Cutoffs and Selectivity: How to estimate the processable content of a deposit in the exploration stage

K. Gerald van den Boogaart[1], Friedrich-Wilhelm Wellmer[2]

[1]TU Bergakademie Freiberg, Germany. E-mail: boogaart@math.tu-freiberg.de
[2] TU Berlin, Germany. E-mail: fwellmer@t-online.de

Abstract

Our aim is to predict from exploration data only the mean grade and the content of the part of a deposit that will actually be mined if the deposit is going into production. Typically the mean grade of the actually mined material is estimated by taking the mean of the exploration samples over cutoff. However this overestimates the actual mean grade in the production for two reasons: 1) a block is mined and processed whenever the measured grade of the block is above cutoff. However the measured grade is not the actual grade of the block. Some blocks slightly below cutoff will be mined because the measurement was by chance above cutoff and some blocks with a mean grade above cutoff will be declared waste, because we tested some internal waste. We propose a paired sampling method to correct for this problem. 2) During a campaign several potential deposits are explored. Mining is only started for those apparently profitable, because the deposit is big enough and the mean grade of the portion above cutoff is high enough to ensure the return of the investments. However this implies that deposits are mined on the condition that the observations are above cutoff. This again implies that a deposit of low grades will be mined if the mean grade or the total content has been overestimated by chance. We propose a first order bias correction for this problem.

Keywords: spatial statistics, grade, bootstrap.

1. INTRODUCTION

The prediction of the amount of ore that will actually be produced if a deposit is mined is closely related to the decision in the lifecycle of a mine. This paper is based on the following simplified logic and decision rules:

1. The potential deposit is sampled by a small scale exploration e.g. delivering a small number (e.g. 30) of grade measurements from drill cores distributed systematically or

randomly over the area of the potential deposit. For mathematical simplicity we will assume a random sampling of block centers, to ensure unbiasedness

2. Based on the estimated mean grade and estimated portion of the deposit a decision is taken whether or not the deposit to mine. The paper focuses on statistical prediction methods, which can be used in this step to predict the final results in the last steps unbiasedly.
3. During production each potential mining block is sampled. We assume that these samples have the same support as samples in the exploration phase and that there is one sample per mining block. Similar effects show up with other decision schemes such as using multiple samples per block in production phase, but corresponding bias correction need to be subject to further research.
4. Each mining block is mined and processed if and only if its measured grade is above some profitability cutoff c. For the sake of simplicity we will assume for this paper, that the decision of processing is done separately and with the same cutoff for each mining block.
5. Processing costs are affine linear dependent of the tonnage processed, i.e. capacity of the mine. Processing will produce revenues of the raw material affine linearly dependent on the average grade of the material processed.
6. The actual final cash flow of the mine consists of the initial investments, the revenue for every ton of the final raw material produced minus the costs for every ton of ore mined and processed .

In the exploration phase of a deposit it is quite problematic to predict the total amount of the raw material that would actually be mined later. A first approximation is to multiply the total size of the deposit by the mean of the grades of the exploration samples. This is an unbiased predictor of the actual amount in the deposit. However due to the cutoff not everything is processed. Thus this first predictor overestimates the actual production. This problem can be adjusted for, by multiplying with the portion of material above cutoff. However this predictor underestimates the production because the mean grade of the actually mined material is higher than the average material in the deposit, as intended by applying the cutoff. It is thus common practice to multiply the tonnage above cutoff with the mean of measurements above cutoff, to predict the production in early exploration (e.g. Wellmer 1997). However surprisingly it has been repeatedly observed that predictions often overestimates the actual grade encountered in production (King 1982).
We have to identify two simple mathematical reasons why this predictor applied to a real world mine is still biased:

1. **Block selection bias**
   It does not respect the fact that a block is not mined, if it is above cutoff, but if it looks being above cutoff. Actual production will thus process material below cutoff and miss material above cutoff.

2. **Deposit selection bias**
   The second reason is linked to the fact that a mine is only going into production if it looks being profitable. If we start out exploring ten possible mining sites, all equal and slightly under the profitability limit, one of them might by chance look profitable.

Figure 1 demonstrates a simulated example setting: Panel (a) shows a simulated concentration field. Blue areas are below the cutoff 1. Red areas are above cutoff. The brighter the colors the higher is the values. The simulation is a log-Gaussian random field with a spherical variogram for the logs with nugget=0.06, sill=0.2, and range 1 with log-mean=0. All simulations are done in R (R 2010) using a direct Cholesky-based simulation method (Cressie 1993).

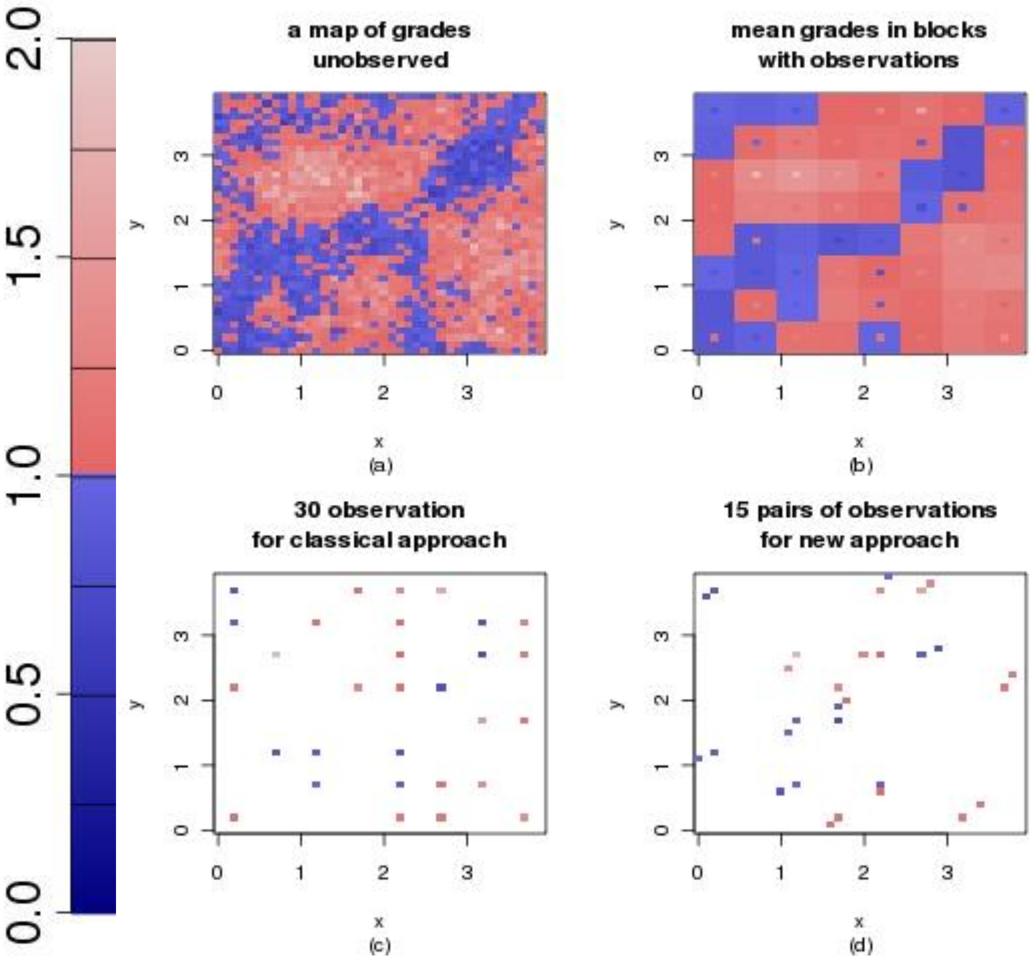

Fig.1: An simulated example situation. Grades are color coded. Blue colors are below the cutoff of
(a) Map of a simulated example deposit.
(b) corresponding mean block grades and local measurements
(c) 30 measurement locations in a random sampling of block centers,
(d) 30 measurement locations from 15 sampling pairs use

Panel (b) shows the mean grade in the blocks and in the center of the block the grade which would be measured in the middle of the block. During production a block will be mined and processed if and only if a sample taken in the center of the block has a grade above the example cutoff c=1. Due to the random variation of values within the block for some blocks the measurement the mean grade of the block and of the measurement might be on different sides of the cutoff. In the example five blue measurements in red blocks would prevent the mining of these blocks that could have been mined profitable if we would have known their true mean grade. The three red observations in otherwise blue blocks correspond to blocks that would be mined despite their low values simply because we would believe them to be above cutoff.

Panel (c) shows the measurement locations in a random selection of block centers simulating a classical exploration campaign of 30 samples. Only exploration observations are available to decide on the value of the mine and the start of production. Panel (d) shows the measurement locations of pairs of randomly selected block centers and a second measurement randomly distributed in the same block, simulating a paired sampling exploration campaign as proposed later in this paper. The color coding in panels (c) and (d) show the actually measured values.

For the evaluation of the deposit before we have actually decided on starting to mine the deposit we would need to predict the mean grade of the material that would be mined and the total content of the deposit that will be produced.

2. The Block Selection Bias

Figures 2 shows for the same simulated example a scatterplot the mean block grades against the grades we would measure taking a sample in the middle of each potential mining block.

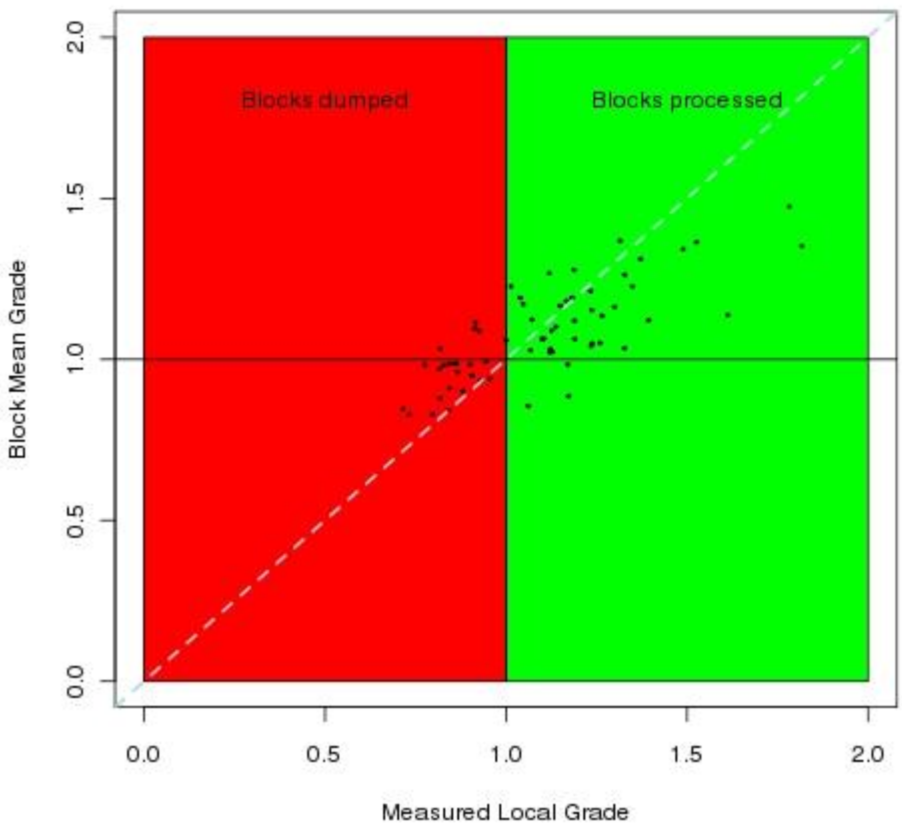

Fig.2: The reason for the bias introduced by selecting the blocks according to a cutoff.

Due to spatial variation and measurement errors these two values are not the same and do not follow precisely the y=x line, which is drawn as a gray dashed line, separating the blocks that will be processed (green) from those that will not (red). Only blocks with measurements

above the cutoff will be send to processing. Although most blocks are correctly classified, some blocks with below cutoff average grade will be processed and we will loose some blocks with average grades slightly above cutoff simply because the measurements were below cutoff. Obviously the expectation of the average grades of the blocks processed is thus lower than the average observation above cutoff but higher than the average grade of the mine.
Figure 3 demonstrates the results we would get for different grade fields. 1000 simulated realizations of grade fields are drawn. For each grade field six parameters are computed.

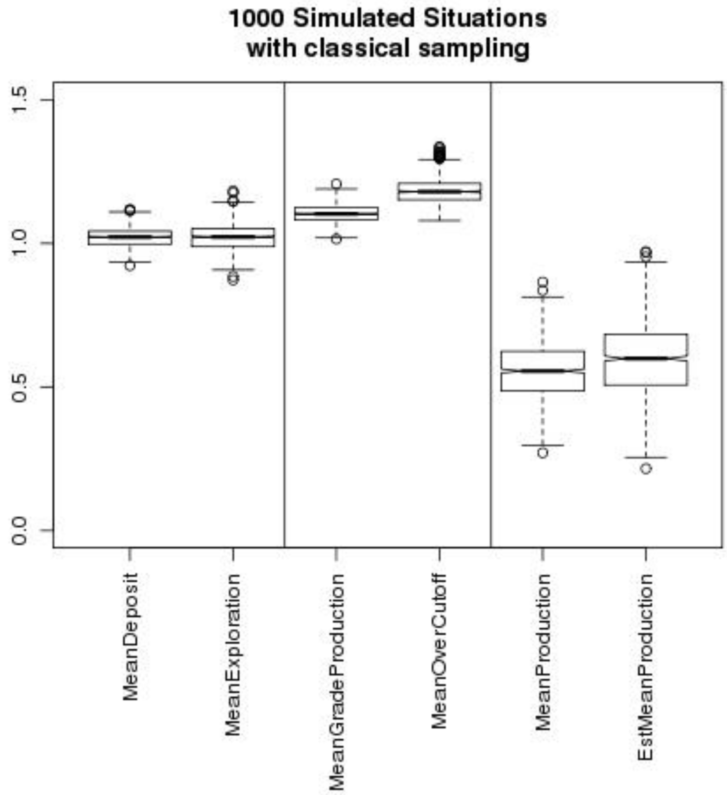

Fig.3: The bias of the classical estimators

Each of the six notched boxplot shows the 1000 different results in the 1000 simulations. Be reminded that the notch is something like a confidence interval for the median, meaning that non overlapping notches prove things to be shifted (Chambers et al. 1983). Each of the boxplot is symmetric such that the median can be taken as a robust estimate of the mean.
The parameters are:

- "Mean Deposit" corresponds to the true mean in the actual area of the deposit. This is never observed in real live, but can easily be computed for simulated deposits and is the quantity unbiasedly estimated by the mean of the measured values.
- "Mean Exploration" corresponds to the mean of the measured values in the deposits. The unbiasedness is exemplified in the similar median of the first two boxplots.
- "Mean Grade Production" is the mean of the grade of the blocks that would be mined based on their center observation being over cutoff. This information is typically only available after the deposit has been completely mined. It is obviously higher than the mean grade of the deposit due to the selection of blocks processed.

- “Mean Over Cutoff” is the mean of the subset of exploration observations above cutoff. It is the usual estimate for “Mean Grade Production” but obviously biased in favor of too higher grades. In the example the bias has a similar magnitude as the underestimation bias of using the unconditional average grade of the exploration samples “Mean Exploration”.
- “Mean Production” is “Mean Grade Production” multiplied with the portion of blocks actually mined. Multiplied with the actual total tonnage of the explored volume it would give the total economically usable content of the deposit.
- “Estimated production” is the observed “Mean Over Cutoff” multiplied with the portion of the exploration measurements above cutoff, which is the classical prediction for “Mean Production”. The not overlapping notches of the symmetric boxplot shows that this quantity overestimates the “Mean Production”

3. CORRECTING FOR THE BLOCK SELECTION BIAS

To provide an unbiased prediction of the mean grade of the blocks processed we need to find out the average of the block average of the blocks which would be selected. This sounds next to impossible since it is unobservable and depends on the full distribution of the random grade field and is not a function of its mean or covariance function. A state of the art approach might be to fit a random fields model and to compute the conditional mining result by conditional simulation (compare e.g. Dimitrakopoulos&Ramazan 2008).
The block average of the blocks mined based on a grade measurement at a specific location of the block can however be predicted unbiasedly by the grade at a random location in this block. We thus propose to consider each exploration measurement as a potential measurement of the production phase used to decide whether or not the block is processed. In each of these imaginary sampled blocks we need a second observation at a random location. Conditionally on the first measurement being above cutoff the conditional expectation of the second measurement is now the conditional expectation of the block average. The mean of these observations thus unbiasedly predicts the average grade of the mining blocks which will be processed. If the distribution of the relative location difference of the production phase measurement and a random location in the block is symmetric or if the random field is assumed to be isotropic, the roles of the two measurements can be exchanged and we can add to the conditional dataset those first measurements where the second measurement is above cutoff. Although the two parts of the dataset would not be independent, averages are still unbiased. This approach calls for a different strategy of sampling, where pairs of two measurements with a specific distribution of relative locations are taken. We call this the **paired sampling strategy**.
With this new sampling we can predict the average grade of the blocks mined unbiasedly as the average of those exploration observations, for which the other observation in the pair is above cutoff. The expected “Mean Production” of the mine can be estimated unbiasedly by this average multiplied with the portion of these exploration measurements above cutoff. This is indeed the same as assigning zero production to each of the virtual mining blocks with a production measurement below cutoff and averaging over blocks with random location in the mine.
This is exemplified in Figure 4. It is like Figure 3, but using paired sampling and adding the proposed corrected estimators “Mean Other Over Cutoff “ estimating “MeanGradeProdcution” and “Better Estimated Mean Production” for “MeanProduction”. “Mean Other Over Cutoff” is

the mean of the values where the other observations in the pair is over cutoff. The simulation shows clearly that the bias is practically removed compared with the classical “Mean Over Cutoff” estimate. “Better Estimator Mean Production” is the “Mean Other Over Cutoff” estimator multiplied with the average portion of observations in the exploration measurements. Again the bias is visually removed.
These predictions do not include the deposit selection correction and no deposit selection is done in this simulation. Mining only those deposits, where the estimated Mean Production is above cutoff however introduces a second kind of bias.
For the same spatial coverage we need twice as many measurements in a paired sampling campaign than in a classical campaign. We could thus only afford half as many pairs as we had sample before. This does not reduce the number of samples taking part in the averages, as each pair can show up in the average up to two times. However it induces a strong correlation between the observations increasing the variances of the estimates.´ For the sake of comparision the paired sampling simulations has thus been done with only 15 pairs of measurements. Comparison of figure 3 and 4 shows a slight increase in variation, which is however small in comparison to the reduction in bias.

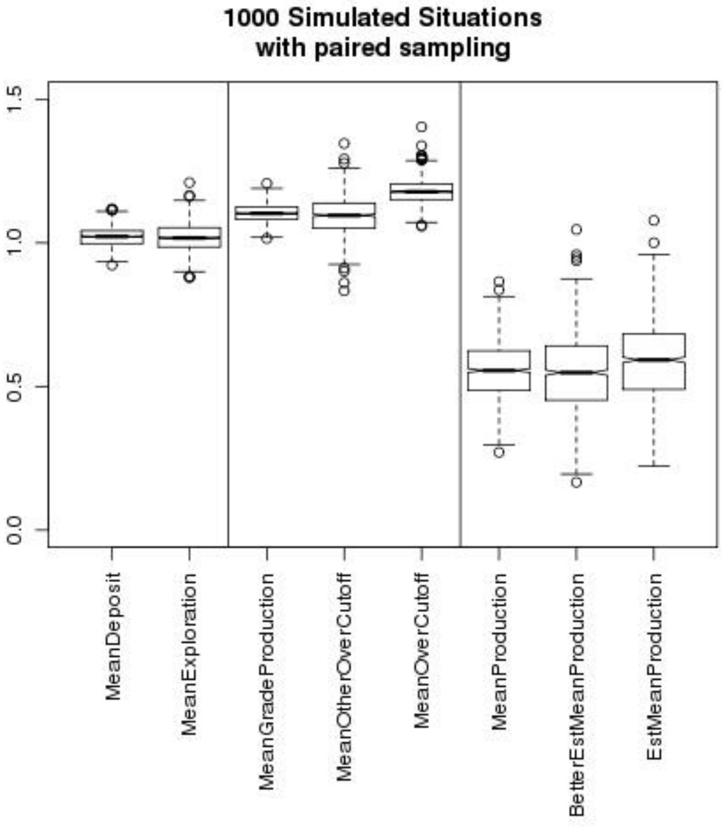

Fig.4: Effects of the correction for block selection bias.

4. THE DEPOSIT SELECTION BIAS

Let us start from a simple mental experiment. Assume we have an big exploration area, where we detected several similar anomalies hinting to potential deposits. All of these subareas are geologically similar and – although we do not know for now- slightly below the limit to support a profitable mining operation. However since the grades on block level vary a considerable portion of paired measurements will show values above our cutoff. For each subarea the exploration sample can be seen as a new sample of the same statistical population of pairs. By chance some of the deposits with thus seem profitable due mean high by chance. The lower the true average grade of the deposits, the lower the portion of these deposits will be. However finally we are only interested in these and for all these the true production is overestimated.

However not all deposits are the same. The true mean production varies and its unbiased predictions vary even more. In principle we would need to know the full joint distribution of the production of a deposit and its prediction. This however cannot be estimated, for example, because deposits are very different and because deposits with low predicted grades are not even mined.

Figure 5 demonstrates the effects of the deposit selection bias and its correction in with the 1000 simulated datasets. Each of the four panels (a)-(d) shows a dot for each of the 1000 simulated possible grade fields. “Content actually mined” forming the y-axis in panels (a)-(c) is the mean content that would be processed if only the blocks with a center measurement over cutoff 1 are mined and processed. The x-axes in these plots show different predictions of the total content. “Predicted minable content” in Panel (a) is the classical estimate of the content as in Estimated Production in figure 4. The choice mining a deposit is done based on such an estimate. The critical content was set to 0.5, such that deposits in the green half of the graphic

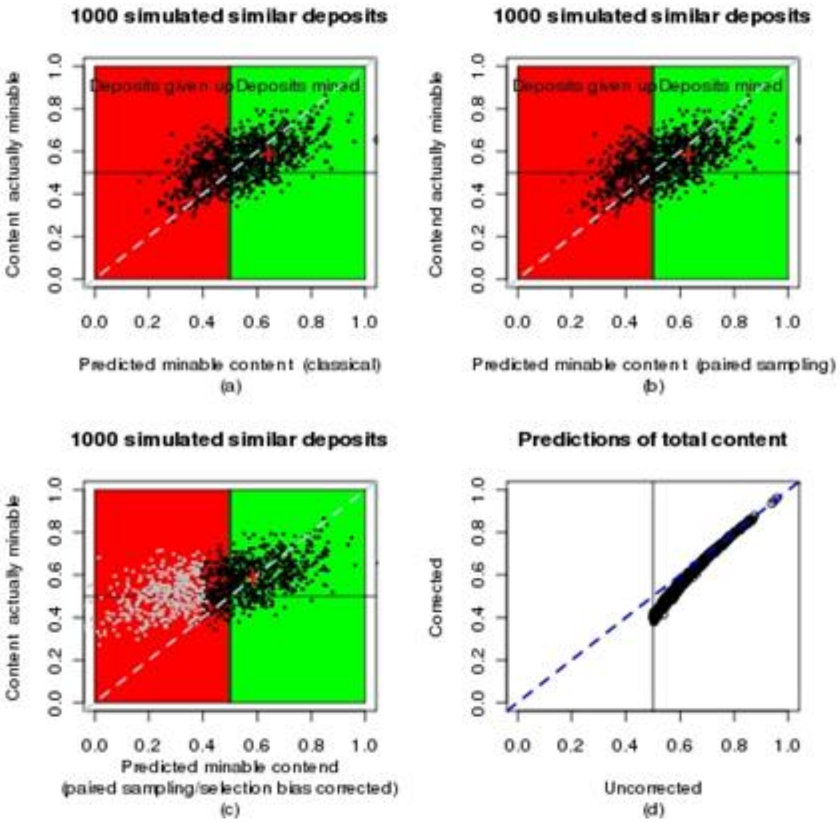

Fig.5: A simulation of 1000 deposits. We assume that only deposits with estimated producible content above a cutoff of 0.5 kg per ton are going into production.

(a) Using the classical estimate
(b) Using the block selection bias corrected estimator from section 3
(c) Using the block and deposit selection bias corrected estimator from section 5
(d) The block selection bias corrected estimator, correct and uncorrected for the deposit selection bias.

would be mined and those on the red half not. The red dot shows the mean over the deposits mined. Each panel has gray dashed x=y line. Panel (a) shows the red dot substantially below the line, implying that we overestimated the total production. Panel (b) of figure 5 shows the same thing using the improved prediction proposed in section 3 instead. The bias is reduced but not completely removed, because it is not corrected for the deposit selection bias. The reason for the bias is similar to the block selection bias. Clearly some deposits with a potential production above the critical content will have a predicted content below this and thus not be mined and vice versa. Unlike for blocks there is however unfortunately no possibility of doing a second conditional sampling of the mine.

It is easy to prove that a deposit selection bias is always present, because conditioned to the actual expected production in the deposit our predictor of the deposits value is unbiased. If now it is conditioned to high values of the estimate the expectation must increase. This increase will be small for high grade deposits which will very likely be predicted being above cutoff and comparably large for deposits near the profitability limit.

The interpretation of this kind of bias is quite complicated: The total production for a deposit can be predicted unbiasedly with the method proposed in section 3. So if we would mine each and every deposit we explored in the long term the mean of productions would converge to the mean of the predictions. However if we mine the deposits only when the predicted content is above some limit a long sum over all these selected deposits productions will fall short of the sum of the predicted productions for these deposits.

5. CORRECTING FOR THE DEPOSIT SELECTION BIAS

We neither have a possibility of repetitive sampling nor a parametric statistical model nor a prior for the true distribution of the mean values of deposits. We don't even have a sample of similar deposits. All we have is

- A small sample of pairs (A_i,B_i) from an otherwise unknown statistical population.
- A nonlinear transformation of these pairs to a variable X_i of which the mean is taken as our unbiased estimate of expected content (generating earnings). The variable is $X_i = 0.5\,(A_i\, 1_{Bi>c} + B_i\, 1_{Ai>c})$, where A_i and B_i are the two measurements, c is the Block cutoff and 1_* is the indicator function of condition *.
- A nonlinear transformation of these pairs to a variable Y_i of which the mean is taken for our unbiased estimated of the expected portion of the deposit to be processed (generating processing costs).The variable is $Y_i = 0.5\,(1_{Ai>c} + 1_{Bi>c})$.
- This results again in a pair of dependent variables (X_i,Y_i).
- In a simplistic economical model the unconditional production value V of the deposit can be predicted unbiasedly as a linear function

  $$V= (p_M (\sum X_i/n - g_r)- p_P \sum Y_i/n)T - p_I$$

  of these pairs, where n is the number of pairs, p_M is the discounted selling price of the produced raw material, p_P are the discounted processing costs per ton ore, T is the total tonnage of the sampled deposit, p_I are the discounted investment costs and g_r is the residual grade that which cannot be recovered during processing.
- The deposit is selected if V is above some profitability cutoff c_p for the whole deposit.

If the statistical population would be the sample, indeed V would be the true value of the mine. However conditioning to those cases with V above cutoff, we introduce a positive bias. The magnitude of the bias however depends on the distribution of the statistical population of pairs itself. Our best estimate for this population is the sample itself. We can draw random samples

(with replacement) from this population to simulate the effect of the deposit selection. Due to the unbiasedness of the estimators we expect estimators applied to this resampled dataset to have the V of the original dataset as expectation. However conditioning now to estimates above cutoff will again introduce the selection bias. The conditional expectation of the newly predicted V minus the original V conditioned to estimated values being above c_p is thus the bias the predictor would have for this smaller but similar statistical population. A mean over such simulated differences can thus be used as an approximate computation of the selection bias in the current situation. Such techniques are commonly known as nonparametric bootstrap (e.g. Davison&Hinkley 1997). Subtracting the bootstrapped bias from the estimate provides a new estimate of the expected value of the mine conditioned to being selected.
A prove for precise properties of this correction is not yet available. Figure 5 shows example calculations for 1000 simulated deposits. Instead of focusing on V, which is unknown because we don't have prices, we only look at the critical part, i.e. the predicted content $\sum X_i/n$.

Panel (c) of Figure 5 shows deposit selection bias corrected estimates on the x-axis. Of course this bias correction does not make sense for the deposits that won't be mined, which are thus plotted in gray. The decision whether or not a deposit is mined is still done based on the paired sampling based content estimation. The bias corrected predictor however predicts some of the deposits now below the 0.5 cutoff, saying that we must expect that the deposits estimated only slightly over cutoff will indeed on average have a content below cutoff. In this panel the red dot showing the average for estimate and true value of the mined deposits is indeed on the x=y line expressing the approximate unbiasedness of the corrected estimator. Panel (d) shows the effect of the correction plotting the deposit selection bias corrected predictions against the predictions not corrected for a deposit selection bias. The correction is stronger for low predictions corresponding to the intuition that for a deposit with observed low values it is more likely to actually have a true mean below the cutoff than for those, for which high values were observed.

7. CONCLUSIONS AND OUTLOOK

The paper explains two sources of bias in mining statistics, which are generated by the rules we apply for exploiting the mine: Mining is only started, if the deposit looks profitable enough, a block is only processed if its concentration is high enough. It provides solutions for a very simple setting, which ignores varying cutoffs, dependent decisions on blocks, problems of block accessibility, the ultimate pit problem and any sort of discounting or spatial structure. If other decision rules are used, the actual production would be different and thus other quantities need to be estimated. However - starting from this first simple approach - estimation rules for more complicated situations might be derived, whenever the precise situation is clarified.

REFERENCES

Chambers, J. M., W.S. Cleveland, B. Kleiner, B., P.A. Tukey, 1983, Graphical Methods for Data Analysis. Belmont, Wadsworth International Group, 395 p.

Cressie, N.C., 1993, Statistics of spatial data: New York, Wiley, 928p.

Davison A.C., D.V. Hinkley, 1997, Bootstrap Methods and their Applications: Cambridge University Press, 582 p.

Dimitrakopoulos, R. and Ramazan, S. (2008): Stochastic integer programming for optimizing long-term production schedules of open pit mines: Methods, application and value of stochastic solutions. IMM Transactions, Mining Technology, v. 117, no. 4, pp.155-160

King HF, McMahon DW, Bujtor GJ (1982) A guide to the understanding of ore reserve estimations. Proc Aust Inst Min Metall Suppl 281

R Development Core Team, 2010, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org

Wellmer, F.-W., 1997, Statistical Evaluations in Exploration for Mineral Deposits: Berlin, Springer, Berlin, 379 p.