Special Publication 12, 2006, pp. 295–310

Chapter 15

Mineral Exploration Using Modern Data Mining Techniques

COLIN T. BARNETT†

Exploration Consultant, 424 Mapleton Avenue, Boulder, Colorado 80304

AND PETER M. WILLIAMS

Statistical Consultant, 28 Eaton Place, Brighton BN2 1EG, United Kingdom

Abstract

Returns from gold exploration have been disappointing over the last 20 years, despite the surge in quality and quantity of exploration data. Historically, major discoveries have occurred in waves following the introduction of new methods. This paper argues that the new methods driving the next wave of discoveries will be found in recent developments in data mining techniques, including visualization and probabilistic modeling.

Visualization techniques present information to the brain in ways that allow patterns to stand out and be more readily perceived by our own human intelligence. Combined with geophysical inversion, these techniques make it easier to integrate multiple data sets and to build geologic models which fit current knowledge and understanding. These models can then be passed among, and visually shared by, workers from all the exploration disciplines.

Probabilistic modeling techniques provide an estimate of the probability that some location with given exploration characteristics hosts a deposit, based on a set of known examples. The weights of evidence approach, which has already been used for this purpose, can provide useful results, but is limited by its basic assumptions. Neural network and kernel methods, on the other hand, are not limited in this way and can extract more meaningful information from data.

The approach is demonstrated by a study of gold exploration in the Walker Lane, a mature mining district straddling the Nevada-California border in the western United States. This study incorporates 25 primary exploration data layers including geology, remote sensing, geochemistry, gravity, aeromagnetic and radiometric surveys, digital terrain and regional structure, together with known gold deposits. Care is needed in presenting data to the model. Geophysical data, for instance, may have little significance as point values, and need an encoding that represents the pattern of data in the neighborhood of a given station. The same is true of regional structure and, to some extent, of geology. The number of inputs to the model can grow in this way into the hundreds, so that efficient optimization and regularization are required.

The model allows the results for individual data sets to be analyzed separately. The geology, for example, shows a strong correlation between the known gold deposits and a Tertiary andesite. The other data sets show similar but not necessarily coincident patterns. The data sets can then be combined to produce an integrated target favorability map. A subarea of the Walker Lane that falls within the Nevada Test Site illustrates the approach. Two specific targets are identified, which would certainly be followed up if this former nuclear weapons testing area was not off-limits to exploration.

Finally, the distributions of favorability scores, over the known gold region and the region as a whole, determine the probability that a location scoring higher than a given threshold hosts a deposit. The distributions of scores also permit the expected costs and benefits of an exploration program to be calculated, and show how improved targeting derived from the model reduces exploration costs and increases the probability of success.

Introduction

FOLLOWING an analysis of recent performance of the gold industry, Schodde (2004) concludes that gold exploration is currently only a break-even proposition. In the last 20 years, the average cost of a new discovery has increased nearly fourfold, and the average size of a deposit has shrunk by 30 percent. The average rate of return for the industry has been 5 to 7 percent, which is the same order as the cost of capital. Why should this be, and what can be done about it?

†Corresponding author: e-mail, colin@bwgmining.com

Paterson (2003) observes that, historically, discoveries have taken place in waves, after the introduction of new methods or advances in the understanding of ore genesis. For instance, discovery rates jumped sharply between 1950 and 1975, following the development of new methods and instruments in exploration geophysics and geochemistry. In the last quarter century, there has been a comparable surge in digital electronics and computing that has resulted in a great increase in the quality and quantity of exploration data. Yet these developments, on their own, evidently were not sufficient to reverse a downward trend in the discovery rate during this period.

So where should we look for new methods to drive the next wave of discoveries? It seems we are now collecting data faster than we can absorb it. But this is also true in bioinformatics with genome sequencing, or in making sense of other huge corpora of data available on the Internet. It is the thesis of this paper that the new methods are to be found in ways currently being developed for extracting meaningful information from data. Specifically, we should look to recent developments in visualization and data mining.

Many data mining techniques are inspired by analogy with human intelligence and suggest a new idea of computing. Conventional computing is restricted to tasks for which a human can find an algorithm. Living creatures, however, are programmed by experience, not by spelling out every step of a process. Data mining is therefore about discovering how machines, like humans, might learn from data.

Machine learning broadly distinguishes between *supervised* and *unsupervised* learning. Supervised learning, or learning from examples, requires sufficiently many labeled cases to be available. These form a set of known input-output pairs, usually called a *training set*, and the task is to learn the true input-output mapping from these examples. In the exploration case, the training set typically consists of a collection of known deposits and known barren regions. For unsupervised learning, we know only the inputs and not the corresponding outputs. The aim, then, is to search for "interesting" features of the data, such as clusters or outliers, or for some latent structure which would account for how they were generated. In this paper only the case of supervised learning is considered, but see Williams (2002) for some further discussion of both approaches.

The paper begins with a review of recent advances in visualization and supervised learning techniques, such as neural network models. The use of these ideas is then demonstrated in a study of gold exploration in the Walker Lane in the western United States. Finally it is shown how the results can be applied to a quantitative analysis of exploration risk, and how improved targeting accuracy can reduce exploration costs and increase the probability of success.

Visualization

An important method of data mining is provided by visualization. The human brain has a great capacity to decipher and make sense of complicated data if the information is presented clearly to the eye. During the last two decades, huge strides have been made in ways of displaying information visually so that it can be more easily assimilated by a trained interpreter.

Two-dimensional imaging

The introduction of color imaging was one of those strides. This enabled geoscientists to display two-dimensional data as colored, shaded-relief images or, in the case of satellite reflectance data, as false-color composites. These color images are far more illustrative, and easier to interpret, than old-fashioned black-and-white contour or posted maps. Also, by simple adjustment of the artificial sun angles, subtle features can be emphasized or picked out in an individual data set.

The next step was to bring multidisciplinary data together in a so-called geographic information system (GIS), which allowed people to integrate two-dimensional data sets, either by laying them out and tracking them side-by-side, or by overlaying them at the same scale. Now commonly used GIS programs, such as ArcView™ and MapInfo™, also have the facility of managing and querying the data bases and of bringing together selected point, line, and area information for the user to visualize.

Three-dimensional interpretation

Major advances have also taken place in the way exploration data are processed before being passed to a visualization system. These include, for example, the inversion of geophysical data from basic field measurements to interpreted physical rock properties using three-dimensional modeling software (Li and Oldenburg, 1996, 1998). So now, instead of simply looking at a gravity or magnetic map presented in the measured units of milligals or nanoteslas, it is possible to view an automated interpretation of the underlying density or susceptibility. Because of the well-known, inherent ambiguity of a potential field geophysical data set, constraints based on any available geological and physical information are normally imposed on the inversion process to obtain a satisfactory result.

The last few years have seen these various technologies come together in advanced three-dimensional visualization packages, such as Gocad™. The term common earth model (CEM) has been coined to describe geologic models that have been built by integration of cross-disciplinary data sets (Garrett et al., 1997; McGaughey and Vallee, 1998). Inputs to a CEM may consist of one-dimensional (or essentially linear) borehole information, two-dimensional plans and cross-sections, and three-dimensional block models of geological, geochemical, and geophysical properties. These require the model to be topologically correct, with closed boundaries and with tied or connected surfaces between different formations.

Figure 1 shows an illustration of a CEM of the Kalgoorlie mining district. From a starting point of the mapped surface and underground geology, this three-dimensional model has been built in Gocad™ to match all the available drilling and geophysical information. The positions of the Moho surface and the prominent Ida fault surface have been interpreted from seismic data. Other formations, such as the granite plutons (shown here in purple), and the banded iron formations (shown in black), have been interpreted from outcrop positions, drill intersections and, just as importantly, from the inversion of gravity and magnetic data (P. Kowalczyk, pers. commun., 2004). Now that this CEM has been constructed, it can be shared and understood by geologists and other geoscientists working together to further explore this important mining district.

Fusion of multiple data sets

The wide variety of data encountered in mineral exploration presents further challenges for intuitive three-dimen-

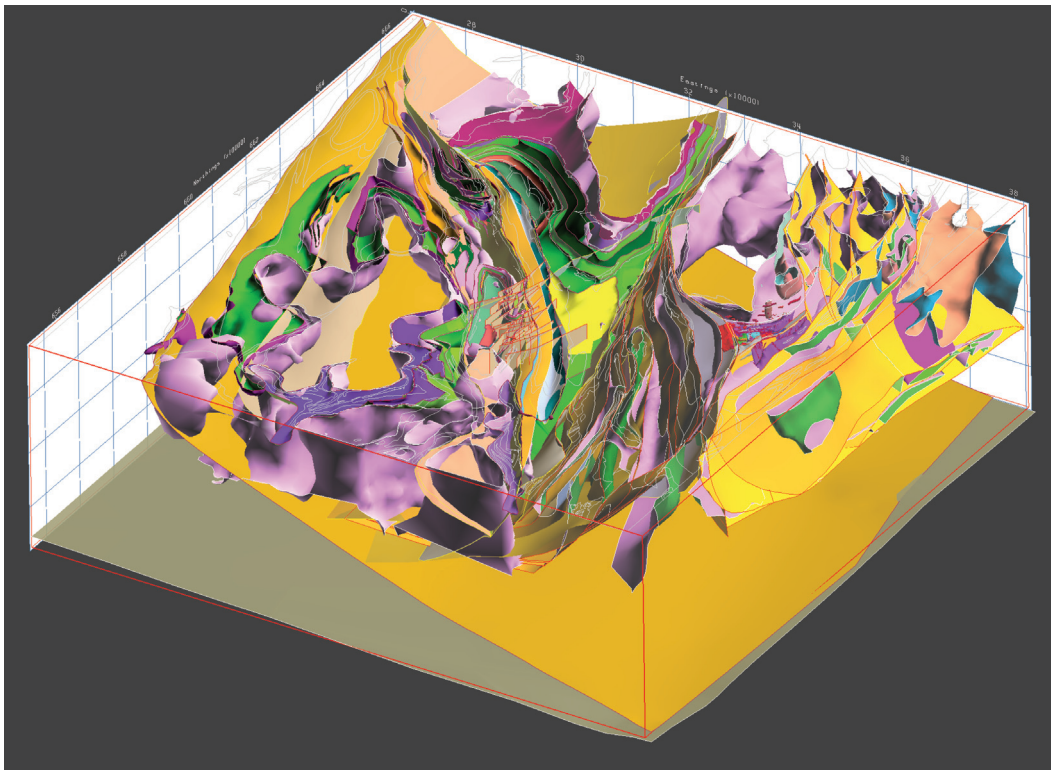

FIG. 1. A three-dimensional geologic model of the Kalgoorlie mining district. The view is looking northwest, and the model dimensions are 112 × 115 × 36 km. The two mines marked in red are Kundana and Kanowna Belle. The gray layer at the base represents the Moho surface, and the tan layer is the Ida fault surface. Granitic plutons are shown in purple. Image courtesy of Placer (Granny Smith) Pty Limited.

sional visualization. Other industries, such as oil exploration and weather forecasting, have led the way in meeting these challenges, using clustered computer systems to create immersive visualizing environments (see Stark et al., 2000). In these room-sized or so-called cave configurations, the interpreter actually enters the environment in person and uses specialized controls and headgear to highlight information in the volume of data that surrounds him.

Because of the cost and complexity of these systems, it may take a few years before they become mainstream mineral exploration tools. In the meantime, other researchers are doing interesting work on the fusion of multiple data sets, which can be implemented on current and therefore more economical computer systems (Treinish, 2001). The approach makes clever use of color and translucency, combined with contour and vector data, to display an impressive number of variables in a single visualization.

Figure 2 shows a three-dimensional visualization of a predicted weather system over the Hawaiian Islands. At the base is a topographic map on which the coastlines of the islands have been draped in black to provide positioning information. A surface field, precipitable water, is displayed as pseudo-colored contour bands. A second surface, the 90 percent isopach of relative humidity, is shown in translucent white to approximate a cloud boundary. Meanwhile, vertical wind speed is displayed as a vertical slice cut in the lee of the large island of Hawaii. Superimposed on all this are vertical profiles showing the direction of the model wind field at two locations. The length and also the color of the arrows and streamribbons represent the horizontal wind speed. Finally, the profile axes are also gainfully employed to display the relative humidity as a pseudo-colored tube.

Data fusion techniques like this could be immediately put to use at little extra cost by the mineral industry to display multiple exploration data sets in three dimensions, and to extract information more readily from the mountains of data that we are accumulating today.

Probabilistic Modeling

Human vision is capable of sophisticated pattern recognition which is often difficult to simulate computationally; for example, the detection of linear patterns in a complex image. Multilayer exploration data sets, however, generally contain subtle numerical patterns of correlation which are unlikely to be discovered by visualization alone. It is important, then, to supplement visualization with statistical data mining and pattern recognition techniques. In the exploration case, the most useful product of a statistical analysis is an estimate of the conditional probability that a deposit occurs at some location within the region of interest, given the sum total of exploration data relating to that location. Symbolically, this can be expressed as $P(D|\mathrm{x})$, where D is the proposition that a deposit occurs, and x is a feature vector expressing the exploration data at or near that location.

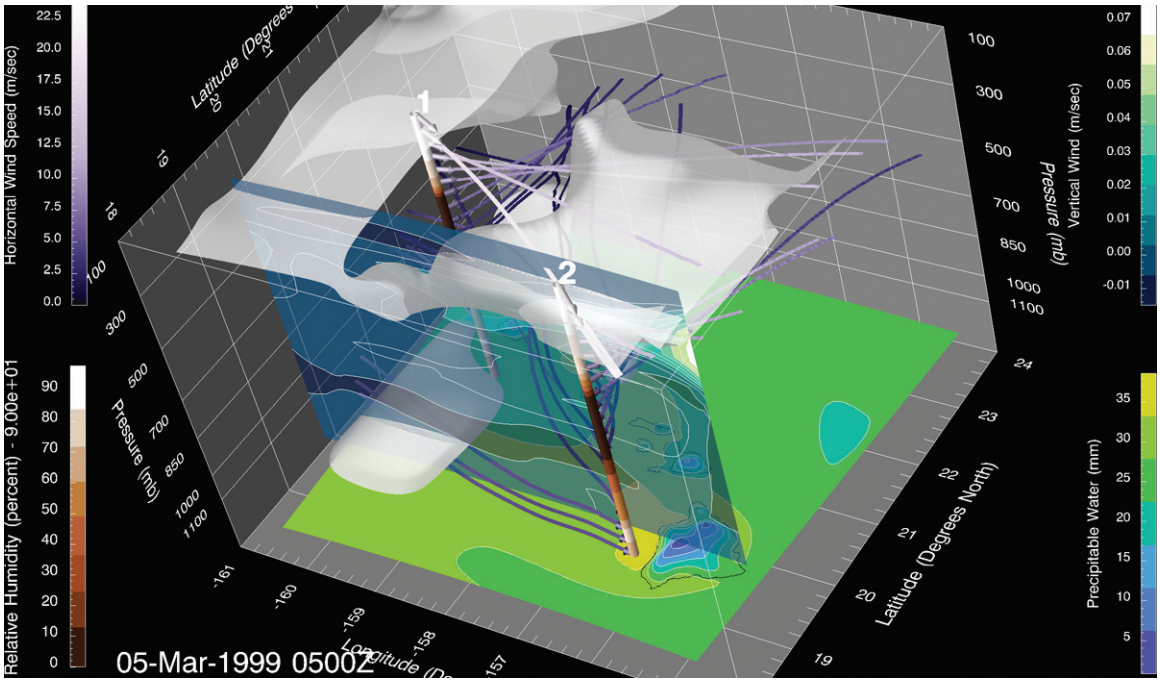

FIG. 2. A fusion of multiple data sets. This is a bird's-eye view of a weather system over Hawaii, showing pressure, relative humidity, precipitable water, horizontal and vertical wind speed, combined in a single image. Image courtesy of Lloyd Treinish, IBM Thomas J Watson Research Center.

Weights of evidence

A straightforward way of estimating $P(D|\mathrm{x})$ is given by the weights of evidence approach. This has been widely used for target favorability mapping in mineral exploration (Bonham-Carter, 1994; Raines, 1999; Mihalasky, 1999; Kemp et al., 2001). Characteristically, it uses a calculus expressed in terms of logarithmic odds, which gives the approach its distinctive name.

The strength of the approach lies in its simplicity. Under suitable assumptions, the probability $P(D|\mathrm{x})$ can be estimated directly from a database by counting and simple arithmetic. The assumptions are that (1) exploration data are conditionally independent and (2) each component of the feature vector x takes only a small finite number of values. These assumptions, however, are only very approximately satisfied by actual exploration data (see Appendix). Despite these limitations, a careful analysis using weights of evidence can provide useful and insightful results. A case in point is the excellent thesis by Mihalasky (1999).

There are other methods, however, which are not subject to these limitations and which are therefore capable of extracting more information from data.

Functional approximation

One characterization of $P(D|\mathrm{x})$ is as the best mean square approximant to the proposition D in terms of the feature vector x. This means that $P(D|\mathrm{x})$ is the function $f(\mathrm{x})$ which minimizes the expected value of $|D - f(\mathrm{x})|^2$, where D is now interpreted as a random variable taking the value 1 if a deposit occurs and 0 otherwise. This function certainly exists, but in general it will be a nonlinear function of x. Indeed, nothing can be said in advance with any certainty about its functional form.

There are several modern mathematical techniques for tackling a problem like this, such as feed-forward neural networks, radial basis function networks, or more general kernel representations. Neural networks, for example, have been used for targeting and prospectivity mapping in Poulton et al. (1992), Brown et al., (2000, 2003) and Bougrain et al. (2003). Radial basis function networks were used by Singer and Kouda (1997, 1999), Harris and Pan (1999), Harris et al. (2003) and Porwal et al. (2003).

Parameter fitting

Whatever functional form is chosen, the exact function to be used, $f(\mathrm{x}, \mathrm{w})$ say, will depend on parameters w which have to be fitted to the data. The aim is to choose w so that $f(\mathrm{x}, \mathrm{w})$ is a good approximation to the conditional probability $P(D|\mathrm{x})$, for realistic values of the exploration data x. This will be true if w is chosen to minimize the expected value of $|D - f(\mathrm{x}, \mathrm{w})|^2$, though other error functions are possible (see the Appendix).

The normal approach is to minimize an empirical estimate of this expectation, using a training set for which the values of x and D are known. Suppose, therefore, that we have a set of N known pairs $(\mathrm{x}_1, d_1), \ldots, (\mathrm{x}_N, d_N)$, where each x_i represents the exploration data at a given location, and $d_i = 1$ or 0 depending on whether or not a deposit occurs at that location. Then the model can be fitted by choosing

w to minimize the empirical error $(1/N)\Sigma_{i=1}^{N} |d_i - y_i|^2$ where $y_i = f(\mathrm{x}_i, \mathrm{w})$.

Capacity control

Minimization of the empirical error, however, provides no guarantee of low error outside the training set, unless some limit is set to the complexity of the model. Indeed, placing bounds on the expected error, in terms of the capacity of the class of approximating functions, is a central concern of learning theory (Vapnik, 1998; Poggio and Smale, 2003). Over-fitting the training set can lead to poor generalization on a test set. This is a special danger in the exploration case, where there may be few known deposits, but a large number of data inputs (van der Baan and Jutten, 2000). It is essential for an analysis to take full account of this. The solution adopted in the case study below, using Bayesian regularization, is described in the Appendix.

Other methods

Various nonprobabilistic methods have been proposed for computer-based prospectivity mapping, including boolean and fuzzy logic methods (Bonham-Carter, 1994). For a useful review and analysis of some of the general issues distinguishing these approaches, see Groves et al. (2000), which includes a case study using the specially devised vectorial fuzzy logic approach of Knox-Robinson (2000) offering a continuous measure of prospectivity and an associated measure of confidence.

The remaining sections of this paper, however, will concentrate on and demonstrate the benefits of, the fully statistical approach provided by neural networks.

Case Study

The Walker Lane

The Walker Lane shear zone, which straddles the border between Nevada and California in the western United States, has a long history of exploration and mining dating back to the discovery of the famous Comstock lode in the late 1850s (see Figs. 3 and 4). Since this is a mature mining district, it provides a good example and illustration of the data mining procedures which are described in the preceding sections of this paper.

Though perhaps not as well known and certainly not as productive as the neighboring Carlin district, the Walker Lane is notable for its numerous occurrences of volcanic-hosted epithermal gold and silver deposits. Aside from the Comstock lode, which in its day produced nearly 200 Moz of silver as well as 9 Moz of gold, and Round Mountain, which has an estimated gold content of 14 Moz, there are at least 10 other deposits with established contents of over one million ounces of gold (see Fig. 4 and Table 1). If all the smaller (that is less than 1 Moz) deposits are taken into account, then to date approximately 50 Moz of gold have been discovered in the area.

The majority of these gold deposits occur in the Walker Lane shear zone, which is a 100-km-wide, northwest-trending structural corridor extending southeast from Reno toward Las Vegas. This strike-slip system contains a series of deep-seated, right-lateral shears and associated normal faults, which presumably provided the channel ways for magmatic and hydrothermal fluids. The Walker Lane shear zone is thought to be related to the Mojave-Sonora megashear, and may have been active since the Mesozoic (Stewart, 1988).

FIG. 3. Map of the western United States showing (in white) the outline of the Walker Lane study area.

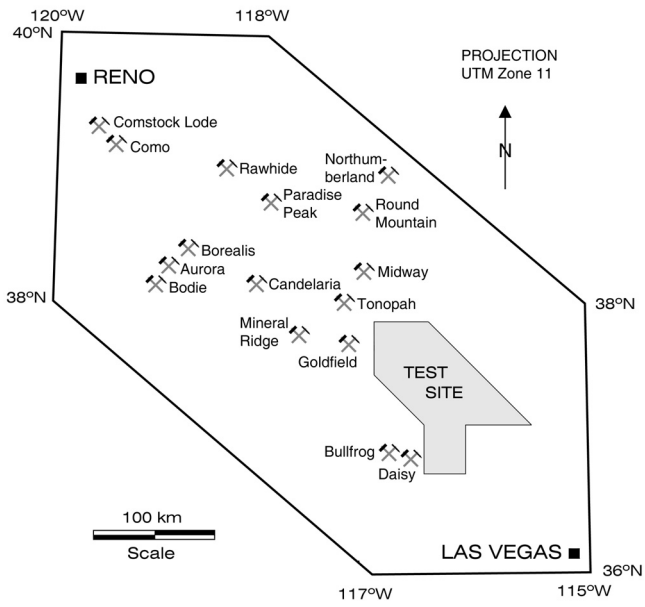

FIG. 4. Locations of known gold deposits within the Walker Lane study area. The area bounded by the hexagon is approximately 100,000 km^2. In addition to the identified deposits, there are about 140 smaller occurrences with recorded production of over 50,000 oz of gold. The area marked in gray contains the Nevada Test Site and other government reserves that are off-limits to mineral exploration.

TABLE 1. Deposits Exceeding 1 Moz of Gold in the Walker Lane

Deposit name	Resource Au (Moz)	Ore grade Au (g/t)	Produced Au (Moz)	Produced Ag (Moz)
Round Mountain	14.1	0.69	6.4	5.1
Comstock Lode	8.6	41.02	8.4	193.0
Northumberland	5.0	3.73	n/a	n/a
Goldfield	4.6	26.63	4.2	1.5
Bullfrog	2.5	8.26	1.9	2.1
Borealis	2.1	0.66	0.6	0.1
Tonopah	2.0	5.61	1.9	174.1
Aurora	1.8	2.33	1.8	20.6
Paradise Peak	1.6	2.07	1.6	24.0
Rawhide	1.5	1.35	1.1	8.9
Bodie	1.5	31.10	1.5	7.3
Daisy	1.5	1.18	n/a	n/a
Totals	46.7	1.56	29.4	436.7

Source: R. Worland, pers. commun., 2003

The Walker Lane is bounded to the southwest by the uplifted Sierra Nevada mountains and to the northeast by the Basin and Range extensional terrain. The rock outcrop in this area is mostly Tertiary volcanics with an assortment of Mesozoic intrusions of mostly monzonitic or granodioritic composition. The basement is made up of strongly folded and thrusted Paleozoic sedimentary rocks. These rocks are exposed in the southeast of the area around Las Vegas, where they contain a high proportion of carbonates.

Exploration data

The two essential requirements for a systematic statistical study of a prospective mineral district are, first, that there be a sufficient number of known deposits to use for a training set and, second, that there be a reasonable collection of modern exploration data sets. The Walker Lane meets both of these requirements. There are many known gold deposits. Also, the area has been the subject of numerous thoughtful investigations by the U.S. Geological Survey, the Nevada Bureau of Mines and Geology, the California Geological Survey, and other groups over the last century. Consequently, there is a huge amount of information accumulated over this region simply waiting to be tapped.

An analytic approach to data integration is obviously highly dependent on the quality and integrity of the exploration data. Great care must be taken in the compilation and registration of the data. A common mistake to watch out for is inadvertently combining data from different map datums. In western Nevada, for example, the difference between the WGS84 and NAD27 datums is about 250 m, which could seriously affect the results if mixed. A GIS can be of great assistance in organizing and visualizing the data, but for statistical representation it is also necessary to have access to the underlying data; for example, to compute derivatives.

The data sets included in this study of the Walker Lane consisted of the regional geology, Landsat thematic mapper (TM), stream sediment geochemistry, isostatic gravity, aeromagnetic and airborne radiometric surveys, digital terrain, and regional structure. The majority of these data sets are in the public domain and can be readily obtained from sources such as the National Geophysical Data Center. The one exception was the regional structure which, as explained below, was based on a proprietary in-house compilation.

The area to be evaluated is roughly 500 × 200 km, or 100,000 km^2. One of the first important decisions concerned the cell size to be used for the data mining exercise. In this case, a cell size of 250 × 250 m was chosen based on the quality and scale of the available data. This was a compromise between the high (28.5 m) resolution of the Landsat data, and the lower resolution of some of the other data sets. In other situations, a larger or smaller cell size might be used.

The known deposits layer is the most critical layer, since it is used for training the network against all the other data sets. In producing this layer, all known gold deposits exceeding 50,000 oz were carefully plotted from air photographs, detailed publications, or field visits with a global positioning system (GPS). About 150 separate deposits were located in this manner. Figure 4 shows the locations of some of the larger deposits in the Walker Lane, marked with traditional crossed hammers. In the data mining study, actual footprints were used wherever possible.

The regional geologic base map used for this study was created from a splice of the geology map of Nevada, drawn at a scale of 1:500,000 (Stewart and Carlson, 1978), and the geology map of California, drawn at a scale of 1:750,000 (Jennings, 1985). These maps are available in digital form from the respective state surveys. Since the Nevada map covered the larger area, both maps were unified to the legend of the Stewart and Carlson map.

The satellite data were picked from the early 1980s to predate any disturbance due to mining in the last 20 years. The data from nine overlapping scenes were carefully brought together to obtain a uniform mosaic of the entire Walker Lane. Then all six high-resolution bands were used in the statistical analysis.

Geochemical data came from the National Uranium Resource Evaluation (NURE) program carried out by the Department of Energy between 1973 and 1984 (Smith, 2001). The data were acquired by several different research groups using different sample and analytical procedures in each of the 1° × 2° quadrangles in the western states. It was therefore necessary first to normalize the data from the 10 quadrangle sheets in the Walker Lane. This was satisfactorily achieved, and a useful set of maps produced for the whole project area.

The elements that were used in this study were not the commonly used pathfinders for gold, such as As and Hg, since the coverage of these elements was incomplete. Instead, the following 11 available elements, Al, Ba, Fe, K, Mg, Mn, Na, Sc, Th, Ti, and V were used. Although some of these may be immobile, the premise was that the arrival of a major gold-bearing system was likely to affect the whole-rock elements in the adjoining country rocks.

The airborne radiometric data also resulted from the NURE surveys. These were wide-spaced, high-level surveys looking for uranium deposits, so the quality of the data was not ideal for gold exploration. A modern, high-resolution

survey would have been far more preferable for mapping lithology and alteration in the young volcanic rocks of the Walker Lane. However, these data were all that were available covering the entire project area.

The regional gravity and magnetic data came from the U.S. Geological Survey and were of reasonably good quality. The magnetic data were assembled from numerous small surveys flown over the years at different line spacings and flying heights, but had been mathematically corrected to a common elevation to produce a uniform map of the Walker Lane. The gravity data also came from many old surveys that were joined together and processed to a complete Bouguer, isostatic residual map. The Bouguer correction reduces the effect of terrain whereas the isostatic correction reduces the effect of mountain roots below areas of high topography.

The digital terrain data were of high quality and were originally assembled to assist in correcting the gravity data. These terrain data, however, also represent local structure in the form of ridge crests and drainage patterns as well as circular features. So these data were included in their own right as a pattern that was input to the neural network analysis.

The final data set that was an input to this study was a map of the regional structure. This was the only proprietary data set (belonging to Newmont), and was produced by an integrated analysis of the geology, Landsat, gravity, and magnetic data. This was not simply a tracing of linear features, but was a geodynamic interpretation which took into consideration regional tectonics such as the thrusting of the Basin and Range, the buttressing effect of the Sierra Nevada batholith, and the strike-slip movements of the San Andreas fault system (B. Davies, pers. commun., 2003).

All these data sets have their own strengths and weaknesses. The geologists may feel that the statewide geology is too broad brush and contains mapping errors; the geophysicists may be concerned about missing data due to the variable station interval of the gravity surveys; the geochemists may worry about the different sample and analytic procedures used to collect the stream sediment data. These are valid concerns. Certainly, the better the quality of the input data the better the results that will be achieved by an analytic process. Nonetheless, these are typical regional exploration data sets that we have learned to deal with in our normal manual interpretations. Furthermore, statistical correlations between multiple data sets provide some redundancy in overall information content, so that local deficiencies of an individual data set are compensated by the combined force of the others.

Data representation

The data just described amounted to 25 primary layers, which is not unusual for a modern exploration data set. Of these, six are Landsat thematic mapper, 11 are geochemistry, and three are radiometric. However, there is little meaning in a single-point geophysical, or for that matter, topographical reading. There is no good reason, for example, that all the gold should occur at a single elevation, such as 3,000 m. The same logic applies to a tranche of gravity readings. It is the pattern or neighborhood around a given station that is important. It may be, for example, that flanks of gravity highs have statistical significance. But it is not necessary to second-guess the neural network by assuming this, only to ensure that the network has the necessary information available to make its own decisions.

Such patterns can be represented by taking the derivatives and integrals of the primary data—for example, gravity—and inputting these as well to the neural network. This is akin to a Taylor Series expansion, with a few extra terms incorporated. For completeness, it is necessary to include both first- and second-order terms, and also to take the horizontal, vertical, and cross derivatives of the data. This results in about a dozen extra secondary layers derived from each primary layer.

The geology also required special treatment. In the Walker Lane, after splicing the Nevada and California geologic maps, there were 69 formations that needed to be represented, resulting in 69 separate inputs to the neural network. However, it was also considered that neighboring formations and geologic contacts were important aspects of the geology. A moving window was therefore used to capture the geologic patterns around each station, expressed by the proportions of the different formations occurring within the window. After various empirical trials, a 5×5 km^2 box was adopted for the Walker Lane.

Regional structure was another special case. The buffer zone procedures commonly used in GIS manipulations were considered inadequate for input to a network. Instead, a geophysical scheme was adopted, in which the faults were assigned arbitrary densities and their gravitational fields were then calculated and summed over the survey area. Since no attitude information had been provided, all the interpreted structures in the Walker Lane were treated as vertical. This process resulted in smoothly varying maps which neatly showed intersections as well as the simple $(1/r)$ fall-off from individual structures. Furthermore, these continuous patterns could be represented by derivatives and integrals, just like any other potential field measurement.

Due to this preprocessing of the data, the number of inputs to the network grows rapidly to exceed the number of primary data layers. In the Walker Lane study, for example, the number of inputs used to accommodate all the data layers was 139. This required the application of specialized regularization procedures, described in the preceding section on probabilistic modeling, and optimization procedures such as those of Williams (1991).

Neural network result

Before embarking on a full integration of the data, it is instructive to examine individual data sets to ensure that the statistical process is working properly and that the results intuitively make sense. This also helps to determine which data sets are contributing most to the final target map.

Figure 5 shows the target favorability map based on the geology layer alone. As mentioned, the geologic map of the

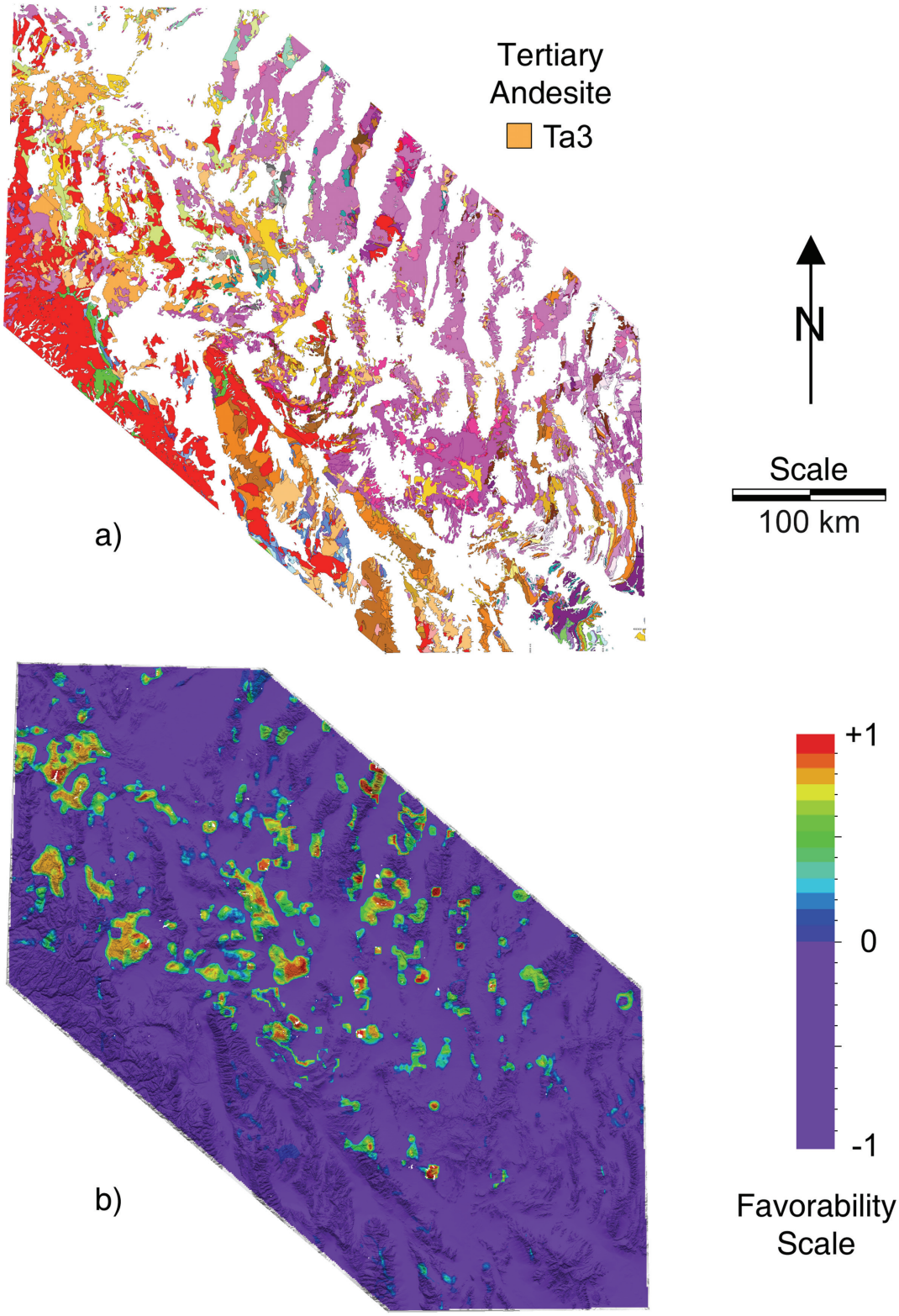

FIG. 5. a) Geologic map of the Walker Lane, in which the post-Cenozoic cover rocks have been left uncolored. b) Target favorability map based on the geology layer alone. The targets are superimposed on a shaded-relief image of the topography, with known deposit footprints marked in white. Note the correlation between the Tertiary andesite *Ta3* and several high-favorability (red-colored) targets.

Walker Lane, shown in the upper half of this figure, contains 69 formations which were input to the neural network. This tally does not include the post-Cenozoic cover rocks, which postdate the gold mineralization and occupy most of the valleys. However, these uncolored areas are represented to the network by default.

The target favorability map shown in the lower half of this figure has been superimposed on a shaded-relief image of the topography, so that the basins and ranges stand out. In order to highlight the better scores, areas with positive favorabilities have been assigned the warmer colors ranging upward from blue to red. Areas with negative favorabilities have been assigned the color magenta. These are areas where the probability of success is estimated to be worse than the prior, or dart-throwing, probability.

It can be seen that, on the basis of geology alone, only a small proportion of the area is considered favorable. The network has clearly recognized, for example, that neither the Sierra Nevada batholith nor the Paleozoic platform sedimentary rocks are good places to look for volcanic-hosted gold deposits. This may be obvious to the geologist, who would no doubt instinctively mask out these areas, but the network had to glean this information purely from the training set.

A quick comparison with the geology map shows that many of the better targets correlate with a formation mapped as *Ta3*. This is a Tertiary andesite that is well known to Nevada geologists for hosting many of the larger volcanic-hosted gold deposits in the Great Basin (see, for example, Mihalasky, 1999). Closer examination at an enlarged scale reveals other formations and geologic contacts in which the majority of gold deposits occur.

All the various data sets, which are too numerous to present here, showed similar but not necessarily coincident target patterns. As one might expect, the Landsat-only target map showed a strong correlation with color TM anomalies, which are commonly associated with high-sulfidation systems such as Goldfield, Nevada. The geochemistry produced a sharply focused target map, which was not intuitively obvious to explain as it was based on a multivariate combination of 11 elements. However, examination of the individual elements revealed familiar patterns such as Na-depletion and K-enrichment in the vicinity of the known gold deposits.

The results become more interesting and revealing when the data sets are combined. Since the Walker Lane is a competitive exploration area, the final, overall target map was deemed too sensitive to be shown in this publication. However, part of the survey area can be shown that falls inside the Nevada Test Site. Not surprisingly, after four decades of nuclear weapons testing, this part of the Walker Lane is permanently off-limits to gold mining.

Figure 6 shows the results that were obtained from an area that is located about 50 km southeast of the historic mining district of Goldfield and covers about 2,500 km^2 in the northwest corner of the Test Site. The target favorability map is based on all the exploration data sets described above except the NURE geochemistry, which were not collected in the Test Site. The color bar for this map is the same as that shown in Figure 5. The Landsat data and the geology are shown at the same scale for direct comparison to the target map in the center.

The data mining exercise produced two interesting targets, which would both certainly be followed up if access was permitted to this area. Target A can be seen to be coincident with a color anomaly in the TM data, indicating the presence of alteration associated with mineralization. Target B coincides with a circular feature that is visible in both the TM and the geology. Further interest is added by the small, orange-colored formation which crops out at the center of this probable caldera. This is the familiar *Ta3*, the Tertiary andesite referred to above. Remember, however, that these targets are based on all the available exploration data—for example, the gravity, magnetic data, and structure—and not just the TM and geology.

These two favorability targets are sufficiently small to be checked out in a matter of days by a preliminary field inspection and grab rock sampling to determine if further exploration is warranted. At this stage, detailed geologic mapping, geophysical surveying, and geochemical sampling would normally be carried out to refine the targets for drill testing.

It is noteworthy that, together, these two targets cover less than 2.5 percent of the area shown (based on a favorability threshold of +0.5). This percentage is representative of the combined-data target map produced for the Walker Lane as a whole. This considerably reduces the area to be explored and, as is shown in the next section, significantly improves the probability of success.

Economic Analysis

The results of the statistical analysis can also be considered from the economic point of view. This involves analyzing the potential costs and benefits of an exploration program, and the added value of accurate targeting (Mackenzie and Woodall, 1987; Green, 2004).

Probability of success

Suppose that potential targets have been ranked according to a scoring function S, and that an exploration program consists of investigating N targets scoring higher than some threshold s. First, what is the probability that any one of these, individually, hosts an economic deposit?

The answer depends on the value of s and on how well S discriminates. Ideally, S should sharply separate deposits from barren regions though, in practice, some overlap is likely. Figure 7 shows the distribution of neural network scores for the Walker Lane study. The right-hand distribution includes only known gold deposits. The left-hand distribution includes any as yet undiscovered deposits, which are expected to be located in its right tail. The left tail of the right-hand distribution comprises known deposits which the network considered to be untypical.

The distributions in Figure 7 yield an estimate of the probability $P(D|S > s)$ that a location in the area to be explored, scoring higher than s, hosts a deposit. This can be expressed, by Bayes' rule, as

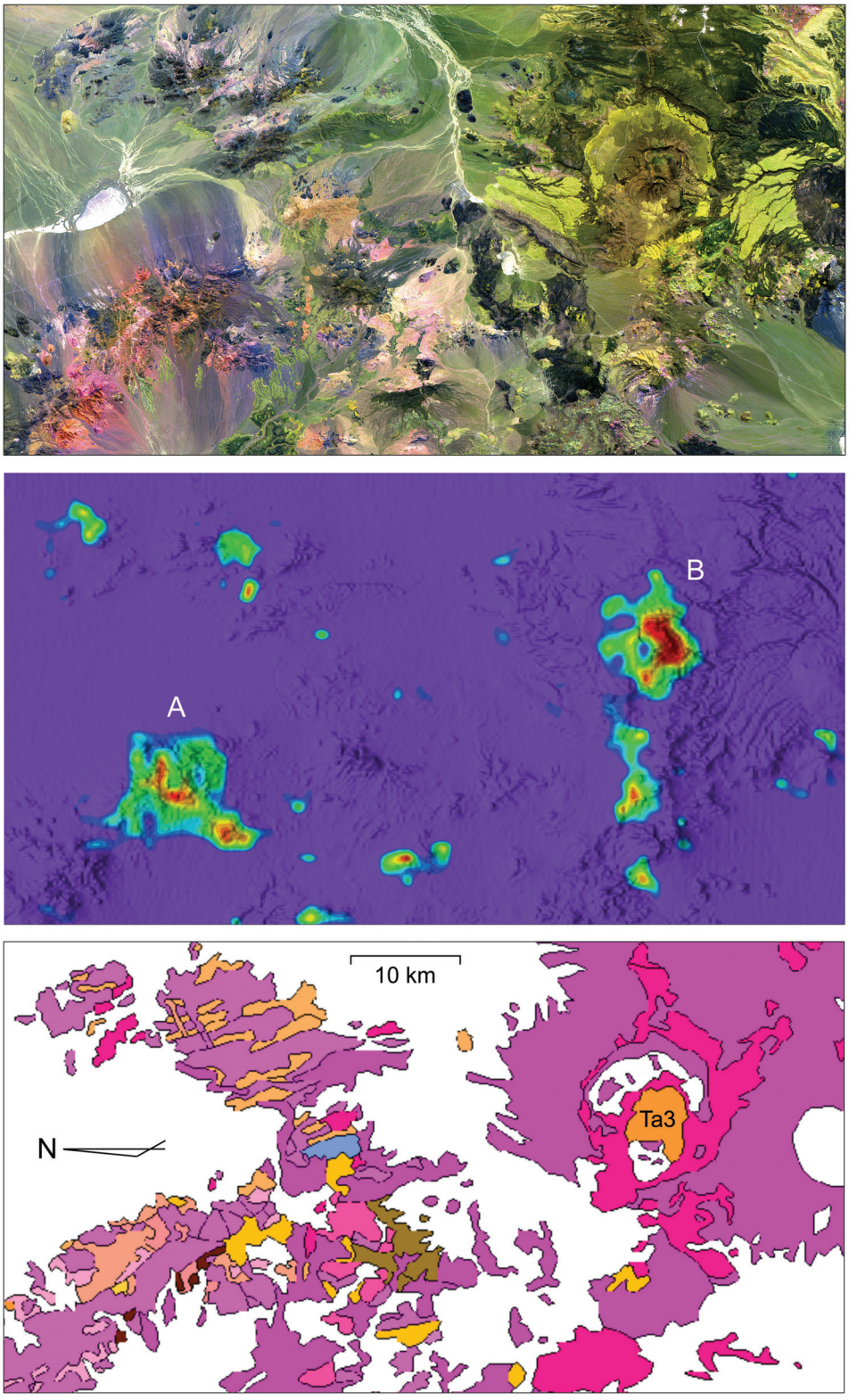

FIG. 6. A close-up of part of the Nevada Test Site, an off-limits area within the Walker Lane. The target favorability map in the middle is compared to the TM image (bands 571) above and the geology below. The targets are superimposed on a shaded-relief image of the topography. Note the circular feature in both the TM and geology, and the outcrop of Tertiary andesite Ta3, in the vicinity of target B.

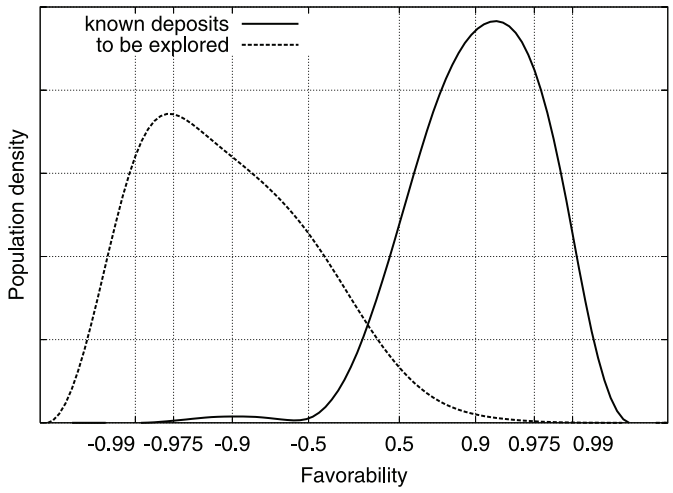

FIG. 7. Figure showing, on the right, the density of the distribution of favorability scores over the known gold deposits and, on the left, the density over the remainder of the region.

$$P(D|S > s) = \frac{P(S > s|D)}{P(S > s)} P(D), \quad (1)$$

where $P(D)$ is again the probability of locating a deposit in the area to be explored by chance. The factor on the right, multiplying the *prior* probability $P(D)$, is called the *improvement factor*. This is the factor by which the probability of finding a deposit is increased by targeting only high scoring locations, to yield the *posterior* probability $P(D|S > s)$.

To calculate the improvement factor, both its numerator and denominator are needed from equation (1). The denominator $P(S > s)$ relates to the distribution of all scores in the area to be explored, so that its value is just the area to the right of s under the left-hand curve in Figure 7. The numerator $P(S > s|D)$ relates to the distribution of scores over the population of deposits in the area to be explored. Unfortunately, by definition, this distribution is unknown. But if it is assumed that the known deposits are representative of the unknown de-posits, the right-hand curve in Figure 7 can be used to estimate the distribution of scores over unknown deposits, using the known distribution of scores over known deposits. This is the area to the right of s under the right-hand curve. The improvement factor, as a function of the threshold score s, can therefore be estimated as the ratio of the areas to the right of s under the two curves in Figure 7. This is shown in Figure 8.

The goal of exploration is to discover economic deposits. In the Walker Lane study, the network was trained to locate gold deposits with resources greater than 50,000 oz. Only a proportion of these will be economic. To calculate the probability that a high-scoring location hosts an economic deposit, it is necessary to multiply the probability that it is a deposit, by the probability that a high scoring deposit is economic. If E expresses the occurrence of an economic deposit, equation (1) implies that $P(E|S > s)$ is given by

$$P(E|D \,\&\, S > s) \frac{P(S > s|D)}{P(S > s)} P(D). \quad (2)$$

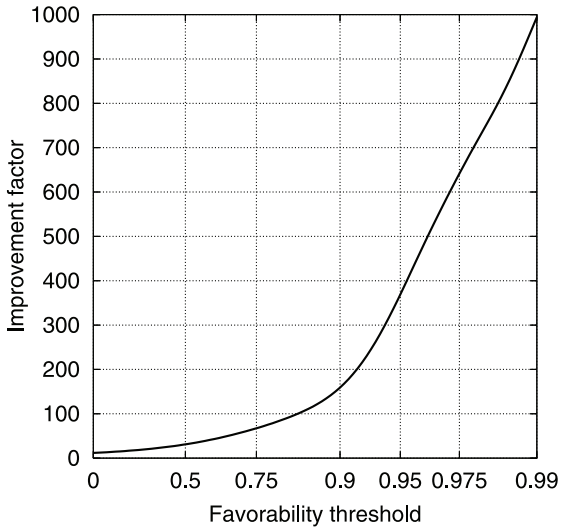

FIG. 8. Figure showing the factor by which the prior odds are multiplied, to yield the posterior odds, for given values of the favorability threshold.

This means that the probability of being an economic deposit is the product of three factors: (1) the probability that a high-scoring deposit is economic, (2) the improvement factor, and (3) the prior probability. As an example, suppose the prior probability is 1/2500 and the conditional probability of being economic is 1/2, so that the product of the first and third factors is 1/5000. If a target is selected at random from those scoring above 0.95, Figure 8 shows that the improvement factor is approximately 375, so that the probability that it is an economic deposit is approximately 375/5000 = 0.075, or better than 1 in 14.

Multiple targets

An exploration program will normally target more than one location. The probability that one or other of these hosts an economic deposit is necessarily greater than the probability that any one does individually. Suppose that N locations, scoring higher than s, are targeted and that these are investigated, in some order, until either an economic deposit is found or the targets are exhausted. If the probability that any individual location hosts an economic deposit is $p = P(E|S > s)$ and the targets are independent, the overall probability of success is

$$p_N = 1 - (1 - p)^N, \quad (3)$$

and the expected number of targets to be investigated is p_N/p. The graph of the probability of success p_N is shown, as a function of N, for various values of p in Figure 9. The example at the end of the last section gave a value of p = 0.075 for a threshold set at s = 0.95.

The largest probability of success would seem to be achieved by setting the threshold s as high as possible, since this will increase the corresponding value of p. But if s is too high, there will be no targets. Estimating the number of targets for a given threshold involves some decisions. For the

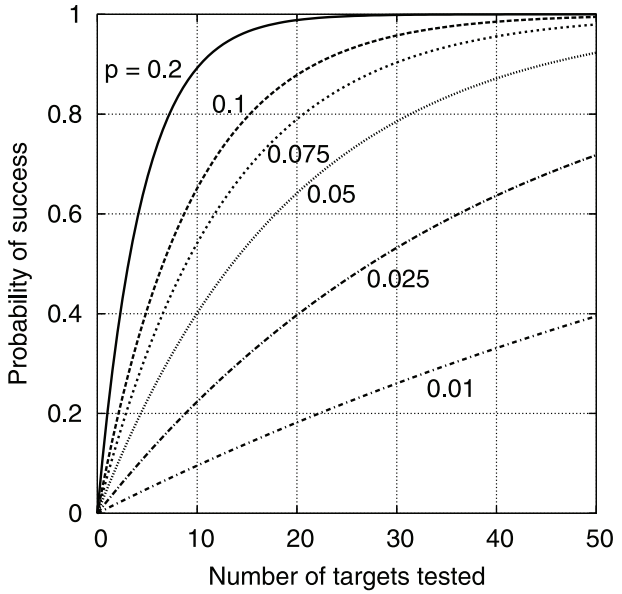

FIG. 9. Figure showing the probability of success when a given number of targets is tested for various values of the probability p of success for an individual target.

present analysis, target locations must be sufficiently separated to ensure the validity of equation (3), which requires the independence of success or failure at all locations. Table 2 shows the number of targets for the Walker Lane as a whole, for various values of the threshold s, and various minimum target separations. For example, the 2-km column shows the number of available targets, on the assumption that targets are required to be separated by at least 2 km, both from each other and from known deposits. This shows that if the region as a whole is to be targeted, an exploration program is more likely to be limited by budget than by the number of targets available, provided the threshold is no greater than about 0.975. In practice, targets are likely to be selected using further information in addition to individual favorability scores. This analysis nonetheless provides an indication of the level of probability to be expected.

Costs and benefits

The analysis can be taken a step further if estimates of exploration costs are available. Suppose there is a cost C_D to determine whether a target location hosts a deposit and, if it does, a further cost C_E to determine whether the deposit is economic. Suppose that the expected value of a discovery is V.

TABLE 2. Table Showing the Number of Target Locations for Various Values of the Favorability Threshold vs. Various Values of the Minimum Target Separation

	Number of targets		
Threshold	1 km	2 km	3 km
0.900	720	320	195
0.950	385	142	82
0.975	140	45	29
0.990	24	10	4

The expected cost of investigating an individual target is $C_D + q\,C_E$, where $q = P(D|S > s)$ is the probability that a sufficiently high scoring target hosts a deposit. This must be multiplied by the expected number of targets that will be investigated, namely p_N / p, to obtain the expected cost of the program. The expected benefit is the expected value of an economic deposit multiplied by the probability of finding it, namely $p_N V$; assuming that up to N targets are to be investigated. It follows that the expected benefit exceeds the expected cost if

$$p > \frac{C_D + q\,C_E}{V} \tag{4}$$

and this is independent of N. This means that the probability that an individual target is an economic deposit must exceed the ratio of the expected cost of investigating it to the expected value of a discovery. In the Walker Lane study it was found that p is better than 1 in 14 for a threshold of 0.95. This is likely to be considerably in excess of the ratio on the right-hand side of equation (4).

There is no difficulty in assessing, in a similar way, the risk of such an exploration program, measured in terms of the expected variation in the gain. But that would go beyond the limits of the paper.

Conclusions

The sheer volume of modern data sets, which simply keep growing with time, dictates a different approach to mineral exploration. Fortunately, new data mining techniques are becoming available to meet this challenge. This paper has reviewed two separate approaches, visualization and probabilistic modeling, which are claimed to have the capability of reversing the parallel trends of falling discovery rates and rising discovery costs that have been observed during the last two decades in the mineral industry.

These new statistical techniques are illustrated by a case study from Nevada. Given 25 layers of exploration data, the neural network procedure produces a highly focused favorability map. The most convincing demonstration that these techniques are effective would of course be a major discovery. The targets are now being carefully followed up, so time will tell if this is the case.

The probability of success in discovering an economic deposit has also been examined by the use of these techniques. Theoretically, it has been shown that a systematic integration of all the exploration data in a given area significantly improves the chances of making a discovery. In the Walker Lane gold district, for example, the prior probability (that is without the use of any exploration technology) was assumed to be 1/2500. The improvement resulting from the neural network application was a factor of 375. This leads to a posterior probability of success of 0.075 for a single target, or 0.8 if 20 targets are tested.

The case study covered a large (100,000 km^2) prospective mineral district. At this scale, the third dimension, depth, is insignificant, so this was essentially a two-dimensional study.

Since the neural networks were not given any information about location and were simply given the patterns associated with the training sites, it would be a straightforward step to extend the method to a three-dimensional situation. The technique should also work at any scale, wherever there is a large volume of data and sufficient knowledge to provide a good training set. In other words, these methods could equally well be used for underground target delineation around a known deposit as for regional exploration for a new deposit.

Although useful in two dimensions, visualization techniques become increasingly important in three-dimensional situations, both for understanding the data and for communicating interpretations to other workers. The authors foresee probabilistic modeling coming together with data fusion as the next important landmark in exploration technology, and fully expect that the combined application of these data mining techniques will reduce the overall cost of exploration and lead to more economic discoveries in the years ahead.

Acknowledgments

The authors would like to dedicate this paper to the memory of David Marr, their close school friend and Cambridge classmate. Among other achievements of a brilliant academic career, sadly cut short by leukemia, D. C. Marr laid the foundations for many of the visualization and neural network techniques used today.

The authors would also like to thank Newmont Mining Corporation for supporting the early development of the methods described here and for permission to show some of the Walker Lane results.

The manuscript benefited from the careful editing and thoughtful suggestions of Jeffrey Hedenquist, John Parry, Michael Doggett, David I. Groves, and one anonymous reviewer.

REFERENCES

Bishop, C.M., 1995a, Neural networks for pattern recognition: Oxford, Oxford University Press, 482 p.

——1995b, Training with noise is equivalent to Tikhonov regularization: Neural Computation, v. 7(1), p. 108–116.

Bonham-Carter, G.F., 1994, Geographic information systems for geoscientists: Modelling with GIS: Tarrytown, New York, Pergamon Press-Elsevier Science Publications, 398 p.

Bougrain, L., Gonzalez, M., Bouchot, V., Cassard, D., Lips, A.L.W., Alexandre, F., and Stein, G., 2003, Knowledge recovery for continental-scale mineral exploration by neural networks: Natural Resources Research, v. 12(3), p. 173–181.

Brown, W.M., Gedeon, T.D., Groves, D.I., and Barnes, R.G., 2000, Artificial neural networks: A new method for mineral prospectivity mapping: Australian Journal of Earth Sciences, v. 47, p. 757–770.

Brown, W.M., Gedeon, T.D., and Groves, D.I., 2003, Use of noise to augment training data: A neural network method of mineral-potential mapping in regions of limited known deposit examples: Natural Resources Research, v. 12(2), p. 141–151.

Buntine, W.L., and Weigend, A.S., 1991, Bayesian back-propagation: Complex Systems, v. 5, p. 603–643.

Chapelle, O., and Vapnik, V., 2000, Model selection for support vector machines, *in* Solla, S. A., Leen, T. K., and Müller, K. B., eds., Advances in Neural Information Processing Systems: Cambridge, Massachusetts, MIT Press, v. 12, p. 230–236.

Cristianini, N., and Shawe-Taylor, J., 2000, An introduction to support vector machines: Cambridge, Cambridge University Press, 279 p.

Garrett, S., Griesbach, S., Johnson, D., Jones, D., Lo, M., Orr, W., and Sword, C., 1997, Earth model synthesis: First Break, v. 15(1), p. 13–22.

Gibbs, M.N., and MacKay, D.J.C., 2000, Variational Gaussian process classifiers: IEEE Transactions on Neural Networks, v. 11(6), p. 1458–1464.

Gold, C. and Sollich, P., 2003, Model selection for support vector machine classification: Neurocomputing, v. 55, p. 221–249.

Good, I.J., 1950, Probability and the weighing of evidence: London, Charles Griffin & Co. Ltd, 119 p.

Green, A., 2004, Exploration, risk and the value of geophysics: Australian Society of Exploration Geophysicists, 17th Geophysical Conference and Exhibition, Sydney, NSW, August 15–19, 2004, Extended Abstracts, 5 p.

Groves, D.I., Goldfarb, R.J., Knox-Robinson, C.M., Ojala, J., Gardoll, S., Yun, G.Y., and Holyland, P., 2000, Late kinematic timing of orogenic gold deposits and significance for computer-based exploration techniques with emphasis on the Yilgarn block, Western Australia: Ore Geology Reviews, v. 17, p. 1–38.

Harris, D., and Pan, G., 1999, Mineral favorability mapping: A comparison of artificial neural networks, logistic regression, and discriminant analysis: Natural Resources Research, v. 8(2), p. 93–109.

Harris, D., Zurcher, L., Stanley, M., Marlow, J., and Pan, G., 2003, A comparative analysis of favorability mappings by weights of evidence, probabilistic neural networks, discriminant analysis, and logistic regression: Natural Resources Research, v. 12(4), p. 241–255.

Hyvärinen, A., Karhunen, J., and Oja, E., 2001, Independent component analysis: New York, John Wiley, 504 p.

Jennings, C.W., 1985, An explanatory text to accompany the 1:750,000 scale fault and geologic maps of California: California Division of Mines and Geology Bulletin 201, Appendix D, p. 125–197.

Kemp, L.D., Bonham-Carter, G.F., Raines, G.L., and Looney, C.G., 2001, Arc-SDM: Arcview extension for spatial data modelling using weights of evidence, logistic regression, fuzzy logic and neural network analysis: http://ntserv.gis.nrcan.gc.ca/sdm/.

Knox-Robinson, C.M., 2000, Vectorial fuzzy logic: A novel technique for enhanced mineral prospectivity mapping, with reference to the orogenic gold mineralisation potential of the Kalgoorlie terrane, Western Australia: Australian Journal of Earth Sciences, v. 47(5), p. 929–941.

Krogh, A., and Vedelsby, J., 1995, Neural network ensembles, cross validation, and active learning, *in* Tesauro, G., Touretzky, D., and Leen, T., eds., Advances in neural information processing systems: Cambridge, Massachusetts, MIT Press, v. 7, p. 231–238.

Li, Y., and Oldenburg, D. W., 1996, 3-D inversion of magnetic data: Geophysics, v. 61(2), pp. 394–408.

——1998, 3-D inversion of gravity data: Geophysics, v. 63, p. 109–119.

MacKay, D.J.C., 1992, A practical Bayesian framework for backpropagation networks: Neural Computation, v. 4(3), p. 448–472.

——2003, Information theory, inference, and learning algorithms: Cambridge, Cambridge University Press, 640 p.

Mackenzie, B., and Woodall, R., 1987, Mineral exploration economics: The search for base metals in Australia and Canada: Kingston, Ontario, Queen's University, Centre for Resource Studies Working Paper 40, 110 p.

McGaughey, W.J., and Vallee, M.A., 1998, Integrating geology and borehole geophysics in a common earth model for improved three-dimensional delineation of mineral deposits: Exploration and Mining Geology, v. 7, p. 51–62.

Mihalasky, M.J., 1999, Mineral potential modelling of gold and silver mineralization in the Nevada Great Basin—a GIS-based analysis using weights of evidence: Unpublished Ph.D. dissertation, University of Ottawa, 360 p. Also available as USGS Open-File Report 01-291.

Paterson, N. R., 2003, Geophysical developments and mine discoveries in the 20th century: The Leading Edge, v. 22(6), p. 558–561.

Poggio, T., and Smale, S., 2003, The mathematics of learning: Dealing with data: Notices of the American Mathematical Society, v. 50(5), p. 537–544.

Porwal, A., Carranza, E.J.M., and Hale, M., 2003, Artificial neural networks for mineral-potential mapping: A case study from Aravalli province, western India: Natural Resources Research, v. 12(3), p. 155–171.

Poulton, M.M., Sternberg, B.K., and Glass, C.E., 1992, Location of subsurface targets in geophysical data using neural networks: Geophysics, v. 57(12), p. 1534–1544.

Raines, G.L., 1999, Evaluation of weights of evidence to predict epithermal-gold deposits in the Great Basin of the western United States: Natural Resources Research, v. 8(4), p. 257–276.

Ripley, B.D., 1996, Pattern recognition and neural networks: Cambridge, Cambridge University Press, 403 p.
Schodde, R.C., 2004, Discovery performance of the western world gold industry over the period 1985–2003: PACRIM 2004, Proceedings of the 2004 Adelaide Congress of the Australian Institute of Mining and Metallurgy, Adelaide, SA, September 19–22, 2004, 14 p.
Schölkopf, B. and Smola, A.J., 2002, Learning with kernels: Support vector machines, regularization, optimization and beyond: Cambridge, Massachusetts, MIT Press, 644 p.
Singer, D.A., and Kouda R., 1997, Use of a neural network to integrate geoscience information in the classification of mineral deposits and occurrences, *in* Gubins, A.G., ed., Proceedings of Exploration 97: Prospectors and Developers Association of Canada, Fourth Decennial International Conference on Mineral Exploration, Toronto, September 14–18, 1997, p. 127–134.
——1999, A comparison of the weights-of-evidence method and probabilistic neural networks: Natural Resources Research, v. 8(4), pp. 287–298.
Smith, S.M., 2001, National Geochemical Database: Reformatted data from the National Uranium Resource Evaluation (NURE) Hydrogeochemical and Stream Sediment Reconnaissance (HSSR) Program, Version 1.30: U.S. Geological Survey Open-File Report 97-492, www release only, http://greenwood.cr.usgs.gov/pub/open-file-reports/ofr-97-0492/index.html.
Sollich, P., 2002, Bayesian methods for support vector machines: Evidence and predictive class probabilities: Machine Learning, v. 46, p. 21–52.
Stark, T.J., Dorn, G.A., Cole, M.J., 2000, ARCO and immersive environments, Part 2: Oil industry experience with immersive environments: The Leading Edge, v. 19(8), p. 884–890.
Stewart, J.H., 1988, Tectonics of the Walker Lane belt, Western Great Basin: Mesozoic and Cenozoic deformation in a zone of shear, *in* Ernst, W.G., ed., Metamorphism and Crustal Evolution of the Western United States: Prentice Hall, Englewood Cliffs, New Jersey, Rubey Volume 7, p. 683–713.
Stewart, J.H., and Carlson, J.E., 1978, Geologic map of Nevada: U.S. Geological Survey and Nevada Bureau of Mines and Geology Map, scale 1:500,000.
Stone, M., 1974, Cross-validatory choice and assessment of statistical predictions: Journal of the Royal Statistical Society, B, v. 36(1), p. 111–147.
Treinish, L.A., 2001, How can we build more effective weather visualizations? Proceedings of the Eighth European Centre for Medium-Range Weather Forecasting (ECMWF) Workshop on Meteorological Operational Systems, Reading, England, November 12–16, 2001, p. 90–99.
van der Baan, M. and Jutten, C., 2000, Neural networks in geophysical applications: Geophysics, v. 65(4), p. 1032–1047.
Vapnik, V., 1998, Statistical learning theory: New York, John Wiley, 768 p.
Williams, C.K.I., and Barber, D., 1998, Bayesian classification with Gaussian processes: IEEE Transactions on Pattern Analysis and Machine Intelligence, v. 20(12), p. 1342–1351.
Williams, P.M., 1991, A Marquardt algorithm for choosing the step-size in backpropagation learning with conjugate gradients: University of Sussex, Cognitive Science Research Paper CSRP 229, 8 p.
——1995, Bayesian regularization and pruning using a Laplace prior: Neural Computation, v. 7(1), p. 117–143.
——1999, Matrix logarithm parametrizations for neural network covariance models: Neural Networks, v. 12(2), p. 299–308.
——2002, Probabilistic learning models, *in* Corfield, D., and Williamson, J., eds., Foundations of Bayesianism, Dordrecht, Kluwer, p. 117–134.
Williams, P.M., Wu, S., and Feng, J., 2005, Improving the performance of the support vector machine: two geometrical scaling methods, *in* Wang, L., ed., Support vector machines: Theory and applications: Springer, Berlin, p. 205–218.

APPENDIX

This appendix provides further technical details of issues raised in the section on Probabilistic Modeling.

Weights of evidence

The aim of a probabilistic analysis is to estimate the conditional probability that a deposit occurs at a location exhibiting exploration features x, namely $P(D|\mathrm{x})$. The weights of evidence approach calculates $P(D|\mathrm{x})$ by using Bayes' rule to express

$$P(D|\mathrm{x}) \propto P(\mathrm{x}|D)P(D), \qquad \text{(A1)}$$

where the constant of proportionality depends on x, but not on D. $P(D)$ is the prior probability of a deposit, which can be thought of as the probability that a dart thrown randomly at the map would strike the interior of a deposit outline. $P(\mathrm{x}|D)$ is the likelihood of observing the features x in the presence of a deposit.

The reason for the inversion in equation (A1) is that, under suitable assumptions, the right-hand side, together with the normalizing constant, can be estimated directly from the data. The assumptions are that (1) exploration data are conditionally independent and (2) each component of the feature vector x takes only a small finite number of values.

The limitations of the approach are that exploration data rarely satisfy either of these assumptions. Geochemical data, for example, can show high correlations between elements, especially in the vicinity of a deposit. Dependencies between other data sets may be more subtle, but a complete analysis should aim to exploit, rather than avoid (or ignore) them. Data can be preprocessed using principal components, but this leads to independent components only for Gaussian data. For non-Gaussian data, independent component analysis (Hyvärinen et al., 2001) may be more promising. But when more advanced, especially nonlinear, preprocessing techniques are used, the approach loses its simplicity and distinctiveness. The second requirement is also a handicap. Most exploration data sets relate to continuous quantities, and any finite partitioning of their ranges entails a loss of information. Further discussion and comparisons with other methods can be found in Singer and Kouda (1999) and Harris et al. (2003).

Neural networks

A feed-forward neural network in the form of a multilayer perceptron consists of a sequence of linear transformations, interleaved with coordinatewise nonlinear transformations of the sigmoid type. In the absence of such nonlinearities, the input-output mapping defined by the network reduces, by composition, to a single linear transformation. In their presence, a neural network has the universal approximation property, that is, the capability of approximating an arbitrary continuous function uniformly on compact sets (Bishop, 1995a; Ripley, 1996).

If there are n input units and a single output unit, the network implements a mapping $y = f(\mathrm{x})$, where y is a scalar output and x is an n-dimensional input vector. The actual

mapping depends on how the hidden units are connected and the values of the weights attached to each connection. Once the underlying architecture is fixed, the function computed by the network depends only on the weights w, so that y can be written as $y = f(\mathrm{x}, \mathrm{w})$.

Error functions

The aim is to choose w so that $f(\mathrm{x}, \mathrm{w})$ is a good approximation to the conditional probability $P(D|\mathrm{x})$, for realistic values of the exploration data x. For sufficiently large networks, this will be true if w is chosen to minimize the expected value of $|D - f(\mathrm{x}, \mathrm{w})|^2$. Calculating this expectation directly, however, requires knowledge of the joint distribution of x and D, which amounts to already knowing $P(D|\mathrm{x})$. The normal approach therefore is to minimize the empirical estimate of this expectation, taken over a training set for which the values of x and D are known. Suppose, therefore, that we have a set of N known pairs (x_1, d_1), $\ldots,(\mathrm{x}_N, d_N)$, where each x_i represents the exploration data at a given location, and $d_i = 1$ or 0 depending on whether or not a deposit occurs at that location. Then w can be chosen to minimize the empirical error $(1/N)\Sigma_{i=1}^{N} |d_i - y_i|^2$, where $y_i = f(\mathrm{x}_i, \mathrm{w})$.

As an alternative to the sum-of-squares error, $P(D|\mathrm{x})$ can also be characterized as the function $p(\mathrm{x})$, with values lying between 0 and 1, which minimizes the expected value of $-\log(1 - |D - p(\mathrm{x})|)$. This is the approach adopted here. It is convenient then to apply the logistic sigmoid $g(y) = e^y/(1 + e^y)$ to the network output $y = f(\mathrm{x}, \mathrm{w})$ so that $g(y)$ necessarily lies between 0 and 1. This leads to the empirical error function

$$E(\mathrm{w}) = -\frac{1}{N} \sum_{i=1}^{N} \log(1 - |d_i - p_i|), \qquad \text{(A2)}$$

where $p_i = g(f(\mathrm{x}_i, \mathrm{w}))$. The terms in the summation are $\log p_i$ when $d_i = 1$ and $\log(1 - p_i)$ when $d_i = 0$.

The validity of this approach, as it stands, depends on sampling the population at random. In the exploration case, this would lead to a substantial preponderance of negative instances, since the prior probability of a deposit is usually low. In practice it is better to train the network using equal numbers of positive and negative instances. A correction must then be made to the output probability. It can be shown that, if the network is trained with equal numbers of positive and negative instances, the estimate of the true posterior probability $P(D|\mathrm{x})$ can be obtained from the network output $f(\mathrm{x})$ by means of the logarithmic odds relation,

$$\log \frac{P(D|\mathrm{x})}{P(\overline{D}|\mathrm{x})} = f(\mathrm{x}) + \log \frac{P(D)}{P(\overline{D})}, \qquad \text{(A3)}$$

where $\overline{D} = 1 - D$. This implies that the network output $f(\mathrm{x})$, before applying the logistic sigmoid, is the weight of evidence (Good, 1950) determined by the sum total of data x. Multiplication by its exponential transforms prior odds to posterior odds.

In practice, it is not necessary to exhibit $P(D|\mathrm{x})$ in order to indicate target priorities. Since $P(D|\mathrm{x})$ is an increasing function of $f(\mathrm{x})$, whatever the value of $P(D)$, the network output $f(\mathrm{x})$ itself or any increasing function of $f(\mathrm{x})$ will serve as a discriminant function. In the case study below the value of $\tanh(f(\mathrm{x})/2)$ is used, since this conveniently lies between –1 and +1.

Capacity control

Minimization of the empirical error provides no guarantee of low error outside the training set, unless some limit is set to the complexity of the model. This problem has been well understood in practice in the context of neural networks. Over-fitting the training set can lead to poor generalization on a test set.

Several methods have been used to control the capacity of neural networks. These include limiting the number of hidden units, early stopping during training, addition of noise to training data, and explicit Tikhonov-style regularization. These methods normally require the selection of optimal parameter values by monitoring performance on a test set. Performance on a test set, however, can be a noisy quantity. The trained network may depend significantly on which test set was chosen. Cross-validation (Stone, 1974) avoids some of these difficulties and is more parsimonious with scarce data, but can be time consuming, especially if there is more than one parameter to set.

An alternative is to use Bayesian techniques to set parameters adaptively (MacKay, 1992) or to eliminate them by integration (Buntine and Weigend, 1991). In the Walker Lane case study described above, the integration approach of Williams (1995, 1999) was used. Interpreting $E(\mathrm{w})$ in equation (A2) as the per-sample negative log likelihood of the data, and using a Laplace prior over weights, whose scale parameter is integrated out using a non-informative Jeffreys prior, it can be shown that the negative log posterior for w is given by

$$L(\mathrm{w}) = NE(\mathrm{w}) + W \log ||\mathrm{w}||, \qquad \text{(A4)}$$

where W is the total number of weights and $||\mathrm{w}||$ is the L_1 norm of w. Minimizing $L(\mathrm{w})$ corresponds to penalized maximum likelihood, where the balance between likelihood and penalty depends on the ratio of the number of parameters to the number of data points. Note that equation (A4) contains no adjustable parameters. Furthermore, at any minimum of $L(\mathrm{w})$, the weights separate into two classes: (1) those with a common data-misfit sensitivity $|\partial E/\partial w|$, and (2) those failing to achieve this common sensitivity and which are automatically set to zero, effectively pruning the network of redundant connections (Williams, 1995).

Committees of networks

A combination of several predictors can improve accuracy. Suppose that M networks are trained to yield separate estimates $P_j(D|\mathrm{x})$ $(j = 1,\ldots, M)$ of the conditional probability of D. Then the average estimate

$$P(D|\mathrm{x}) = \frac{1}{M} \sum_{j=1}^{M} P_j(D|\mathrm{x}), \qquad \text{(A5)}$$

has lower expected error than the average of the expected errors of the individual networks. The improvement depends on the extent of correlation between network errors (Krogh and Vedelsby, 1995). Ideally, the errors should be independent. One way to promote independence is to use different architectures for different networks. Another is to use different training sets. Varying the training set is straightforward for favorability mapping. Barren regions comprising negative instances are often chosen by selecting locations at random, away from the known deposits. If several networks are to be trained, and their results averaged, a different set of random negative instances can be used for each network. To a limited extent, positive instances may also be chosen independently for each network by small perturbations in known deposit coordinates. Both variations will tend to reduce correlations between the errors of different networks and hence, the expected error of their average prediction. This approach has similarities to the addition of synthetic noise to training inputs (Bishop, 1995b; Brown et al., 2003) except that, in this case, the noise is generated by the data.

Kernel methods

There are many ways of fitting a nonlinear function, such as $P(D|\mathrm{x})$, other than using a multilayer perceptron. Most can broadly be grouped under the heading of kernel methods where the regression, or classification, function takes the form

$$f(\mathrm{x}) = \sum_{i=1}^{N} c_i K(\mathrm{x}, \mathrm{x}_i), \qquad \text{(A6)}$$

where K is some positive-definite symmetric kernel, and summation is over the training inputs $\mathrm{x}_1,\ldots,\mathrm{x}_N$. The coefficients $c_1,\ldots,c_N$ are determined by fitting $f(\mathrm{x})$ to the corresponding target outputs $y_1,\ldots,y_N$ in one of several ways (Poggio and Smale, 2003). Radial basis function networks are of this type, with the kernel taking the form of a spherical Gaussian.

Another kernel method, widely considered to be the method of choice in areas of pattern recognition such as bioinformatics, is the Support Vector Machine (SVM). In its simplest form, where each $y_i = \pm 1$, this is a wide-margin binary classifier. Coefficients are fitted by solving a quadratic programming problem. Typically, this approach leads to a sparse representation, with relatively few nonzero coefficients, implying tighter bounds on the misclassification error (Vapnik, 1998; Cristianini and Shawe-Taylor, 2000; Schölkopf and Smola, 2002). A deterministic classifier, however, where only the sign of $f(\mathrm{x})$ is significant, is not ideal for favorability mapping since it does not provide a target ranking. An alternative is to use a regression form of the SVM, or an SVM variant explicitly designed to obtain predictive class probabilities (Sollich, 2002). All SVM methods, however, require careful attention to model selection in the form of choice of kernel type, kernel parameters, and misclassification penalties (Chapelle and Vapnik, 2000; Gold and Sollich, 2003; Williams et al., 2005).

Kernel methods are also used in Gaussian process models (MacKay, 2003). These can be thought of as a form of kriging in feature space, with the kernel corresponding to an analytic parameterized covariogram. Gaussian processes have also been adapted to perform classification tasks (Williams and Barber, 1998; Gibbs and MacKay, 2000). It is certainly worth investigating what these and the SVM methods have to offer for favorability mapping.