Budget Allocation and the Stopping Problem in Mineral Exploration

Andy Green
OTBC Pty ltd
8 Lawley Cres. Pymble NSW Australia, 2073
andy.green@ozemail.com.au

SUMMARY

Most greenfields mineral exploration projects involve a process of testing targets that have been selected on the basis of geoscientific data. Although this data can be used to rank targets, questions still arise as to how many targets should be tested before the area is dropped. This paper addresses this question with a probabilistic model of the exploration process and illustrates the method with a geophysical example.

The model is governed by the target and background distributions of the variable used to rank targets, various economic parameters and two geologically determined probabilities. It generates Expected Value, Probability of Success and Return on Investment (ROI) for a range of possible project budgets showing where each is maximized. It is argued that a lower and upper limit for the number of tests to be undertaken can be defined in terms of this model. The lower limit, called the Equal Opportunity Truncation point, occurs where the ROI is maximized and is relevant when other equally attractive prospects are available. The upper limit, called the Economic Truncation point, occurs when the Expected Value is maximized.

The kimberlite exploration case study illustrates a new method of estimating the target distribution using magnetic modelling and shows how the probabilistic model could have been used to budget this exploration project.

Key words: Mineral exploration. Expected Value, Return on Investment, ROC curve.

INTRODUCTION

How long should mineral exploration persist in an exploration licence area? The longer exploration continues, the greater the probability of success, but this does not guarantee the venture will be profitable. Eventually, there must come a point where the incremental cost of testing will be greater than the expected value of the result. Moreover, usually well before this point, a decision must be made as to whether it is better to terminate a current exploration programme and start anew somewhere else. Such decisions can only be made by considering how the financial variables (deposit value and exploration costs) interact with the effectiveness of the detection method and probabilities related to the prospectivity of the area.

Over the past 10 years there has been increasing recognition of the need to impose geoscientific and financial rigour on the planning and execution of the mineral exploration process. This has resulted in an increased focus on probabilistic models and their incorporation into an economic analysis of exploration programs. Two broad themes have emerged under the general title of Mineral Prospectivity Analysis (Porwal and Kreuzer, 2010). They are generally applied in the early stages of mineral exploration and can be described as Data-driven methods and Process-based methods.

Data-driven methods, such as the well-established Weights-of-Evidence method (WoE, Bonham-Carter, 1994), developed from early studies of statistical modelling of the spatial relationship between known mineral deposits and geoscientific data, exploit the capabilities of modern Geographic Information Systems to integrate exploration data. Subsequently, Neural Net methods (Singer and Kouda, 1999, Barnett and Williams, 2006), which are less constrained by assumptions of conditional independence, have been used for the same problem. Both are "data-driven" in the sense that they require substantial training information in well-explored areas to characterize the joint distributions of the exploration data in the presence *and* in the absence of known deposits. Through the application of Bayes Rule this dual focus naturally gives rise to an explicit consideration of the false positive rate and a definition of the conditional probability P(M|T) that a selected target (T) will be found to be mineralized (M) after testing.

More recently (Kreuzer et. al., 2008), there has been increased focus on comparing staged, process-based models of ore formation with observed data, to establish the overall probability of mineralization. Such methods are ideal in greenfields terrains where there are no known deposits to train the data-driven methods. However, the necessary trade-off for this flexibility means the probabilities associated with the component parts of these sophisticated models are determined more subjectively than with the data-driven methods. Moreover it is still difficult to characterize the data that will lead to false positives in subsequent target detection.

Both methods can include an economic analysis of the exploration program. The various cost and value parameters, which can usually be defined much more confidently than the geological probabilities, are used to estimate Expected Costs and Expected Values for the program as a whole. Such valuations usually encompass the full scope of an exploration program from area selection to deposit delineation.

In this paper I have chosen to narrow the focus by concentrating on the planning of the detection phase of exploration. Here costs increase rapidly and budget decisions become important. Hronsky and Groves (2008) make this important point:

> In addition, it is at this scale that the company has to make a serious commitment of resources to the project, in terms of time, people and money, to advance to the direct detection phase of exploration. Where direct detection technologies are available at the camp scale, they are invariably associated with large false positive rates (soil geochemistry grids, detailed geophysical surveys such as IP, EM, ground magnetics, and gravity). Greenfields exploration is a classic example of a low base rate probability of success scenario, which is therefore highly sensitive to the false positive rate of any targeting or detection technology. Future research needs to focus on addressing this predictive–detective switchover point

To achieve this focus, I will assume that the area selection is complete and that I can express downstream costs and valuations as simple expected values. I will then look in detail at how the trade-off between detection and false positives determines the Expected Value (EV), the Expected Return on Investment (ROI) and the probability of success, P_S, of a project.

A SIMPLIFIED MODEL FOR EXPLORATION

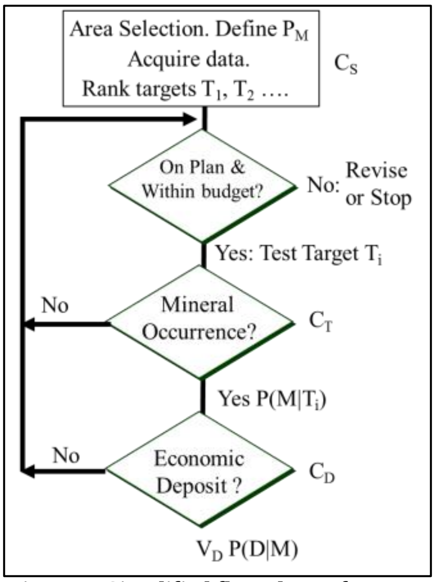

Figure 1. Simplified flow chart of an exploration program.

Figure 1 illustrates the flow chart of the simplified model of an exploration program. The critical decision point focuses on the detection of a mineral occurrence. The definition will be project dependent. Makenzie (1989) suggests "indications of mineralization of potential economic grades across mineable widths … obtained by drilling" but these are not very common and in some circumstances "indications of mineralization" might be sufficient to motivate more drilling in the area.

The parameters of the model are:

- N — Number of possible targets (e.g. 100,000)
- C_S — Cost to set up program (e.g. 500 K)
- C_T — Cost of testing a target (e.g. \$10 K)
- C_D — Cost of delineating a mineral occurrence. (e.g. \$200 K)
- V_D — Expected Value of a deposit (e.g. \$100 M).
- P_d — Probability of detecting a mineral occurrence = P(T|M)
- P_{fp} — Probability of a false positive = $P(T|\bar{M})$
- P_M — Probability of a mineral occurrence from untargeted drilling. (e.g. $P_M = 10^{-3}$)
- P(D|M) Probability that a mineral occurrence will be economic. (e.g. $P(D|M) = 10^{-2}$)

It is important to note that P_M is defined by the upstream planning and research. It is equal to the expected number of mineral occurrences divided by the number of possible targets. It is assumed that P(D|M) has also been defined on the basis of assumed or established size and grade distributions for the style of deposit in question (Singer and Kouda 1999b). P_d and P_{fp} are the important probabilities for this model. They are functions of the two key probability distributions for target detection

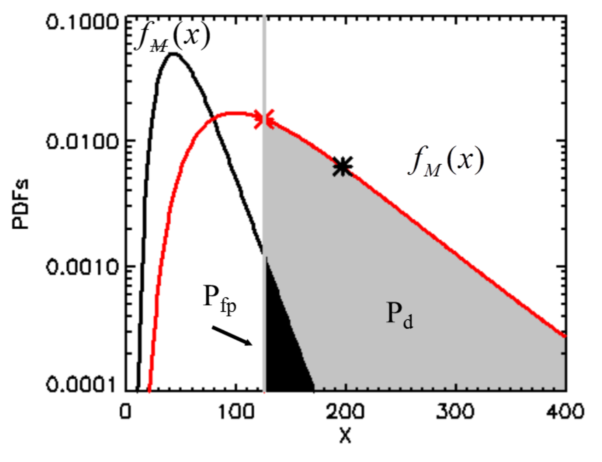

Figure 2 Target and Background distributions for the example exploration program. Note the log scale. Stars indicate two cut-off points defining possible exploration programs.

Figure 2 illustrates typical distributions. The exploration data generates a variable x which is used to rank targets and define a cut-off value x_0 (here ~125). x could be as simple as a single element geochemical measurement or as complex as a favourability index from a neural net algorithm which uses many different data inputs. In the presence of a mineral occurrence, x is distributed as $f_M(x)$ and in the absence of a mineral occurrence, as $f_{\bar{M}}(x)$. Here $f_M(x)$ will be called the target distribution and $f_{\bar{M}}(x)$ the background distribution. Once the cut-off is selected, the project is defined and the budget set because all the targets with $x > x_0$ will be tested. Now P_d will be the proportion of the area under $f_M(x)$ with $x > x_0$ and P_{fp} is similarly defined in terms of $f_{\bar{M}}(x)$.

In this simplified model, the selection of a cut-off value for x determines the budget and is directly related to the number of targets that can be tested. Clearly the nature and separation of $f_M(x)$ and $f_{\bar{M}}(x)$ determine how successful detection will be. Both distributions shown here are Lognormal and will be used as an example to illustrate the technique. One way to estimate these distributions will be discussed later but, for the moment, given this simple model and the parameter settings above, the following financial and probability results follow.

Expected number of false positives	$N_{fp} = N\,(1-P_M)\,P_{fp}$
Expected number of mineral occurrences	$N_d = N\,P_M\,P_d$
Expected Expenditure	$E = N_{fp}\,C_T + N_d\,(C_T + C_D) + C_S$
Expected Value of a Mineral Occurrence	$V_M = V_D\,P(D\|M) - (C_T + C_D)$
Expected Value	$V = V_D\,P(D\|M)\,N_d - E$
Return on Investment	$ROI = V / E$
Probability of Success (at least one deposit).	$P_S = 1 - \{ 1 - P_M\,P_d\,P(D\|M) \}^N$
Probability the k[th] target is a mineral occurrence	$P(M\,\|\,T_k) = P(M\,\|\,x = x_k) = \frac{l_k\,P_M}{l_k\,P_M + (1-P_M)}$ where $l_k = \frac{f_M(x_k)}{f_{\not M}(x_k)}$
Expected Expenditure and Value on the kth target	$E_k = C_D\,P(M\|T_k) + C_T$, $V_k = \{V_D\,P(D\|M) - C_D\}P(M\|T_k) - C_T$
Expected ROI of the k[th] target	$ROI_k = V_k / E_k$

The Receiver Operating Characteristics (ROC) curve, Figure 3(a), plots P_d against P_{fp} and summarizes the critical information from Figure 2 (Scharf, 1990). Each point on the curve corresponds to a different cut-off. The stars correspond to the two cut-off values shown in Figure 2. When these probabilities are used with the example parameters, the results for the economic variables are plotted as a function of possible budget allocation. These are shown in Figure 3 (b), (c) and (d).

Figure 3(b) shows how the Expected Value varies as a function of expenditure. It reaches a maximum after which the rate at which false positives are encountered so exceeds the rate at which mineral occurrences are detected that overall value declines. Smith (2005) calls this point Economic Truncation (ET) and suggests that exploration should not continue after this point. The cut-off for the ET point is shown on Figure 2 at ~125 defining the shaded areas for P_d and P_{fp}.

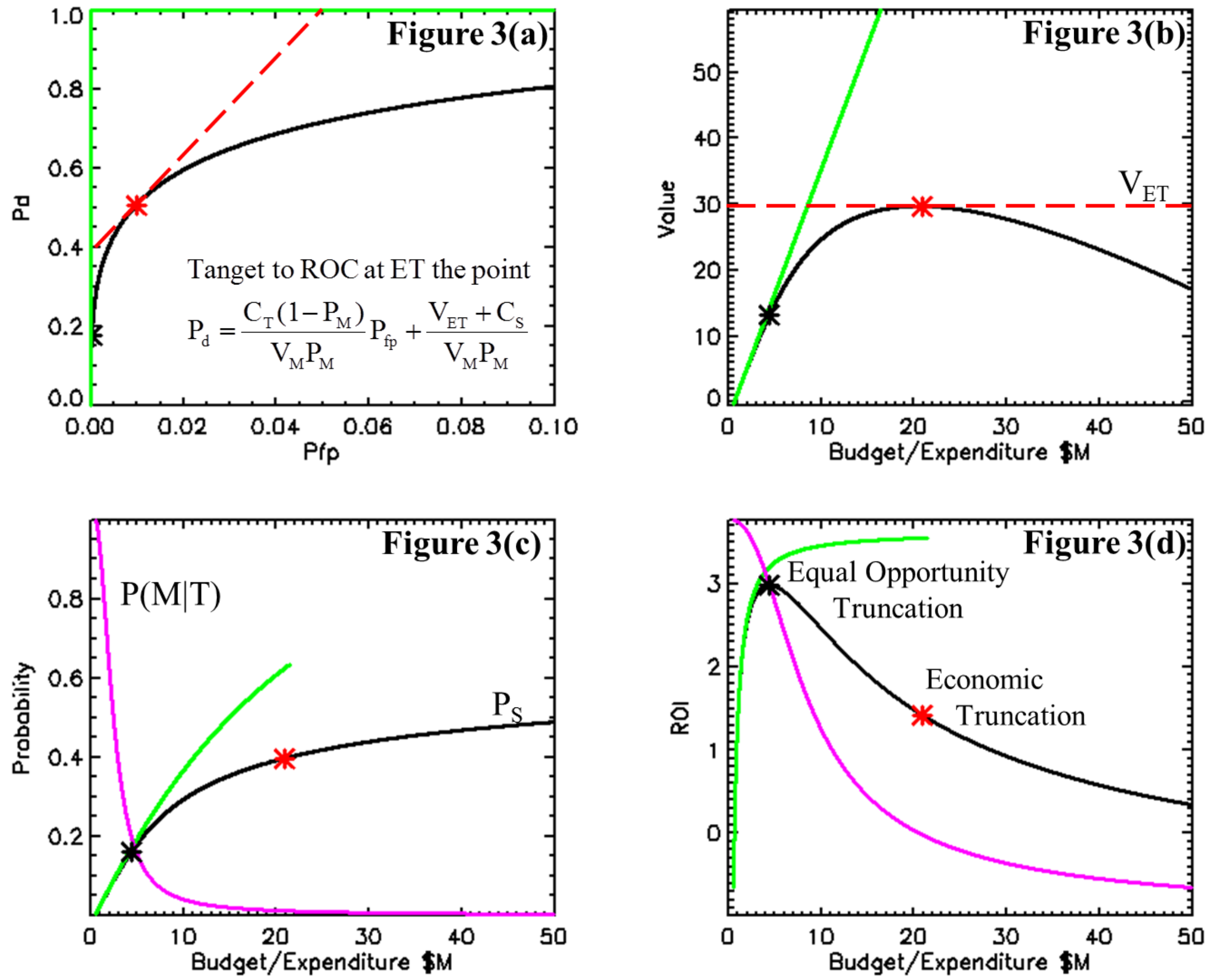

Figure 3. Results derived from the distribution shown in Figure 2. (a) ROC curve (b) Expected Value. (c) Probability of success (black) and P(M|T) mauve (d) Return on Investment (black) and Target ROI (mauve). The green curves in each case show the result if the detection is perfect ($P_d = 1$ for all P_{fp}).

Figure 3(c) shows how the probability of success (P_S) changes as a function of expenditure. Clearly most of the increase in P_s is complete by the time Economic Truncation is reached and there is probably little point in proceeding further. The corresponding curve for the perfect detection case reaches ~0.65 where it terminates because all the expected mineral occurrences in the EL have been found. P(M|T), the probability a target will be a mineral occurrence at each stage of the program, falls rapidly as exploration proceeds and is negligible by the Economic Truncation point.

Figure 3(d) shows the ROI. This reaches a maximum well before the Economic Truncation point where it is only slightly less than the ROI for perfect targeting. I have called this the Equal Opportunity Truncation (EOT) point because, if an opportunity exists to invest in a series of small projects with the same characteristics as this one, you should do so and restrict each to an expenditure of ~$5M. On Figure 2 the cut-off for EOT is shown where x~200 and corresponds to the black stars on Figure 3(b)-(d).

Given P(M|T) (Figure 3(c)) it is possible to calculate the expected ROI of a target at each stage of the detection program. This is shown in purple in Figure 3(d). It is interesting to note that this target ROI curve *must* intersect with the project ROI curve at the EOT point. This means that, if equal opportunities are still available, testing targets beyond this point cannot be justified by arguing that the project set-up costs are sunk and "one more test will be neither here nor there".

It is interesting to consider the effect of changes in the various economic parameters on the three output variables (V, P_S and ROI). Changes in the set-up cost, C_S, displace but do not affect the shape of the Value and probability curves. The Value curves are translated equally down and to the right as C_S increases and the probability curves shift to the right by the same amount. The position of the ET point on the ROC curve is solely determined by the slope of the tangent, which is independent of C_S. This means that as C_T declines, the ET point moves progressively to the right and V_{ET} increases because the Y-intercept of the tangent increases.

The behaviour of the ROI is driven primarily by the set-up costs. When C_S is zero, the EOT point occurs at the very start of target testing and the ROI falls steadily with increasing expenditure. This is because the project ROI at the start is the same as the ROI of the first test (the peak of the mauve curve in Figure 3(d)). As C_S increases, the EOT point moves towards the ET point until they are coincident when the value and ROI are both zero.

Once a time scale and discount rate is defined, it is relatively simple to modify this analysis to provide discounted expenditures, values and internal rate of return measures. The general characteristics of this model as described above hold for a wide variety of distributional and economic parameters. Assuming P_M and P(D|M) have been well chosen, the major difficulty that remains is to characterize f_M and $f_{\bar{M}}$ reliably.

The model can be extended to multivariate exploration data (**X**) when the joint distributions of this data is available for both targets $f_M(\mathbf{X})$ and backgrounds, $f_{\bar{M}}(\mathbf{X})$. In this case, the *x* in the above discussion is replaced by the likelihood ratio $l = f_M(\mathbf{X}) / f_{\bar{M}}(\mathbf{X})$ (Scharf, 1990). Alternatively (Barnett and Williams, 2006), neural nets can be used to define f_M and $f_{\bar{M}}$ where *x* is a favourability index. However, whatever method is used, large numbers of representative samples of target and background data must be available to generate these distributions.

CASE STUDY

The following case study uses magnetic data acquired for kimberlite exploration. The area has ten kimberlites with normal magnetic response, eight with reversed magnetic response and twenty five with no observable magnetic response. The data was resampled onto a grid with 190 m line spacing and 33m along-line sampling and was reduced to the pole. After pre-whitening (Green and Hunter, 2004) the grid was filtered with a Constant False Alarm Ratio (CFAR) filter (Scharf, 1990) which was optimized for detecting kimberlites with a 50-100m radius. Having no information about diamond grade, I stretch the definition of a mineral occurrence to define a kimberlite as a mineral occurrence.

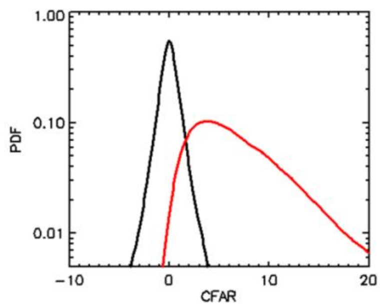

Figure 4 Target (red) and Background (black) distributions for the case study.

In order to define the filter and to generate $f_M(x)$, 10,000 semi-infinite cylinder models of kimberlite responses (Singh and Sabina, 1978) were calculated using lognormally distributed radius and susceptibility values. The background distribution $f_{\bar{M}}(x)$ was estimated with magnetic data from a similar area without known kimberlites. Figure 4 shows the two (smoothed) distributions.

The background distribution is almost symmetrical and is distinctly heavy-tailed. Here large positive values are likely to be false positives when filtering for normally magnetized targets and large negatives for the reversely magnetized targets. Because the data has been reduced to the pole, the target distribution for reversely magnetized targets (not shown) is merely the reflection of the normal curve in the ordinate axis.

The area of the survey is ~1183 km^2 and I obtain the cell size of 0.16 km^2 by assuming it is governed by the way the flight-line grid samples the survey area. Together, these give the number of possible targets in the survey area (N) as 7394. Because the model assumes all kimberlites are potentially detectable with magnetics, the true number of kimberlites is N_t = 18 and this gives P_M = 18/7392 = 0.0028. Assuming the other economic information is the same as in the example above, the curves shown in Figure 3 can be reproduced for this survey (Figure 5).

In addition, using knowledge of the true kimberlite locations, it is possible to simulate the progress of this exploration program assuming the CFAR filtering method was used. This simulation produces a series of targets for testing with successively smaller CFAR values. Each target is either a successful detection or a false positive. At each stage the number of detections (N_d) and false positives (N_{fp}) is retained and the True Positive Rate (TPR = N_d/N_t) and False Positive Rate (FPR = $N_{fp}/(N-N_t)$ are calculated. The

TPR and the FPR are the practical realization of the P_d and P_{fp} predicted by the model. The realized ROC curve (in red) is plotted with the modelled ROC curve in Figure 5(a), as are the corresponding realizations of Value, P_S and ROI are in Figure 5(b) – 5(d).

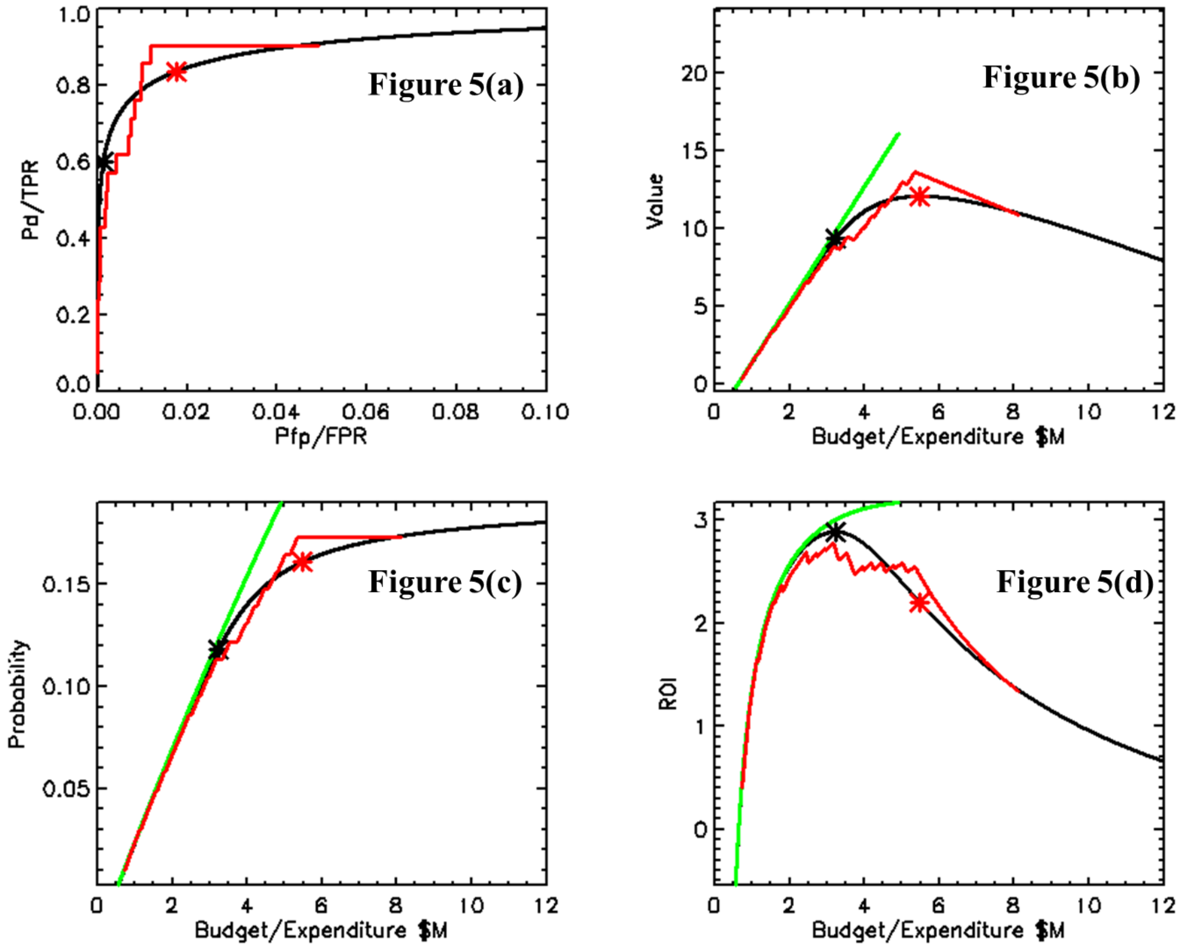

Figure 5. Results derived from the distribution shown in Figure 4. In each case the model results are in black, simulation results in red, perfect detection results in green. (a) ROC curves (b) Expected Value. (c) Probability of success (d) ROI.

DISCUSSION

In the kimberlite example, the modelled ROC curve provides a reasonable match to the simulated realization of the exploration program. However, the model under-estimates the rate at which false positives are encountered in the middle stages of the program and over-estimates at the latter stages. In the end, the two errors cancel so that the ET point would have been a good place to stop the detection phase.

The method used to establish f_M and $f_{\bar{M}}$ is a hybrid between the purely "data-driven" methods like WoE and the staged, process-based geological models of Kreuzer et. al. (2008). $f_{\bar{M}}$ was established in a purely data-driven way using areas thought to be kimberlite-free and otherwise similar in background response. However, because the target distributions were generated by geophysical modelling, multiple instances of known kimberlites were not needed - a major advantage in greenfields exploration.

The method has other potential advantages:

- The method could be used to compare different exploration strategies. It could be used to decide if adding AEM, with its increased costs, would improve the attractiveness of the kimberlite project.
- If necessary, the geophysical models that generate f_M can be refined mid-project. For example if a kimberlite is found in the early stages of exploration, the size and susceptibility distributions might be revised to make the ROC curve more accurate.
- Modelling could also be used within the program to check that the initial assumptions are still relevant. In the example, if the expected number of kimberlites was predicted correctly but they were all assumed to be detectable by magnetics, P_M would have been estimated to be 0.006. This would have predicted a V_{ET} of $27M with a P_S of 0.33 for an expenditure of $11M. Halfway through the program it would have become obvious that things were not going as hoped and the model revised to terminate the program with about half the planned expenditure and only slightly lower ROI.

More generally, the above analysis suggests we should consider using the EOT and ET points as the lower and upper bounds for the detection phase of an exploration program. Clearly we might expect to fund a smaller project that is closer to the EOT point when there are other, equally-attractive projects available. In the example shown in Figure 3, four $5M projects would produce a combined Expected Value about twice V_{ET} with ~50% increase in P_S for the same $20M expenditure. This result confirms the view (Kreuzer et. al., 2008) that:

> …there are three principal levers that control the return on exploration investment at the portfolio and program levels: (1) the number of projects effectively tested and turned over, (2) the average expenditure per project (especially on those that failed), and (3) the average probability of success across the portfolio.

In this case, the improved performance comes from testing only the highest ranked anomalies – the ones least likely to generate false positives – before moving on to an equally prospective area. However, as the kimberlite case study cautions, allowance must be made for the errors in the modelling process. In that example, truncating too early would have missed kimberlites that were found quite easily.

CONCLUSIONS

The modelling strategy discussed here formalizes part of what an exploration manager does implicitly when assessing a project. A decision to proceed means that it is expected that the rate at which mineral occurrences will be found will be sufficiently greater than the false positive rate to produce an acceptable Expected Value. Implementing this formal process is not easy because the distributions f_M and $f_{\not{M}}$ are hard to define and the advantages of the strategy can only be achieved if the input parameters are correctly tuned to the realities of the exploration problem.

The difficulty in selecting prospective ground (and thus estimating P_M) is well recognized and the subject of continuing research. On the detection side, while there is considerable research to understand what deposits look like from a geological, geochemical and geophysical viewpoint, there is less effort put into understanding how they and their smaller cousins, mineral occurrences, are distributed in the data. However, we put even less research to understanding how the background geology within which we explore can appear to be a mineral occurrence to our detection technologies. While defining the $f_{\not{M}}$ is perhaps somewhat unglamorous, it is just as critical as the defining f_M.

If these difficulties can be overcome we should be able to make mineral exploration more cost effective through improved planning and execution of the detection phase. In this context the Economic Truncation and Equal Opportunity Truncation points are a good starting point for deciding when to terminate exploration. Between the EOT point and the ET point, the trade-off is between ROI and Expected Value. How this is best resolved will depend on how the project sits within a larger portfolio of projects and the management overheads required by multiple small projects.

ACKNOWLEDGMENTS

I would like to express my gratitude to the sponsors of AMIRA Project P845 who supported the early stages of this project.

REFERENCES

Barnett, C.T., and Williams, P.M., 2006, Mineral exploration using modern data mining techniques: Society of Economic Geologists, Special Publication 12, 295–310.

Bonham-Carter, G., 1994, Geographic information systems for geoscientists: Pergamon.

Green. A. and Hunter, D., 2004, AEM Target detection in geological noise: Extended Abstracts, 17th ASEG Conference Sydney, August 2004.

Green. A., 2004, Exploration, risk and the value of geophysics: Extended Abstracts, 17th ASEG Conference Sydney, August 2004.

Hronsky, J.M., and Groves, D.I., 2008, Science of targeting: definition, strategies, targeting and performance measurement: Australian Journal of Earth Sciences, 55, 3–12.

Kreuzer, O.P., Etheridge, M.A., Guj, P., McMahon, M.E., and Holden, D.J., 2008, Linking mineral deposit models to quantitative risk analysis and decision-making in exploration: Economic Geology, 103, 829–850.

Makenzie, B. 1989, Economic guidelines for gold exploration: Australian Mineral Foundation Course Notes 628/89.

Porwal A.K. and Kreuzer, O.P., 2010, Introduction to the special issue: Mineral prospectivity analysis and quantitative resource estimation: Ore Geology Reviews, 38, 121–127.

Scharf, L.L., 1990, Statistical signal processing: Addison-Wesley.

Singer, D.A., and Kouda, R., 1999a, A comparison of the weights of evidence method and probabilistic neural networks: Natural Resources Research, 8, 287–298.

Singer, D.A., and Kouda, R., 1999b, Examining risk in mineral exploration: Natural Resources Research, 8, 111-122.

Singh, S.K. and Sabina, J.S., 1978, Magnetic anomaly due to a vertical right cylinder with arbitrary polarization: Geophysics, 43, 173-178.

Smith, J.L., 2005, Petroleum prospect valuation: The option to drill again: The Energy Journal, 26, 53-68.