the concentrations themselves. The following 28 features were used: $TiO_2$, La, Ce, Pr, Nd, Sm, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu, Sc, Y, Zr, Nb, Hf, Ta, Pb,

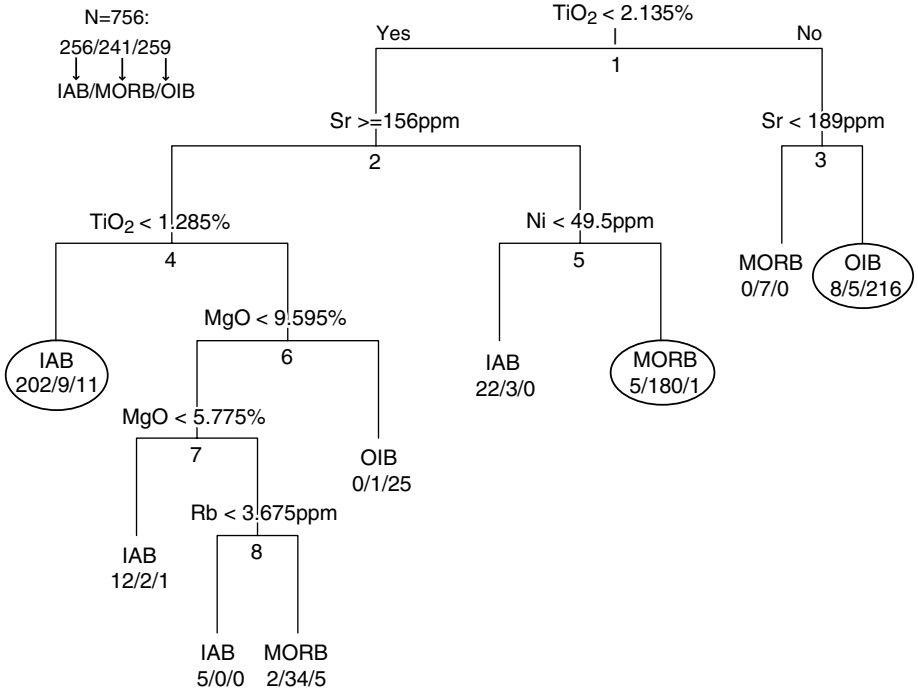


Fig. 4. The optimal classification tree, based on a training set of 756 geochemical analyses of at most 51 elements and isotopic ratios. The "heaviest" terminal nodes are encircled.

Table 1
Primary and surrogate splits for the nodes of Fig. 4

| Split number | IAB/MORB/OIB[a] | Primary split | Surrogate 1 | Surrogate 2 |
|---|---|---|---|---|
| 1 | 256/241/259 | $TiO_2 < 2.135\%$ | $P_2O_5 < 0.269\%$ | Zr < 169.5 ppm |
| 2 | 248/229/43 | Sr ⩾ 156 ppm | $K_2O \geqslant 0.275\%$ | Rb ⩾ 3.965 ppm |
| 3 | 8/12/216 | Sr < 189 ppm | — | — |
| 4 | 221/46/42 | $TiO_2 < 1.285\%$ | $Al_2O_3 \geqslant 15.035\%$ | $SiO_2 \geqslant 46.335\%$ |
| 5 | 27/183/1 | Ni < 49.5 ppm | Cr < 82 ppm | $TiO_2 < 0.71\%$ |
| 6 | 19/37/31 | $MgO < 9.595\%$ | $SiO_2 \geqslant 46.605\%$ | $Al_2O_3 \geqslant 13.945\%$ |
| 7 | 19/36/6 | $MgO < 5.775\%$ | $Al_2O_3 \geqslant 17.03\%$ | $CaO < 10.02\%$ |
| 8 | 7/34/5 | Rb < 3.675 ppm | $Na_2O \geqslant 4\%$ | — |

[a] The number of IABs, MORBs, and OIBs from the training data. The nodes without surrogates do not have any alternative variables that do better than a "go with the majority" decision. Split number 8 has only one worthwhile surrogate.

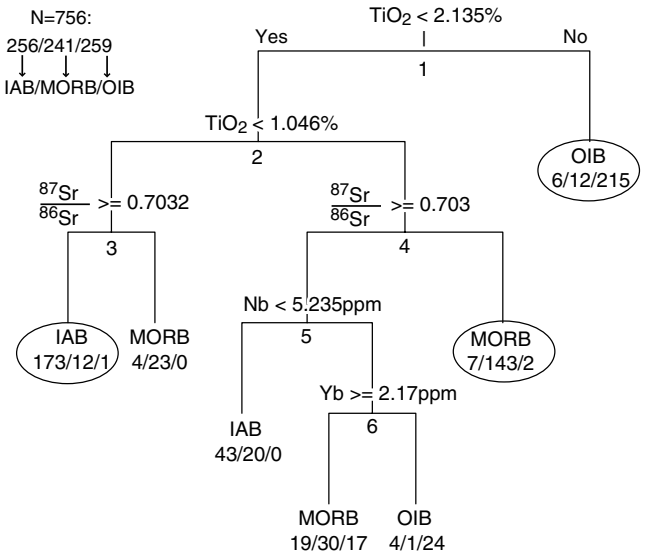


Fig. 5. The optimal classification tree using only HFS cations and isotopic ratios. The encircled terminal nodes contain the bulk of the training data.

Th, U, $^{143}Nd/^{144}Nd$, $^{87}Sr/^{86}Sr$, $^{206}Pb/^{204}Pb$, $^{207}Pb/^{204}Pb$, and $^{208}Pb/^{204}Pb$. The resulting pruned tree is shown in Fig. 5, while its surrogate splits are given in Table 2. The first two splits, both on $TiO_2$, contribute 85% of the discriminative power of the tree. The cross-validated misclassification error of the optimal tree is 16%, i.e., there is 84% chance of a correct classification.

## 4. Testing the trees

First, we will use the trees given by Figs. 4 and 5 to classify some fresh, modern basalts of known tectonic affinity, but from field areas that were not included in the training dataset. Then, the classification tree method will be applied to a published dataset of unknown tectonic affinity.

### 4.1. Classifying rocks of known tectonic affinity

Both trees were tested on three suites of samples that had not been used in the tree construction. The test data (electronic annex EA-1) include:

- 67 IABs from the Aleutian Arc.
- 55 MORBs from the Galapagos Ridge.
- 60 OIBs from the Pitcairn Islands.

First, these geochemical analyses were classified using the classic Ti–Zr–Y diagram of Pearce and Cann (1973). The results are shown on Fig. 6 and Table 3. A large subset of the data could be not classified, because (1) either Ti, Zr, or Y had not been analysed, or (2) because the data do not plot inside any of the labeled areas of the ternary diagram.

A substantial portion of the remaining data plots in field "B" of the discrimination diagram, which is of mixed tectonic affinity, although further classification can be done using the Ti–Zr diagram (Pearce and Cann, 1973). For the samples that plot in fields "A," "C," and "D," the classification seems to be quite successful, although it is hard to assess the misclassification risk because the number of "classifiable" points is so small.

It might not seem fair to the discrimination diagram method to only compare our classification trees with the method of Pearce and Cann (1973). Although this diagram has great historical significance and is still used a lot, it suffers from many of the wrong statistical assumptions that have plagued the analysis of compositional data, and have

Table 2
Surrogate splits for the HFS tree shown in Fig. 5

| Split number | IAB/MORB/OIB | Primary split | Surrogate 1 | Surrogate 2 |
|---|---|---|---|---|
| 1 | 256/241/259 | $TiO_2 < 2.135\%$ | Zr < 169.5 ppm | — |
| 2 | 250/229/44 | $TiO_2 < 1.0455\%$ | Zr < 75.5 ppm | Y < 22.9 ppm |
| 3 | 177/35/1 | $^{87}Sr/^{86}Sr \geqslant 0.703175$ | — | — |
| 4 | 73/194/43 | $^{87}Sr/^{86}Sr \geqslant 0.703003$ | $^{143}Nd/^{144}Nd < 0.5130585$ | Nd ⩾ 12.785 ppm |
| 5 | 66/51/41 | Nb < 5.235 ppm | $TiO_2 < 1.565\%$ | Zr < 100.5 ppm |
| 6 | 23/31/41 | Yb ⩾ 2.17 ppm | Lu ⩾ 0.325 ppm | Sm ⩾ 3.705 ppm |

Table 3
Test of the Pearce and Cann diagram (Fig. 6)

| True affinity | Missing data | Out of bounds | Predicted tectonic affinity | | | |
|---|---|---|---|---|---|---|
| | | | MORB or IAB[a] | IAB | MORB | OIB |
| IAB | 22 | 6 | 31 | 5 | — | 3 |
| MORB | 44 | 3 | 8 | 0 | — | 0 |
| OIB | 32 | 2 | 2 | 1 | — | 23 |

[a] The Ti–Zr–Y plot does not discriminate MORBs from IABs (both plot in field "B").

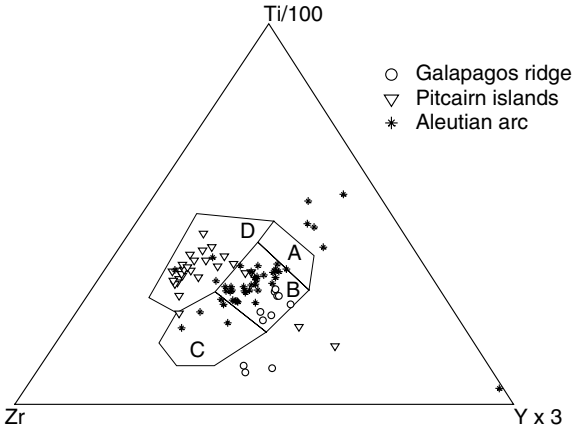


Fig. 6. The test data plotted on the Ti–Zr–Y discrimination diagram of Pearce and Cann (1973). More than half of the test data could not be plotted on this diagram because at least one of the three elements was missing. A—island arc tholeiites, C—calc-alkali basalts, D—within plate basalts, B—MORB, island-arc tholeiites, and calc-alkali basalts. A and C are the IAB fields, D the OIB field, and B a mixed field of MORBs and IABs.

been discussed elsewhere (e.g., Aitchison, 1986). The Ti–V diagram of Shervais (1982) largely avoids these problems, because it only uses two variables, and does not rescale them to a constant sum, as is the case for the ternary Ti–Zr–Y diagram of Pearce and Cann (1973). Furthermore, the training data of Shervais (1982) do not consist of averages of multiple samples, but of individual geochemical analyses. The Ti–V diagram can distinguish between all three tectonic affinities, so there is no field of "mixed affinity" like field "B" of Fig. 6. Fig. 7 shows the test data plotted on the Ti–V diagram. Table 4 summarizes the performance of this classification. The decision boundaries of all the tectonic discrimination diagrams discussed so far were drawn by eye. Vermeesch (2006) revisited these and other diagrams and recalculated the decision boundaries using the statistically more rigorous technique of discriminant analysis. Besides revisiting the diagrams of Pearce and Cann (1973), Shervais (1982), and others, Vermeesch (2006) also performed an exhaustive exploration of all possible binary and ternary combinations of 45 elements, based on the same training data used in the present paper. Here, only two of these diagrams will be discussed. The best overall linear discrimination diagram uses the combination of Si, Ti, and Sr (Fig. 8 and Table 5). The best quadratic discrimination diagram of only relatively immobile

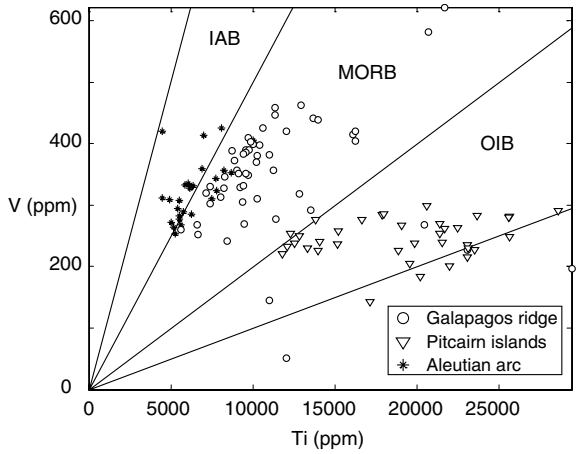


Fig. 7. The test data plotted on the Ti–V discrimination diagram of Shervais (1982).

Table 4
Test of the Ti–V discrimination diagram (Fig. 7)

| True affinity | Missing data | Out of bounds | Predicted tectonic affinity | | |
|---|---|---|---|---|---|
| | | | IAB | MORB | OIB |
| IAB | 40 | 0 | 17 | 10 | 0 |
| MORB | 2 | 3 | 0 | 48 | 2 |
| OIB | 24 | 7 | 0 | 1 | 28 |

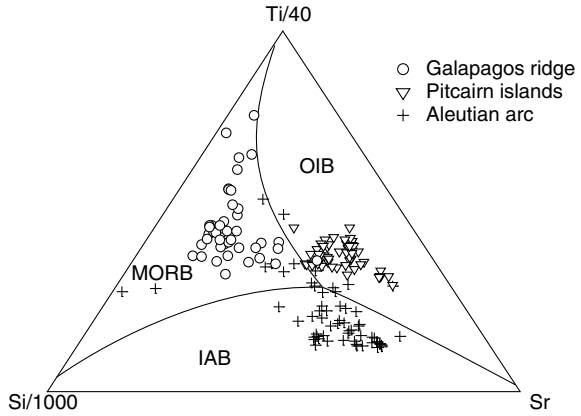


Fig. 8. The test data (164/182 used) plotted on the Si–Ti–Sr linear discrimination diagram (redrawn from Vermeesch, 2006).

Table 5
Test of the best linear discriminant analysis of Vermeesch (2006), using Si, Ti, and Sr (Fig. 8)

| True affinity | Missing data | Predicted tectonic affinity | | |
|---|---|---|---|---|
| | | IAB | MORB | OIB |
| IAB | 6 | 45 | 9 | 7 |
| MORB | 9 | 0 | 45 | 1 |
| OIB | 3 | 0 | 0 | 57 |

elements uses Ti, *V*, and Sm (Fig. 9 and Table 6). The same data were also classified with the trees of Figs. 4 and 5. The results of this experiment are shown in Tables 7 and 8. Contrary to the discrimination diagrams, the classification trees managed to assign a tectonic affinity to all 182 test samples. The HFS tree misclassifies quite a few more IABs and MORBs than the full tree. For the MORBs, it is probably not surprising that 14 out of 54 Galapagos ridge samples were misclassified as OIBs, considering the possible presence of plume-ridge interactions near the Galapagos hot spot. The higher misclassification risk of the HFS tree reminds us of the fact that unless rocks are obviously altered, it is better to use the full tree, which includes Sr. Performance of the Ti–V diagram is remarkably good and comparable to that to the full tree with a misclassification rate of 13/93 for the former and 23/182 for the latter (Tables 4 and 7). The Si–Ti–Sr (17/164 misclassified, Table 5) and Ti–V–Sm (8/85 misclassified, Table 6) diagrams even seem to perform better than the classification trees. However, this is likely to change for new trees created from larger sets of training data. Discriminant analysis does not gain much from excessively large databases, whereas classification trees keep improving. And again, neither the Si–Ti–Sr nor the Ti–V–Sm diagram succeeded in classifying all the test data, in contrast with the classification trees.

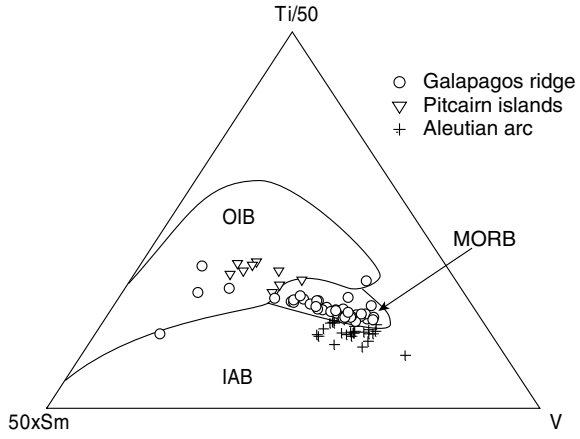


Fig. 9. The test data (85/182 used) plotted on the Ti–V–Sm quadratic discrimination diagram (redrawn from Vermeesch, 2006).

Table 6
Test of the best quadratic discriminant analysis of Vermeesch (2006), using only the relatively immobile elements Ti, V, and Sm (Fig. 9)

| True affinity | Missing data | Predicted tectonic affinity | | |
|---|---|---|---|---|
| | | IAB | MORB | OIB |
| IAB | 41 | 24 | 2 | 0 |
| MORB | 5 | 1 | 44 | 5 |
| OIB | 51 | 0 | 0 | 9 |

Table 7
Test of the full tree (Fig. 4)

| True affinity | Predicted tectonic affinity | | |
|---|---|---|---|
| | IAB | MORB | OIB |
| IAB | 54 | 8 | 5 |
| MORB | 4 | 49 | 2 |
| OIB | 2 | 2 | 56 |

Table 8
Test of the tree using only HFS elements (Fig. 5)

| True affinity | Predicted tectonic affinity | | |
|---|---|---|---|
| | IAB | MORB | OIB |
| IAB | 44 | 17 | 6 |
| MORB | 3 | 38 | 14 |
| OIB | 0 | 1 | 59 |

### 4.2. Predictions for rocks of unknown tectonic affinity

We will now apply the classification tree method to a published dataset of 20 basalts from the Pindos Basin in Greece (Saccani et al., 2003). This exercise will serve as an illustration of the way the tree method works in practice, and of its limitations. The features relevant to the classification tree analysis are shown in Table 9. As a first example, consider sample GR 47b. It travels through the full classification tree of Fig. 4 as follows: $TiO_2 < 2.135\% \rightarrow Sr \geqslant 156$ ppm $\rightarrow TiO_2 \geqslant 1.285\% \rightarrow MgO < 9.595\% \rightarrow MgO \geqslant 5.775\% \rightarrow Rb \geqslant 3.675$ ppm. The same sample travels through the HFS tree of Fig. 5 along the following path: $TiO_2 < 2.135\% \rightarrow TiO_2 \geqslant 1.046\% \rightarrow Nd < 12.785$ ppm. Note that for the last step, the second surrogate variable was used (Table 1), because no isotopic ratios were measured for these samples. No rare earth elements were measured for sample GR 56c. Therefore, its path through the HFS tree ends at node 4, where a ''follow the majority'' decision must be made. Since the distribution of training data in this node is IAB/MORB/OIB = 73/194/43 (Table 2), GR 56c is classified as a MORB, albeit not with the greatest confidence.

There is agreement between the full tree and the HFS tree for only half of the samples. Samples GR 50d, 51a, and 51b were classified as IAB by the full tree, and as MORB by the HFS tree. The distinction between IAB and MORB is the hardest one to make. IABs have a much greater compositional diversity than both MORBs and OIBs. This is also reflected in most discrimination diagrams (see for example, Fig. 6). Furthermore, it might be possible that Mg was lost during the greenschist metamorphism that affected the Pindos ophiolites (Saccani et al., 2003). This would have caused sample GR 50d to be sent left, rather than right at node 7 of Fig. 4. Likewise, it is possible that Sr-loss caused samples GR 71a–e and 195a to be sent left, rather than right at node 3 of the full tree. Finally, sample GR 181b was classified as an IAB by the HFS tree (Fig. 5). However, its terminal node (left branch of node 5) is not very pure: IAB/MORB/OIB = 43/20/0, once again illustrating the difficulty of distinction between MORB

Table 9
Summary of geochemical analyses of Triassic basalts of the Pindos Basin, from Saccani et al. (2003)

| Sample name | $TiO_2$ (%) | MgO (%) | Ni (ppm) | Rb (ppm) | Sr (ppm) | Nb (ppm) | Nd (ppm) | Yb (ppm) | Tectonic affinity[a] | | Comment[b] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | Full tree | HFS only | |
| GR47b | 1.41 | 6.6 | 81 | 17 | 163 | 10.6 | 12.3 | 2.87 | MORB | MORB | |
| GR50c | 1.6 | 8.36 | 61 | 7 | 225 | 9.69 | 13.1 | 3.16 | MORB | MORB | |
| GR50d | 1.52 | 5.69 | 56 | 18 | 176 | — | — | — | IAB | MORB | Mg loss? |
| GR50e | 1.56 | 5.6 | 66 | 20 | 138 | 9.38 | 13.1 | 3.31 | MORB | MORB | |
| GR51a | 1.18 | 7.22 | 104 | 4 | 242 | 6.87 | 9.5 | 2.17 | IAB | MORB | |
| GR51b | 1.17 | 7.05 | 97 | 5 | 180 | — | — | — | IAB | MORB | |
| GR56a | 1.68 | 6.57 | 85 | 4 | 98 | 5.87 | 14 | 3.15 | MORB | MORB | |
| GR56b | 1.61 | 6.77 | 72 | 5 | 200 | 4.23 | 11.8 | 3.4 | MORB | MORB | |
| GR56c | 1.68 | 6.9 | 65 | 5 | 81 | — | — | — | MORB | MORB | |
| GR56d | 1.52 | 6.33 | 72 | 7 | 247 | 3.94 | 11.8 | 3.37 | MORB | MORB | |
| GR71a | 2.82 | 6.87 | 76 | 4 | 114 | 3.74 | 19.8 | 5.96 | MORB | OIB | Sr loss? |
| GR71b | 2.79 | 6.82 | 73 | 3 | 115 | — | — | — | MORB | OIB | Sr loss? |
| GR71c | 2.36 | 7.24 | 80 | 3 | 166 | 3.27 | 17.7 | 5 | MORB | OIB | Sr loss? |
| GR71d | 2.8 | 7.12 | 75 | 5 | 125 | 3.66 | 19.7 | 5.92 | MORB | OIB | Sr loss? |
| GR71e | 2.89 | 6.66 | 71 | 3 | 124 | — | — | — | MORB | OIB | Sr loss? |
| GR181a | 1.38 | 6.61 | 83 | 8 | 176 | — | — | — | MORB | MORB | |
| GR181b | 1.5 | 7.32 | 88 | 4 | 152 | 4.18 | 21.8 | 6.55 | MORB | IAB | 43/20/0 |
| GR181c | 1.35 | 6.64 | 90 | 8 | 366 | 1.65 | 10.1 | 2.91 | MORB | MORB | |
| GR181d | 1.3 | 6.21 | 85 | 4 | 76 | — | — | — | MORB | MORB | |
| GR195a | 2.43 | 6.55 | 63 | 4 | 102 | 4.35 | 17.6 | 5.37 | MORB | OIB | Sr loss? |

[a] Most of the analyses point towards a MORB affinity for these rocks. There is agreement between the full classification tree and the one only using HFS elements for half of the samples (all agreeing on a MORB affinity).
[b] The cases where there is disagreement between the two classifications might be caused by the selective loss of Mg and Sr. Sample GR 181b was classified as an IAB with a probability of only ~2/3.

and IAB affinities, which is caused by the complicated petrogenesis of the latter.

## 5. Conclusions

It was not the purpose of this paper to claim that discriminant analysis or discrimination diagrams are either bad or obsolete. It merely suggests a completely different statistical approach to tectonic classification by rock geochemistry. Classification trees are presented as a simple yet powerful way to classify basaltic rocks of unknown tectonic affinity. Some of the strengths of the method are:

- Classification trees approximate the feature space by a piecewise constant function. This non-parametric approach avoids the statistical quagmire of discriminant analysis on compositional data (e.g., Aitchison, 1986; Woronow and Love, 1990).
- Classification trees are insensitive to outliers. In other words, if some of the training data were misidentified or accidentally very inaccurate (e.g., misplaced decimal point), trees remain almost unaffected.
- They can be used for the classification of highly multivariate data, while preserving the possibility of a simple, two-dimensional visualization. Therefore, trees are extremely easy to use. The trees presented in this paper were based on a database of moderate size (756 analyses). If a much larger database were compiled, the trees would grow and their discriminative power increase, but they would still be easy to interpret. It should also be easy to extend the trees given in this paper to more tectonic affinities, such as active continental margins, continental within-plate basalts, or different lithologies, simply by adding data to the training-set. In principle, there is no upper limit to the number of "class labels" that the method can discriminate, provided enough training data are available.
- Trees do not discriminate according to some complicated decision boundary (e.g., multivariate discriminant analysis) or a black box process (e.g., neural networks), but split the data space up one variable at the time, in decreasing order of significance. Therefore, the split variables have geochemical significance. For example, if $TiO_2$ and Sr contribute 87% of the discriminative power, there likely is a real geochemical mechanism that causes this to be so.
- Although the trees presented in this paper were built from as many as 51 different variables, we can still use them if some of these variables were not measured for the unknown sample that we want to classify. This can be done with the surrogate split variables of Tables 1 and 2.
- A rough idea can be had of the statistical uncertainty of the classification by looking at the purity of the terminal nodes. For examples, see samples GR 181b and 56c of Section 4.2. This is not possible for discrimination diagrams because the decision boundaries of the latter are drawn as hard lines, and not as the "fuzzy" zones which they really are.

On the other hand, trees are not perfect, and also have a few problems:

- Classification trees are biased because the piecewise constant approximation which they implement is an oversimplification of the feature space. Increasing the size of the training dataset will alleviate this problem. Alternatively, we could also allow linear combination splits instead of only splits on a single variable, but this would hurt interpretability.
- Trees also suffer from large variance. In other words, they are unstable. A different set of training data could result in very different looking trees. This somewhat limits the interpretability of the split variables, which was hailed before. More importantly, small errors in one of the split variables of the unknown dataset are propagated down to all of the splits below it. ''Bagging'' is a way to solve these problems by collecting a large number of ''bootstrap samples'' of the training data, and building a large number of trees from them (Hastie et al., 2001). The unknown data are then sent through all these trees and the results averaged. This provides a more robust classification algorithm, but again at the expense of interpretability, because bagged trees can no longer be easily plotted as a simple two-dimensional graph.
- One of the properties of many data mining algorithms, including classification trees, is the ''garbage in, garbage out'' principle. There is no field of ''ambiguous'' tectonic affinity, which is effectively the output of geochemical compositions that plot ''out of bounds'' on traditional discrimination diagrams (e.g., Fig. 6 and Table 3). Any rock will be classified as either IAB, MORB or OIB, also when it really is a continental basalt, granite or even sandstone! Therefore, one might treat compositions that plot far outside the decision boundaries of the traditional discrimination diagrams with extra caution.

Most importantly, as was illustrated by the examples of Section 4, no classification method based solely on geochemical data will ever be able to perfectly determine the tectonic affinity of basaltic rocks (or other rocks for that matter) simply because there is a lot of actual overlap between the geochemistry of the different tectonic settings. Notably IABs have a much wider range of compositions than either MORBs or OIBs. Therefore, geochemical classification should never be the only basis for determining tectonic affinity. This is especially the case for rocks that have undergone alteration. In such cases, mobile elements such as Sr, which have great discriminative power, cannot be used. If in addition to this, some other features have not been measured (such as isotope ratios and rare earths in some of the samples of Table 9), then one might not be able to put much faith in the classification.

## Acknowledgments

The author would like to thank Richard Arculus, John Rudge, and Steve Galer for extremely constructive and insightful reviews.

*Associate editor:* Stephen J.G. Galer

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gca.2005.12.016.

## References

Aitchison, J., 1986. *The Statistical Analysis of Compositional Data*. Chapman and Hall, London.

Breiman, L., Friedman, J.H., Olshen, R., Stone, C., 1984. *Classification and Regression Trees*. Wadsworth, Belmont, CA.

Carreras, X., Marquez, L., 2001. Boosting Trees for Anti-Spam Email Filtering. In: Proceedings of RANLP-01, 4th International Conference on Recent Advances in Natural Language Processing, Tzigov Chark, BG.

Chayes, F., 1971. *Ratio Correlation; a Manual for Students of Petrology and Geochemistry*. Chicago University Press, IL.

Chayes, F., Velde, D., 1965. On distinguishing basaltic lavas of circum-oceanic and oceanic-island type by means of discriminant functions. *Am. J. Sci.* **263**, 206–222.

Hastie, T., Tibshirani, R., Friedman, J.H., 2001. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, Berlin.

Kushiro, I., Kuno, H., 1963. Origin of primary basalt magmas and classification of basaltic rocks. *J. Petrol.* **4** (1), 75–89.

Pearce, J.A., Cann, J.R., 1971. Ophiolite origin investigated by discriminant analysis using Ti, Zr and Y. *Earth Planet. Sci. Lett.* **12** (3), 339–349.

Pearce, J.A., Cann, J.R., 1973. Tectonic setting of basic volcanic rocks determined using trace element analyses. *Earth Planet. Sci. Lett.* **19** (2), 290–300.

Pearce, J.A., 1976. Statistical analysis of major element patterns in basalts. *J. Petrol.* **17** (1), 15–43.

Ripley, B.D., 1996. *Pattern Recognition and Neural Networks*. Cambridge University Press, Cambridge, MA.

Rollinson, H.R., 1993. *Using Geochemical Data: Evaluation, Presentation, Interpretation*. Longman, London.

Saccani, E., Padoa, E., Photiades, A., 2003. Triassic mid-ocean ridge basalts from the Argolis Peninsula, Greece: new constraints for the early oceanization phases of the Neo-Thetyan Pindos basin. In: Dilek, Y., Robinson, P.T. (Eds.), *Ophiolites in Earth History*, Vol. 218. Geological Society, pp. 109–127.

Shervais, J.W., 1982. Ti–V plots and the petrogenesis of modern ophiolitic lavas. *Earth Planet. Sci. Lett.* **59**, 101–118.

Vermeesch, P., 2006. Tectonic discrimination diagrams revisited. *Geochem. Geophys. Geosyst.* (in press).

Wood, D.A., 1980. The application of a Th-Hf-Ta diagram to problems of tectonomagmatic classification and to establishing the nature of crustal contamination of basaltic lavas of the British Tertiary volcanic province. *Earth Planet. Sci. Lett.* **50** (1), 11–30.

Woronow, A., Love, K.M., 1990. Quantifying and testing differences among means of compositional data. *Math. Geol.* **22**, 837–852.