Title: Using GIS to Estimate Lake Volume from Limited Data

Hollister, J.W. and W. B. Milstead 2009. Using GIS to Estimate Lake Volume. Lake Reserv Manage. 00:00-00.

Abstract

Estimates of lake volume are necessary for calculating residence time and modeling pollutants. Modern GIS methods for calculating lake volume improve upon more dated technologies (e.g. planimeters) and do not require potentially inaccurate assumptions (e.g. volume of a frustum of a cone), but most GIS methods do require detailed bathymetric data which may be unavailable. GIS technology cannot correct for a lack of data; however, it can facilitate development of methods that better use the relatively simple, and more widely available measurements of lake shape and maximum depth. In this research note we describe a method to model bathymetry and estimate the volume of a lake with a limited set of data that consists only of a maximum depth measurement and a GIS layer of lake shoreline. Using a simple linear transformation, we estimate depth as a function of distance from shoreline and with the resultant information estimate lake volume. We applied and compared this method with estimates derived from field bathymetry data of 129 lakes in New Hampshire. In New Hampshire lakes, the assumption of depth as a function of distance is appropriate and the simple GIS method has lower overall error than simply using the formula for volume of a cone to estimate lake volume. This approach has broad implications in the assessment of lake condition from national surveys (e.g. USEPA’s National Lakes Assessment) and should improve upon models of nutrients, contaminants, and hydrology even in the absence of detailed bathymetric data.

Key words: bathymetry; geographic information systems (GIS); lake morphometry; mean depth; New Hampshire; volume

Scientist and managers need information on lake volume to estimate lake residence time, model concentrations of pollutants and nutrients, and calculate lake productivity. Inaccuracies in the estimation of lake volume will therefore impact our ability to fully understand and manage lakes. In an ideal situation, methods used to calculate lake volume to keep pace with current technologies that have been assessed for accuracy and precision, and allow for reproducibility of results. However, this is not always the case.

Lake volume is still commonly estimated by calculating the area of depth contours from paper maps with a planimeter; a method that has not changed significantly over the last century (see Welch 1935). For each contour slice, the volume is estimated by applying the formula for volume of the frustum of a cone; total volume is the sum of the volumes of individual slices (Kalff 2002, Wetzel and Likens 2000). While this method has worked well, it assumes that the lake basin is shaped as multiple conic frusta (Figure 1a), requires a complete bathymetric survey to generate contour maps from which to calculate area of depth contours and it does not take advantage of advances in technology such as Geographic Information Systems (GIS). A more modern approach and one often used in terrain modeling is to use GIS to estimate the volume of a Triangulated Irregular Network (TIN) created from the bathymetry points in the bathymetric survey (e.g. Zhou et al. 2008). This method also requires a complete bathymetric survey, but is based on a more realistic model of the actual lake basin (Figure 1b). Given the existence of a detailed bathymetric survey, an analyst or lake manager need go no further than calculation of a TIN; however, detailed bathymetric data are expensive to collect and are not always available. This is especially true when a large number of lakes are involved. In these instances, a GIS may be used to facilitate development of methods that better use simple and limited data, such as lake shape and maximum depth.

The recently completed National Lakes Assessment (NLA) is a good example of this situation. The only data available for all lakes included in the NLA were GIS layers of lake shoreline, and thus lake area, and a field measured estimate of maximum depth. Given this limited set of data there are only a couple of possible conceptual models for estimating lake volume. The simplest is to assume a conical volume (Figure 1c). While this assumption is quick and easy to apply, it lacks realism. Alternatively, one could assume that depth is a function of distance from shore (Figure 1d). This assumption is easy to implement in a GIS, requires very little data, and incorporates a higher degree of realism.

We developed a simple GIS method that uses available data (i.e. lake polygons and a single estimate of maximum depth) to improve the accuracy of lake volume estimates over those based on the traditional conic volumes. Our goal for this research note is to describe the method we developed to estimate the volume of a lake with this assumption, apply it on a subset of New Hampshire lakes and then assess the methods accuracy at estimating lake volume as compared to volume calculated with the formula for a cone.

Materials and methods:

Study Site:

This is a “Green Open Access” copy of the manuscript. The final formatted article is available from: http://dx.doi.org/10.1080/07438141.2010.504321

The methods we describe in this note are applicable for all lakes where data on shoreline location and maximum depth are available. For accuracy assessment, we limited the analysis to 129 NH lakes for which we have detailed bathymetric data (Figure 2). The lakes range in size from approximately 2.8 ha to 3147 ha with an average size of approximately 870 ha. Field measured maximum depth for these lakes ranges from 1 m to 38 m with an average field measured maximum depth of 10 m. The lakes span the entire state of New Hampshire and represent a wide range of geophysical settings.

Estimating volume using depth as a function of distance:

If we assume a linear increase in depth with distance from the shore then we can estimate the lake depth at any point with the following simple linear transformation:

$$Z = \frac{D * Z_{max}}{D_{max}} \qquad (1)$$

Where Z is the depth for any given location, D is the Euclidean distance from the shoreline, including islands, Z_{max} is the measured maximum depth for a given lake and D_{max} is the maximum distance from the shoreline of a given lake. To apply this formula across an entire lake requires the following steps: 1) convert polygon lake data to raster (a standard procedure in raster-enabled GIS packages, such as ArcGIS 9.3 ™ with the Spatial Analyst ™ extension), 2) calculate distance of each cell in lake to lake shoreline, and 3) use formula 1 to transform distance to depth. The result of these 3 steps is a GIS raster layer with each cell (of known area) representing an estimate of depth at that point.

This is a "Green Open Access" copy of the manuscript. The final formatted article is available from: http://dx.doi.org/10.1080/07438141.2010.504321

Estimating lake volume is simply a matter of calculating (i.e. cell area x depth) and summing the volume across all cells.

$$LakeVolume = \sum_{i=1, j=1}^{n} CellArea * Depth_{i,j} \qquad (2)$$

Based on these formulas, we developed a script in the R language for statistical computing that uses R tools for spatial data handling and ArcGIS 9.3 ™ scripting objects via the R library RPyGeo for the spatial data analysis that implements these formulas (Supplement 1; Brenning 2009, R Development Core Team 2009). We refer to this method as the "distance method."

Assessment data and methods:

Bathymetry data for 129 lakes in New Hampshire were acquired from the NH Department of Environmental Services (Robert Estabrook and Scott Ashley, pers. comm.). For each bathymetry point in a lake the distance to the shore was calculated using a Euclidean distance function in the GIS. We correlated distance to shore with depth to test the assumption that depth is a function of distance from shore; significant, positive Pearson's correlation coefficients support the assumption.

Maximum depth values from the field bathymetry survey were used to estimate volume for both the conical formula and the distance method. For each lake the bathymetric data density was sufficient to create a TIN that provided the most accurate representation of the three dimensional structure of the lake basin that was available to us. A "true" lake volume was then calculated from the TIN and compared to the conical and distance method volume estimates.

Four methods were used to assess accuracy. First we calculated the percent difference for each lake between each method and the true volume (i.e. TIN volume);

$$PD = \frac{EstimatedVolume - TrueVolume}{TrueVolume} \quad (3)$$

the second metric we used was root mean square difference (RMSD) to compare the differences between methods, and;

$$RMSD = \sqrt{\frac{\sum(EstimatedVolume - TrueVolume)^2}{N_{Lakes}}} \quad (4)$$

the third metric was the probability that a given method more accurately estimated the "true" volume. The formula for calculating this for the distance method is

$$P(MoreAccurate) = \frac{N_{|DistVolume - TrueVolume| < |ConicalVolume - TrueVolume|}}{N_{Lakes}} \quad (5)$$

To estimate confidence intervals for all three metrics we used 1000 bootstrapped samples and used the 0.975 and 0.025 percentiles as estimates of the upper and lower confidence limits (Hollister et al. 2009, Manly 2007).

Lastly, we used linear regression (estimate vs. true) to assess accuracy (Hollister et al. 2004). As a measure of accuracy, two values would be in perfect agreement when the regression of those two values have a an R^2 equal to one, a slope (β_1) equal to one, and an intercept (β_0) equal to zero. The volume estimator that has the R^2 and β_1 nearest to one and β_0 nearest to zero is the most accurate.

Results and Discussion

The distance method volumes ranged from 0.456 km^3 to 116 km^3 with a mean value for all lakes of 3.62 km^3. The conical volumes ranged from 0.420 km^3 to 161 km^3 and had an average value of 4.21 km^3. The "true" volumes ranged from 0.347 km^3 to 81.0 km^3 with a mean of 3.60 km^3 (Supplement 2). A student's t-test on the differences revealed no significance difference (α = 0.05) between either the distance method estimate and the "true" value or the conical estimate and the "true" value.

The assumption that depth is a linear function of distance is reasonable. Using a simple ratio of the Z_{max}:D_{max} results in a linear transformation that is applicable to lakes with widely varying Z_{max}:D_{max} ratios (Figure 3). Also, of the 129 lakes tested, 123 (95.3%) had significant (α = 0.05), positive correlation coefficients (mean= 0.606; Supplement 2). For the remaining 6 lakes the distance method was a better estimator of volume than the conical formula, but distance and depth were not significantly correlated (3 lakes) or the correlation coefficient was negative (3 lakes). The results support the assumption that depth increases as distance from shore increases.

The mean PD was slightly lower for the conical method than for the distance method (0.008 vs. -0.03); however the range and standard deviation of the PD was larger for the conical method (range = -0.46 – 2.93, stand. dev = 0.46) than for the distance method (range = -0.44 – 2.62, stand. dev = 0.37). The distance method outperformed the conical method 59% of the time and had a lower RMSD (3.36 km^3 vs. 7.15 km^3), and when compared to the "true" volume, the distance method also more closely approximates the expected one-to-one relationship, although the slope estimates were not significantly different at an $\alpha = 0.05$ (Figure 4). Lastly, the differences between the error measures for each method are not statistically significant as estimated by bootstrapped confidence limits (Table 1) and there was no apparent pattern in lake morphology that would predict which method performed better.

All metrics indicate that differences between both estimates and the true volume were slight (Table 1 and Figure 4).   This suggests that for cases where a GIS is unavailable the conical volume, based on an estimate of maximum depth and lake area, should provide reasonable estimates. Although the conical and distance methods were similar in their estimates of volume, in instances where a GIS may be used, the distance method would be preferable because of the fact that for 59% of the lakes the distance method estimate was closer to the "true" value. Although not significant at an $\alpha = 0.05$, the probability is significant at an $\alpha = 0.07$. This suggests that the higher probability of distance method more accurately estimating lake volume is likely real, albeit slight.

This approach has broad implications in the assessment of lake condition from national surveys (e.g. USEPAs National Lakes Assessment) and should improve upon models of nutrients, contaminants, and hydrology even in the absence of detailed bathymetric data. This is especially important because broad-scale models that incorporate hydrology have often reduced lakes into flat surfaces or center lines. Ignoring the in-lake processes that are so closely linked to volume and residence time (i.e. productivity, nutrient cycling, etc.) in these models results in less realistic, and ultimately, less useful predictions. Simple GIS methods that are based on a very limited set of data can possibly improve upon these predictions.

One key limitation to this approach, especially if it is to be applied at broad scales, is the availability of field measured depth data. In cases where field measurements are unavailable, an important addition to this process will be to develop GIS methodologies based on publically available data to estimate maximum lake depth. In fact some work has been done across a wide range of lakes in Europe to predict mean lake depth (Pistocchi and Pennington 2006). Similar approaches might prove useful in predicting maximum lake depth that could then be used as an input to the lake volume estimation methods described in this note.

In summary, we have presented a simple GIS based method for estimating volume when the only data available are spatial data on the shape of the lake shoreline and an estimate of maximum depth. This method is based upon a more realistic conceptual model and tends to be a more accurate estimate of lake volume that a simple conical volume using the same data. This method does provide a reasonable estimate of lake volume that can be used to estimate mean depth and, if flow data are also available, hydraulic residence time. However, the method does not replace the need for traditional bathymetry surveys when greater detail is required.

Authors’ Names: Jeffrey W. Hollister and W. Bryan Milstead

Address: U.S. Environmental Protection Agency, Office of Research and Development, National Health and Environmental Effects Research Laboratory, Atlantic Ecology Division, 27 Tarzwell Drive, Narragansett, RI 02882, email: Hollister.Jeff@epa.gov, phone: 401.782.9655

Acknowledgements:

We would like to thank Bob Estabrook and Scott Ashley from New Hampshire’s Department of Environmental Services for providing us with the bathymetry data we used to assess our methods described in this research note. Additionally, we would like thank Anne Kuhn-Hines, John Kiddon, Henry Walker, and the anonymous reviewers for the comments and criticisms that greatly improved the manuscript. The research described in this paper has been funded by the US Environmental Protection Agency. This paper has not been subjected to Agency review. Therefore, it does not necessary reflect the views of the Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use. This is contribution number AED-10-018 of the Atlantic Ecology Division, Office of Research and Development, National Health and Environmental Effects Research Laboratory.

References:

Brenning, A. 2009. RPyGeo: ArcGIS Geoprocessing in R via Python. R package version 0.9-1. Accessd online April 14, 2010: http://CRAN.R-project.org/package=RPyGeo

Hollister JW, Gonzalez ML, Paul JF, August PV, and Copeland JL. 2004. Assessing the accuracy of the National Land Cover Dataset area estimates at multiple spatial extents. Photogramm Eng Remote Sens. 70:405–414.

Hollister JW, Walker HA, Paul JF. 2008. CProb: a computational tool for conducting conditional probability analysis. J. of Environ. Qual. 37:2392–2396

Kalff J. 2002. Limnology: Inland Water Ecosystems. Upper Saddle River (NJ): Prentice Hall.

Manley BFJ. 2007. Randomization, bootstrap, and Monte Carlo methods in Biology, 3rd edition. Chapman Hall/CRC, Boca Raton, FL, 480 pp

Pistocchi A, and Pennington D. 2006. European hydraulic geometries for continental SCALE environmental modeling. J Hydrol. 329(3-4):553-567.

R Development Core Team (2009) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Accessed online April 14, 2010: http://www.R-project.org

Welch, PS. 1935. Limnology. New York(NY): McGraw-Hill.

Wetzel RG, and Likens GE. 2000. Limnological Analyses. 3rd ed. New York(NY): Springer.

Zhou Q, Lees B, and Tang G. 2008. Advances in Digital Terrain Analysis. New York(NY):Springer-Verlag.

This is a “Green Open Access” copy of the manuscript. The final formatted article is available from: http://dx.doi.org/10.1080/07438141.2010.504321

Figure Captions:

Figure 1. Comparison of the realism and data requirements for various conceptual models of lake volume.

Figure 2. Map of New Hampshire lakes with bathymetry data.

Figure 3. Scatterplots of depth vs. distance for three lakes with the minimum, mean, and maximum Z_{max}:D_{max} ratios. Points are depth and distance from bathymetry surveys and the dark grey line represent the predicted depths using the Z_{max}:D_{max} ratio.

Figure 4. Scatterplot of ln-ln relationship between estimated conical volumes (km^3), distance method volumes (km^3) and “true” volumes (km^3). Black line represents perfect agreement.

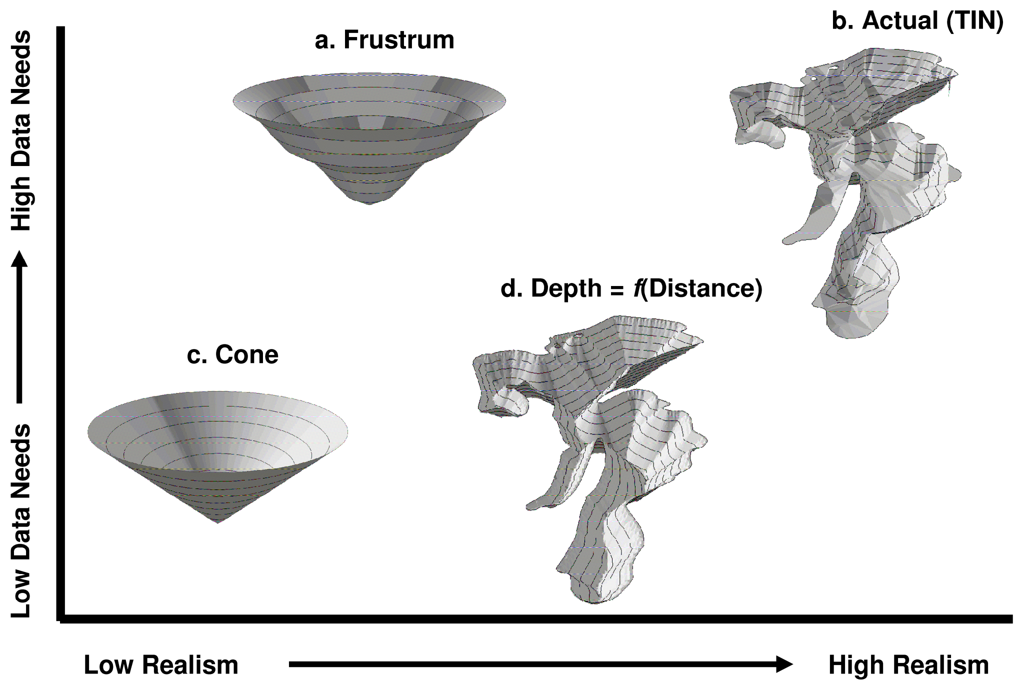
a. Frustrum
b. Actual (TIN)
d. Depth = *f*(Distance)
c. Cone
High Data Needs
Low Data Needs
Low Realism
High Realism

Figure 1

This is a “Green Open Access” copy of the manuscript. The final formatted article is available from: http://dx.doi.org/10.1080/07438141.2010.504321

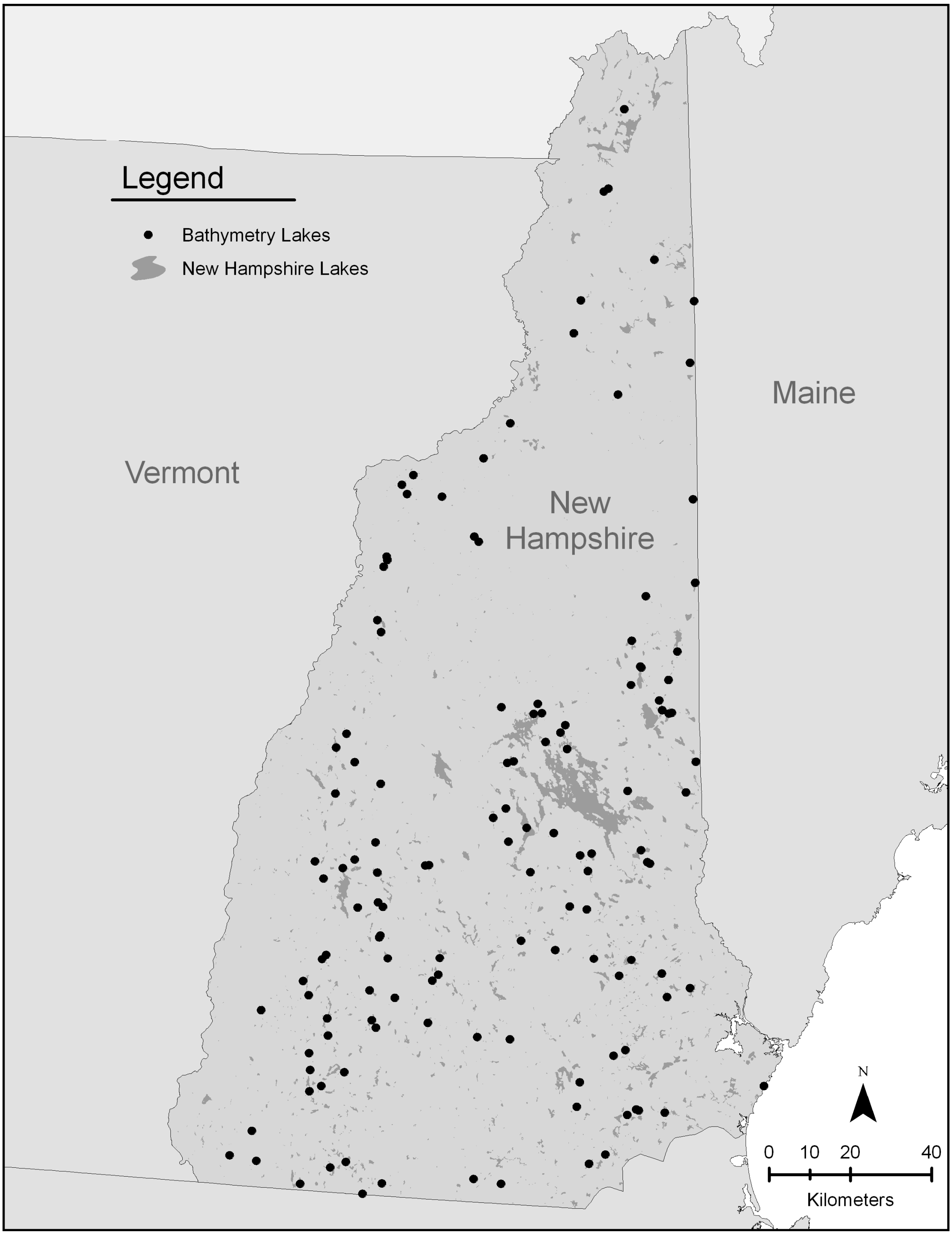

Figure 2

This is a “Green Open Access” copy of the manuscript. The final formatted article is available from: http://dx.doi.org/10.1080/07438141.2010.504321

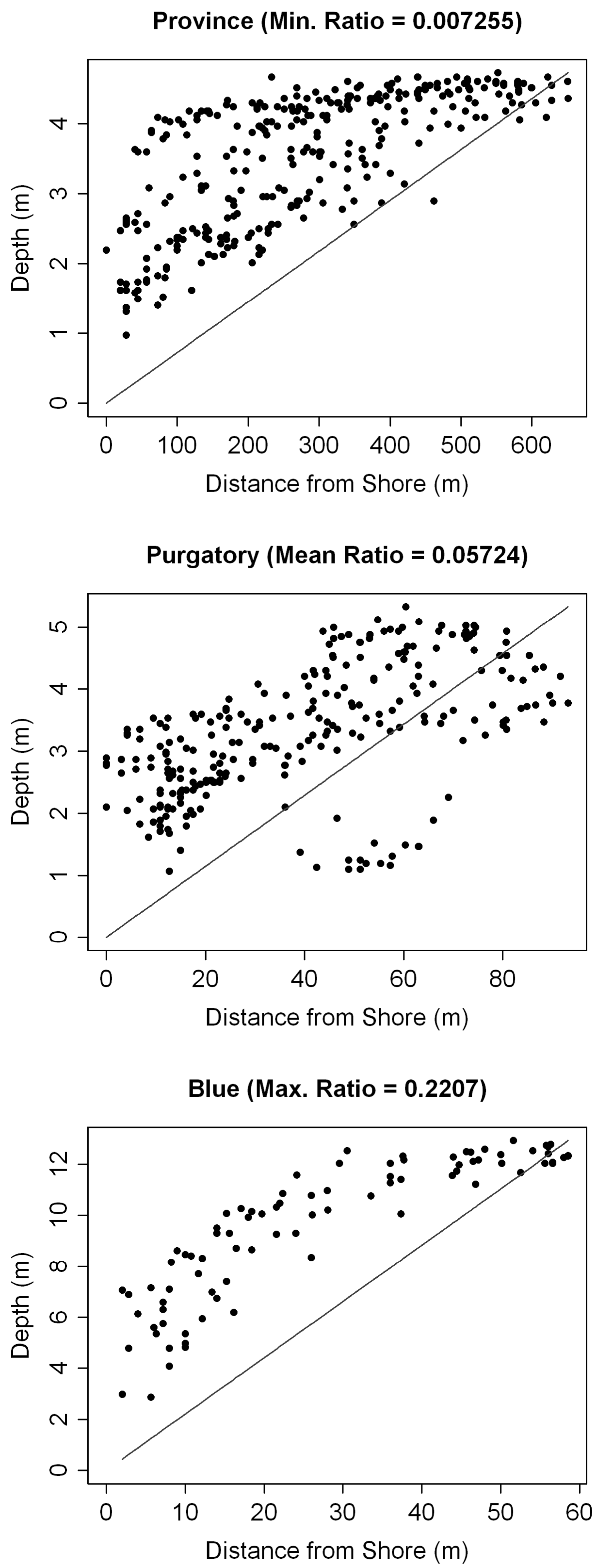

Figure 3

This is a “Green Open Access” copy of the manuscript. The final formatted article is available from: http://dx.doi.org/10.1080/07438141.2010.504321

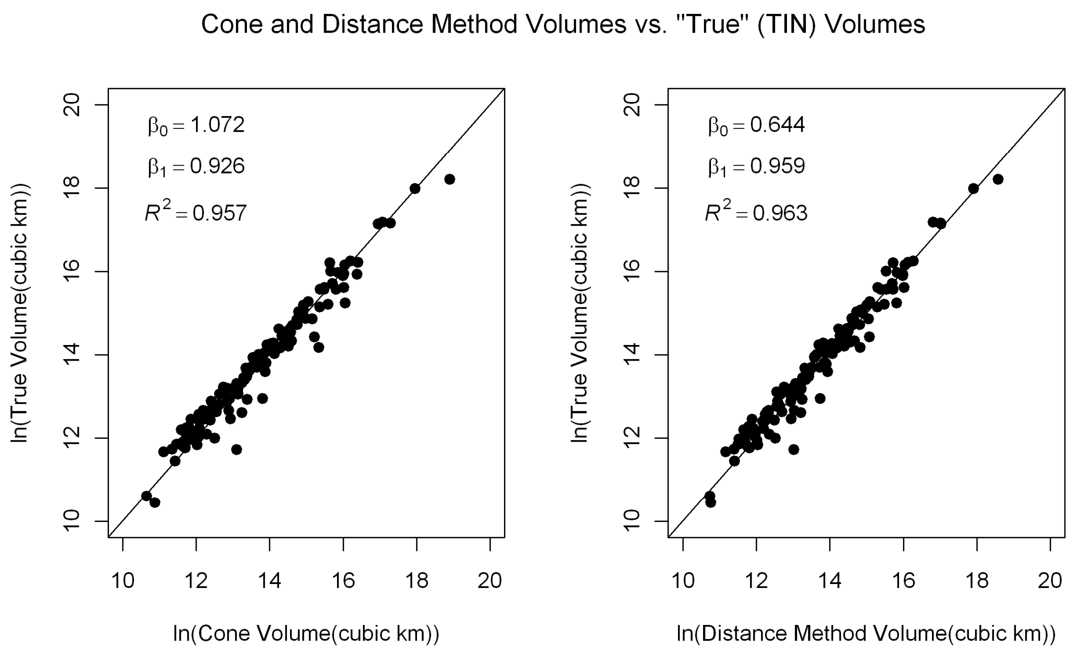

Figure 4