Assessing the impact of program selection on the accuracy of 3D geologic models

Kelsey E. MacCormack1 and Carolyn H. Eyles2

1School of Environmental Sciences, University of Guelph, 50 Stone Road East, Guelph, Ontario N1G 2W1, Canada
2School of Geography and Earth Sciences, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada

ABSTRACT

As the field of three-dimensional (3D) subsurface geological modeling develops at an increasingly rapid rate, so too does the number of available software programs catering to these applications, most of which offer very similar ensembles of algorithms for interpolating data. A few studies have analyzed the effect of algorithm selection on the accuracy and uncertainty of subsurface geologic models, but little consideration has been given to the uncertainty and variability introduced into the model by software program selection. In this study, inverse distance weighting (IDW) and ordinary kriging (OK) algorithms are used to interpolate identical data sets by three different software programs (ArcGIS, ROCKWORKS 2006, and VIEWLOG). The results indicate that the output of the IDW and OK interpolation algorithms are inconsistent between programs and that this variability should be considered when assessing the uncertainty associated with subsurface model results. Program selection appears to have a significant influence on model output results when modeling complex subsurface geological environments, particularly when interpolating clustered data, which are most commonly used in geological and environmental applications.

INTRODUCTION

As the demand for three-dimensional (3D) models increases, so too does the number of software programs used for the generation of such models. The majority of these 3D modeling programs offer similar methods for data interpolation. The two most commonly used interpolation algorithms for subsurface modeling applications are inverse distance weighting (IDW) and ordinary kriging (OK; Kravchenko and Bullock, 1999; Johnston et al., 2001; Jones et al., 2003; Mueller et al., 2004). Each of these two algorithms has a very different strength for spatial data processing; IDW is often favored for being computationally “quick and easy,” whereas kriging is favored for its ability to provide the best linear unbiased estimates (Weber and Englund, 1992; Mueller et al., 2004). Several studies have evaluated the effectiveness of these algorithms in producing accurate models (Tabios and Salas, 1985; Weber and Englund, 1992; Weber and Englund, 1994; Brus et al., 1996; Walker and Loftis, 1997; Nalder and Wein, 1998; Zimmerman et al., 1999; Schloeder et al., 2001; Jones et al., 2003; Kravchenko, 2003; Dille et al., 2003; Lapen and Hayhoe, 2003). However, no studies to date have compared the effectiveness of the IDW and OK algorithms to create accurate models when run by different software programs. It is commonly assumed that data modeled using either algorithm in one program would produce identical results, if modeled using the same algorithm in another program, providing that both programs were supplied with identical input data sets. A study by Lelliott et al. (2009) indicated that program selection can have an impact on model uncertainty; however this hypothesis was not tested.

The goal of this study is to compare the output of models run using the IDW and OK algorithms using three different software programs commonly used in 3D subsurface investigations (ArcGIS, ROCKWORKS 2006, and VIEWLOG) in order to quantify the impact of program selection on model accuracy. Models are created with each of the three programs using identical data sets extracted from synthetic grids of variable complexity and with varying numbers and distributions of data points. This will allow evaluation of the performance of IDW and OK under each program and assessment of the impact of program selection on model accuracy. The purpose is not to show that any one program is “better” than another, but to identify the degree and nature of differences in the model outputs from each of the programs. The three software programs are not identified by name in the results, but are referred to as programs X, Y, and Z in order to conceal their identities.

METHODS

To effectively test differences in the output models of the three 3D modeling programs selected for study, it was necessary to develop synthetic grids with known values at all point locations from which to extract input data. Four geologically realistic synthetic grids (Fig. 1) were created to allow the extraction of specific numbers and distributions of data in a controlled manner. The advantage of using synthetic data sets to conduct this evaluation is that the point values for each grid being modeled are known at every location, allowing quantitative analysis of the variability between actual and interpolated values across the entire grid. The accuracy of modeling natural surfaces has been tested elsewhere using previously interpolated grid surfaces (such as a digital elevation models [DEMs]) that may themselves include unknown errors (Zimmerman et al., 1999). These errors could then be propagated through all subsequent analyses, making it difficult to discriminate variations in the results produced by the processing mechanisms used by the various software programs, from those due to errors propagated from the original model (Burrough, 2001).

The synthetic grids were sampled using three different sampling patterns (clustered, random, and regular) to determine if data distribution had any impact on the ability of each program to produce an accurate model. These sampling patterns were selected to represent the types of data distributions that may be encountered in various geoscientific and environmental applications. To determine the influence of data quantity on the output models from the three programs, the number of data points used for interpolation was also varied (100, 256, and 676 points) and modeled independently. This created 96 data sets that were modeled by all three programs using

Geosphere; April 2012; v. 8; no. 2; p. 534–543; doi:10.1130/GES00732.1; 6 figures; 5 tables.
Impact of program selection on 3D model accuracy

Geosphere, April 2012 535

both the IDW and OK algorithms. The data subsets extracted from the four synthetic grids were interpolated, and then re-imported into MATLAB to allow comparison of the interpolated results with the original synthetic models. The accuracy of the modeled grids generated by each software program was quantified using a variety of statistical measures including root mean square error (RMSE) and bias error (BE). The RMSE results were used to determine how accurately each of the programs was able to interpolate the original grids, and the BE was used to show where the models created underestimations or overestimations of the original data.

To assess the relative influence of each variable (i.e., number and distribution of data points, grid complexity, and algorithm and program selection) on model accuracy, a multi-way (n-way) analysis of variance (ANOVA) was calculated using MATLAB. The results obtained from the ANOVA tests are used to quantify the influence of program selection relative to grid complexity, number and distribution of data points, and algorithm selection in producing the most accurate 3D subsurface model (Tables 1–5).

Grid Creation

Four synthetic grids were created to represent realistic geological environments of variable spatial complexity using ROCKWORKS 2006 software (Fig. 1). The method used to create the synthetic grids is described by MacCormack (2010) and was developed to test the impact of data quantity, distribution, and algorithm selection on the accuracy of 3D subsurface models. Each synthetic grid was constructed using identical 80 × 80 grid dimension templates. The grid spacing was set to 1 arbitrary unit, which generated 6400 grid cells for each model. This allows each grid to have sufficient complexity to test each interpolation process while not being computationally exhausting. Each grid cell was assigned a value (thickness/elevation) between 1 and 9 to create the topographic variation shown in each of the four synthetic models. The 3D models are shown here with flat lower surfaces for ease of illustration (Fig. 1).

The first synthetic grid (grid 1) was created to represent a simple, gently sloping surface with lateral continuity in the direction perpendicular to the slope (grid 1; Fig. 1A). The second synthetic grid surface is slightly more complex, consisting of a linear trough between areas of relatively high elevation (grid 2; Fig. 1B). This surface shows undulating topography with alternating highs and lows. The third synthetic grid surface shows more spatial and topographic complexity, and consists of a series of interconnected troughs separated by irregularly spaced “highs” (grid 3; Fig. 1C). The fourth grid is characterized by a sinuous trough, with a high degree of directional variability, cut into a flat surface (grid 4; Fig. 1D); grid 4 is anticipated to be the most difficult for the various software programs to accurately model. For more detailed descriptions of the four synthetic grids, refer to MacCormack (2010).

Data Extraction

To allow quantitative analysis of differences between the model outputs from the three modeling software programs, data were extracted from the four synthetic grids in a consistent and unbiased manner. A study by Bond et al. (2007) showed that data selection can be unintentionally biased by prior knowledge of the user.

Figure 1. Synthetic grids created with 6400 known data points represent realistic geologic environments from which the data points for modeling were extracted. (A) Grid 1—gently sloping surface representing one element of a basin system; (B) grid 2—consists of two linear “highs” separated by a central trough representing a simple valley form; (C) grid 3—series of interconnecting troughs separated by linear “highs” representing an eroded bedrock surface or a braided river system; and (D) grid 4—flat surface incised by a highly sinuous channel representing an incised meandering river system.
MacCormack and Eyles

Hence, the points used for interpolation were selected using MATLAB scripts to eliminate the introduction of user bias into the analysis. The MATLAB scripts were designed to extract specified quantities and distributions of data points from each of the four synthetic grids to assess whether the quantity of data points, distribution of the data, and/or complexity of the modeled surface had any impact on the ability of each of the software programs to consistently produce an accurate model. Four separate data-point data sets were created, each containing 100, 256, 676, or 1600 cells (representing 1.6%, 4%, 10.5%, and 25% surface coverage, respectively). The points included in each data-point data set were extracted in three common sampling distribution patterns: clustered, random, and regular (Zimmerman et al., 1999; Davis, 2002; Fig. 2).

Random sample distributions (Fig. 2B) were created for the desired quantity of data points by repeating computer-generated random assignment of x and y grid locations without replacement on each synthetic surface. Regular sample distributions (Fig. 2C) were produced by imposing a square-grid of equally-spaced sample points on the synthetic grids. The spacing between sample points was universally adjusted to accommodate the specified number of data points; this ensured maximum spatial coverage of the surface, while preserving the equal spacing and distribution (number of rows and columns) of sample points. The clustered sampling distributions (Fig. 2A) were generated by establishing data clusters of sampling points by randomly assigning “cluster centers” on the synthetic grid, and then equally distributing the desired number of sample points between each cluster.

Data Formatting

Each of the extracted data subsets was exported from MATLAB as text document (.txt) files in a three-column format, containing x and y coordinates, and z values, respectively. These files were copied three times so that each software program received the same data for interpolation. Each data set was then reformatted to meet the specific requirements of data input for each program. Once all the data subsets were formatted appropriately and saved, they were imported into each software program (ArcGIS, ROCKWORKS 2006, and VIEWLOG) for interpolation.

Establishing Unbiased Interpolation Settings

Each of the data subsets was modeled using both IDW and OK gridding algorithms available in all three software programs. Each software program provided numerous options for selection or adjustment of parameters that provide the user some control over the interpolation process. Deciding how to include or adjust these parameters requires user input and expert knowledge, and can introduce user bias and uncertainty into the model output results (Englund, 1990; Bond et al., 2007). To minimize the impact of user bias on the model output results, settings were applied consistently for each program. For IDW, the number of user options was relatively small and only required determination of the number of points utilized for interpolation, set to 8 for all models run in this study. Some programs offer additional data-point search options; however these were not applied to the data sets because they were not consistently offered amongst all the programs. When interpolating using the OK algorithm, some programs offer additional options (i.e., maximum distance, spoke increment, etc.) to alter how the data are processed, and/or adjustment of parameters that provide the user some control over the interpolation process.
to 8 (to be consistent with the IDW parameters). Although adjusting the modeling parameters within the program is the preferred approach to data exploration, maintaining consistency of the modeling parameters was the only way to minimize the impact of external variables in this particular study. Minimizing external inputs into the modeling process allowed any deviations in the output results to be considered as a result of the software program and not confused with user bias and/or influence.

Comparing the Output Models

Once the 96 data subsets were interpolated by each of the three programs, they were converted from Excel files into text document (.txt) files and re-imported into MATLAB to allow comparison of results. To identify the differences between the models interpolated by the three programs, the interpolated model outputs were compared with the original synthetic grids. This was done by creating custom MATLAB functions developed to provide a quantitative comparison of each interpolated grid with the original synthetic grid. Assessment of the differences between the interpolated and original grids was then possible using a number of statistical measures such as RMSE, relative RMSE (rRMSE), mean average error (MAE), and the correlation coefficient (r2). The BE was also calculated to determine whether the interpolated grids were underestimating or overestimating the original synthetic values due to the various sampling treatments. Although all of these statistical measures were compared during initial assessment of the impacts of software selection, RMSE (Equation 1) was determined to provide the best overall comparative statistic because it provides an unbiased indication of how similar the interpolated values are to the original values from the synthetic grids. When analyzing the RMSE statistics, a small RMSE value indicates that the interpolated values for the output model are more similar to the original synthetic values, whereas a large RMSE value indicates that the interpolated model produced by the software program is less similar to the original synthetic grid. Thus, RMSE values are used here to determine the accuracy of the model output with low RMSE values indicating a high degree of model accuracy (Zimmerman et al., 1999; Davis, 2002; Dille et al., 2003; Jones et al., 2003; Mueller et al., 2004).

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left( \hat{z}(s_i) - z(s_i) \right)^2}
\]

where \(\hat{z}(s_i)\) is the interpolated value at the point \((s_i)\), \(z(s_i)\) is the observed (true) value from the synthetic data set at that same location, and \(n\) is the number of points within the input data set.

The RMSE results produced by all 96 models were visually compared using a series of graphs, and through a multi-way (n-way) ANOVA for each variable (grid complexity, distribution of data, number of data points, algorithm, and program selection). ANOVA is often used to quantify the differences between results of multiple trials in which one variable is altered at a time to assess its singular influence on the results (Carr, 2002; Davis, 2002; Borradale, 2003). The benefit of using a multi-way ANOVA is the ability to determine if and/or how the results differ with respect to the influence of individual variables, or a combination of variables (Davis, 2002; Borradale, 2003).

The ANOVA was performed in MATLAB to statistically assess which variables had the most influence on the accuracy of the synthetic grids modeled by the three software programs (Tables 1–5). ANOVA results were based on a significance of 95% (\(\alpha = 0.05\)), which is a commonly used confidence level (Isaaks and Srivastava, 1989; Cressie, 1993; Carr, 2002; Davis, 2002; Haneberg, 2004).

Results and Discussion

The model outputs from the three software programs under investigation (ArcGIS, ROCKWORKS 2006, and VIEWLOG) of the four synthetic grids using the OK and IDW algorithms were quantitatively assessed using RMSE and BE. Any differences between the RMSE and BE results produced by the three programs were analyzed graphically, and the significance of RMSE differences was quantified in the multi-way ANOVA (Tables 1–5). The statistical tests were used to establish the amount of influence that each of the variables (i.e., data-point distribution, grid complexity, number of data points, algorithm selection, and program selection) have on the accuracy of models created by each of the three programs.

Influence of Grid Complexity

To evaluate the impact of grid complexity on the ability of the three programs (X, Y, and Z) to produce accurate models, RMSE values were compared for models of the four synthetic grids produced by interpolation of 100, 256, and 676 regularly distributed data points (Fig. 3). The RMSE results clearly show that the value and range of RMSE values increase with greater grid complexity for all programs and algorithms (Fig. 3). For grids 1 and 2, the RMSE values are all very similar to one another regardless of the number of data points used to generate the model or which software program and interpolation algorithm was used (Fig. 3). RMSE values for the more complex grids (grids 3 and 4) are higher than those for grids 1 and 2 but decrease...
as more data points are used to create the models (Fig. 3). When modeling the more complex subsurface environments represented by grids 3 and 4, program X typically produced the lowest RMSE using either IDW or OK, followed by program Y, and then Z (Fig. 3). Program X also generated the most accurate model of grid 4 (lowest RMSE values) when both 256 and 676 data points were interpolated using the OK algorithm (Figs. 3B and 3C).

The ANOVA based on the RMSE results showed that grid complexity had by far the greatest influence on how accurately the synthetic grids were modeled and accounted for most of the variation in the RMSE results (Table 1). Data-point distribution and number of data points accounted for the second and third highest amounts of variability, respectively, followed by algorithm and program selection, which both had relatively low influences on model accuracy (Table 1). Grid complexity had such an overwhelming influence (60.81%) on the RMSE variance, that in order to better assess the influence of the other variables, four separate n-way ANOVAs were run on the RMSE results for each of the four synthetic grids (Tables 2–5).

For grids 1 and 2, representing relatively simple subsurface conditions, data distribution had the greatest individual influence (highest S$^2_T$) on RMSE, followed by number of data points, algorithm selection, and program selection, respectively (Tables 2 and 3). For grid 3, the order of influence was the same as grids 1 and 2, with the exception that algorithm and program selection had essentially equal influence on model accuracy (Table 4). Interestingly, the relative influence of each individual variable was different for the grid 4 ANOVA, which showed that the number of points used for interpolation had the greatest influence on model accuracy, closely followed by data-point distribution, program selection, and algorithm selection (Table 5). In addition, when the combination of variables was analyzed for grid 4 (representing an extremely complex subsurface unit), the distribution and number of data points (two variables shown to have a substantial impact on the model accuracy of complex grids individually; Table 5) ranked only slightly higher than the combination of program selection and data-point distribution in controlling model accuracy (Table 5). This is an important finding because it shows that when interpolating complex environments, the combined influence of program selection and data distribution can have as significant an influence on the accuracy of the model output as the combined influence of number and distribution of data points used for modeling. These results also indicate that program selection has a greater influence on model accuracy than algorithm selection when interpolating complex environments.

These results indicate that program selection has little influence on model accuracy when modeling relatively simple grids (e.g., grids 1 and 2; Figs. 3, 4B, and 4E), but has a stronger impact, noted by the greater differences in RMSE values between the three programs (e.g., grids 3 and 4; Figs. 3, 4H, and 4K), when modeling more complex grids. The ANOVA results support these conclusions by showing that the influence of program selection increased with increasing grid complexity (Tables 2–5). This is especially true in the case of grid 4, where program selection was shown to have a greater influence on model accuracy than algorithm selection (Table 5). Therefore, when modeling complex subsurface environments, program selection should be carefully considered as it can have a statistically significant impact on model accuracy.

**Figure 3. Root mean square error (RMSE) results for each software program (X, Y, Z) using algorithms inverse distance weighting (IDW) and ordinary kriging (OK), as increasing numbers of regularly distributed data (A) 100, (B) 256, and (C) 676 were modeled.**
Impact of program selection on 3D model accuracy

Influence of Data-Point Distribution

Data were extracted from the original synthetic grids in three distribution patterns (clustered, random, and regular) and interpolated using OK and IDW by the three software programs. Analysis of the results from the models run with 256 data points showed that grids 1 and 2 were modeled most accurately (lowest RMSE values) by randomly and regularly distributed data using OK by all three programs (Figs. 5A and 5B). However, when more complex grids were modeled using random and regularly distributed data, the RMSE values show considerable increase, regardless of algorithm or program used (grids 3 and 4; Fig. 5). Clustered data produced the least accurate results (largest RMSE values) when modeling all four grids, regardless of program or algorithm selection (Fig. 5C). It is interesting to note that the most variation in the RMSE results between program and algorithm selection occurred when clustered data were used to model grids 1 and 2, and became more similar when interpolating grids 3 and 4 (Fig.

Figure 4. Root mean square error (RMSE) values produced by software programs (X, Y, Z) as increasing numbers (100, 256, 676) of data points in random (A, D, G, J), regular (B, E, H, K), and clustered (C, F, I, L) distributions are used to interpolate synthetic grids 1 (A, B, C), 2 (D, E, F), 3 (G, H, I), and 4 (J, K, L).
Program Z produced the highest RMSE values when the OK algorithm was used, but was the least impacted by the spatial distribution of the data (Fig. 5). This contrasts with the performance of programs X and Y, which showed high levels of accuracy (low RMSE) when modeling random and regularly spaced data, but performed with similar accuracy to program Z when modeling clustered data (Fig. 5).

Of all distributions, the RMSE results for randomly distributed data sets were the least susceptible to variation between the three programs, followed by regular and clustered distributions (Fig. 4). The increased range of RMSE results for clustered data sets may reflect how the individual programs manage the irregular distributions of high data-point densities, which can produce large amounts of variability over short distances, and how this variability is propagated through the model. This is an important finding because most geological studies use clustered data (Paulen et al., 2006; Bond et al., 2007; Keefer, 2007). Overall, program X produced slightly lower RMSE results, followed by programs Y and Z (Figs. 4C, 4I, 4F, and 4L).

The ANOVA results showed that overall, data distribution had the second greatest influence on RMSE (model accuracy) when all models were analyzed (Table 1). When the influence of data-point distribution was analyzed independently for each of the four grids, the ANOVA results showed that data-point distribution had the greatest influence of any variable when interpolating grids 1, 2, and 3 (Tables 2–4). When modeling complex surfaces (grid 4), data-point distribution had the second-most influence after the number of points being modeled (Table 5).

These results indicate that data-point distribution has a significant influence on the accuracy of model results. In particular, program selection should be carefully considered when interpolating clustered data to ensure that all sources of uncertainty and variability to the model output are appropriately identified.

Impact of Data Quantity

A common assumption made by 3D modelers is that incorporating more data in the interpolation process will produce a more accurate model. To determine whether the number of data points used in the modeling process had any impact on the effectiveness of the different programs to create accurate models, the RMSE values were graphed based on the number of data points used for interpolation (Fig. 4). Overall, the RMSE values drop as the number of data points used for interpolation increases, with the exception of the clustered data sets modeled for grids 1, 2, and 3 (Figs. 4C, 4I, and 4F).

The OK algorithms applied within the X and Z software programs produced the lowest RMSE values (i.e., the most accurate model results) when grids 1 and 2 were modeled using a large number of data points (676; Figs. 4A, 4B, 4D, and 4E). Differences in RMSE values computed for grids 1 and 2 modeled by the three programs also decreased with the addition of data points, and models run with 676 points show remarkably similar RMSE results for all three programs (Figs. 4A, 4B, 4D, and 4E). In contrast, the RMSE results for grids 3 and 4 modeled using random and regularly distributed data...
Influence of Algorithm Selection

The influence of algorithm selection on the accuracy of the modeled grids using clustered, random, and regular data sets of between 100 and 676 points is shown in Figure 4. Overall, the OK algorithm produces the lowest RMSE results regardless of data distribution (Fig. 4). When the OK algorithm was used to interpolate grids 1 and 2, the RMSE values produced by programs X, Y, and Z were very similar to one another. This indicates that program selection has little impact when using OK to interpolate relatively simple models (Fig. 4A–4F). However, when OK was used to interpolate the more complex grids (3 and 4), greater differences in the RMSE values were produced by the three programs (Figs. 4G–4L). Overall, the variability of RMSE values for models interpolated with IDW was fairly consistent between the three programs, indicating that IDW is less susceptible to program selection than OK (Fig. 4). When modeling grid 4, algorithm selection had little impact because there were minimal differences between the RMSE values produced by models interpolated with either IDW or OK (Figs. 4J–4L). The ANOVA results reveal that algorithm selection was the third most influential variable on model accuracy for grids 1 through 3, and was the least important for grid 4 (Tables 2–5).

These results imply that the program-specific processes operating within the three programs are affected more strongly by the OK algorithm than IDW. These program-specific influences are shown here to cause variability in model accuracy, and should be considered when attempting to quantify model variability and uncertainty. This is especially important when modeling complex grids, for which program selection can have more influence on model accuracy than the algorithm chosen for interpolation.

Bias Error Analysis

Analyzing the bias errors associated with the interpolated results is useful for determining whether the predictions typically represent underestimations or overestimations of the actual surface. Ideally, the bias errors should be as close to zero as possible, although the amount that they deviate from zero and whether they are positive or negative can provide valuable information about the accuracy of the predictions (Elith et al., 2002; Hengl et al., 2004; Equation 2).

\[ BE = \frac{1}{n} \sum_{i=1}^{n} (\hat{Z}(s_i) - z(s_i)), \]

where \( \hat{Z}(s_i) \) is the interpolated value at the point \( s_i \), \( z(s_i) \) is the observed (true) value from the synthetic data set at that same location, and \( n \) is the number of points within the input data set. If the bias error is negative, the interpolated results are biased toward the production of underestimations when compared to the observed values. A positive bias error reveals that the interpolated results are typically producing overestimations when compared to the observed values (Elith et al., 2002; Mueller and Pierce, 2003).

In this study, bias errors were analyzed and graphed in two ways by calculating the average of the bias errors, and the absolute sum of the bias errors produced by the three software programs using different numbers of points for all four synthetic grids (Fig. 6). The ABBE and AVBE results for program Z followed similar trends to those for programs X and Y, but the actual values were quite different (Fig. 6). These results indicate that programs X and Y may follow similar processes when interpolating data under certain conditions because they tend to produce similar errors. Program Z is likely interpolating data using a different set of procedures because it produced distinctly different bias errors than programs X and Y (Fig. 6). Unfortunately, the software codes for each program are proprietary and are therefore not available for comparison in order to determine exactly why the output results are different. However, the purpose of this paper was not to evaluate the efficacy of the software’s interpolation codes, but to quantify the impact of program selection on model accuracy. These results exemplify the importance of considering program selection when quantifying variability or uncertainty, because program selection alone can impact whether a model represents an overestimation or underestimation of the actual values.

CONCLUSIONS

This paper compares the ability of three commonly used 3D modeling programs to accurately interpolate grids of variable complexity, using different numbers and distributions of data points. The results presented here show...
that, although certain software programs offer the same interpolation algorithms, they do not necessarily provide the same output results. In most 3D subsurface investigations, a substantial amount of time and effort is spent collecting and analyzing data, as well as assessing data parameters to ensure the most accurate model possible is produced. The results of this study indicate that program selection should be seriously considered as a possible source of model uncertainty, the effects of which should be considered especially when modeling complex subsurface geological environments, interpolating with clustered data, or when relatively large quantities of data (more than 4% data coverage) are used for interpolation.

This study demonstrates that the output of the OK and IDW interpolation algorithms is not consistent when used by different software programs, and that this variability should be considered when assessing the uncertainty associated with subsurface model results. The purpose of this study was not to suggest that any particular program is better than another, but to show that software selection can impact the accuracy of the output model in a similar manner to that documented for algorithm selection (Tabios and Salas, 1985; Weber and Englund, 1992; Weber and Englund, 1994; Brus et al., 1996; Walker and Loftis, 1997; Nalder and Wein, 1998; Zimmerman et al., 1999; Schloeder et al., 2001; Jones et al., 2003; Kravchenko, 2003; Dille et al., 2003; Lapen and Hayhoe, 2003). In some instances, program selection can actually have a greater impact on model accuracy than which algorithm is used to perform the interpolation process.

Given that 3D modeling is increasingly used as an analytical tool for numerous applications in geological and environmental sciences, and may form the basis of large-scale, multimillion-dollar decisions, serious attention should be paid to the many factors that control model accuracy. The results of this study indicate that program selection can have a significant influence on the accuracy of model results and therefore should be seriously considered as a potential source of uncertainty in 3D geologic modeling.

ACKNOWLEDGMENTS

We would like to thank Jason Brodeur, Dr. Jeff Harris, Dr. Antonio Paez, and the two anonymous reviewers for their helpful comments and suggestions pertaining to this research. Paul Durkin, Prateek Gupta, and Riley Mulligan are thanked for their assistance with data entry and manuscript preparation. We would also like to thank Natural Sciences and Engineering Research Council of Canada for their assistance in providing support by means of a Doctoral Postgraduate Scholarship (PGS D) scholarship to Kelsey MacCormack and a Discovery grant to Carolyn Eyles.

REFERENCES CITED


Dille, J.A., Milher, M., Groeteke, J.J., Moretensen, D.A., and Williams, M.M., 2003, How good is your weed...
Impact of program selection on 3D model accuracy


MANUSCRIPT RECEIVED 14 JUNE 2011
REVISED MANUSCRIPT RECEIVED 22 NOVEMBER 2011
MANUSCRIPT ACCEPTED 13 DECEMBER 2011

Geosphere, April 2012 543