Revealing Uncertainty in Maps of Glacial Lake Algonquin

Scott Drzyzga, Ashton Shortridge and Randall Schaetzl

Michigan State University, Department of Geography, 314 Natural Science Building, East Lansing, Michigan 48824 USA

Abstract

This paper examines the effects that uncertainty in digital elevation data has on results obtained from a spatial model. Ongoing analysis of glacial Lake Algonquin in northern Michigan has used a spatial model to interpolate unknown shoreline locations between known shoreline positions. Extant shoreline features were surveyed with global positioning technology and ancient shorelines were reconstructed using a DEM and a statistical model of isostatic rebound. However, shoreline data contain relatively small measurement errors and relatively large errors are associated with the DEM. These errors, when propagated through a series of GIS operations, may render uncertain results. This research recognises and attempts to assess these errors in order to produce a new map of Lake Algonquin shorelines - one that illustrates shorelines and areas of positional uncertainty. Results indicate that even small errors in input data can contribute uncertainty to model output. Understanding these uncertainties can be valuable to further research concerning these and other ancient shorelines.
Keywords: error propagation, spatial modelling, simulation, Lake Algonquin

1 Introduction

According to Heuvelink (1998, p.3), most work on spatial modelling "has been concentrated on the business of deriving computational models that operate on spatial data, on the building of large spatial databases, and on linking computational models with GIS". However, some serious attention has also been paid to the issue of spatial data quality. Goodchild and Gopal (1989) concluded that virtually all spatial data stored in a GIS are, to some extent, contaminated by error. Subsequently, Heuvelink (1998) has argued that a potential danger exists in GIS because such errors can be propagated through GIS operations, corrupt results and render interpretations meaningless.

In work about mapping glacial shorelines, Schaetzl *et al.*, (in press), reported results obtained from a spatial model. These results contain an unknown quantity of uncertainty resulting from errors propagated from input data and through the GIS operations used. Therefore, the purpose of this research is to re-evaluate the spatial model employed by Schaetzl *et al.*, and account for the effects that input data errors had on results.

1.1 Mapping Lake Algonquin

Lake Algonquin was a prominent proglacial lake in the Great Lakes basin (Fig. 1). It maintained a relative high stage from approximately 11,200 to 10,400 BP and developed conspicuous shoreline features (Larson and Schaetzl, 2001). As glaciers receded from the area lower outlets were uncovered, the land surface rebounded as the weight of ice was lifted, and lake levels fell in stages. Locations in northern Michigan have since rebounded at a faster rate than locations in southern Michigan (Futyma, 1981). This crustal uplift process is called isostatic rebound; Fig. 2 illustrates the differential effects this process has had on the highest known Lake Algonquin shoreline.

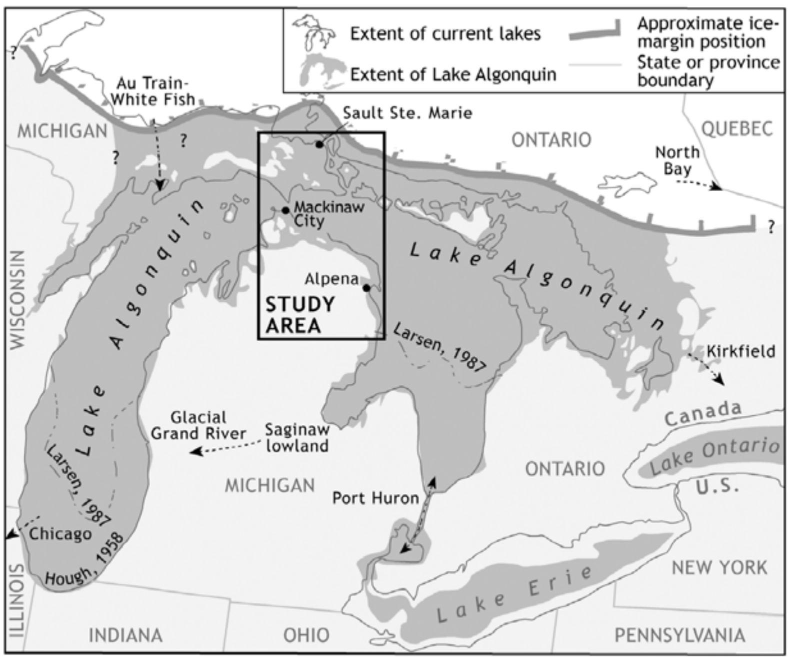


Fig. 1. Hypothesized extents of Lake Algonquin circa 11,000 BP and known extent of study area (after Schaetzl *et al.*, [in press.]). Study area is approximately 175km by 250km.

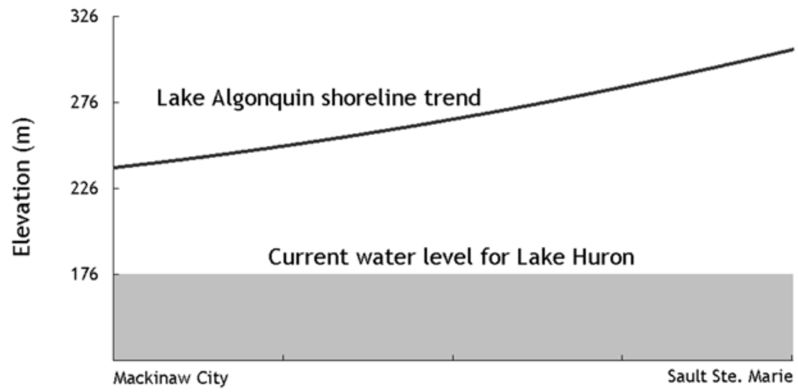


Fig. 2. Differential effects of isostatic rebound on the Algonquin shoreline in Michigan. Trend elevations estimated along an 85km transect from Mackinaw City to Sault Ste. Marie. Elevation values are in meters above sea level.

Much has been written about how Lake Algonquin was formed (Spencer, 1891; Gilbert, 1898; Goldthwait, 1908; Leverett and Taylor, 1915; and Stanley, 1936), but little agreement exists about the exact spatial extent of the water surface (Hough, 1958; Futyma, 1981; Larsen, 1987, see Fig. 1). Little agreement exists because the terrain and some shoreline features have been altered over the past 10,000 years by natural or human processes, and other evidence has been buried or reworked by later lake transgressions. Uncertainty about the extent of this lake prompted an extensive survey by Schaetzl *et al.*, (in press). The field element of the survey employed global positioning technology to measure positions {x,y,z} along the highest known Lake Algonquin shoreline. In the laboratory, a GIS was used to conflate and analyse the shoreline data and a digital elevation model (DEM) of the region in order to interpolate unknown shoreline locations and generate a map of results. In sum, a GIS was used to evaluate a computational model - a simple representation of a complex, crustal deformation process.

1.2 Statement of Problem

The certainty with which one can assume a DEM represents true surface form is a function of the quality of the elevation data provided. Kyriakidis *et al.*, (1999, p. 678 after Bolstad and Stowe, 1994) maintain, "that DEM source data and production methods are not perfect, and consequently mismatch errors occur between actual and DEM-reported elevation values." GIS operations can propagate error in the DEM through processing and analysis, thereby producing uncertain results. Therefore, a need exists to quantify and represent uncertainty in results obtained from the spatial model used to interpolate Lake Algonquin shorelines.

The objectives of this research are to quantify and model errors in Schaetzl's shoreline data and DEM (a mosaic of USGS 3 arc second DEMs resampled to a

90 meter ground resolution), re-evaluate the shoreline interpolation model, and represent uncertainties in a map of results. Six related questions were posed in an effort to achieve the research objectives:

1. How much elevation error is contained within the shoreline data set?
2. How much elevation error is contained within the DEM?
3. What influence do elevation errors in shoreline data have on interpolated shoreline locations?
4. What influence do elevation errors in DEM data have on interpolated shoreline locations?
5. What influence do combined elevation errors have on interpolated shoreline locations?
6. Where, if anywhere, do uncertainties exist in a map of results?

2 Methods used for Summarising Error

The area selected for study is located in Michigan's northern Lower and Upper Peninsulas (Fig. 1). This region contains many conspicuous Lake Algonquin bluffs. Local relief varies within the region, which includes moraines, steep bluffs and relatively flat outwash plains.

2.1 Measurement Errors in the Shoreline Data

While Schaetzl *et al.*, (in press) collected position data, efforts were also made to assess the accuracy of GPS-measured elevation data. Positions were measured for eight different benchmarks of known elevation (one was measured twice, but during different trips to the field). Measured benchmark elevations were compared to actual elevation values published by the National Geodetic Survey for the same monuments (NGS Information Services Branch, 1999). A root mean squared error (RMSE) of 1.20 meters represents the average magnitude of difference between measured and published values. In addition, RMSE values for eastings and northings were not greater than 0.8 meters (not shown). Because the methods used to gather and process shoreline data were the same as those used to gather and process benchmark data, the average vertical error in the shoreline data was inferred to be similarly small. Summary statistics were calculated to describe the distribution of error within the benchmark data (Table 1). Although the sample size is small (n=9), the statistical evidence in Table 1 does not indicate that the error distribution is not normal. Hence, we assumed the distribution of error to be normal. Further assessment indicated that error did not appear to be a function of space, position, or some form of landscape covariation process. Therefore, we believe these errors are measurement errors and error magnitudes are associated with the limits of equipment precision. We assume this error process is random and spatially independent.

Table 1. Summary statistics for elevation error distributions

Statistic	Shoreline data	DEM
RMSE	1.20	5.09
Min	-1.63	-21.20
Max	1.75	22.10
Mean	0.00	0.01
Variance	1.61	25.94
Skewness	0.25	0.24
Kurtosis	-1.64	2.23

2.2 Errors in the DEM

During their survey, Schaetzl *et al*. collected additional position data for shorelines that existed later and lower than the Main Algonquin shoreline. These ancillary position data and National Geodetic Survey benchmark data (NGS, 2001) were used to assess the accuracy of DEM elevations at 1353 locations. All data sets were referenced to the same coordinate system, and horizontal and vertical control networks. A root mean squared difference of 5.09 meters (see Table 1) indicates a larger average magnitude of error exists in the DEM than in the shoreline data set. Therefore, it is reasonable to suspect that larger errors in the DEM contribute more uncertainty to results than smaller errors in the shoreline data.

Summary statistics were calculated to describe the distribution of error within the DEM (Table 1). The Kolmogorov-Smirnov test was used to test the cumulative frequency distribution for non-normality. Significance test results did not provide evidence to reject the standard null hypothesis; *the sample distribution fits an expected normal distribution*, using a 90 percent confidence interval. Hence, we assumed the distribution of error in the DEM to be normal.

Although we assumed errors to be normally distributed, we did not assume errors to be spatially independent. Bolstad and Stowe (1994), Brown and Bara (1994) and Kyriakidis *et al.* (1999) have demonstrated reasons for why errors in USGS DEM products exhibit spatial autocorrelation and not spatial independence structures. Isotropic and anisotropic semivariograms were calculated and analysed to characterise the distribution of errors; each indicated an isotropic and positively autocorrelated covariance structure exists. However, unlike Bolstad and Stowe (1994), further analysis did not reveal any linear relationship between error values (signed or unsigned) and position values (x or y or z). Hence, we treated the covariance structure associated with DEM errors as a spatially dependent random process of unknown origin.

2.3 The Spatial Model

The spatial model used by Schaetzl *et al.* (Fig. 3a) is relatively simple, but it simulates a complex, crustal deformation process. In essence, the model reconstructs the ancient shoreline by adjusting a DEM of modern terrain to account for isostatic rebound. First, former water surface elevations are modelled as a function of location using the shoreline data and a second-order polynomial trend surface with parameters estimated via ordinary least squares. This trend surface represents the upwarped nature of the ancient landscape sans local relief (e.g., Fig. 2).

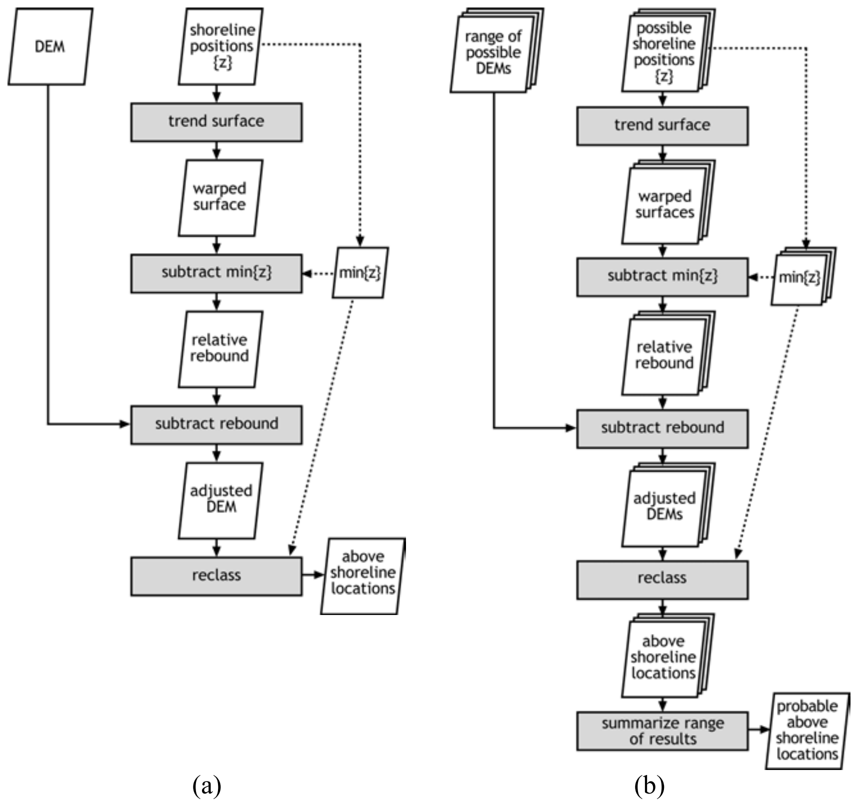


Fig. 3. The spatial models used by **(a)** Schaetzl *et al.*, and **(b)** this research

Second, the minimum elevation value in the shoreline data set is identified then subtracted from all elevation values in the warped surface. The resulting surface represents isostatic rebound relative to the lowest known shoreline elevation. This step is necessary because much disagreement exists in the literature regarding the actual lake level and controlling outlet elevations (e.g., Hough, 1958 and Larsen, 1987). Therefore, the isostatic rebound surface is calculated relative to the lowest known elevation and not a controlling outlet.

Third, the isostatic rebound surface is subtracted from a surface model of modern elevations, a.k.a. the DEM, thereby producing an isostatically depressed elevation model with local relief. Finally, classifying all elevation values with respect to the minimum shoreline elevation value cartographically floods the adjusted elevation model. Nodes with elevation values greater than the minimum value are classified as *above shoreline*, while the balance are classified as *not above shoreline*.

3 Methods Used for Modelling Uncertainty

Monte Carlo simulation is a computer intensive technique used to reveal the contributions of error to results from input data, as well as differences across a geographic area (Burrough and McDonnell, 1998). In essence, a typical simulation consists of evaluating a spatial model many times, each time using inputs that have been modified. Individual input values are modified randomly via some model that specifies the error distribution of the input and makes assumptions about the behaviour of error. Parameters for the model are derived with respect to known or assumed information about errors in the input data. After a large number of realisations are generated, the set of spatial model results is summarised to determine the range of possible outcomes.

The goal of a Monte Carlo simulation is not to determine a single correct answer, but rather to define probable limits associated with a large number of possible outcomes given random errors in data. The Monte Carlo simulation technique is appropriate for this analysis because it can reveal contributions of error to results from input data, which is the purpose of this research.

The error model employed during this work adopts four assumptions. First, the distribution of error associated with each input value in each data set is normal. Second, because each error may be treated as a random variable, every measured value comprises one out of an infinite set of possible yet, equally valid representations. The third assumption is employed with respect to the shoreline data set only; the error associated with any one input value is independent, random, and cannot be systematically corrected. Last, the fourth assumption is employed with respect to the DEM only; the errors associated with input values are spatially structured, random, and can be simulated via the sequential Gaussian simulation technique described below (see Section 3.2 and Goovaerts, 1998).

3.1 Simulation One: Uncertainty from Shoreline Data

During the first simulation, 100 realisations of shoreline elevations were generated. Each realisation consisted of original shoreline elevation values that were perturbed independently by error values drawn randomly from the normal probability distribution function with mean = 0 and variance = 1.61 (see Table 1). The adjusted spatial model (Fig. 3b) was run 100 times, each time using one of the perturbed data sets, rather than the original, as input. The DEM values were held constant. This simulation produced a set of 100 binary grids, each containing cells that represented *above shoreline* (1) and *not above shoreline* (0) locations. Using grid algebra, the set of 100 binary grids were summed and divided by the scalar 100 to produce a single and final grid. Final grid cells with a zero value indicate locations that were classified *above shoreline* zero percent of the time. Final grid cells with a unity value indicate locations that were classified *above shoreline* 100 percent of the time. Values ranging between zero and one indicate the proportion of results that were above the minimum shoreline elevation value.

3.2 Simulation Two: Uncertainty from the DEM

During the second simulation, 100 realisations of the DEM were generated. Each realisation consisted of DEM values that were perturbed by random and spatially structured errors, which were made conditional to ancillary shoreline and NGS benchmark data.

Realisations were generated using a two-step procedure. First, the spatial structure of DEM error was estimated by employing the geostatistical software program Gstat (Pebesma and Wesseling,1998) to fit a compound exponential variogram model [Eq. 1] to an empirical variogram of measured DEM errors. And second, we employed sequential Gaussian simulation (sGs and succinctly described by Goovaerts, 1998:4-5) to account for the spatially dependent error process, and condition the DEM to higher accuracy spot data. According to Goovaerts (1998:4),

> "sequential simulation amounts to modelling the conditional cumulative distribution function (ccdf)... ...then sampling it at each of the grid nodes visited along a random sequence. To ensure reproduction of the *z*-semivariogram model [Eq. 1 - authors inclusion], each ccdf is made conditional not only to the original *n* data but also to all values simulated at previously visited locations."

In sum, we used sGs to develop 100 models of spatially structured DEM errors that were used to perturb the original DEM.

$$\gamma(h) = 1 + 8(1 - e^{-3h/270}) + 13(1 - e^{-3h/3800}) \qquad (1)$$

where: nugget = 1,
$sill_1$ = 9,
$range_1$ = 270 m,
$sill_2$ = 22, and
$range_2$ = 3800 m.

The adjusted spatial model (Fig. 3b) was run 100 times, each time using one of the perturbed data sets, rather than the original, as input. Original shoreline data values were held constant. This simulation produced a second set of 100 binary grids, which were summarised in the same manner described in Section 3.1. It was expected that this simulation would produce a wider range of results, and therefore more uncertainty, than did the first simulation because the average magnitude of error associated with the DEM is greater than that associated with the shoreline data set.

3.3 Simulation Three: Uncertainty from Error Interaction

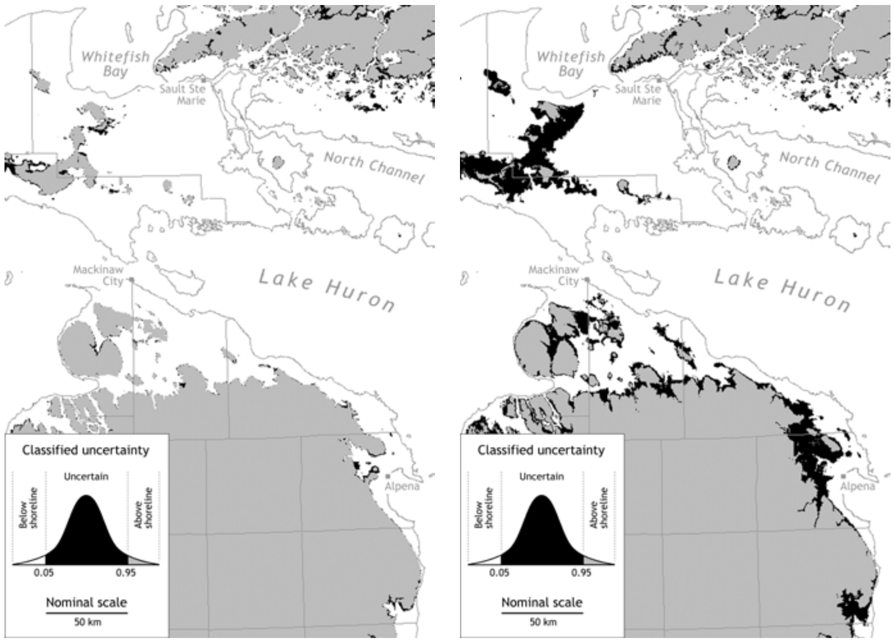
The interaction between input errors can take many forms. For example, errors in one data set may cancel errors in the other - a best-case scenario. A worst-case scenario would reveal a compounding effect. Moreover, the degree of interaction may vary over space, thereby perturbing results in some locations differently than others. The third simulation, which incorporated 100 realisations each of shoreline data and the DEM, was used to reveal influences that the interaction of errors in both data sets had on results. Accordingly, the adjusted spatial model (Fig. 3b) was run 100 times to produce a third set of 100 binary grids. Again, binary grids were summarised in the same manner as describe in Section 3.1.

4 Mapping the Results

Maps were prepared to illustrate the spatial distributions of shoreline locations associated with the three final grids. For display purposes, grid cells with values less than 0.05 were reclassified as *below shoreline* and illustrated with white. Grid cells with values greater than 0.95 were reclassified as *above shoreline* and illustrated with a gray tone. Values including and between 0.05 and 0.95 were reclassified as *uncertain* and highlighted with black.

4.1 Influence of Elevation Errors in the Shoreline Data Set

Relatively small measurement errors inherent to shoreline data contributed some uncertainties to results. Fig. 4a shows the probable distribution of shoreline locations and areas over which the position of the Lake Algonquin shorelines remains in doubt. Much of the land and water interface is distinct but some sections are peppered with relatively small areas of uncertainty. Several larger areas of uncertainty exist at locations apparently reworked by fluvial activity (e.g., the Thunder Bay River system west of Alpena) and spaces possibly associated with emergent wetlands (e.g., islands in the Upper Peninsula). However, some of these uncertainties may also be products of data distribution. Much of the shoreline survey data was collected in the Lower Peninsula. Many but fewer positions were measured in the Upper Peninsula and none were gathered in Canada. Given what is commonly known about the behaviour of trend surfaces calculated beyond the extent of input data (i.e., for locations in Canada), we hypothesise that some of the uncertainties illustrated in Fig. 4a were contributed by trend surface flutter in the adjusted spatial model. However, that is a hypothesis to be entertained during future research.



(a) (b)

Fig. 4. Shoreline uncertainties propagated from, **(a)** measurement errors in shoreline elevation data, and **(b)** spatially structured errors in the DEM. Modern water body names, county boundaries, and places mentioned in the text were added as references.

4.2 Influence of Elevation Errors in the DEM

Fig. 4b shows the distribution of above shoreline locations and areas over which the position of the Lake Algonquin shoreline remains uncertain. Although the larger errors in the DEM were suspected to contribute more uncertainty to results, distributions in the output are surprisingly large. While similarities exist between the spatial distributions of uncertainty in Figs. 4a and 4b, some striking differences are present. In some areas, the width of the band of uncertainty is greater than 10 km. For example, the peninsula located north of Alpena (Fig. 4a) was cartographically reduced to two islands (Fig. 4b). Much of the land area in the Upper Peninsula (Fig. 4a) has also been reduced by uncertainty (Fig. 4b). However, modelling DEM elevation errors has served to reveal some new information. For example, results from simulation two show a massive spit complex located approximately halfway between Alpena and Mackinaw City. This complex was not apparent in Schaetzl's original findings or after simulation one. Large errors in the DEM may have propagated into broad uncertainties, but when rigorously modelled, they may also contribute to discovery.

4.3 Influence of Interacting Errors

The interaction between errors during simulation three manifested into uncertainty distributions nearly identical to those produced during simulation two. Hence, a separate figure is not included here. Shadows of uncertainty caused by errors in the shoreline survey were masked by shadows caused by errors in the DEM. No compounding effect was evident. Most of the slight differences occurred in the island and spit complex located just south of Mackinaw City and the island complex located north of the North Channel. As hypothesized above, however, the observable differences in Canada may have been induced by trend surface flutter in the adjusted spatial model.

5 Discussion

The shoreline survey data used in this research consist of position measurements made along relict, and thus static, lacustrine features formed by Lake Algonquin. Yet, active coastal environments are dynamic and their associated shorelines can change quickly via the combined influences of erosion and deposition processes. For that reason, the exact position of any shoreline, whether modern or ancient, cannot be known with absolute certainty. However, based on arguments put forth by Schaetzl *et al*. (in press), we accept the shoreline survey data set as the set of best proxy indicators of Lake Algonquin shoreline locations.

Burrough and McDonnell (1998:243) consider the question, "if a new attribute U is defined as a function of inputs $A_1, A_2,...A_n$, we want to know what is the error associated with U, and what are the contributions from each A_n to that error?" In this research, the new attribute U represents Lake Algonquin shoreline position, and we want to know the contribution of error in shoreline position from the shoreline survey data and the DEM, A_1 and A_2 respectively.

Results from this research illustrate the significant impact that even small errors in spatial data can have on model results. Recall that elevations in the shoreline data set contain an average error magnitude of only 1.2 meters. These small errors, after being propagated through several GIS operations, manifested into hundreds of meters of positional uncertainty in some locations.

The DEM contained elevation values with an average error magnitude greater than 5 meters. This error range is larger than the range associated with the shoreline data, and subsequently, was propagated into larger uncertainties.

5.1 Uncertainty Distributions in Maps of Results

Obvious spatial patterns are evident in results when errors in the DEM are propagated. Review of Fig. 4 reveals several locations that are sensitive to the larger errors inherent to the DEM. A re-examination of field notes, USGS topographic maps, and the original DEM provided grist for an explanation. Relatively flat areas, like those formed by alluvial outwash processes, contain little local relief and subtle changes in elevation. Should the interpolated position of a shoreline vary along the vertical dimension of these relatively flat locations, even by only a few meters, then the intersection of the water surface and the land surface can vary widely over the horizontal plane. Conversely, those locations with relatively steep gradients and significant local relief, like those associated with wave-cut bluffs (e.g. like those located just south of Alpena), will tend to be less sensitive to errors in interpolated shoreline positions. Although differences in shoreline elevation will change the position of the shoreline along the vertical axis, less change would occur along the horizontal plane. In sum, those areas most sensitive to errors tended to exhibit shallow gradients.

6 Conclusions

Lake Algonquin shorelines in Michigan cannot be mapped with the same degree of precision at all locations, at least not with the data sets, spatial model or terrain data currently available. This finding in no way discredits those data sets or the work done by Schaetzl *et al*., because they have delineated the boundaries of Lake Algonquin much more extensively and precisely than any previous work. However, it seems appropriate to suggest limitations on how their results could be used. For example, their results would be quite useful for reference purposes when illustrated at smaller cartographic scales, like the one used for Fig. 1, because the width of the *band of uncertainty* may be concealed within the width of a pen stroke. At larger nominal cartographic scales, like the one used for Fig. 4, their results would not be as useful for determining shoreline position because the exact extent of the former lake boundary remains uncertain in many locations. Yet, future efforts to map the extent of Lake Algonquin and other paleolakes will benefit from the research presented here as uncertainty maps can be used to focus subsequent fieldwork and surveying efforts. An iterative process of planning, fieldwork, spatial modelling, and uncertainty mapping, with refinements made along the way, may be what is required to finally map Lake Algonquin.

Acknowledgements

This work was supported by a NSF (Geography and Regional Science) grant (BCS-9819148) to RJS. We acknowledge the many contributions of students from GEO 809 (Fall 1999) to this project: B. Weisenborn, K. Kincare, X.

Cordoba, K. Shein, C. Dowd, and J. Linker. Also, we thank Beth Weisenborn for sharing with us her template of Lake Algonquin, which we used to create Fig. 1.

References

Bolstad PV, Stowe T (1994) An evaluation of DEM accuracy: elevation, slope, and aspect. Photogrammetric Engineering & Remote Sensing 60(11):1327-1332

Brown DG, Bara TJ (1994) Recognition and reduction of systematic error in elevation and derivative surfaces from 7-1/2 minute DEMs. Photogrammetric Engineering and Remote Sensing 60(2): 189-194

Burrough PA, McDonnell RA (1998) Principles of Geographic Information Systems. New York: Oxford University Press Inc.

Futyma RP (1981) The northern limits of glacial Lake Algonquin in upper Michigan. Quaternary Research 15: 291-310

Gilbert GK (1898) Recent earth movements in the Great Lakes region. U.S. Geological Survey 18th Annual Report, Part 2. Washington D.C.

Goldthwait JW (1908) A reconstruction of water planes of the extinct glacial lakes in the Lake Michigan basin. Journal of Geology 16: 459-476

Goodchild MF, Gopal S (1989) Accuracy of Spatial Databases. London: Taylor & Francis.

Goovaerts P (1998) Impact of the simulation algorithm, magnitude of ergodic fluctuations and number of realizations on the spaces of uncertainty of flow predictions. Stanford Center for Reservoir Forecasting, Stanford University, Unpub. annual report No 11.

Heuvelink GBM (1998) Error propagation in environmental modelling with GIS. Bristol, PA: Taylor & Francis

Hough JL (1958) "Geology of the Great Lakes." Urbana: University of Illinois Press

Kyriakidis PC, Shortridge AM, Goodchild MF (1999) Geostatistics for conflation and accuracy assessment of digital elevation models. International Journal of Geographical Information Science 13(7):677-707

Larsen CE (1987) Geological history of Glacial Lake Algonquin and the Upper Great Lakes. U.S. Geological Survey Bulletin 1801

Larson G, Schaetzl RJ (2001) Origin and evolution of the Great Lakes. Journal of Great Lakes Research 27(4):518-546

Leverett F, Taylor FB (1915) Pleistocene of Michigan and Indiana and the history of the Great Lakes. U.S. Geological Survey Monograph 53

National Geodetic Survey, Information Services Branch. (2001) NGS Datasheets [online]. Available from: http://www.ngs.noaa.gov/datasheet.html (Viewed on March, 2001)

Pebesma EJ, Wesseling CG (1998) Gstat, a program for geostatistical modelling, prediction and simulation. Computers and Geosciences 24(1):17-31

Schaetzl RJ, Drzyzga SA, Weisenborn BN, Kincare KA, Cordoba XD, Shein KA, Dowd CM, Linker J (in press) Correlation and mapping of Lake Algonquin shorelines in Michigan. Annals of the Association of American Geographers.

Spencer JW (1891) Deformation of the Algonquin beach, and birth of Lake Huron. American Journal of Science 4:12-21

Stanley GM (1936) Lower Algonquin beaches of Penetanguishene Peninsula. Bulletin of the Geological Society of America 47: 1933-1960