Author's Response to Independent Peer Review Comments on Bennion, D., 2009, "Statistical and Spatial Analysis of Bathymetric Data for the St. Clair River, 1971–2007", USGS Scientific Investigations Report 2009–5044.

Author's comments in Red

DRAFT

June 2, 2009
To: the International Upper Great Lakes Study St. Clair River Task Team of the International Joint Commission
RE: Review of the Reports to the IJC by Bennion
Review by Colin Rennie, Ph.D., P.Eng.

Summary
The Bennion (2009) study assessed changes in St. Clair River bathymetry based on available survey data since 1971. The Bruxer and Thompson (2008-2009) report utilized a 2D hydraulic model to assess change in conveyance since 1971. In general, both studies were performed well. However, key weaknesses were found in both studies. The Bennion study found no statistically significant change in bathymetric volume between survey years (Table 7). Bennion used a conservative method to assess uncertainty of change in bathymetry between survey years by summing all sources of uncertainty. It is possible that a more statistically valid pooling of uncertainty sources would lead to findings of statistically significant volume change. This would be an important finding, given that a primary objective of the IJC study is to determine whether changes in conveyance have caused reduction in lake water levels. A key weakness of the Bruxer and Thompson study was that the SMS routine of the RMA2 model utilized linear interpolation when generating the model mesh. This resulted in substantial errors in modelled bathymetry for survey years with sparse data, and the authors should have considered other means for generating the model mesh.

Bennion, D., 2009, "Statistical and Spatial Analysis of Bathymetric Data for the St. Clair River, 1971–2007", USGS Scientific Investigations Report 2009–5044.

The author has provided a detailed and thorough review of available bathymetry data, survey and interpolation uncertainty, and estimation of channel change. This reviewer had only a four questions/concerns, as specified below.

Survey data

p.4 It was surprising to read that the author was unable to determine the survey data binning procedure for the 2000 survey. Given that this survey occurred only nine years ago, one would assume that the original surveyors could be contacted to elucidate the survey and binning methods.
Attempts were made to discern the binning method through a review of the literature, but not pursued further because of time constraints. Knowing the method used would do little to refine the dataset, and would not affect the outcomes of analysis. Knowing the method used would still not provide a means to adjust the reported surveyed values back to the "original" surveyed values. Further, the process of binning the 1971 data would likely remain unknown.

Interpolated Bathymetry

p.4 The author utilized ordinary kriging in blocks, such that interpolated sections of river were relatively straight, which probably improved the interpolation because measured variogram anisotropy was consistent for a block. It would have been useful to see figures of measured and modelled variograms (perhaps as appendices), which would assist in the evaluation of the interpolation.
The number of reaches that the analysis areas were divided into and the number of different datasets used would have resulted in approximately 100 measured and modeled variograms for display. Since the variogram alone can not be used to assess interpolation accuracy, and due to the large number of figures already included with the report the author decided to not include the individual variograms as part of the report.

No plots of the interpolated bathymetry were provided. These would have been useful for the reader to assess the geomorphic importance of locations of bathymetric error and locations of volume/elevation change.
The focus of the report is the changes in bathymetry over time. Figure 5 in the report illustrates consistent areas of change in relation to significant river features. Locations of error would be confined to the locations of extracted test points, and would vary if a different test point set were extracted and used for error assessment. Because significant river features are shown in other figures and included kriging prediction standard error maps illustrate spatial locations of estimated interpolator fitness, overlay of this information with the interpolated bathymetry is not displayed.

Statistical Evaluation of Cut/Fill

p.6 The author has utilized an extremely simple uncertainty model to calculate errors in cut/fill estimates. The model can be written as:

$\varepsilon_{cf} = (\varepsilon_b + \varepsilon_a)nA$

where ε_{cf} is the uncertainty of the estimate in the total change of section volume, ε_b is the uncertainty of elevation at a cell in the initial (before) survey, ε_a is the uncertainty of elevation at a cell in the initial (after) survey, n is the number of cells in the comparison, and A is the area of a cell. The cell uncertainties were calculated by adding the mean absolute interpolation error to the average measurement (survey) error for a given survey.

This model adds all errors linearly, which appears to the reviewer to be excessively conservative. A typical error model would pool the variances due to error sources (i.e. take the square root of sum of square errors). This is because a linear combination of random variables has variance

$\sigma^2 = \sum_i a_i^2 \sigma_i^2$ (Harris 1966), where σ_i is the variance of random variable i, and a_i is the coefficient for random variable i in the linear model. Assuming all a_i in the error model equal 1, then the variances of each error term are summed, and the square root is taken to yield a standard deviation (standard error). If errors are not normally distributed, it is also worthwhile to conduct a Monte Carlo simulation with full distributions of each error term to estimate a probability distribution for ε_{cf}. Importantly, the pooled variance results in an error estimate less than simply summing all error sources.

This is an important issue, because no estimates of volume change were found to be statistically significant due to the large uncertainty estimates (Tables 7). If the estimated uncertainty were reduced, possibly statistically significant volume changes would be found. The error analysis was conducted again using the suggested method. The total error of each dataset was represented by the sum of the standard deviations of the errors associated with survey and interpolation uncertainty. These values were then input into the equation: Cell Count * Cell Area * square root(Zbefore^2 + Zafter^2). While this method did lower the error thresholds as the reviewer suggest, still none of the results of the cut/fill analyses exceed (or come close to) these new thresholds and the conclusions reached in the report remain unchanged.

Before Year	After Year	Overall Change (m3) (neg = deposition)	Original Error Threshold (+/-)	New Error Threshold w Interp RMSE (+/-)	New Error Threshold w Interp SD (+/-)
1971	2000	1,018,078	2,115,479	1,766,046.97	1,757,615.06
2000	2002	-262,522	1,120,888	807,773.30	801,764.42
2002	2005	36,392	678,848	424,865.17	410,873.83
2005	2006	-691	726,209	496,277.51	486,681.47
2006	2007	29,254	726,209	551,701.16	497,457.57
1971	2002	755,800	1,626,077	1,609,008.63	1,602,769.20
1971	2005	792,174	1,673,439	1,626,940.26	1,617,158.57
1971	2006	800,248	1,673,439	1,629,321.86	1,623,856.72
1971	2007	784,878	1,673,439	1,644,693.72	1,620,434.20
2000	2005	-226,128	1,168,250	842,925.59	830,155.96
2000	2006	-218,055	1,168,250	847,513.22	843,129.73
2000	2007	-188,798	1,168,250	876,701.98	836,519.04
2002	2006	35,700	678,848	433,895.75	436,492.99
2002	2007	73,722	678,848	488,460.05	423,582.91
2005	2007	37,328	726,209	544,627.46	475,137.04

The author then proceeds to utilize only grid cells that had changes in elevation exceeding the estimated uncertainty (Table 8). While this masking approach reduces the influence of survey noise from the analysis, it is not clear to the reviewer if the observed changes in volume are considered to be statistically significant.
It is noted in the report on p8, column 2 that "The results of these analyses are shown in table 8. It should be noted that the overall magnitude of volumetric change indicated by the masked analyses is still subject to variation within the error range."

Ultimately, the author utilizes direct comparison of surveyed points to identify locations of significant elevation change. In general, as one might expect, there has been degradation on outer bends and aggradation on inner bends in the upper channel. Furthermore, these results suggest that the river bed degraded between 1971 and 2000, and aggraded between 2000 and 2007 (Appendix Figures 31-36).
No response needed

Review by Richard M. Vogel, Tufts University, April 2, 2009

International Upper Great Lakes Study – Sub-Product Reviews, Synthesis Product Reviews, and Draft Final Study Report Reviews Template

Peer Review of Manuscripts

Manuscript: Statistical and Spatial Analysis of Bathymetric Data for the St. Clair River, 1971-2007
Author(s): David Bennion
Name of Reviewer: Richard Vogel

Comments (limit responses to one paragraph for each question; reference pages, charts, and data. Please distinguish if responses are of major or minor concerns.)

A. What is the best/most unique part of the analysis?

The best aspect of the work is that it addresses one of the primary goals by clearly demonstrating in Figure 5 the areas of geomorphic change in the upper St. Clair River where change was consistently indicated in their analysis. This is important, because the author shows over and over, various inconsistencies associated with the conclusions drawn from various statistical analyses, leading him to conclude, for example, that "no clear pattern emerges from the data" regarding changes in channel bed volumes. However, Bennion was able to discern some patterns in elevation changes as shown in Figure 5 as well as to show movement of bottom materials. Importantly, Bennion also documents clearly that if one wishes to determine whether or not significant changes in bed volumes occur, then higher resolution datasets will be required, similar to the 2007 dataset. The lower resolution datasets associated with previous measurement campaigns, prior to 2007, were simply not adequate to provide the type of definitive conclusions regarding channel bed volumetric changes sought in this study. This is an important and useful finding to help target future resources.
No response needed

B. What is the most critical aspect of the study/analysis? Why?

Since the purpose of this study is to determine whether the morphology of the St. Clair River has changed and whether or not zones of active erosion and deposition can be identified, the study hinges upon its findings that: (1) no clear patterns emerged from the analysis of volumetric changes and (2) Figure 5 highlights the areas of consistent changes in channel elevations. Thus the most critical aspects of the work involves the error analyses, comparisons and interpolations performed on the DEM grids. This analysis appeared quite reasonable and generally supports the findings. However, as described below, there are numerous ways in which the author could be more clear in his explanations of the various changes he has observed using summary statistics.
See comments below

C. Which aspect of the analysis/modeling is weakest? Why? How can it be improved?

Throughout the report, the author uses statistical terms like root mean square error, standard deviation, mean, etc. However, he often uses them too loosely, without defining them clearly. For example, all of these (mean, standard deviation, root mean square error) are statistics of a random variable, yet he often does not state clearly which random variable he is referring to. An example of this is in paragraph 1 on page 6 where Bennion states that ‘the root-mean-square error was examined’ and then again he states that “a close relation of the mean standard error value to the root-mean square error value indicates …” What I would ask him to clarify is what exactly he is referring to. That is, ‘root-mean-square error’ of what? Similarly ‘mean standard error’ of what? I believe here he is referring to the root mean square error associated with the elevation errors and the mean standard error associated with the elevation errors. Note that mean square error of a variable is equal to the square of its standard error plus the square of its bias. Thus saying that the mean standard error (of the elevation errors) is related to the root mean square error (of the elevation errors) simply implies that the bias is low. Why not just examine bias and say that bias is low. A reading of the report indicates that the author doesn’t seem to understand the meaning of these basic statistics, yet he reports them over and over again.
Generally, specific labeling of which dataset a statistic is derived from was not done because the report is discusses the different datasets in sections that limit the discussion to one data type (ie: survey data, interpolated data, results of comparisons). That being stated, edits will be made to more clearly label which variable a statistic applies to when the discussion involves statistics from multiple sources.
Bias is not examined directly simply because the other statistics in question (mean standard error of the elevation errors and root mean square error of the elevation errors) are generated as part of the interpolation process and provide an estimation of bias without involving separate testing. Because of the number of interpolations that are conducted in the process of generating a complete surface, separately testing bias on each result was unrealistic in the time frame of the study, and not a standard practice when assessing interpolation accuracy. Language will be added that clarifies that the reason the mean standard error of the elevation errors and root mean square error of the elevation errors is inspected is to provide a qualitative assessment of interpolator bias.
Overall, efforts were made in writing this report to use plain language whenever possible. To someone well versed in statistics this may seem simplistic, but the author feels that keeping the use of specialized terms to a minimum increases the accessibility of the document to a general audience.

Another example of lack of statistical clarity: page 4, column 1, paragraph 2, “Vertical uncertainty associated with… is assumed to be 0.3 m.” What is this uncertainty relating to? Are you quoting a standard deviation? Here and elsewhere you need to be more clear about what you mean by uncertainty.
“Vertical uncertainty associated with the survey process for the 1971 and 2000 datasets is assumed to be 0.3 m.” This is not a standard deviation, but a +/- accuracy value reported

by the agency that conducted the survey. Unfortunately this is the only information available concerning survey uncertainty. The value is reported in the standard way survey uncertainty is conveyed. The text will be edited to include the +/- in front of the uncertainty value.

All random variables can have a mean, a standard deviation (sometime called standard error), a mean square error, and a root mean square error. So one must always follow those terms by the variable of interest. For example, in Table 5 he reports the mean absolute error and the standard deviation of the errors for the 1-meter DEM's.
Again, edits will be made to more clearly label which variable a statistic applies to when the discussion involves statistics from multiple sources.

Another concern I will raise which is minor here, but I raise it in the hopes that you will do better in the future. This concerns your tests of normality in column 2 on page 5. There you quote several previous studies who have suggested how to test whether a sample is normally distributed. Neither of the methods used are efficient or entirely reliable methods because they do not account for differences in sample size which impact the fit of a normal distribution to data. Instead, a preferred approach is to construct a normal probability plot and to evaluate the linearity of the plot using the probability plot correlation coefficient hypothesis test. This method is outlined in Chapter 18 of the *Handbook of Hydrology*, McGraw Hill, 1993.
Not standard practice (see sources in report) for interpolation assessment due to time and computational limitations, but the suggestion is appreciated and will be further researched for practical application.

D. Are there any other suggestions that are related to how this analysis may be used more effectively or the results explicated in a more understandable manner?

Abstract – Bennion says: "and interpolation processes limit the statistically certain results".
I suggest he say "and interpolation processes limit the statistical certainty of the results".
Edited

Abstract – Bennion says: "within the range of uncertainty associated with the datasets". I suggest he say "within the range of measurement and interpolation uncertainty associated with those methods".
Edited

Page 4 – column 1 – paragraph 5 – 'data are predisposed to certain results from comparison". This is very misleading and confusing. Please rewrite.
Edited to say "Inferences can be made about the outcome of comparative analysis based on comparison of the dataset's statistical properties"

Page 5 – column 1 – paragraph 1 – please write out the equation for DEMcell size, because your use of square root * (area / # of points) is not at all clear. The * symbol is usually a multiplication symbol, but I don't think you intend that here. Please use a standard equation.

Edited $\sqrt{area / \# of po\,\text{int}\,s}$

Page 5 – column 1, second to last line, '10 percent of the original dataset that were withheld'. Please clarify how they were withheld.
Edited - test sets were extracted via the use of the "Create Subsets" tool as part of the ArcGIS Geostatistical Analyst extension.

Page 5 – column 2, line 6, "allowing the best 95-percent points'. Please clarify what you mean here. What is a 'best 95-percent point'? This is very nonstandard language. How can one quantile be better than another?
Edited " allowing the best 95-percent *of the* points"

Page 6 – column 1 – paragraph 2, 'maps of prediction errors' please clarify. Prediction errors of what?
Edited – kriging prediction standard errors

Page 6 – column 2 – formula for 'total potential cut/fill error is missing a right hand parenthesis.
Edited - Cell Count * Cell Area * (Zbefore – Zafter)