Some Thoughts on the Value of Incremental data in Geostatistical Modeling

Karl P. Norrena (knorrena@ualberta.ca) and Clayton V. Deutsch (cdeutsch@ualberta.ca)
Department of Civil & Environmental Engineering, University of Alberta

Abstract

Reservoir models are updated as new well data become available. These updated models provide the basis for future decisions such as modifications to the locations of wells that have not been drilled. We perform numerical studies to quantify the value of incremental information for decision making.

The first numerical study shows that incremental data does not always improve decision making. In particular, there is no improvement when the incremental data only reduce local uncertainty; new wells are often drilled outside the range of correlation of previous wells, thus those previous data do not help much. A second numerical study shows that incremental data can be very valuable if it changes the conceptual model of the reservoir or provides significant new data about the geostatistical modeling parameters.

Introduction

A critical step in geostatistical modeling is to devise a modeling approach and required parameters that are consistent with all sample data. The modeling approach includes decisions regarding rock type modeling and the suitability of Gaussian techniques for continuous variables. Important parameters include the histogram of each variable, the spatial correlation of each variable, and the relationship between variables.

Modeling errors affect decision making. A lack of data is often the root cause for inappropriate conceptual models. The best way to proceed with a modeling exercise is to use all of the data at hand to revise the conceptual model and the geostatistical model at each stage of the reservoir life cycle. As reservoir development proceeds, the additional data can be used to incrementally update the model and aid in future decisions. It is difficult to keep track of ever-changing models and how the decision making evolves with incremental data. We aim to understand more clearly how incremental data improve the model with no change to the conceptual model and how incremental data help when the conceptual model changes.

We are concerned with the placement of wells in reservoir development. The idea is to place the wells using information from previous wells to better select the next well location. Proceeding in this fashion should result in well locations that improve recoverable reserves. Evaluating the value of added information requires a model for what that new data might be. We simulate a set of possible "truth" models and extract incremental data from this set of truth models. The value of using incremental data for decision making can be assessed relative to decision making without using the incremental information and the possible truth model. Of course, the simulated truth model provides no new data; it simply provides a means to assess incremental information. The proposed procedure is described in detail.

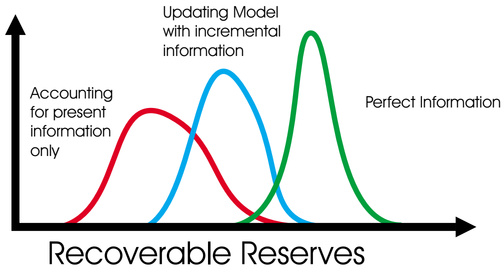

Figure 1: Schematic illustration of histograms of uncertainty in recoverable reserves accounting for exploration data (red on left side), updating the model and well locations with incremental information (blue in the middle). The distribution of reserves given "perfect" information is the green distribution on the right.

The paper starts with an example where we show that simply adding incremental information to the model with no updating of the conceptual model or parameters does not yield better decisions. Then, another example illustrates that a change in the conceptual model improves decision making.

Methodology 1

Consider the problem of selecting N wells in petroleum reservoir development. Choosing N positions all at once with exploration data is not expected to be as good as using the well data incrementally. The idea is to start with a model that represents our current state of uncertainty and proceed by adding information to update this uncertainty. At each step a decision is made and the value of the added information is assessed. Of course, the very best decisions are made with the inaccessible true reservoir description. We have color-coded these three cases (see Figure 1):

1. The *red* case (histogram on the left) is the decision using only exploration data for all wells. It is important to note that all wells are placed together to be jointly optimal.

2. The *blue* case (histogram in the middle) shows the distribution of reserves where each well location is chosen using the exploration data and all previous development wells. This is the best approach in practice and should be better than the red case.

3. The *green* histogram (on the right) shows the distribution of reserves if the true reservoir description were available. This will be better than the red or blue case.

True reserves are unique and would appear as a "spike" on a histogram. The true reserves are not accessible in practice, however, and we substitute the model of uncertainty as the "true" reservoir and suppose that each realization represents a plausible truth. The distribution of reserves is found by measuring the reserves obtained by optimally locating the required number of wells on each realization.

This simulation approach to predicting the value of incremental information is shown schematically in Figure 2, and explained in more detail below:

1. The starting point is a set of truth reservoirs, R_l, $l = 1, \ldots, L$. These truth reservoirs are the set of geostatistical realizations that represent the current state of uncertainty, that is,

they reproduce all available exploration data. Note that each reservoir model R_l is a full specification of the reservoir structure and all internal heterogeneity.

2. Apply the best available algorithm to determine the optimal number of wells, N, and their locations over the set of true reservoir models $\{u_i, i = 1, \ldots, N\}$. Note that there is only one optimal number of wells and set of locations. It is unreasonable to determine the best set of wells on each realization because there is no practical way of reconciling differences in the L different well plans afterwards.

3. Calculate the recoverable reserves on each of the L realizations and construct a distribution of recoverable reserves given the initial state of uncertainty. This is the *red* histogram in Figure 1. Set $M = N$, where M will be the remaining wells to drill and N is the optimal number considering the present state of uncertainty.

4. The following steps are to be considered for each of the $l = 1, \ldots, L$ true reservoir models.

   a. Choose one of the M wells to drill based on logistical drilling considerations or simply the one with the largest expected recoverable reserves. Extract the reservoir properties and structure at that location from the current (l) realization to be considered in building a new set of realizations with the added information. Reset M to $M - 1$ once one of the wells is chosen.

   b. Construct another suite of geostatistical realizations conditional to all available information including the original data and the data from all previously drilled $(N - M)$ simulated wells. Clearly, this procedure must be automated in a script or in some commercial software such as $Jacta^{TM}$.

   c. Run the optimal well placement to refine the position of the remaining M wells using the latest suite of geostatistical realizations. These new locations will be considered to select the next one. Loop back to b if $M > 1$; otherwise, proceed to e and then keep looping over L.

   d. Calculate the recoverable reserves using the updated well locations and the current simulated true reservoir R_l. This is one number to go into the blue histogram of Figure 1. The set of updated reserves make up the blue histogram.

5. Find the unique optimal well locations for the N well on each of the L realizations. Use these locations to calculate the production for each realization. This set of L numbers make up the green histogram.

This simulation exercise provides a measure of how incremental information improves the ultimate reservoir decision-making or profitability. We acknowledge that no new information is being considered in this exercise; in that sense it is similar to the bootstrap technique.

First Example

The setting of the first example is a model of uncertainty for a synthetic reservoir on a regular $50x50$ grid. The model consists of porosity and porosity-derived geo-bodies. Porosity is conditionally simulated using sequential Gaussian simulation (SGS) and the four conditioning data shown in

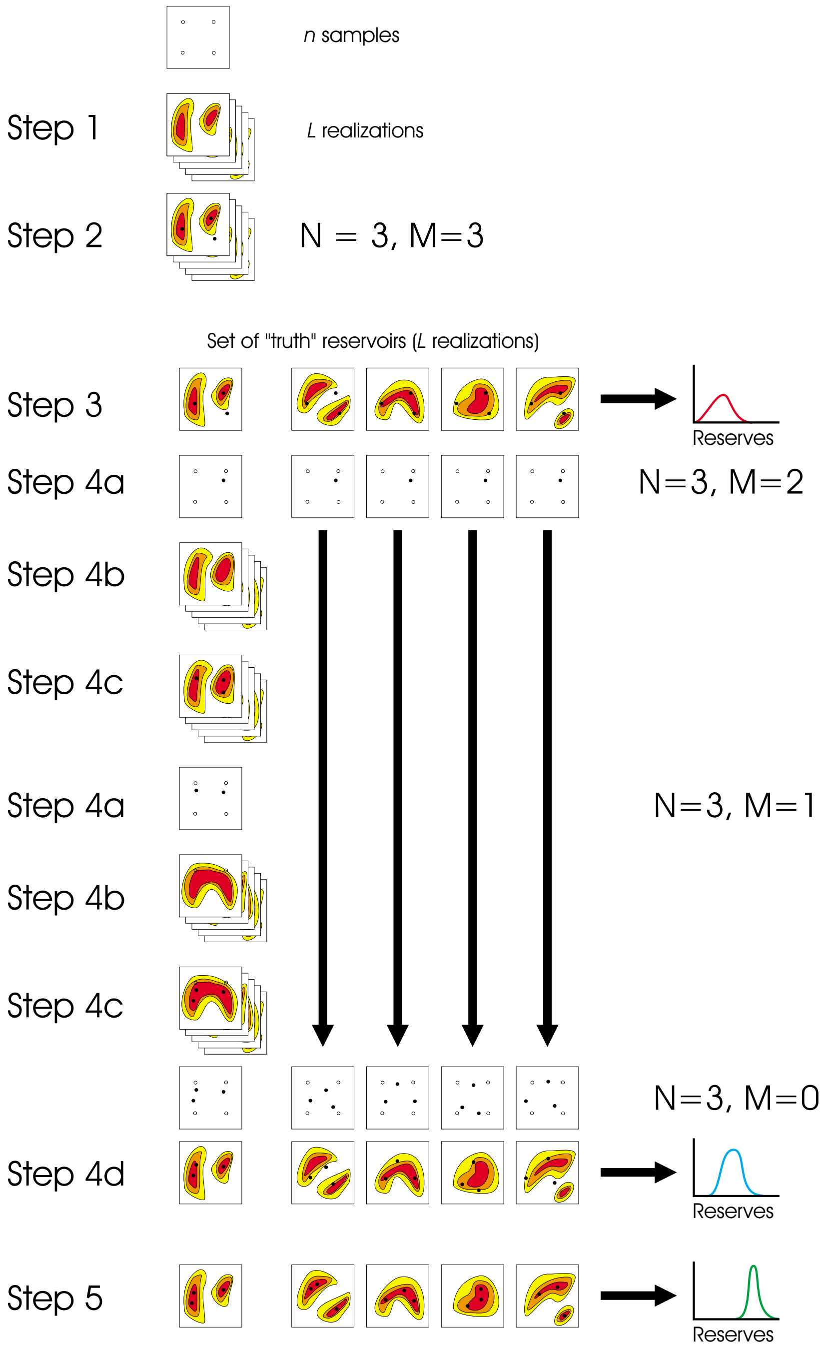

Figure 2: Schematic illustration of a procedure to assess the value of incremental data.

Figure3. Porosity is first simulated as a normally distributed variable and later transformed to porosity using the simple linear transform to avoid issues surrounding the Gaussian transform: $z = 5 \cdot y + 10$, where y is the simulated Gaussian value and z is the porosity value. A histogram of porosity over 100 realizations is shown in Figure 4. The geo-bodies represent a set of connected blocks having porosity greater than 7.5md. Figure 5 shows a simulated porosity map and the corresponding geo-body map. We select optimal well locations by calculating a well location performance metric. For each node in the grid we calculate a quality parameter:

$$quality(\mathbf{u}) = \phi(\mathbf{u})(1 - S_w) \cdot C$$

where S_w is the water saturation and C is related to the selling price per unit hydrocarbon. For simplicity we keep S_w and C constant. The performance of a well locations is evaluated by calculating the sum of all the quality parameters falling within the drainage radius and belonging to the same geo-body intersected by the well path. In Figure 6 we show a geo-body map with four well locations superimposed on the map. The well path is indicated by a dark pixel. The drainage radius is indicated by a shaded circular region around the well path. The well located in the bottom right of the map would have the highest quality because, as indicated in the corresponding porosity realization shown in Figure 5, it is in a location of high porosity and the drainage radius is almost completely occupied by a geo-body. The well location of the lowest quality is shown in the top right.

We submitted 100 realizations of our synthetic reservoir to the incremental data algorithm with the goal of sequentially placing four wells. Figure 7 shows the results for steps 3 through 5 for a single realization. The map at the top of the figure corresponds to step 3: a realization of a geo-body map with four optimally placed wells superimposed. The set of maps below the step 3 map are each one realization of the updated model of uncertainty with the updated optimal well locations corresponding to step 4c. Starting from the right, the maps show the first through the fourth fixed wells. A heavy circle is drawn around the fixed wells. The map on the far left is the map for step 4d: the updated well locations superimposed on the geo-body realization in step 3. Below the set of step 4 maps is the map for step 5: the four well locations optimised to the realization.

The Red, Blue and Green Histograms for 100 realizations are shown in Figure 8. The mean of the Red Histogram is 7476, the Blue Histogram is 7302, and the Green Histogram is 10596. Over 100 realizations updating does not lead to better well locations. The results were verified and extensively tested.

Comments

The results of the first experiment are an apparent violation of common sense: using incremental data should lead to better decision making. We showed that incremental information did not help in the decision making. The primary reason for this is that the model is only refined within the range of correlation which is about the same as the drainage radius of the well. The updated portion of the model is never available for decision making because it is already occupied by a well.

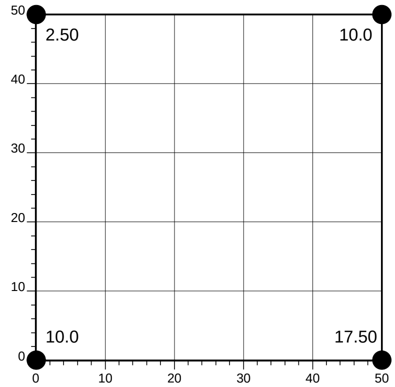

Figure 3: The four conditioning data are positioned at the corners of the grid.

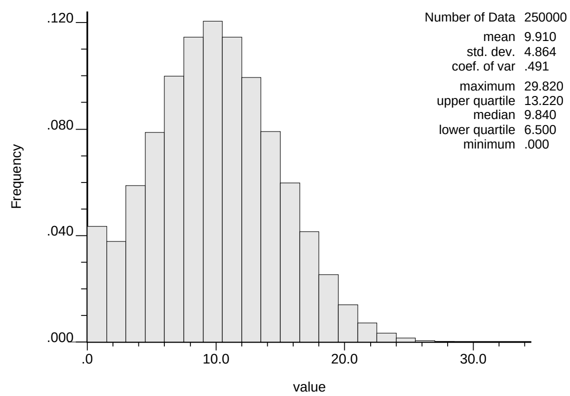

Figure 4: The distribution of transformed porosity for 100 realizations.

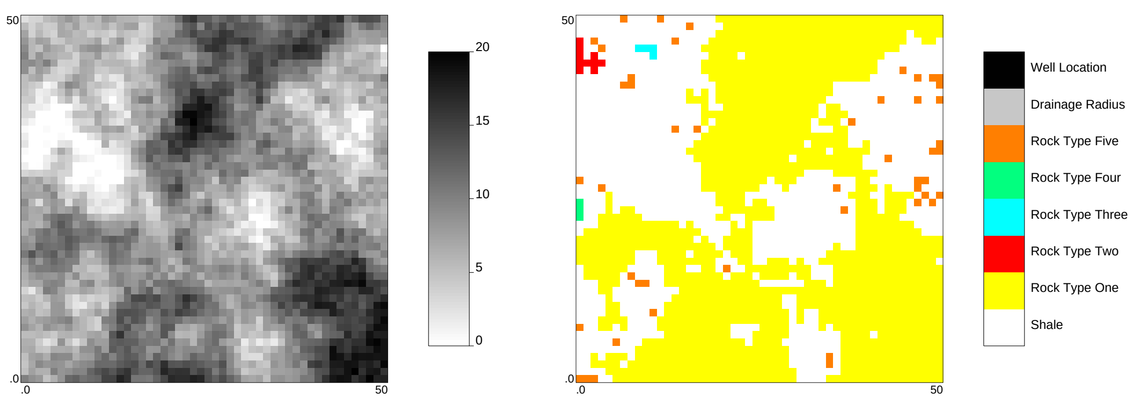

Figure 5: One porosity realization and the corresponding geo-object realization.

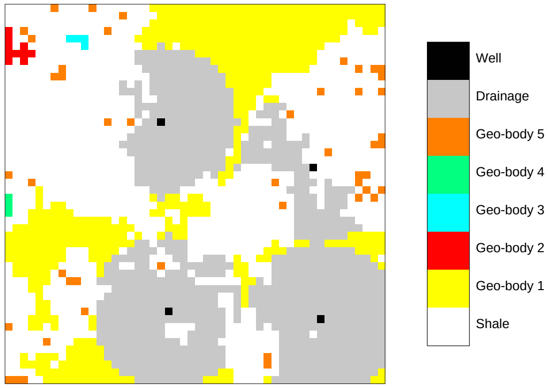

Figure 6: A geo-body map with four well locations superimposed. The well locations are denoted as a single dark pixel surrounded by a shaded circular drainage radius. Referring to the porosity map in Figure 5, the well location on the bottom right corner has the highest quality, the well on the top right has the lowest.

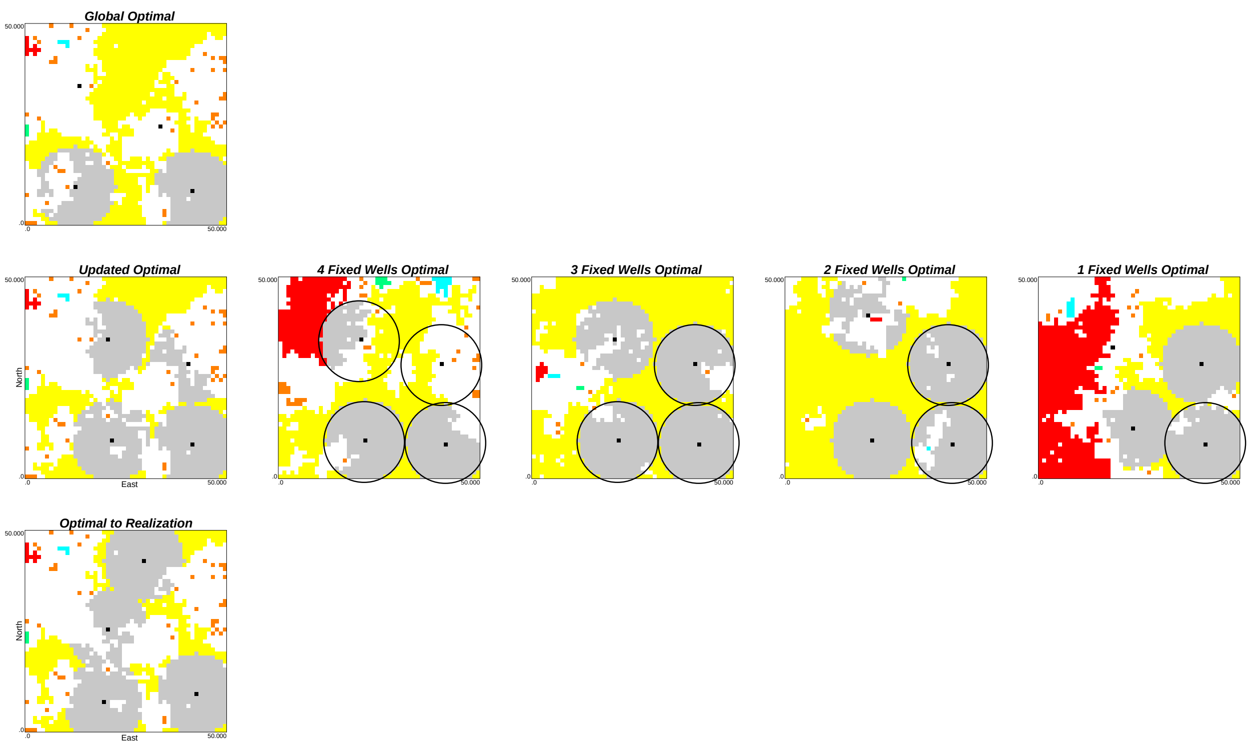

Figure 7: Together the maps represent one leg from Figure 2. The maps on the left are the geo-object maps and the placed wells for one realization for step 3, step 4d, and step 5. The maps on right from the map showing step 4c show 1 to 4 fixed wells as indicated by the heavy circles.

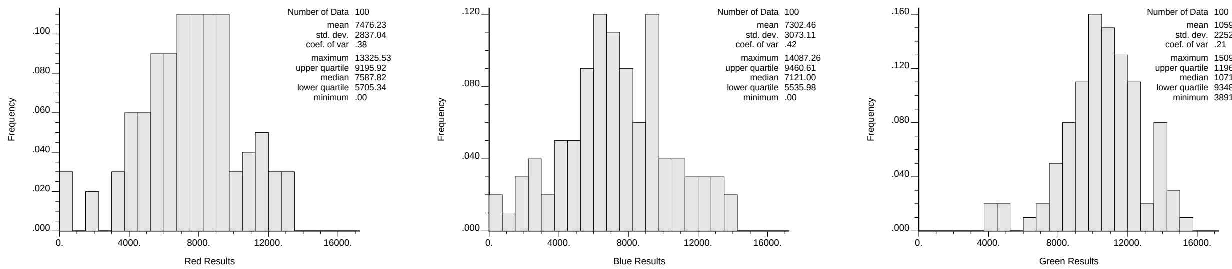

Figure 8: The red histogram is on the left, the middle one is the blue histogram, and the right is the green histogram. Note that over 100 realizations updating does not lead to better decisions.

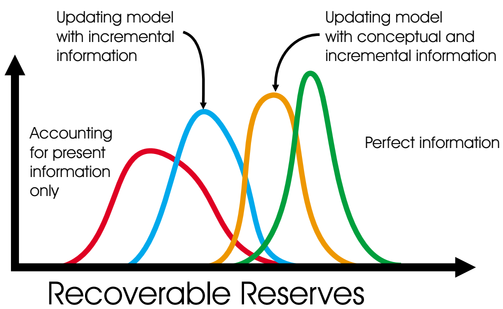

Figure 9: Schematic illustration of histograms of uncertainty in window quality accounting for exploration data (red on left side), updating the model and window locations with incremental information (blue second from the left), updating the model and window locations using the conceptual information and the incremental data (orange second from the right). The distribution of reserves given "perfect" information is the green distribution on the right.

Methodology 2

The results of the first experiment motivate a second more simplified experiment where we examine the impact of changing the conceptual model and using incremental data versus using only the incremental data. We set the goal of this experiment to be the sequential selection of two optimal locations or "windows" that maximize quality Q. We quantify the quality of a window location by taking the sum of all the values falling within the window. This is similar to optimally placing two wells of maximum quality except that in this experiment a square window is used instead of a drainage radius.

We retain the color coding scheme outlined previously but add one more histogram: the histogram that represents the use of a revised conceptual model and incremental information for decision making. We expect that this distribution will have a mean greater than that of using only incremental information. Thus we expect that the distribution will fall between the blue distribution and the green distribution. We call this the orange histogram and it is shown in Figure 9.

The methodology for this experiment is a variation of the algorithm described above. We start by constructing a "true" conceptual model consisting of a map of secondary data (Y_{truth}), a map of primary data (Z_{truth}) and a randomly selected correlation coefficient (ρ^l) correlating the primary and secondary data. The Y_{truth} is intended to mimic the type of information that would be provided by seismic data. The pseudo-seismic information and the correlation coefficient are used to construct a map of a primary reservoir parameter. We proceed as follows:

1. Draw a random correlation coefficient ρ^l either +/-, where l is the l^{th} trial for $l = 1, \ldots, L$. We alter the correlation between the primary and the secondary information over a large number trials to observe the effect of using an inappropriate conceptual model.

2. Simulate Y unconditionally and call this the true secondary information Y_{truth}.

3. Simulate a Z using y as collocated with ρ^l. Call this Z_{truth}.

4. From the corners of the Z_{truth} map collect four conditioning data. This data is considered exploration data. At this point the conceptual model consists of the exploration data and the same variogram as used in the construction of Z_{truth}.

5. Simulate 20 realizations of the Z variable using the four conditional data. At this stage we proceed with the construction a geostatistical model with all available data.

   - Find the two best locations, $i(ix, iy)$ and $j(jx, jy)$, for the windows that maximize Q^*, , where Q^* is the window quality at locations i, j given one piece of information: the exploration data. The two best locations are found by searching the grid exhaustively. Overlapping windows are accounted for by assigning the blocks in the overlapping area to only one window.
   - Using the two optimal locations $i(ix, iy)$ and $j(ix, iy)$, calculate Q^l_{red} from the truth model. The distribution of the Q^l_{red} values is the red histogram.

6. Randomly select one of the two locations i or j. Go back to the truth model Z_{truth} and extract a new conditioning data.

7. Simulate 20 Z realizations using only the 5 conditioning data.

   - Find the second location that maximizes Q^{**}, $i(ix, iy)$, $j(jx, jy)$ respectively, where Q^{**} is the quality for the window locations using two pieces of information: the exploration data and the incremental information. This step represents decision making using only incremental data.
   - Calculate Q^l_{blue} from the truth model.

8. Cosimulate 20 Z realizations using the 5 data, Y_{truth}, and ρ^l.

   - Find the second location that maximizes Q^{***}, $i(ix, iy)$, $j(jx, jy)$ respectively, where the third piece of information is the conceptual model. This step represents decision making using both the conceptual model and the incremental data.
   - Calculate O^l_{orange} from the truth model. The distribution of the O^l_{orange} values

9. Using Z_{truth} find the two locations i, j that maximize O_{green}. The distribution of O_{green} values is the green histogram in Figure 9.

10. Repeat over L trials to observe the distributions of uncertainty.

Second Example

The experiment occurs on a 20 x 20 grid. We randomly select the correlation between the primary and the secondary information to be either 0.75 or -0.75. The secondary map is constructed unconditionally using SGS and a Gaussian variogram having a long range and minimal nugget effect. The primary map is also constructed unconditionally but uses an exponential variogram with moderate nugget effect. The simulated values for both maps are normally distributed: no transform was applied. Figure 10 shows a map of Y_{truth} on the left and Z_{truth} on the right. The corresponding histograms are shown below. The observed correlation coefficient is 0.75.

We show the mapped results for one trial and for both of the optimal locations selected in the Red case. Collectively the following figures represent a single trial in the second experiment. For each of the figures, the histogram of Gaussian values over the 20 realizations is shown on the left. The map in the middle is the e-type estimate over 20 realizations, and the map on the right is a single realization. The optimal window locations are superimposed on the maps. We show the histograms to illustrate the effect of bias sampling. The optimal window locations are preferentially located in high valued areas. Using the data extracted from Z_{truth} conditions the realizations and artificially increases the mean. The e-type map is shown because it illustrates what the optimisation routine "sees" when searching for the two optimal locations.

Figure 11 shows the results for the Red case. Two optimal locations are in the bottom left and bottom right corners of the maps. The mean of the realizations is not the same as the mean of the Z_{truth} map (-0.236 vs. 0.027, respectively). In practice there is no way to know if the sample mean is equivalent to the true mean, but the sample mean represents one aspect of our conceptual model. Figure 13 shows the results after updating using the optimal window location on the bottom right in the Red case. The updated location is the same as in the Red case. The mean is 0.07. The mean is much higher in the Blue case because of the high values conditioning data. Figure 12 shows the results after using the conceptual model. The second window location is better than the selection made without the conceptual model. The mean is -0.009. Figures 14, and 15 are the same as above but use the incremental data from the bottom left window location in Figure 11. The respective means are -0.001, and 0.032. Note that all of the updated means differ significantly from the mean in the Red case. Despite the fact that they are closer to the mean of Z_{truth} and all represent differences in the conceptual model. Figure 16 shows the optimal location for the Green case: perfect information.

Comments

The resulting histograms for 100 trials are shown in Figure 17. The mean quality of the Red Histogram is 50.6, the Blue Histogram is 61.0, the Orange Histogram is 88.4, and the Green Histogram is 113.7. In this case using only the incremental information leads to better optimal location selection, but only marginally. Using both the conceptual model and the incremental information lead to significantly better location selection on average.

A summary of results is shown in Table 1. Note the number of times the Red case exceeded the Blue and Orange cases ($Red > Blue$, $Red > Orange$), the Red case was equivalent to the Blue and Orange case ($Red < Blue$, $Red < Orange$), the Blue and Orange cases were equivalent to the Green case ($Green = Blue$, $Green = Orange$). The tabulated results show that using just the incremental information does not clearly lead to better decision making even though the Blue

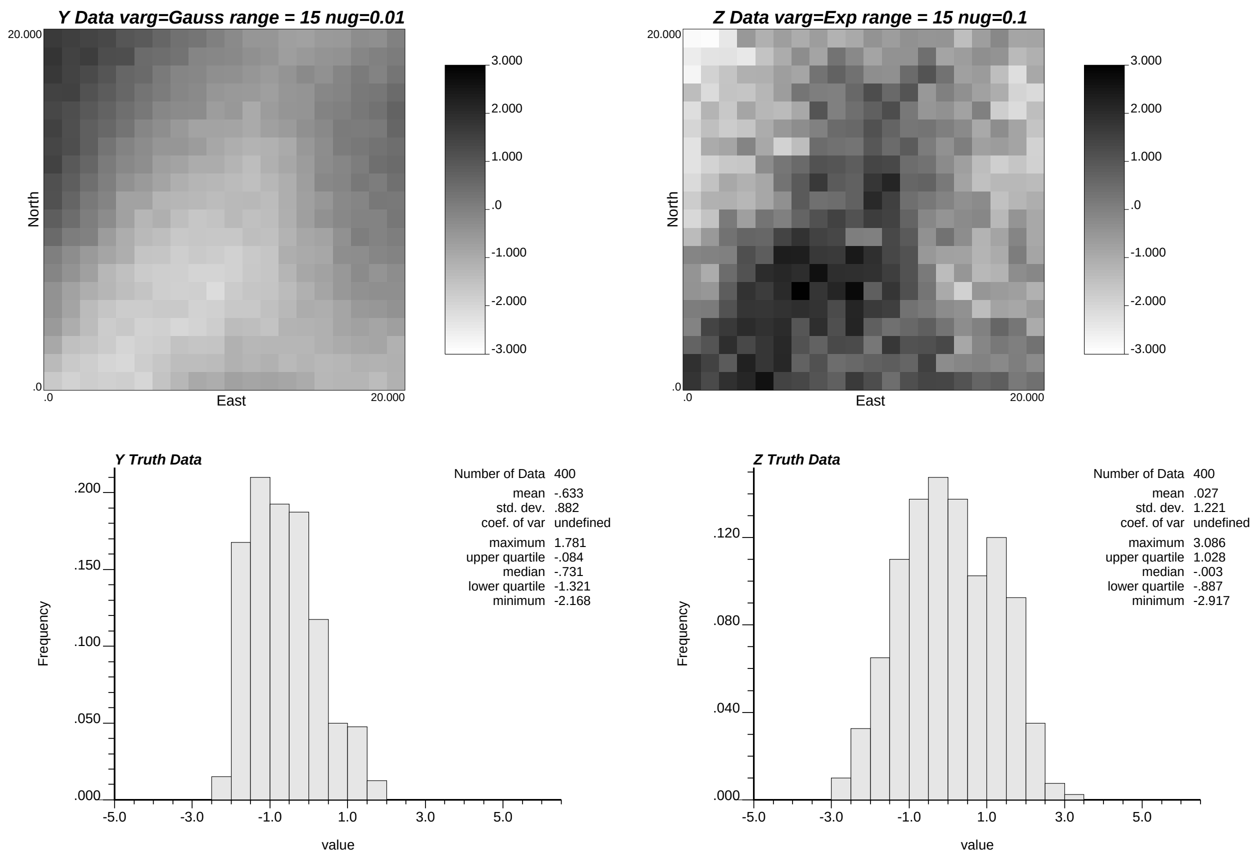

Figure 10: The map on the left is the secondary information, the Y information. The corresponding histogram is below. The map on the right is the Z map. The secondary information was used to construct this map. It corresponding histogram is on the right. The correlation coefficient is -0.75. The mean of Z_{truth} is 0.027

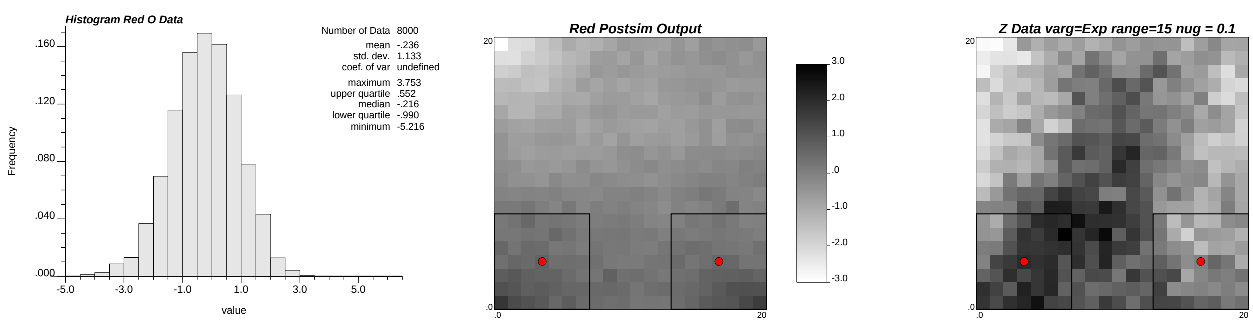

Figure 11: The histogram for the 20 realizations is shown on the left. The mean is -0.236. The map in the middle is the e-type map conditional to four data taken from the corners of the Z_{truth} data shown in Figure 10. The selected locations are shown as a small circle with a square window surrounded it. The map on the right shows the locations superimposed onto the Z_{truth} map.

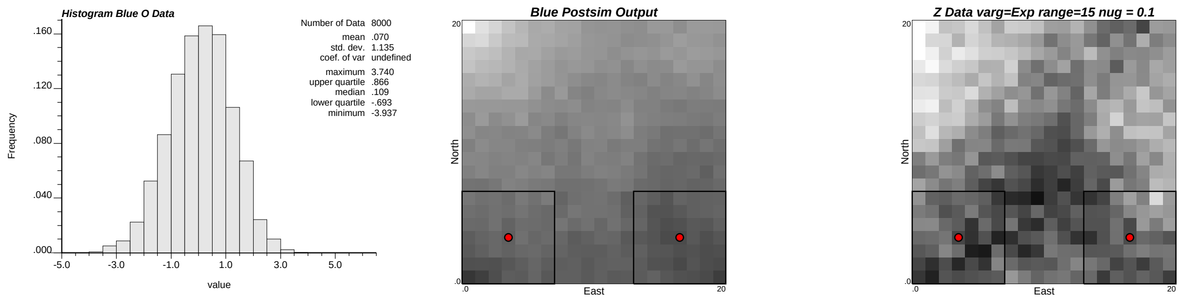

Figure 12: The results using the incremental data from the Red case. The location on the right was selected. On the right is the histogram of the realizations constructed with the 5 data. The mean is 0.07. The map in the middle is the e-type map over the 20 realizations and using the 5 conditioning data. The map on the left shows the locations superimposed on the Z_{truth} map.

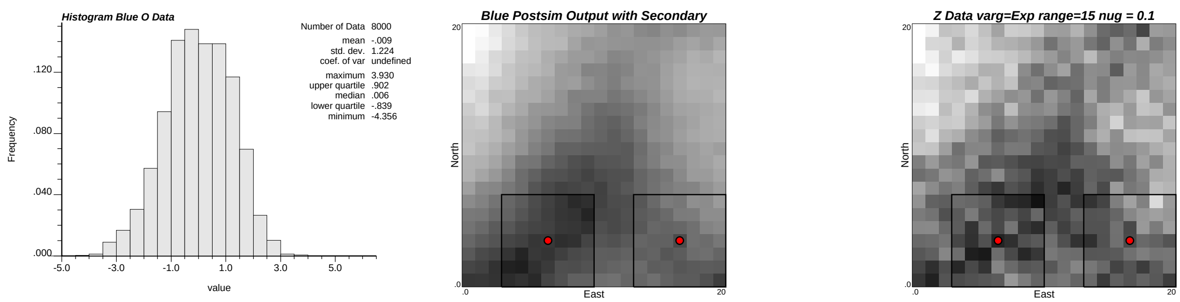

Figure 13: The results for the Orange case: the incremental data from the bottom right corner of the Red case and the conceptual model was used. The mean is -0.009.

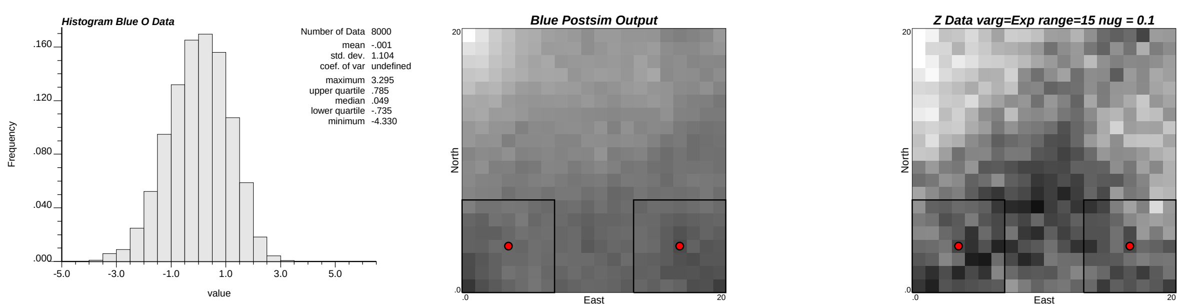

Figure 14: The results using the incremental information provided by the window located on the bottom left corner of the Red case. The mean is -0.001. Note that the incremental information does not change the selected location.

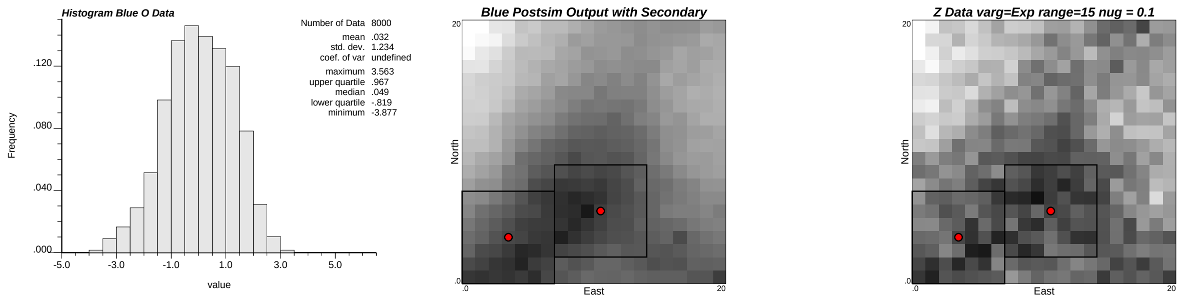

Figure 15: The results using the incremental information from the bottom left corner of the red case and the conceptual model. The mean is 0.032.

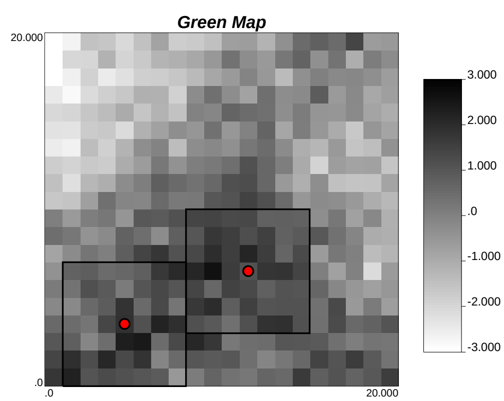

Figure 16: The green results.

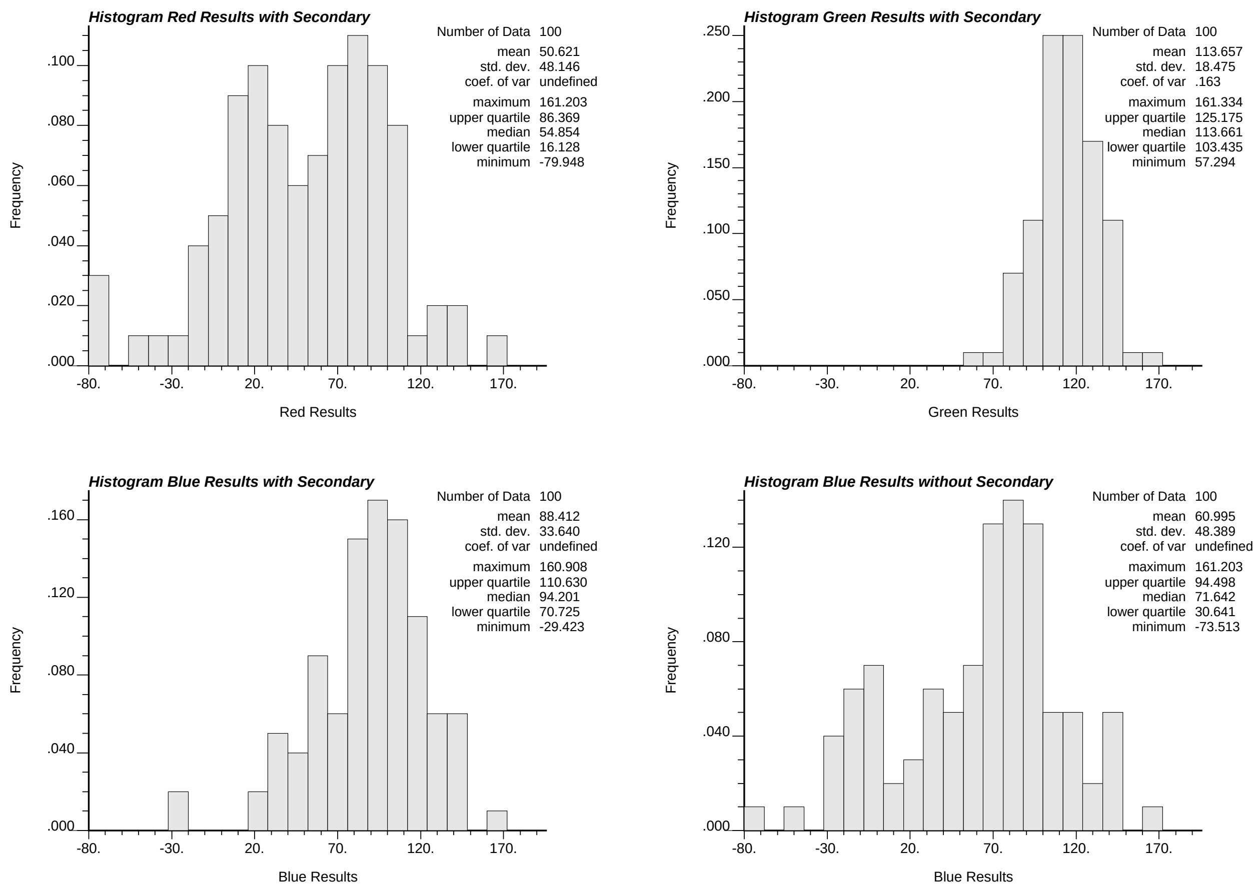

Figure 17: The histogram of results.

	Blue (Percent)	Orange (Percent)
Red >	22	3
Red =	37	19
Red <	41	78
Green =	5	10

Table 1: Summary of results for 100 trials in the second experiment.

Histogram showed higher average quality than the Red Histogram: only 41% of the time the Blue case exceeded the Red case. Comparing the Blue and Orange cases, the Orange case yielded better decisions 78% of the time.

Summary

The results of the first experiment show an apparent violation of common sense: using incremental well data to update a reservoir for subsequent decision making does not always lead to better well locations. The results of the first experiment can be explained: (1) due to the range of correlation the new data succeeded in updating the model near the existing well and provided little additional information for the rest of the model, and (2) the well locations are preferentially located in high valued areas and the extracted samples lead to bias in the simulations.

The results from the second experiment show that although the incremental information did lead to better decision regarding optimal window placement the improvement was not clear. Using conceptual information and incremental information clearly led to better window locations.

In general these findings should not be extended to other decision making problems without further investigation. The problem of selection well locations is peculiar because the region of interest is not allowed to contribute to the next decision for a future well location.

One useful lesson that could be extended to almost any other decision making problems is that of revising the conceptual model as more information becomes available. The conceptual model is the foundation of the entire model; stochastic or deterministic. As we have shown, revisions to the conceptual model has greater impact than that of simply adding data.

References

[1] Beckner, B. L. and Song X.,: Field Development Planning Using Simulated Annealing - Optimal Economic Well Scheduling and Placement, SPE 30650, Annual SPE Technical Conference and Exhibition, Dallas, October 22-25, 1995.

[2] Deutsch, C.V. and Journel, A.G.: *GSLIB: Geostatistical Software Library and User's Guide*, New York, Oxford University Press, (1992), 340 pages.

[3] Deutsch, C. V.: Fortran Programs for Calculating Connectivity of three-dimensional numerical models and for ranking multiple realizations, Computers & Geosciences, 24(1), p. 69-76.

[4] Gawith, D.E., Gutteridge, P.A., and Tang, Z.: Integrating geoscience and engineering for improved field management and appraisal. In *1995 SPE Annual Technical Conference and Exhibition: Formation Evaluation and Reservoir Geology*, pages 11- 23, Dallas, TX, October 1995. Society of Petroleum Engineers, SPE Paper Number 29928

[5] Goggin, D.J., and Gidman, J., and Ross, S.E., Optimizing Horizontal Well Locations Using 2-D Scaled-Up Geostatistical Reservoir Models, *SPE Annual Conference and Exhibition*, Dallas, TX, October 1995, Society of Petroleum Engineers, SPE Paper Number 30570

[6] Hird, K. B. and Dubrule, O.: Quantification of reservoir Connectivity for Reservoir Description Applications, SPE 30571, 1995 SPE Annual Technical Conference and Exhibition, Formation Evaluation and Reservoir Geology, Dallas, TX.

[7] Seifert, D. and Lewis, J.J.M. and Hern, C.Y. and Steel, N.C.T.: Well Placement Optimisation and Risking using 3-D Stochastic Reservoir Modelling Techniques, SPE 35520, presented at the NPF/SPE European Reservoir Modelling Conference, Stavanger, April 1996.

[8] Wang, B., *Well Site Selection Algorithm Considering Geological, Economical and Engineering Constraints.* Master's thesis, University of Alberta, Edmonton, AB, 2000.