Journal of Geochemical Exploration 145 (2014) 12–24

Contents lists available at ScienceDirect

Journal of Geochemical Exploration

journal homepage: www.elsevier.com/locate/jgeoexp

Using geochemical proxies to model nuggety gold deposits: An example from Sunrise Dam, Western Australia

E. June Hill [a,*], Nicholas H.S. Oliver [b,c], Louise Fisher [a], James S. Cleverley [a], Michael J. Nugus [d]

[a] CSIRO, P.O. Box 1130, Bentley, WA 6102, Australia
[b] Holcombe Coughlin Oliver, P.O. Box 3533, Hermit Park, QLD 4812, Australia
[c] Economic Geology Research Unit, James Cook University, Australia
[d] AngloGold Ashanti, 44 St Georges Tce, Perth, WA 6000, Australia

ARTICLE INFO

Article history:
Received 12 March 2014
Accepted 7 May 2014
Available online 24 May 2014

Keywords:
Nugget effect
Gold
Geochemistry
Conditional probability

ABSTRACT

Gold distribution in vein-hosted hydrothermal ore deposits is commonly nuggety (i.e. occurs as very localised concentrations of gold). In these cases samples for gold assay from diamond drill core may be too small to model the underlying heterogeneity of gold distribution and result in poorly constrained ore body models and underestimated gold resources. Hence, it is common practice to use more spatially continuous proxies for mineralisation to help define the boundaries of mineralised regions. We present a method for automating the use of geochemical proxies for nuggety gold ore bodies.

Sunrise Dam Gold Mine, in Western Australia, is a world-class gold deposit with a very high nugget effect. Multi-element geochemical data has been collected at this site in order to improve prediction of mineralised regions. Suitable proxy elements have been selected from this data set, in particular, those that are spatially related to gold mineralisation but do not display nuggety distribution, such as Sb, Rb and Cr.

We applied a probabilistic approach to the problem of quantifying the relationship between gold assay values and geochemical elements. It is shown that a kernel density estimator and Bayes conditional probability can provide an effective method for calculating the probability of a sample having elevated gold content and that this measure will be more spatially continuous than gold assay values if the appropriate geochemical proxies are selected. Using conditional probability and suitable cut-off values, we reclassified approximately 27% of samples as mineralised which returned low Au assay results. When plotted on drill holes conditional probability values provided a much more spatially continuous guide to mineralised regions than Au assay values alone.

1. Introduction

Gold deposits associated with vein-hosted hydrothermal mineralisation are commonly nuggety, meaning that the gold distribution is highly spatially discontinuous ("geological nugget effect" of Dominy et al., 2003). This makes it very difficult for geologists to establish boundaries of continuous sections of mineralised rock using drill core assays and to manually connect these boundaries from drill hole to drill hole in order to create a 3D model.

Geologists attempt to overcome this problem of separating continuous sections of mineralised and unmineralised rocks by incorporating geological information (e.g. Dominy and Johansen, 2004). The information is generally incorporated manually, on an ad hoc basis, which is not repeatable; meaning that different geologists would produce different spatial models for the ore body. The types of geological information geologists incorporate may include (1) presence and intensity of alteration styles known to be associated with mineralisation; (2) presence of a lithological unit known to be favourable for mineralisation; and/or (3) presence of structural features which may indicate a zone favourable for the transport and deposition of minerals (e.g. presence of fault, shear zone or veins). These favourable geological features are called proxies for mineralisation. The ideal proxy is one which not only is coincidental with gold mineralisation but also is more spatially continuous in nature, i.e. less nuggety. This means that the mineralised zones defined by proxy can be interpolated between drill holes with much greater confidence than the raw gold assay values.

If a single geochemical element is used as a proxy it is straightforward to model the distribution of the element. However, if it is desirable to include more than one element in the proxy (i.e. multivariate data), then we need a method which is able to combine the effect of all these elements into a single informative numerical value. For example, it may be useful to combine an element that indicates alteration with an element that indicates primary lithological variation.

In order to make the incorporation of multivariate geological information a more rigorous process we attempt to automate the use of proxies using probabilistic methods. Automation also has the advantage

* Corresponding author at: CSIRO, 26 Dick Perry Ave, Kensington, Western Australia 6151, Australia. Tel.: +61 8 6436 8651.
E-mail address: June.Hill@csiro.au (E.J. Hill).

http://dx.doi.org/10.1016/j.gexplo.2014.05.008
0375-6742/

that large data sets can be processed rapidly. This method was attempted in Hill et al. (2013c) using logged geological information from diamond drill core. Whilst the method was successful in local regions, it was found that the logging of the core over a larger mineralised region was very inconsistent as it was logged by many different geologists (Hill et al., 2013a) and this was reflected as undesirable artefacts in parts of the resulting spatial models. In this paper we present a new method of automating proxies using geochemical data instead of logged data, because the geochemical data, if rigorously measured, should provide a more consistent and quantitative data set. The main difference between using geochemical data and logged data is that we are dealing with continuous numerical data rather than categorical data. The method given in Hill et al. (2013c) was specifically developed for categorical data; this paper presents alternative methods which are suitable for compositional data.

Sections 2 and 3 of this paper describe the geological setting and geochemical analysis of the samples used in this study. The test data set only sampled a small proportion of the ore body, so whilst we cannot use it to help construct a model of the whole ore body, it is of sufficient size to test that the method works. In Sections 4–6 we present, apply and discuss the method used to quantify the geochemical proxies as a probability value. In these sections we also compare some alternative popular classification methods for analysing the proxies and we demonstrate their inadequacy in this case.

2. Geological setting and mineralisation at the Sunrise Dam Gold Mine

The Sunrise Dam Gold Mine (SDGM) in Western Australia (Fig. 1) is a useful case study for proxies for Au mineralisation as its exceptionally high-nugget Au has made it very difficult to define the boundaries of the mineralised zones and estimate the resources available for mining (Nugus et al., 2009). At SDGM Au assay values are collected from samples of NQ diamond drill core (47.6 mm diameter, typically 0.3–1.0 m in length). The unreliability of Au assays as a representation of the distribution of mineralisation was demonstrated in a trial on repeatability conducted by Carswell and Clark (2013). In that study, intervals of core were cut in half lengthways and the two samples were analysed separately; the pairs of half-core samples showed low correlation. Bulk mining and recent experiments with large-diameter RC drilling of several lodes have uncovered a higher calculated grade than that estimated from the distribution of Au from diamond core assays (Carswell and Clark, 2013), supporting the hypothesis that the nugget effect typically results in underestimates of the mined grade. This is anecdotally considered to reflect that lode continuity is greater than that suggested by the diamond core assays.

Due to the nuggety nature of the gold, it has been widespread practice at SDGM to use proxies to help define ore body boundaries. Documented uses of proxies for gold mineralisation at SDGM include those from Haren and Williams (2000), who used the presence of shear zones and sedimentary banded iron formation, and Nugus et al. (2009) who used structural features and alteration as proxies to map zones of high grade mineralisation. Mine geologists estimating grade underground at SDGM have been able to use visual proxies to estimate gold using the intensity of alteration, intensity of deformation and the abundance of veins (Hill et al., 2013c). These manually applied proxies have provided smoother and less nuggety ore domains than the assays themselves for volumes of rock greater than $5 \times 5 \times 5$ m. However, they are prone to error due to inconsistencies between different geologist observations and personal biases.

SDGM ore bodies are hosted in a sequence of Archaean metasedimentary and metavolcanic rocks, intruded by dolerite and porphyritic microgranite and granite. The region has undergone lower- to mid-greenschist metamorphism, and multiple deformation events. An early stage of extensional deformation was followed by several phases of compressional deformation. Gold mineralisation appears to be largely associated with shear zones and breccia zones associated with the later stages of compressional deformation (Blenkinsop et al., 2007; Nugus et al., 2009). Although the relative timing of the two events is unclear, gold mineralisation is also closely spatially associated with porphyritic microgranite intrusion. Most gold is vein-hosted and there is a complex relationship between alteration and shearing and between faulting and veining which has proven difficult to unravel in detail and which varies from ore body to ore body (Hantler, 2009; Hill et al., 2013c).

Fig. 1. Location of the Sunrise Dam Gold Mine in Western Australia.

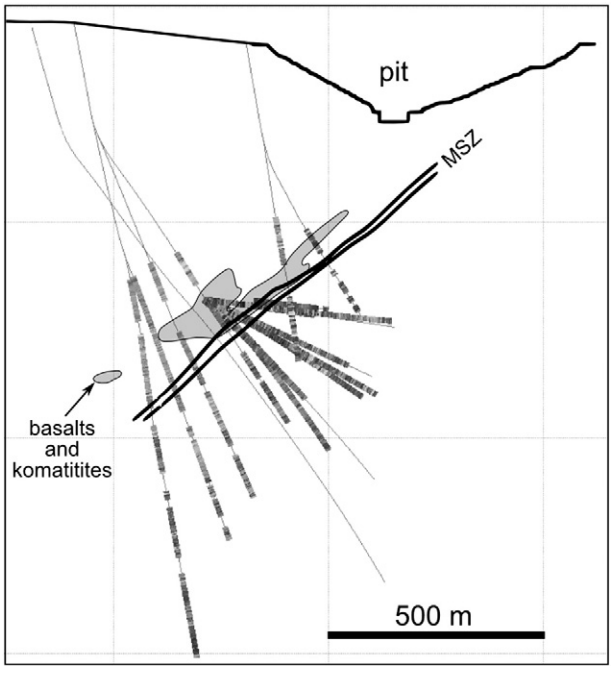

Fig. 2. W–E cross-section showing the extent of geochemical sampling for the Vogue ore body. Diamond drill holes that were sampled (thin grey lines) are projected onto the section. Location of geochemistry samples on drill holes is shown as thicker lines with grey-scale colour according to Rb composition, a proxy for sericitic alteration; darker colour = higher Rb. MSZ = midway shear zone.

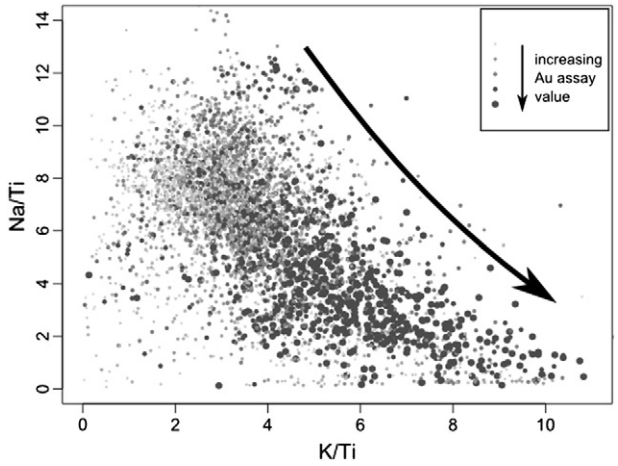

Fig. 3. Degree of sericitic alteration in intermediate composition samples can be represented by decreasing Na/Ti ratio and increasing K/Ti ratio. The plot illustrates that, although the general trend is for higher Au content in more altered rocks (large arrow); even strongly altered samples may have low Au assay values and some high Au samples are apparently not very altered.

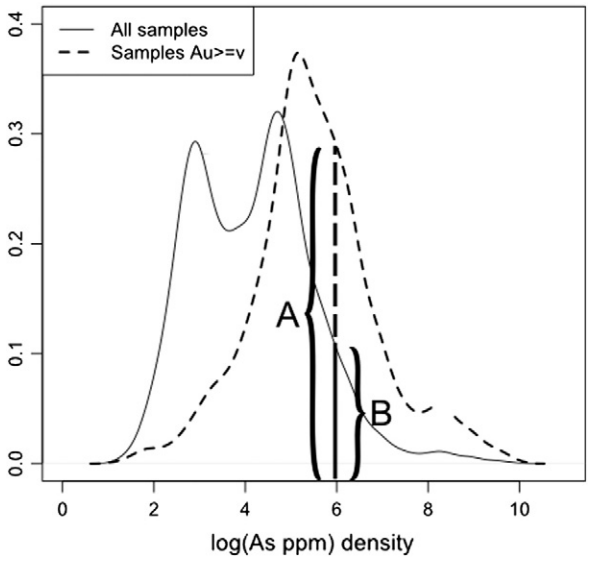

Fig. 5. Plot shows the density plots of log (As) for all samples (solid line) and for samples with elevated Au values (dashed line). The density ratio is A/B. The larger this value, the higher the probability.

The Vogue ore body, which is the subject of this study, is located several hundred metres below the SDGM open pit (Fig. 2) and shows some differences in mineralisation style to the higher level ore bodies. The Vogue ore body lies largely within intermediate composition volcanic and volcaniclastic rocks, with minor metasedimentary rocks. The ore body is overlain by high-Mg basalts and komatiites which host comparatively small amounts of gold mineralisation. Porphyritic microgranite intrusions are narrow and rare in this region. Deformation within the Vogue ore body region is generally weaker and more localised than in the higher levels of SDGM. A west-dipping foliated zone, the Midway shear zone, bounds the top of the ore body. Weaker fabrics parallel to this shear zone occur throughout the Vogue area with heterogeneous intensity. In addition, steep zones of breccia veins, with or without an enclosing strongly foliated zone, define high grade but narrow and discontinuous (cm- to m-scale) shoots. Gold mineralisation appears to be spatially related to zones of stronger foliation development, especially where steep zones and west-dipping foliation zones intersect. Gold mineralisation is also known to be spatially associated with sericitic alteration and with the presence of sulphides and carbonates in veins or wall rock haloes around veins. Alteration and high vein abundances are generally broadly distributed in lodes at higher levels in the SDGM system. However, in Vogue high gold grades occur in very narrow (<5 cm) veins with comparably narrow alteration haloes, resulting in individual enriched zones in drill core that are smaller than or similar in size to the minimum core sampling length (30 cm).

Geological features that are considered useful as proxies for mineralisation include: rock type (intermediate composition volcanogenic rocks are most favourable), alteration (sericitic alteration in particular), sulphide minerals, veins (quartz–carbonate), strong foliation and brecciation.

3. Geochemistry data

Geochemical data were collected from diamond drill cores from Vogue as an experiment to determine their usefulness as proxies for mineralised rocks (Fig. 2). Whole rock powders were prepared for assay from 18 diamond drill cores, comprising 0.3 to 1 m intervals of pulped half-core (NQ) from selected intervals that intersected the Vogue lode. Assay interval boundaries were chosen by geologists on the basis of comparison with the core logging information; these boundaries typically reflect the contact between different logged rock types, or the contact between packets of strongly altered and veined material and the less altered surrounds. Where boundaries could not readily be ascertained in drill core and the rock showed little visible evidence of alteration, samples of 1 m length were taken.

Multi-element analyses were collected by Intertek Genalysis in Western Australia. Major elements were determined by X-ray fluorescence (XRF) on lithium metaborate fusions of whole-rock powders. Trace elements were subsequently determined by inductively coupled

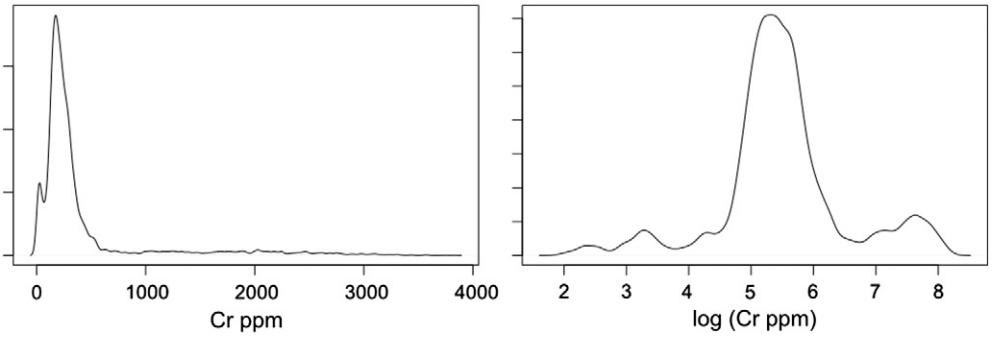

Fig. 4. Probability density plots for Cr and log (Cr) demonstrating that neither distribution is Gaussian. The main peak on the plots represents the andesitic rocks; the lower Cr peaks are tholeiitic basalts and felsic rocks and the higher Cr peaks are komatiites and high-Mg basalts.

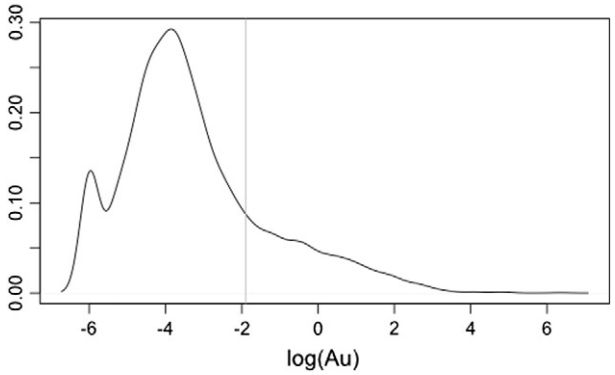

Fig. 6. Density plot of log (Au). The vertical line indicates 80 percentile for Au samples. The small peak at log(Au) ~ −6 is an artefact due to assignment of a fixed value (0.0025 ppm) to samples below detection limit (0.005 ppm).

plasma mass spectrometry (ICP-MS) or inductively coupled plasma optical emission spectroscopy (ICP-OES) following a four acid digest.

Our preliminary work prior to probability analysis involved the recognition of (a) hydrothermally immobile elements that are best used to define lithogeochemical rock types with Ti, Zr and Cr most effective for discriminating lithological boundaries, and (b) a range of elements found in veins and alteration haloes around those veins that correlate both with logged alteration types and appeared to show some sort of down-hole spatial association with Au. These include major elements such as K and Mg and elements such as As, S, Cu and Sb which are associated with sulphides and sulphosalts in the alteration assemblage.

Using geochemistry data alone is an indicator of primary rock type and subsequent alteration and cannot provide structural and textural information. Hence, our proposed method may not provide the best estimate of the broader Au distribution if structural and/or textural features are the dominant control on Au grades, without a corresponding geochemical signature.

4. Method of selecting and quantifying proxies

This study is based on the assumption that the Au grades measured from samples from diamond core are not representative of their host rocks because of the nugget effect. Furthermore, when using chemical compositions to indicate Au grades we must take into account that there is no one-to-one relationship between Au-rich samples and their chemical composition. For example, not all Au bearing veins

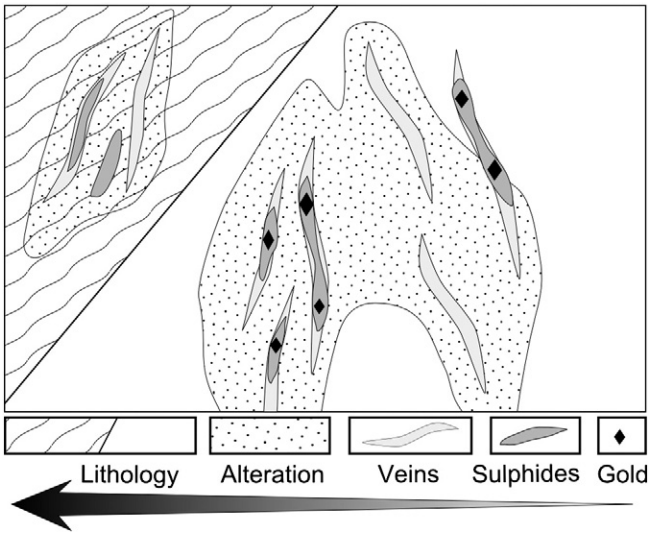

Fig. 7. A schematic diagram summarising the spatial relationships between various geological features and Au mineralisation at SDGM. This expert information helps us to select appropriate geochemical elements as proxies. "Veins" refers to quartz–carbonate veins. Legend indicates increasing spatial continuity of features from right to left.

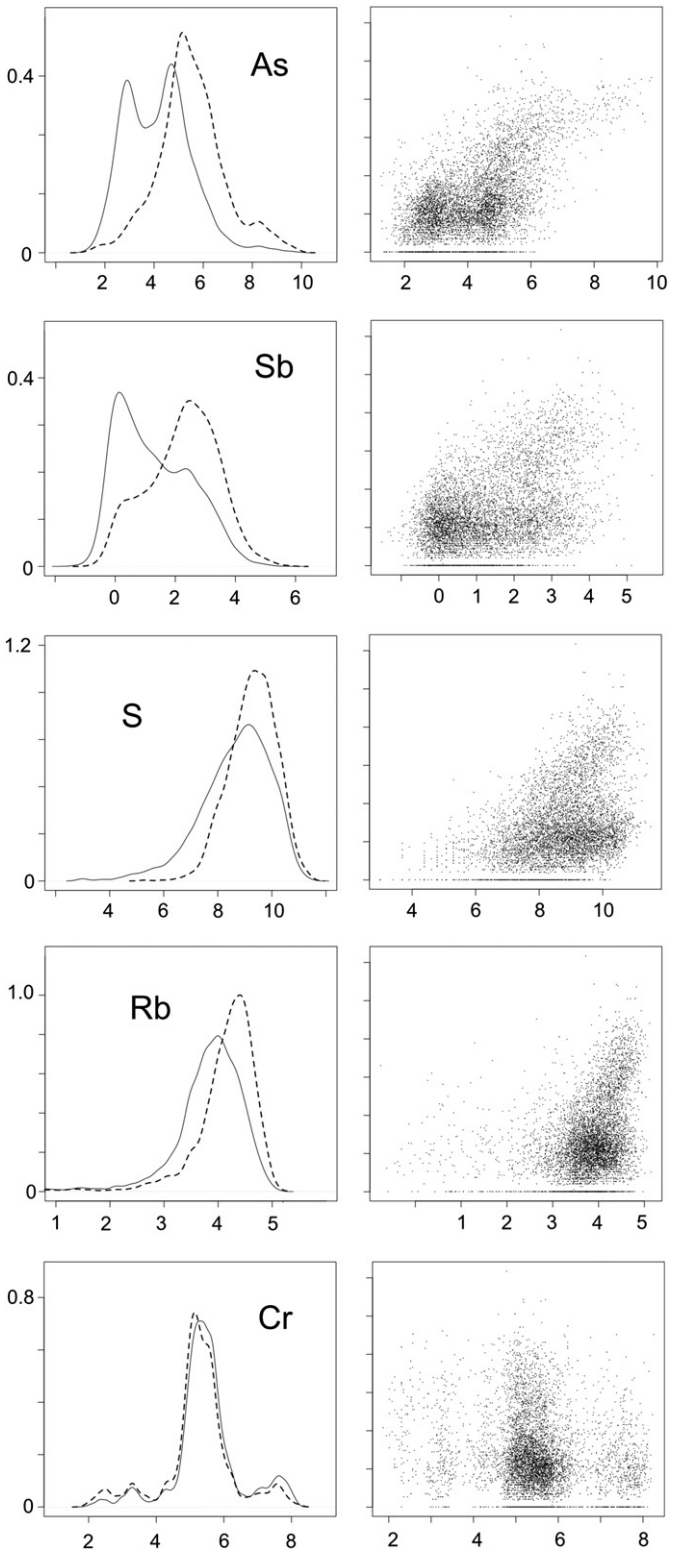

Fig. 8. Density plots (left) and scatter plots (right) are useful for determining if there is a relationship between Au and other geochemical elements. The density plots show a comparison between all samples (solid line) and elevated Au samples (dashed line). X-axes show the log of the value of each element in ppm. Y-axes of right-hand plots are log(Au).

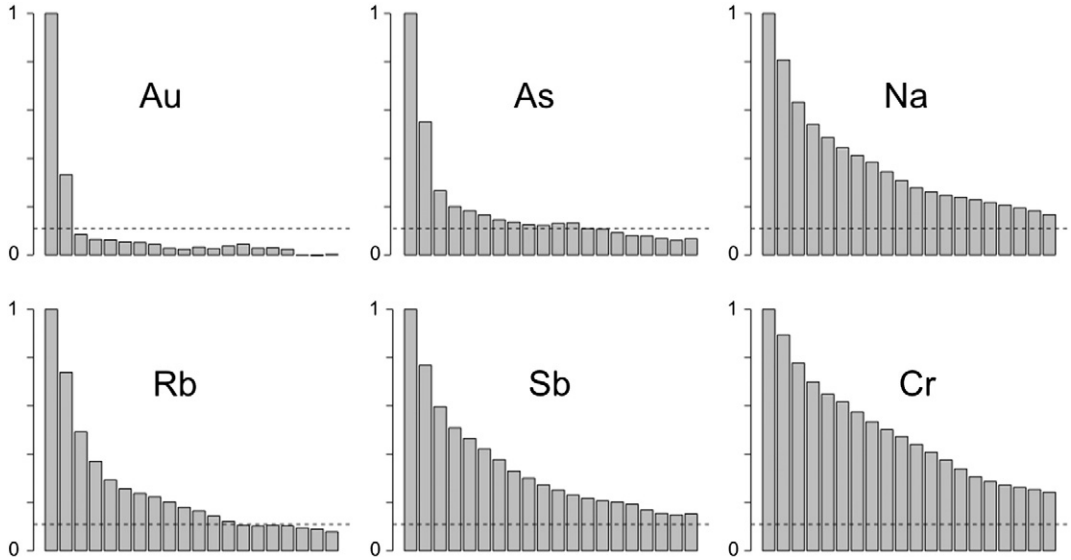

Fig. 9. The plots show the mean autocorrelation values for all the drill holes in the data set. Minimum lag distance is 1 m and maximum lag shown is 20 m. The dotted line is the mean 95% confidence interval.

have an observable sericitic alteration halo, although the majority of veins do have a halo (see Fig. 3); and not all sulphide rich veins contain Au, although this is a commonly observed relationship. Therefore it is useful to use a probabilistic approach which will give us a measure of the likelihood that a sample is mineralised given its chemical composition. From this information we can use geochemical data to determine the regions where mineralisation is most likely to occur, whether or not the samples contain elevated Au grades.

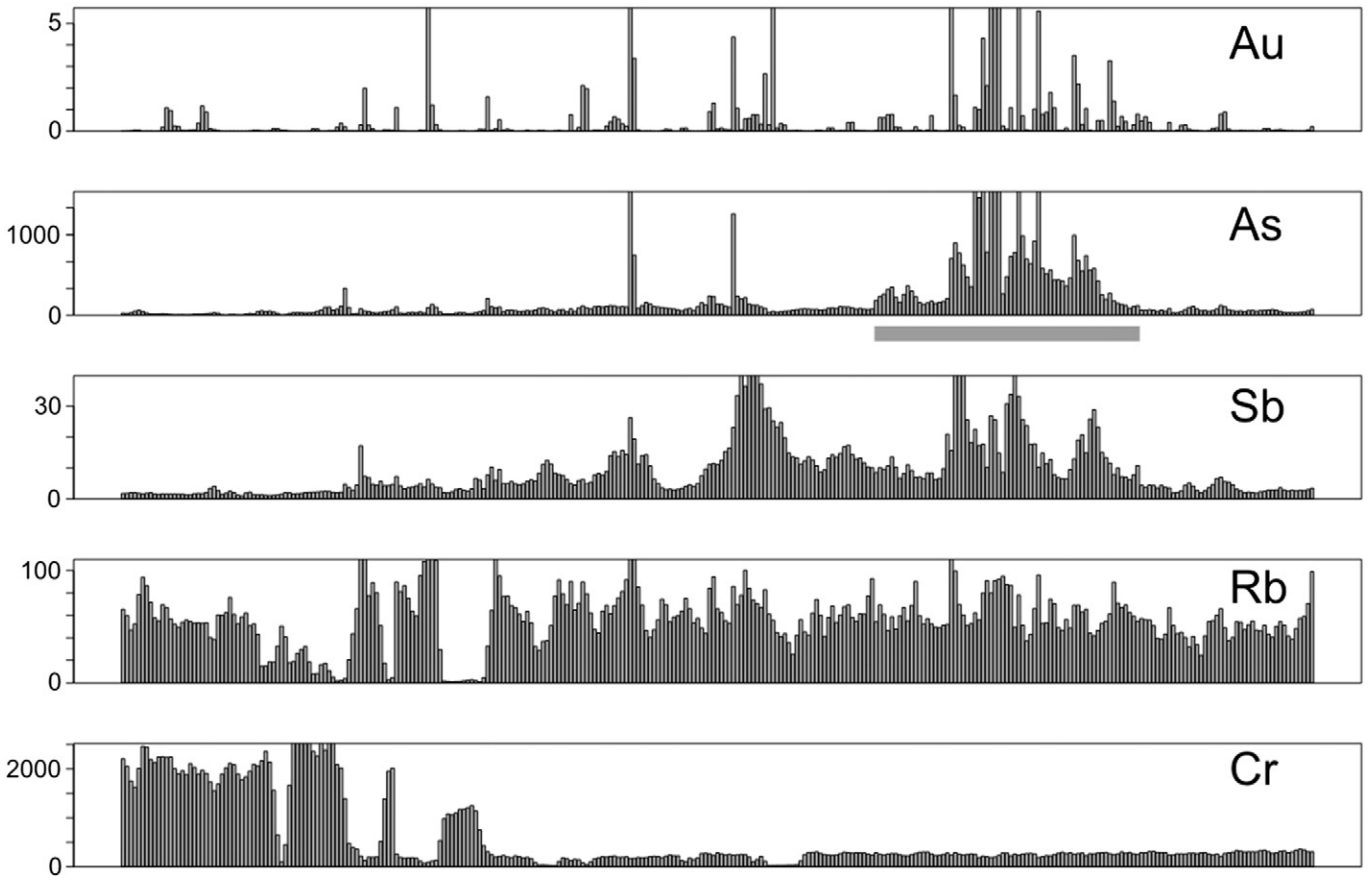

Fig. 10. Bar plots of Au, As, Sb, and Rb (ppm) against depth down drill hole for UGD1944, total drill hole depth is 307 m. Top plot shows the nuggety nature of Au assay values. The plots illustrate the spatial relationships between the elements and Au distribution. Note that As can also be quite nuggety in places. A zone of high As is highlighted by the grey bar beneath the plot (see text for discussion).

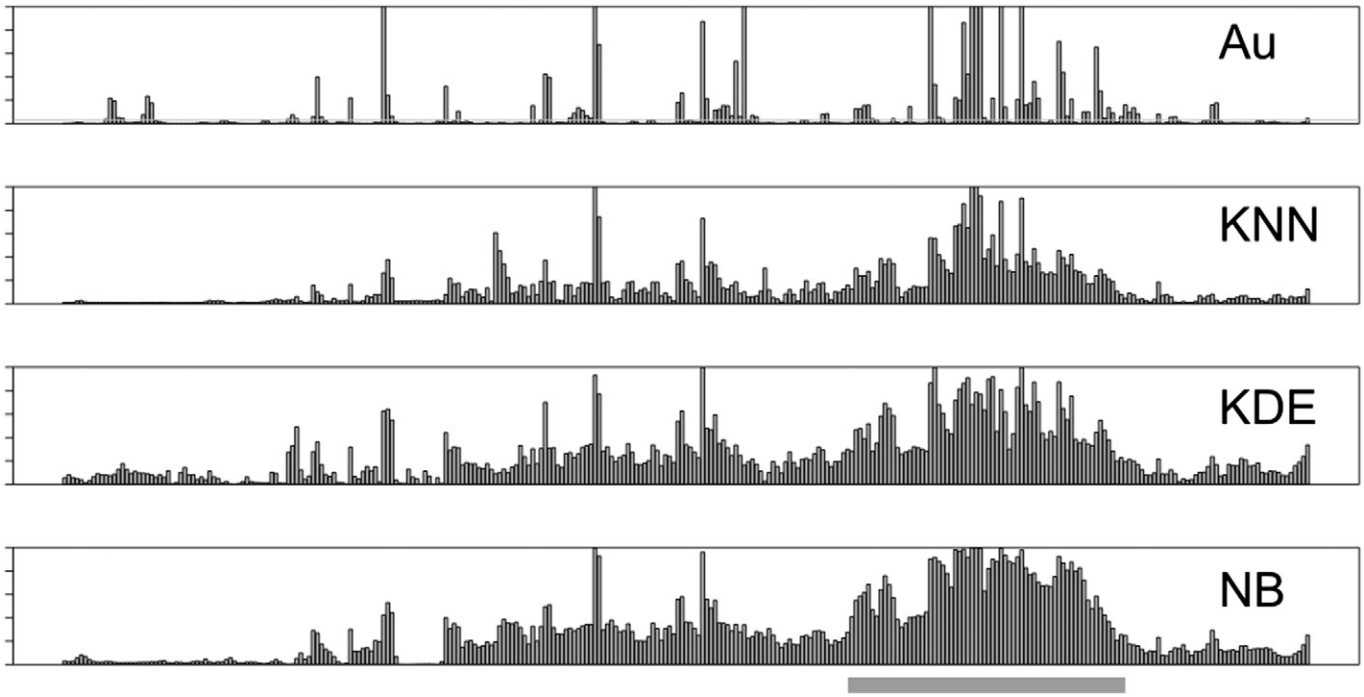

Fig. 11. Bar plots comparing (a) Au assay values for UGD1944 with results of the three algorithms using Cr–Rb–As: (b) KNN (k = 501), (c) KDE probability, (d) NB probability. A zone of high As is indicated by the grey bar beneath the last plot (see also Fig. 10); this zone has significant influence on the results.

The probabilistic method described here is also a good choice because it uses a non-parametric approach, i.e. the method assumes nothing about the shape of the underlying data distribution. For example, many parametric methods assume a Gaussian or log Gaussian distribution of data (all log plots shown in this article use natural log). This is often not true for geochemical elements; for example, Cr shown in Fig. 4. Furthermore, it is worth noting that all the methods described here are entirely data driven and do not take geological knowledge on ore genesis into account; however, this information is implicit in the selection of geochemical elements.

Categorical data is relatively easy to deal with as each sample belongs to a category which can be characterised by all the other samples which fall into the same category (e.g. logging data, Hill et al., 2013a). Geochemical data, however, must cover the full range of real numbers from zero to the total (e.g. 100% or 10^6 ppm). Therefore to estimate the probability for any sample we need to consider the samples which

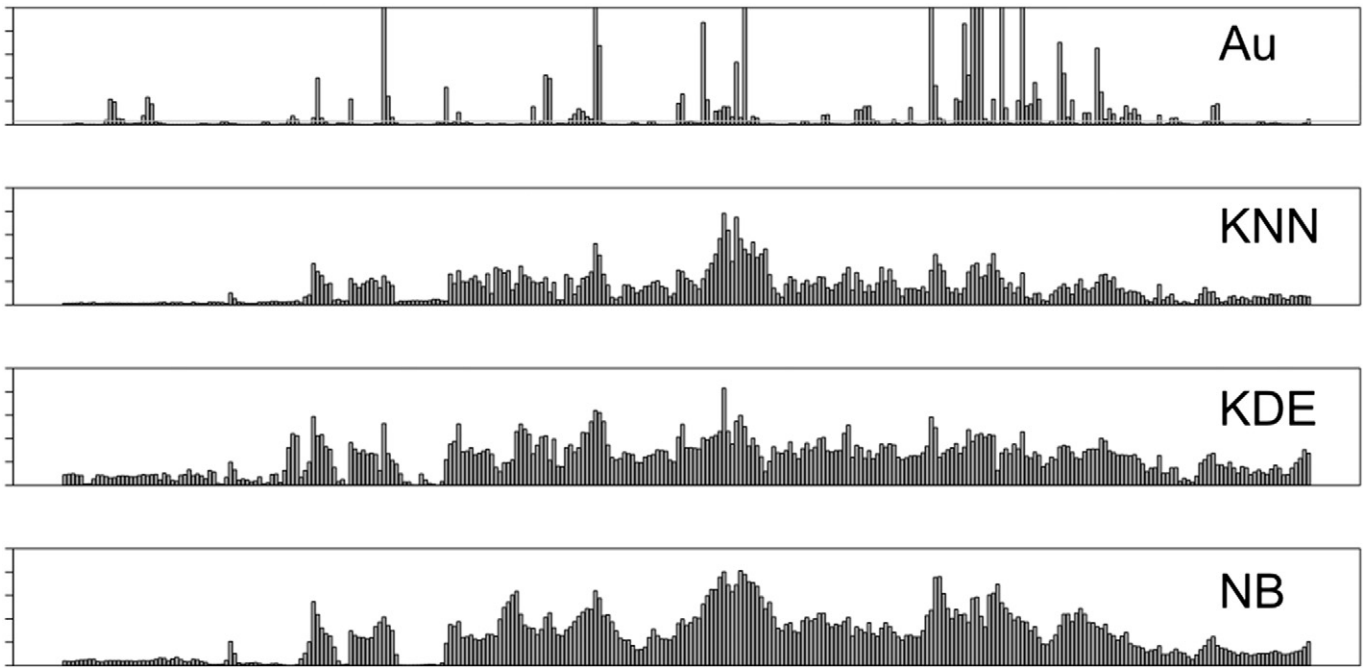

Fig. 12. Bar plots comparing (a) Au assay values for UGD1944 with results of the three algorithms using Cr–Rb–Sb: (b) KNN (k = 501), (c) KDE probability, (d) NB probability.

are similar, i.e. other samples which lie within the neighbourhood of the sample of interest in feature space. A neighbourhood can be defined by using a window of specified size and shape; this is called a kernel method. The probability value that is calculated depends on the shape and size of the neighbourhood we choose to use.

An alternative method for selecting a neighbourhood is to choose a fixed number of neighbouring samples. This method may overcome problems associated with variable data density discussed below. The best known and most popular is the k-nearest neighbour method (KNN). We compare results using KNN to those achieved using the probabilistic approach to determine if there are benefits to using this alternate approach.

Most of the data manipulation was performed using statistical software environment R, the details of which are summarised in Appendix A.

4.1. Kernel methods

The problem can be expressed as a conditional probability (Hill et al., 2013c) and solved using Bayes' theorem. The probability that Au is greater than or equal to a cut-off value given a specified geochemical composition is given by:

$$p(Au \geq v \,|G_1, G_2, G_3 \ldots) = \frac{p(G_1, G_2, G_3 \ldots | Au \geq v)\; p(Au \geq v)}{p(G_1, G_2, G_3 \ldots)}$$

where $p(Au \geq v)$ is the prior probability, i.e. it is the probability that the Au assay value of any sample exceeds a cut-off value v; and $p(G_1, G_2, G_3 \ldots)$ is the probability of a sample having the chemical composition G_1 and G_2 and G_3 etc., where G_1, G_2, G_3 … represent the chemical elements of interest.

For continuous numerical data the probabilities can be estimated from the density of data:

$$p(Au \geq v \,|G_1, G_2, G_3 \ldots) = \frac{density(G_1, G_2, G_3 \ldots | Au \geq v)\; p(Au \geq v)}{density(G_1, G_2, G_3 \ldots)}.$$

Fig. 5 illustrates how the density ratio for a single element can be visualised in a data density plot; when multiplied by the prior $p(Au \geq v)$ this gives the conditional probability:

$$p(Au \geq v |\ \log(As)) = \frac{A.p(Au \geq v)}{B}.$$

To estimate the density of data we use the kernel density estimation (KDE) method of Duong (2007), see Appendix A for details.

4.2. Naïve Bayes method

The kernel density estimator is a form of data smoothing; the amount of smoothing is dependent on the size (also called the bandwidth) of the kernel (i.e. the size of the neighbourhood). The fixed size of the kernel means that it will perform better in regions with high data density than in regions of low data density. This becomes a problem when we want to increase the number of features we input into the probability estimate because, as the dimension of the feature

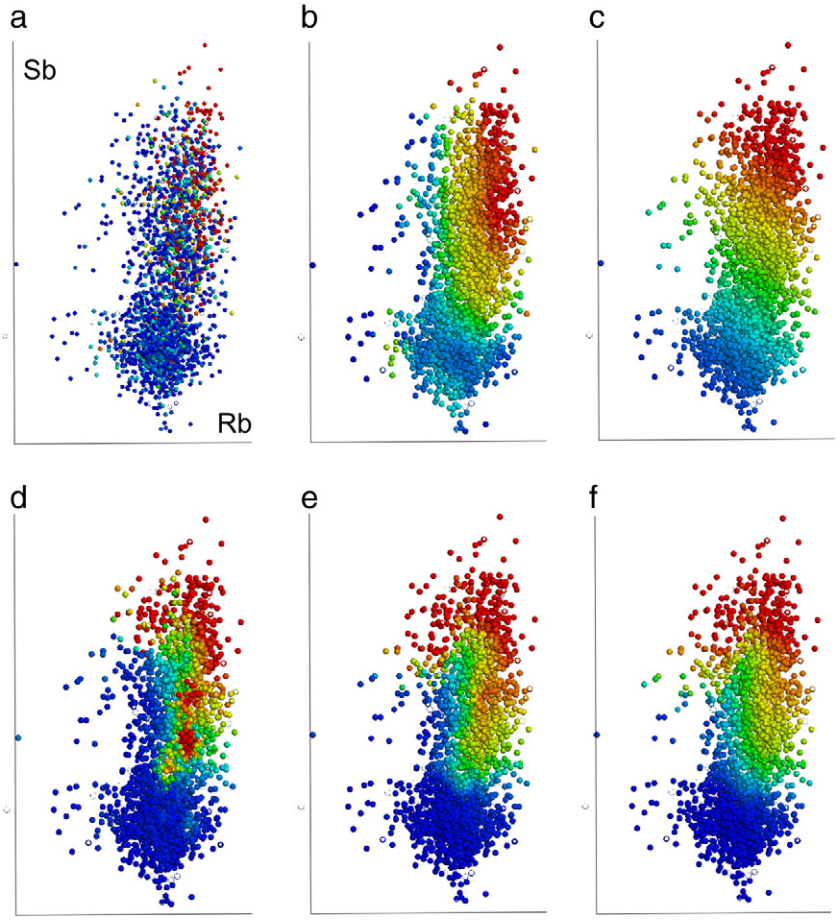

Fig. 13. Scatter plots of Sb vs Rb coloured by: (a) Au assay value, (b) KDE probability, (c) NB probability, (d) KNN k = 101, (e) KNN k = 501, (f) KNN k = 1001. For clarity, only mid-range Cr values are shown (i.e. andesitic composition). Sb and Rb values have been transformed using clr transform. Increasing values are coloured from blue through to red.

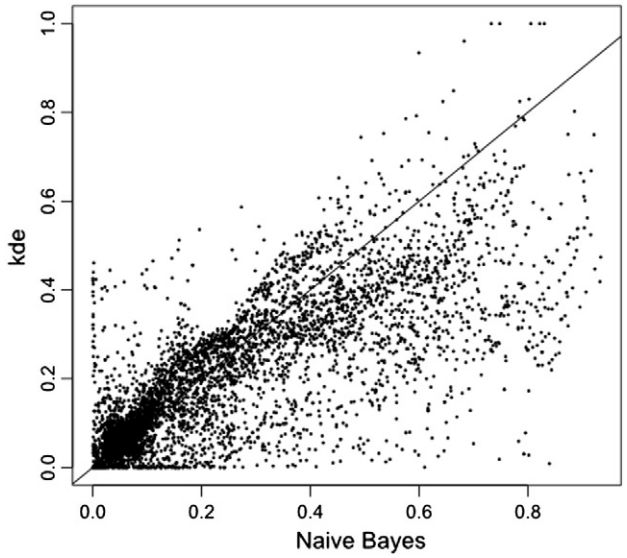

Fig. 14. Scatter plot of KDE vs NB probability values. For points above the diagonal line NB underestimates the probability compared to KDE; for points under the line NB overestimates the probability.

space increases, the data density will decrease exponentially, this problem is known as the "curse of dimensionality".

One way to get around the dimensionality problem is to assume that the features are conditionally independent, and then we can use the naïve Bayes approach (NB). If the features are conditionally independent then the probability of each element can be calculated individually and then multiplied together, for example:

$$p(G_1, G_2|Au \geq v) = p(G_1|Au \geq v).p(G_2|Au \geq v).$$

This is a strong assumption; however, NB is a popular form of classification because it often works well despite conditional dependence between input features (e.g. Domingos and Pazzani, 1997; Garg and Roth, 2001). Under certain strict conditions it can be shown to be an optimal classifier (e.g. Domingos and Pazzani, 1997; Zhang, 2004), this implies that although the probability values would be incorrect, their relative values should be correct. We can test this with our experimental data set to determine how closely the NB results compare to ordinary Bayes. We use the NB algorithm of Weihs et al. (2005).

4.3. K-nearest neighbour method

A popular method for classification of high-dimensional data is the k-nearest neighbour method (KNN). The variable k refers to the number of neighbouring data points used to estimate the class (or value) at the new data point. This method should overcome the problems for low density data as it always uses the same number of data points to estimate the new value. In this paper we test the weighted KNN method of Hechenbichler and Schliep (2004). Similar to weighted KNN is adaptive Gaussian filtering, which adapts the bandwidth of the kernel to the density of data points; it is claimed to be faster than KNN (Mills, 2011) so may be good for large data sets; we have not used adaptive Gaussian filtering here since our test data set is quite small. The KNN method is different conceptually from the conditional probability methods described above, as it does not return a probability value. Instead KNN returns an estimate of the Au content as a weighted average of samples of similar geochemical composition. Of course, this estimate will be an underestimate of the true Au content since it is based on samples which underestimate the Au content due to the nugget effect. Therefore only the relative values are of significance.

4.4. Selection of gold cut-off value

The probabilistic methods described in the preceding section require a cut-off value for Au grade. The cut-off value is used to classify samples into two categories based on assay values. When deciding whether the host rock from which a sample is extracted is mineralised or unmineralised, it may be tempting to use a cut-off value of the Au assay that has economic significance. However, an important factor to bear in mind is that we are using proxies to determine continuous regions of elevated Au content, not to measure the grade of a mineral resource. Therefore the cut-off value is of no economic significance, what we want is a value which separates background Au grades from elevated Au grades that were accompanied by detectable chemical (metasomatic) changes in the host rocks. It is essential that a cut-off value is selected such that there are sufficient samples to characterise rocks which have been mineralised, irrespective of the degree of mineralisation.

In the experiments described here we use 80 percentile as a cut-off value for elevated Au, i.e. the assay value for which 80% of samples fall below the cut-off value. Fig. 6 shows the location of the 80 percentile on a density plot of Au assay values. It is encouraging to note that this value falls approximately at the location where the density curve becomes strongly skewed to the right which can be taken to indicate the point separating the background population of Au samples (to the left) from the population of enriched Au samples (to the right). The

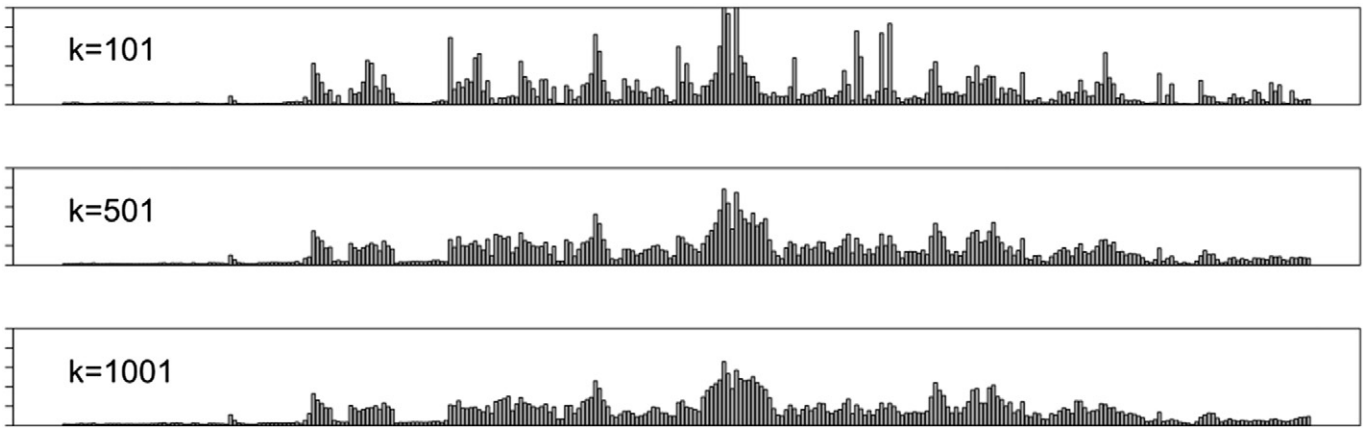

Fig. 15. Bar plots of KNN using Rb–Sb–Cr for UGD1944 for various values of k. Larger values of k reduced the variance of the results.

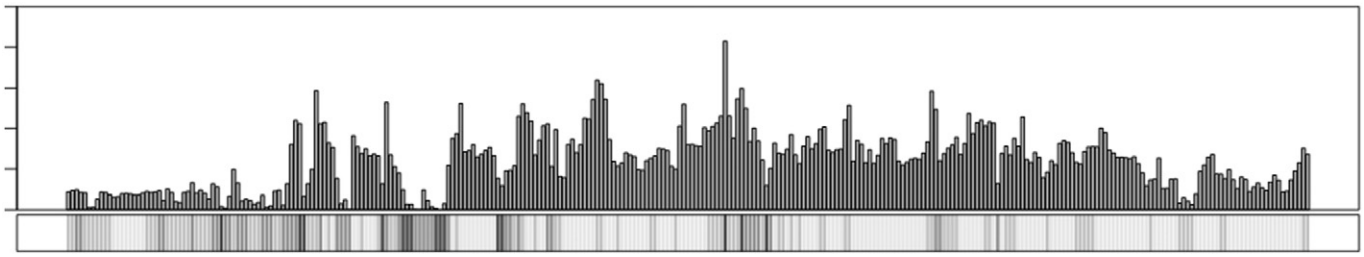

Fig. 16. Bar plot of KDE probability for UGD1944. The lower bar indicates the density of data points in feature space: dark colours = low density (high uncertainty); light colours = high density (low uncertainty).

effect of changing this value is discussed in Section 6. In this article we refer to samples containing Au above the selected cut-off value as elevated Au.

4.5. Selection of geochemical elements

An important first step before attempting to use any form of classification algorithm is to reduce the number of features as this generally results in better prediction performance as well as being less computationally intensive (Guyon and Elisseeff, 2003). Where expert knowledge of the data domain is available, then the best method is to use this knowledge to reduce the number of features, as illustrated in Fig. 7.

The ideal geochemical proxies for Au mineralisation are a balance between elements which have good spatial continuity and correlate well with Au assays. We can determine the best proxies from detailed geological observations and understanding of geological processes (i.e. expert knowledge, Fig. 7) and we can confirm these using simple exploratory statistical techniques. For example, primary lithological types can be separated using the immobile elements, in particular, Cr, Zr and Ti; sericitic alteration can be traced by increasing K or Rb content; quartz–carbonate veins have proved difficult to distinguish in these geochemical samples as they are generally volumetrically small even in very high Au content samples; sulphides can be detected by using S or one of the other elements which are major parts of the sulphide minerals, such as As or Sb. Fig. 8 shows the relationship between elevated Au assay values and As, Sb, S, Rb and Cr content. The figure illustrates that, although elevated Au values are associated with high As, Sb and Rb content, samples with high As, Sb and Rb more frequently return low Au values. This observation would result partly from the nugget effect (i.e. host rock is mineralised but no nuggets are present in the sample) and partly because only a proportion of altered rocks are mineralised. These plots show clearly that there is no simple linear relationship between Au and the geochemical elements which are believed to be associated with Au. However, they do show that samples with higher values of these geochemical elements are more likely to have high Au, supporting the practicality of a probabilistic approach to quantifying these complex relationships.

Autocorrelation plots can be used to help distinguish which elements are more spatially continuous than Au. The plots in Fig. 9 show the mean autocorrelation values for all the drill holes in the data set. Au has very poor autocorrelation, dropping away to very low values for lags greater than 2 m. Arsenic has better autocorrelation than Au, but still drops off rapidly with increasing lag. Na, S, Rb and Sb appear significantly better than As, whilst Cr (representing primary rock type) has very good spatial continuity.

Although S shows reasonably good spatial continuity, we decided not to use it as input to Au prediction as the density plot (Fig. 8) clearly shows that elevated Au values are associated with moderate S values (~5 000–35 000 ppm, log(S) ~8.5–10.5) rather than with high S values (for log(S) >10.5 ppm the lines for all samples and elevated Au sample are almost coincidental, Fig. 8); which makes S not a straightforward predictor of Au content. We chose three elements to test each method.

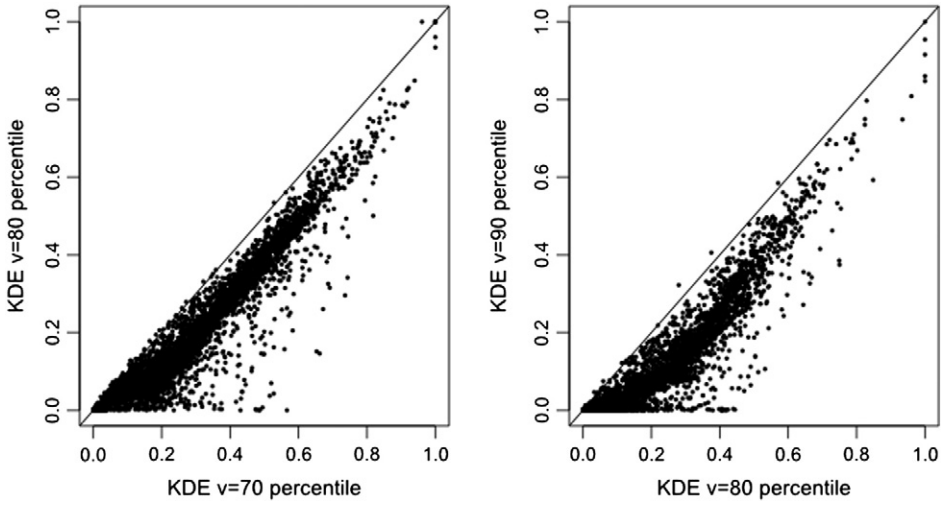

Fig. 17. The effect in the calculated probability of changing the cut-off value, v, for elevated Au. Lower cut-off value means a higher probability, as expected. More noise occurs for very high cut-off values (e.g. 90 percentile) where there is less data to characterise the higher Au samples.

The reason for this is that we wished to include the maximum number of elements where we could still easily verify the results by visualisation and three elements could easily be plotted and viewed in 3D space. Choosing more than three elements could also affect the reliability of our results by significantly reducing the data density. We experimented with two sets of three elements: (1) Cr–Rb–As and (2) Cr–Rb–Sb, being two ways of measuring the effect of primary lithology + sericitic alteration + sulphide alteration.

It is necessary to perform a log ratio transform on the compositional data before applying the kernel density estimators or KNN algorithm. In the examples illustrated in this article we have used the centred log ratio transform (clr, Aitchison, 1986). However, we could also have used a ratio with a common denominator, such as Ti, if we wanted to use the ratios in Fig. 3, for example. This is called the additive log-ratio transform (alr, Aitchison, 1986).

5. Results and comparison of methods

We have selected one drill core, UGD1944, to illustrate the results. Fig. 10 shows the distribution of the selected elements along the core. The high Cr near the top of the hole reflects the presence of komatiite and high-Mg basalts.

Firstly, we compare the effect of changing the set of geochemical elements selected for the proxy. Figs. 11 (Cr–Rb–As) and 12 (Cr–Rb–Sb) show the comparison between the distribution of Au assay values and the results for predicting the presence of elevated Au values using the three methods outlined above. When As is replaced by Sb in the input data the results are much less spiky because Sb is more spatially continuous than As (i.e. higher autocorrelation values). Note that the interval in the deeper part of the drill hole indicated by a grey bar in Fig. 12 is much reduced in amplitude when As is replaced by Sb. The input data in Fig. 10 indicates that this feature can be attributed to the presence of a zone of high As. The better spatial continuity of Sb compared to As makes it more desirable as a proxy element.

Secondly, we compare the results of the two types of probability estimate: KDE, which makes no assumptions about conditional independence, and NB, which assumes that all the proxy elements used are conditionally independent. To understand the reasons for the different results for the different methods it helps to view a plot of the data in feature space. Fig. 13a–c shows point data plotted in Sb–Rb space and illustrates clearly that the NB probability method results in extreme smoothing of the data and loss of much internal detail. The figure also shows that outlying data points in the KDE probability plot are heavily influenced by the Au value of the individual sample because the geochemically similar neighbours are too far away to influence the prediction. Fig. 14 illustrates how the assumption of conditional independence of data affects the results. The point plot of conditional probability from KDE versus NB shows a wide scatter of data and a strong tendency for NB to overestimate the probability, indicating that the NB does not perform well against KDE for this data set.

Finally, we illustrate how the KNN method performs against the probability method. The results produced by the KNN method depend on the value of k (i.e. number of neighbouring points used in the calculation). Fig. 15 illustrates how increasing the neighbourhood size has the effect of decreasing the variance of the results: for k of 11, 101, 501 and 1001 the corresponding variance of the results is: 11.0, 2.1, 0.7 and 0.5, respectively. In this method there is a trade-off between obtaining a smooth result at the cost of loss of detail as the neighbourhood increases in size (Fig. 13d–f). The anticipated result, that the outlying data points more closely reflect the values of their neighbours, has been achieved. However, for low values of k there is over-fitting of results in the high density regions (Fig. 13d). It also appears that for higher values of k the lower density regions may be over-smoothed, resulting in the very high KNN values observed for low Rb + high Sb in contrast to the relatively low values provided by the conditional probability method; compare Fig. 13e–f with b.

6. Discussion and conclusions

6.1. Preferred method

The strong smoothing of data in Fig. 13c and the high level of scatter illustrated in Fig. 14 indicate that NB probability method is not a good substitute for KDE probability for this particular data set (i.e. the conditional independence assumption is not a good one) and we reject it. However, NB may provide a reasonable estimate for projects where the data is insufficient to provide a suitable data density for the KDE method.

The KNN method for small neighbourhoods (e.g. k = 101) appears to provide a better estimator for data in very low density regions of feature space compared to the probability method at the cost of increased detail in the high density regions (Fig. 13d). Higher values of k provide a smoother result in the high density regions (Fig. 13e–f) but this affects the values in low density regions particularly for low Rb. It

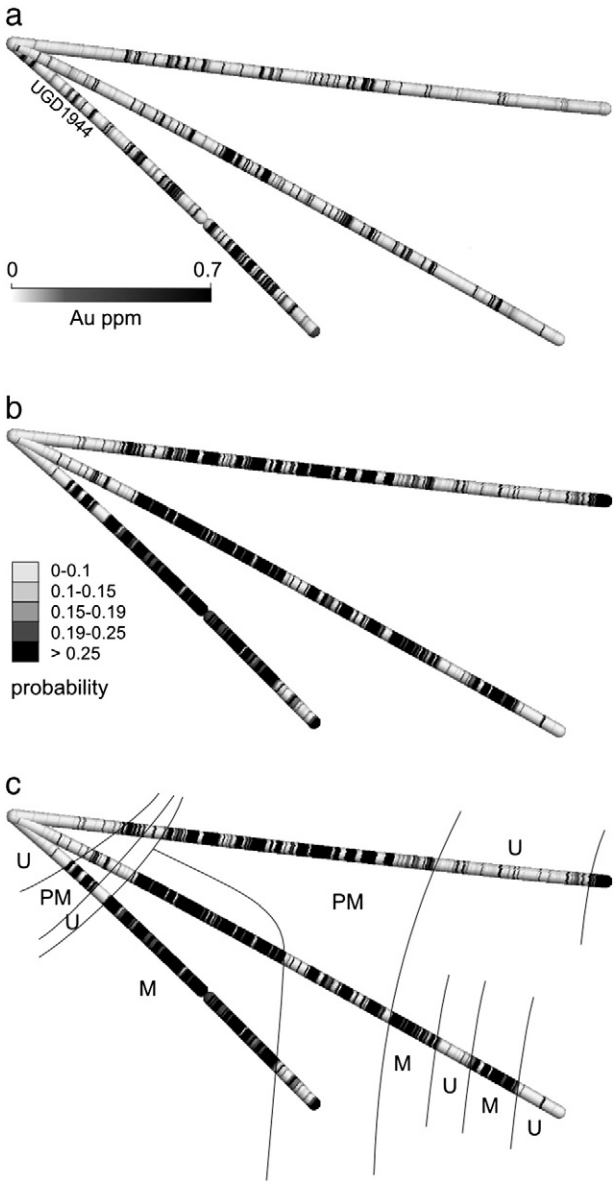

Fig. 18. Comparison of the (a) spatially discontinuous nature of Au assays and (b) more spatially continuous conditional probability values for Rb–Sb–Cr proxy plotted on a drill hole fan. (c) An interpretation of mineralised zones using probability values: M = mineralised zone, PM = patchy mineralised zone, U = unmineralised zone. View is looking approximately to the north on a west to east cross-section; the deepest hole is UGD1944.

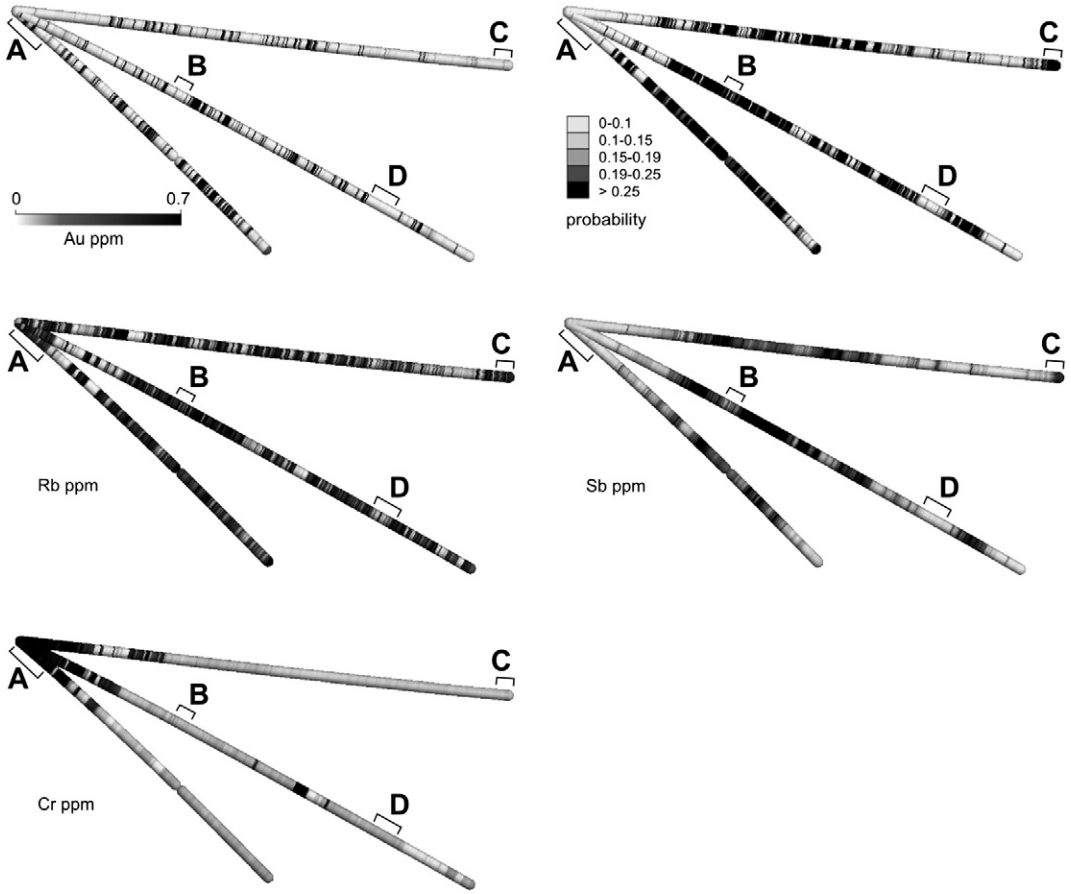

Fig. 19. Results can be evaluated visually by comparing input data (Rb, Sb, Cr) to the probability values. For example, at location A probability is low because high Au assays are rare in samples with high Cr, irrespective of Sb or Rb content. At location **B**, high Rb and moderate Sb resulted in a high probability although high Au assays are rare in this interval. Location **C** also has very few high Au assays but has high probability due to high Sb and Rb. Location **D**, on the other hand, returned a low probability despite favourable Cr composition because Sb was low and Rb was only moderate. Low values are shown in light grey and high values in dark grey to black.

is more important to have a good estimator for the high to medium density regions as these contain most of the data; variation in the low density regions will have comparatively little effect on the results. In order to visualise the reliability of the results based on data density, we can plot data density adjacent to the results; Fig. 16 shows an example of this. If we can take into account the uncertainty associated with low density data then the probability method is preferred over the KNN method.

6.2. Cut-off value for gold

As discussed previously, the cut-off value for Au used to indicate the presence or absence of mineralisation must be selected such that a sufficient proportion of data points will be assigned to the elevated Au category. The choice is a compromise between having insufficient data to characterise the geochemical composition of elevated Au samples against including samples whose geochemical signatures should belong to the background Au category. The effect of changing this cut-off value is illustrated in Fig. 17. The graphs show how the cut-off value will affect the calculated values of the probability. The strongly linear distribution of data points indicates that the relative probability values show little change with changing cut-off value and it is these relative values that we are interested in. What is more important is that as we increase the cut-off value (to 90 percentile, for example) the noise (scatter) in the results increases and this is undesirable. We chose the 80 percentile as a suitable cut-off value as it falls at a natural break in the trend of the density curve and allows a significantly large proportion (20%) of the total data set to fall into the elevated Au category.

6.3. Assessing the conditional probability method for automating proxies

When assessing how well the conditional probability method works for geochemical proxies we have to consider two points separately: (1) how well does the conditional probability method perform and (2) how well does the selected proxy perform.

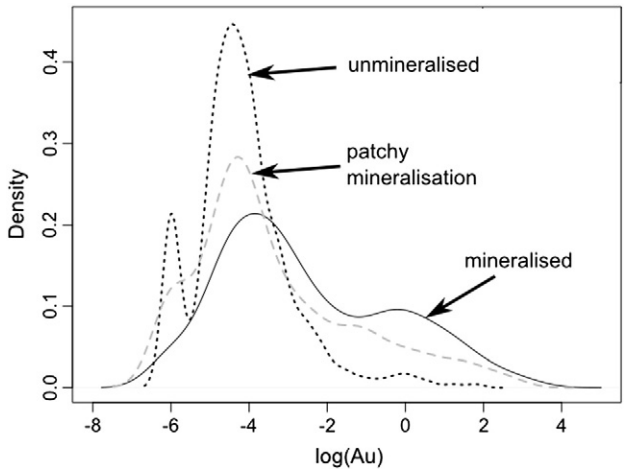

Fig. 20. Density of Au assay values for the 3 spatial categories interpreted in Fig. 18c.

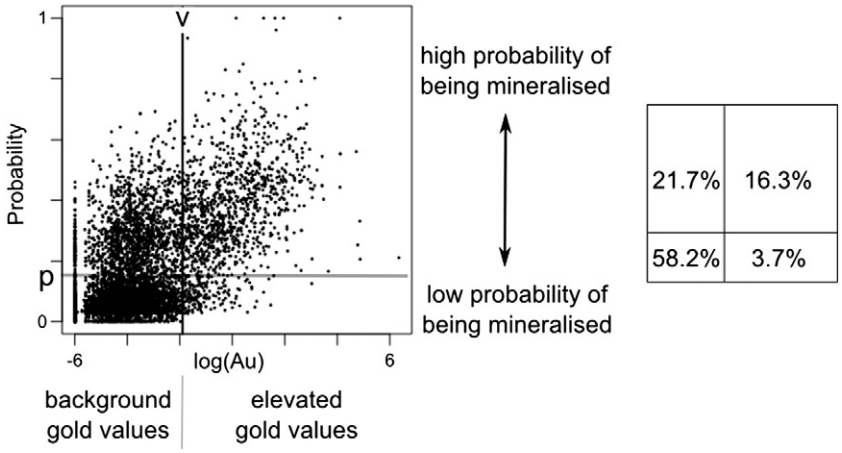

Fig. 21. The plot shows probability plotted against log (Au); the top-left quadrant is most interesting as it contains samples which had low Au assays but a high probability that the host rocks are mineralised. The vertical line represents the Au cut-off value (v). The horizontal line represents the probability cut-off value (p). Note that a few higher grade samples are now classified as unmineralised (lower right quadrant); these would be samples whose compositional type is rarely mineralised. The samples from the lower left quadrant are almost certainly waste: low Au assay and low probability that their host is mineralised. The percentage of samples which fall into each quadrant is shown on the right.

As discussed in the preceding section, the success of the probability method can be evaluated using plots like those of Fig. 13, which show how well the geochemical data is generalised as a predictor for elevated Au grades for the various methods. For example, if only a small data set is available for analysis then plots like those in Fig. 13 would be very useful when making a decision whether the KDE or NB approach would give better results.

There are two aspects to consider when determining how well the geochemical elements we chose as proxies performed: (1) the proxy provides a more spatially continuous predictor of mineralisation than the Au assays and (2) the proxy is a good indicator of Au.

To determine if the proxy is spatially continuous we plot the probability values on the drill holes, see Fig. 18. It is clear from this figure that the Rb–Sb–Cr proxy is much more spatially continuous than the Au assay values. The use of a proxy results in more clearly defined mineralisation boundaries than using even quite low Au assay values, and interpolating boundaries between drill holes using the probability values is less problematic than using Au assays alone.

We can also use plots of data on the drill holes to assess how each element in the proxy contributed to the resulting probability value. For example, in the drill hole fan illustrated in Fig. 19, the resulting probability reflects our initial hypothesis based on expert knowledge that high Cr rocks are unlikely to contain high Au values, whilst high Rb and Sb samples are favourable for Au mineralisation. These observations give us confidence in the result.

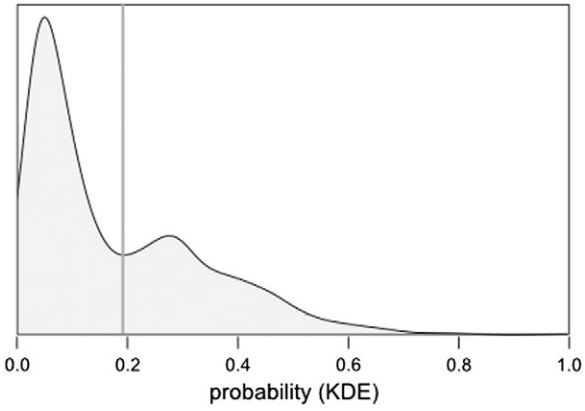

Fig. 22. The density plot for probability is bimodal. The vertical line is plotted at p = 0.19, the location of the minimum point between the two peaks.

When selecting proxies we chose geochemical elements that showed a clear relationship to Au content, i.e. the proportion of the proxy element is related to the proportion of Au, although they do not need to be linearly related, Fig. 8. However, we can also assess the how well the combined elements capture the elevated Au values by plotting our interpretation from Fig. 18c on a density plot, see Fig. 20. This plot shows that very few high grade samples are included in the unmineralised zone and that this entire zone could be confidently classified as waste. This leads to the question of how to use the probability values to help define the mineralised zone, which will be discussed below.

6.4. Using the conditional probability values

We expect that the number of samples that would have been classified as mineralised from the Au assays alone has been underestimated because of the nugget effect. The aim of applying the conditional probability method to the data is to decide which samples should be reclassified from unmineralised to mineralised. If we have confidence in our geochemical proxies, any sample that returned a low assay but has a geochemical composition similar to the samples with elevated assay values should be classified as mineralised, i.e. the samples in the top left quadrant of Fig. 21. The probability values do not tell us which samples to include as mineralised, but they provide a clear guide. For example, using the cut-off values illustrated in Fig. 21, approximately 27% of samples which returned background-level Au assay values could be classified as likely to be mineralised.

A degree of judgement is required to use the probability values effectively. A cut-off value for probability may be derived from statistical analysis. For example, Fig. 22 shows a bimodal distribution of probability values which may be interpreted as representing two distributions: mineralised and unmineralised samples. In this case the obvious cut-off value for probability would be the at minima between the two peaks, i.e. 0.19. A probablility of 0.19 means that approximately 1 in 5 samples with that composition returned Au grades above the Au cut-off value. This is the probability value that was used in the interpretation in Fig. 18c. Interestingly, the probability values in the upper drill hole are still quite discontinuous in nature. This is a result of the distribution of the geochemical elements used in the proxy and probably reflects geological observations that, in parts of the ore body, alteration and deformation are localised to very narrow zones.

Mining methods may also be an important consideration when determining the boundaries of mineralisation. Where the zones of mineralisation are narrower than the minimum mineable width then the probability values can be used as a guide to select the most

appropriate adjacent material to include in the mine blocks, i.e. that which has the highest probability of containing elevated grades.

One of the advantages of conditional probability modelling is that we derive a single meaningful value from the geochemical data that can be interpolated using 3D modelling software to rapidly generate a first-pass 3D mineralisation model (Hill et al., 2013b). It is our intention to use this method to model the Vogue mineralisation once a complete data set has been collected.

7. Summary

The conditional probability method (using KDE) proved to be a successful method for generalising the relationship between Au assay results and the geochemical composition of samples. The exception is that it works less well in regions of low data density, where the naïve Bayes method (assumption of conditional independence of features) and the k-nearest neighbour methods (for low values of k, i.e. small neighbourhood) may be preferable. However, it is more important to have a good estimator for the high to medium density data regions as these contain most of the data. So, except in the case where the data set is very small, the KDE method is preferable.

When choosing a cut-off value to represent samples with elevated Au content in the probability calculation it is more important to choose a value that allows sufficient samples to represent the elevated Au category rather than taking any economic implications of Au grade into consideration.

The number of geochemical elements considered in the proxy can be reduced using expert knowledge and exploratory statistical analysis. For the test data set we found the combination of Sb, Rb and Cr to provide a good compromise between elements that were significantly more spatially continuous than Au but were still good proxies for elevated Au assays.

Acknowledgements

This work is the result of a joint research project between CSIRO, AngloGold Ashanti and James Cook University. We thank Sunrise Dam Gold Mine Exploration Department for providing data, allowing us to publish results, and for contributing to the understanding of the geology. The authors would like to thank Thomas Poulet and Michael Gazley for helpful suggestions on improving the manuscript.

Appendix A. Technical information

Sampling for geochemical analysis was taken from 18 drill holes, and covered an accumulated drill hole length in excess of 6000 m. Most samples were taken from drill hole intervals of 1 m, but in areas of high grade the intervals may be shorter. The intervals were re-composited to 1 m using Leapfrog Mining software (www.leapfrog3d.com). There were 6216 re-composited intervals.

Figs. 2, 13, 18 and 19 are images taken from Leapfrog Mining 3D models. All other data manipulation and plotting were performed using statistical software environment R (www.r-project.org), as follows:

- Imputation of missing multi-element geochemistry data using impKNNa function from the robCompositions package (Hron et al., 2010).
- In order to avoid any issues that may arise from applying multivariate statistical techniques to compositional data sets (closure problems), all multi-element geochemistry data transformed using the centred log-ratio transform (Aitchison, 1986) using the cenLR function from the robCompositions package.
- Kernel density estimates used the Hpi and kde functions from the ks package (Duong, 2007). The kernel is a symmetric normal probability density function. We used the bandwidth estimator recommended by Duong (2007): SAMSE (sum of asymptotic mean square error) pilot bandwidth criterion and sphering pre-transformation of data.
- Naïve Bayes was calculated using the naïve Bayes and predict. Naïve Bayes functions of the klaR package with default arguments (Weihs et al., 2005).
- Weighted nearest neighbour regression was performed using kknn function of the kknn package with a triangular kernel (Hechenbichler and Schliep, 2004).
- All plots showing log values use natural log.

For more information on the use of R in the geosciences see Grunsky (2002).

Appendix B. Supplementary data

Supplementary data associated with this article can be found in the online version, at doi: http://dx.doi.org/10.1016/j.gexplo.2014.05.008. These data include Google maps of the most important areas described in this article.

References

Aitchison, J., 1986. The Statistical Analysis of Compositional Data Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London, UK, (416 pp.).
Blenkinsop, T., Baker, T., McLellan, J., Cleverley, J., Nugus, M., 2007. Sunrise Dam Gold Mine, Project G15, Final Report. Predictive Mineral Discovery Cooperative Research Centre, Australia.
Carswell, J.T., Clark, F., 2013. Underground reverse circulation grade control drilling in a high nugget gold deposit — Sunrise Dam, Western Australia. Exploration, Resource and Mining Geology Conference, Cardiff, Wales, UK, October 2013.
Domingos, P., Pazzani, M., 1997. Beyond independence: conditions for the optimality of the simple Bayesian classifier. Mach. Learn. 29, 103–130 (Bottom of Form).
Dominy, S.C., Johansen, G.F., 2004. Reducing grade uncertainty in high-nugget effect gold veins-application of geological and geochemical proxies. PACRIM 2004 Congress, 19–22 September 2004, Adelaide, SA, Australia.
Dominy, S.C., Platten, I.M., Raine, M.D., 2003. Grade and geological continuity in high-nugget effect gold-quartz reefs: implications for resource estimation and reporting Applied Earth Science. Trans. Inst. Min. Metall. B 112, 239–259.
Duong, T., 2007. ks: kernel density estimation and kernel discriminant analysis for multivariate data in R. J. Stat. Softw. 21, 1–16.
Garg, A., Roth, D., 2001. Understanding probabilistic classifiers. In: Raedt, L., Flach, P. (Eds.), Proceedings of the 12th European Conference on Machine Learning. Springer, pp. 179–191.
Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157–1182.
Grunsky, E., 2002. R: a data analysis and statistical programming environment — an emerging tool for the geosciences. Comput. Geosci. 28, 1219–1222.
Hantler, A., 2009. Paragenesis of Mineralized Veins at Sunrise Dam Gold Mine, Laverton, Western Australia. (M. Sc. thesis) Economic Geology Research Unit, James Cook University, Australia.
Haren, E., Williams, P., 2000. Mine geology practices at the Sunrise Open Pit. Proceedings of the 4th International Mining Geology Conference, Coolum, Queensland, Australia, May 2000.
Hechenbichler, K., Schliep, K.P., 2004. Weighted k-nearest-neighbor techniques and ordinal classification. Discussion Paper 399, SFB 386. Ludwig-Maximilians University Munich.
Hill, E.J., Nugus, M.J., Clark, F., Oliver, N.H.S., Cleverley, J.S., Carswell, J.T., 2013a. Predictive three-dimensional modelling of nuggety gold orebodies, Sunrise Dam, Western Australia. Proceedings of the World Gold Conference, Brisbane, Australia, September 2013, AusIMM.
Hill, E.J., Oliver, N.H.S., Cleverley, J.S., Nugus, M.J., 2013b. Modelling ore bodies of high-nugget gold using conditional probability. Proceedings of the 15th Annual Conference of the International Association for Mathematical Geosciences, Lecture Notes in Earth System Sciences. Springer.
Hill, E.J., Oliver, N.H.S., Cleverley, J.S., Nugus, M.J., Carswell, J., Clark, F., 2013c. Characterisation and 3D modelling of a nuggety, vein-hosted gold ore body, Sunrise Dam, Western Australia. J. Struct. Geol. http://dx.doi.org/10.1016/j.jsg.2013.10.013.
Hron, K., Templ, M., Filzmoser, P., 2010. Imputation of missing values for compositional data using classical and robust methods. Comput. Stat. Data Anal. 54 (12), 3095–3107.
Mills, P., 2011. Efficient statistical classification of satellite measurements. Int. J. Remote Sens. 32, 6109–6132.
Nugus, M., Biggam, J., Clark, F., Erickson, M., 2009. Development and application of macroscopic geological proxies for gold mineralisation in high nugget environments — an example from the Sunrise Dam Gold Mine, Western Australia. Proceedings of the Seventh International Mining Geology Conference 2009, Perth, Western Australia, pp. 309–318.
Weihs, C., Ligges, U., Luebke, K., Raabe, N., 2005. klaR analyzing German business cycles. In: Baier, D., Decker, R., Schmidt-Thieme, L. (Eds.), Data Analysis and Decision Support. Springer-Verlag, Berlin, pp. 335–343.
Zhang, H., 2004. The optimality of naive Bayes. Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference FLAIRS 2004. AAAI Press.