

# Predicting bathymetric features of lakes from the topography of their surrounding landscape

Adam J. Heathcote, Paul A. del Giorgio, and Yves T. Prairie

**Abstract:** Estimating the distribution of water across the landscape is critical to understanding how local aquatic biogeochemical processes may upscale to regional or continental scales. Whereas technology to estimate the areal extent of inland waters has improved dramatically, predictions of lake bathymetry still rely primarily on correlations with lake area that lack a mechanistic underpinning. Using topographically diverse regions of Quebec (Canada), we developed a model for predicting lake volume and depth that relies on geographic data that are widely available, which can be easily adapted to other regions. We found that the average change in relief between the surrounding terrestrial landscape and the lake surface to be the best predictor of bathymetric properties (lake volume, lake depth). Unlike previous models, our method provides a clear mechanistic link between relief outside and within the perimeter of a lake that is supported by basic geographic principles. This model will be useful in estimating the volumetric distribution of water as well as the distribution of lake depth, which are integral to our understanding of biological and geochemical lake processes.

**Résumé :** L'estimation de la répartition de l'eau dans le paysage est essentielle à la compréhension de la transposition de processus biogéochimiques aquatiques locaux à des échelles régionales ou continentales. Malgré l'amélioration marquée des techniques d'estimation de l'étendue spatiale des eaux intérieures, les prévisions de la bathymétrie des lacs reposent toujours principalement sur les corrélations avec la superficie des lacs, corrélations sans fondement mécaniste. En nous servant de régions du Québec (Canada) présentant des reliefs variés, nous avons élaboré un modèle pour prédire le volume et la profondeur des lacs qui repose sur des données géographiques largement accessibles et qui s'adapte facilement à d'autres régions. Nous avons constaté que le changement moyen du relief entre le paysage terrestre environnant et la surface du lac constituait le meilleur prédicteur des propriétés bathymétriques (volume et profondeur du lac). Contrairement aux modèles antérieurs, notre méthode offre un lien mécaniste clair entre le relief à l'extérieur et à l'intérieur du périmètre d'un lac, qui est appuyé par des principes géographiques de base. Ce modèle sera utile pour l'estimation de la répartition volumétrique de l'eau, ainsi que la répartition de la profondeur des lacs, des caractères centraux pour comprendre les processus biologiques et géochimiques associés aux lacs. [Traduit par la Rédaction]

## Introduction

There is an increasing recognition of the global biogeochemical role of inland waters, driven in part by reassessments of both their areal extent and the intensity of local aquatic processes (Downing et al. 2006; Cole et al. 2007). Although we continue to improve our knowledge of ecosystem-level processes in lakes, upscaling these results is dependent upon accurate estimates of lake size and shape (Bastviken et al. 2004; Downing 2010; Ferland et al. 2012; Raymond et al. 2013). Additionally, key lake ecosystem functions, such as trophic state and primary production, are not well correlated with lake size, but instead are best predicted by lake depth (Vollenweider 1975; Canfield and Bachmann 1981), and consequently upscaling of these properties without lake-specific knowledge of bathymetry is unreliable. Unfortunately, even modern lake survey methods to measure lake bathymetry (e.g., SONAR echosounding) become impractical and cost-prohibitive at larger scales, as even modest regional upscaling exercises may incorporate tens of thousands of lakes (Emmertson et al. 2007).

Techniques to estimate the areal extent of inland waters have improved global estimates of the size distribution of lakes (Lehner and Döll 2004; Downing et al. 2006; McDonald et al. 2012); how-

ever, the prediction of bathymetric properties (e.g., lake volume, maximum depth, mean depth) often still relies on correlations with lake area (Håkanson and Karlsson 1984). While these models are simple to replicate, they lack a mechanistic underpinning (i.e., large lakes need not be deep) and ignore other geographic information (e.g., topography) despite a long-standing assumption that the lake and terrestrial environment were often formed by a common physical processes, especially in glacially influenced landscapes (Hutchinson 1957).

Assuming the lakebed and the surrounding environment were formed by similar geological processes (e.g., glaciation), the topographical relief below the surface of the lake should be an extension of the surrounding environment. Consequently, lakes surrounded by landscapes with large changes in elevation may have similarly large changes in elevation below their surface, leading to an increased potential for deeper lakes with larger volumes of water. Following this assumption, more recent models have begun to include relief of the surrounding environment into predictions of lake bathymetry. Håkanson and Peters (1995) proposed a model that predicted lake volume and maximum depth from the relief of the drainage area using a set of Swedish lakes,

Received 26 August 2014. Accepted 18 February 2015.

Paper handled by Associate Editor David Brickman.

A.J. Heathcote,\* P.A. del Giorgio, and Y.T. Prairie. Groupe de Recherche Interuniversitaire en Limnologie – UQAM, Département des sciences biologiques, Université du Québec à Montréal, Case postale 8888, Succursale Centre-ville, Montréal, QC H3C 3P8, Canada.

**Corresponding author:** Adam J. Heathcote (e-mail: [adam.heathcote@gmail.com](mailto:adam.heathcote@gmail.com)).

\*Present address: St. Croix Watershed Research Station, 16910 152nd Street North, Marine on St. Croix, MN 55047, USA.

and Hollister et al. (2011) proposed a model for predicting maximum depth from the median slope in the surrounding lake catchment from lakes in the northeastern United States of America. Both examples, however, required knowledge of the individual lake catchment, which is often unavailable and time-consuming to estimate for a very large number of lakes. Thus, though useful, these models lacked a simple and universal application that could be used at very large spatial scales or in regions where geographic information systems (GIS) resources are relatively scarce.

In acknowledgment of this problem, Sobek et al. (2011) developed a model to predict lake volume and depth from the maximum slope of the surrounding land within a static buffer (50 m wide) using a large set of lakes from the Swedish lake register. This model overcame the aforementioned problems by requiring only data on topographic relief and hydrographic information that are available for most of the planet (see [asterweb.jpl.nasa.gov](http://asterweb.jpl.nasa.gov) for coverage and data availability). Sobek et al.'s multiple regression model for lake volume, based on lake area and topographic parameters, performed better than lake area alone ( $R^2 = 0.92$  versus 0.89 for lake area alone), but these variables explained 36% or less of the variation in lake depth.

In this regard, although the prediction of lake volume itself is extremely useful, particularly for the determination of water retention time and for upscaling purposes, the inability of previous models to accurately predict maximum depth directly or mean depth (defined as lake volume/lake area) is problematic because many key limnological, ecological, and biogeochemical parameters of interest are strongly dependent on depth. Lake depth is critical to the physical structure of the water column and consequently on processes such as thermal stratification, light penetration, and gas exchange (Gorham and Boyce 1989). Lake depth and shape also strongly influence sediment dynamics, including the arrival of materials to sediments and the mean sediment temperature, which in turn influence sediment metabolism and carbon accumulation (Gudasz et al. 2010; Ferland et al. 2014). There are also examples where depth is either the single best predictor (e.g., lake productivity; Canfield and Bachmann 1981) or a necessary part of a more complex lake model for limnological processes (e.g., lake anoxia; Nürnberg 1995).

Given the limited success of current models in predicting both lake volume and depth across heterogeneous landscapes, we set out to further explore the relationships between lake bathymetry and the landscape and attempt to find satisfactory and easily reproducible models for predicting lake volume and maximum depth. We hypothesize that these variables, which describe the shape of a lake, will be related to differences in topography of the surrounding landscape. Additionally, we propose a different method for delineating the area surrounding a lake used in the model that is scaled to lake size rather than using a static buffer or requiring knowledge of the catchment.

## Methods

### Study area

For this study we selected 433 natural lakes of glacial origin located in boreal and sub-boreal Quebec (Canada). Lakes were divided among five distinct regions of Quebec (Fig. 1): Abitibi ( $n = 134$ ; 46°N–49°N, 75°W–80°W), Chicoutimi ( $n = 16$ ; 47°N–49°N, 71°W–72°W), Eastmain ( $n = 11$ ; 51°N–52°N, 75°W–76°W), Eastern Townships ( $n = 47$ ; 45°N–46°N, 70°W–72°W), and the Laurentians ( $n = 225$ ; 45°N–47°N, 74°W–76°W). Regions were selected to cover a wide range in topographical and bathymetric features and also based on the availability of bathymetric data (Table 1). Mean and maximum depth of the lakes ranged from 0.6 to 50.2 m and 1.5 to 125 m, respectively. Lake area ranged from 0.005 to 1030 km<sup>2</sup>, and lake volume spanned over six orders of magnitude, from 5000 to  $11.6 \times 10^9$  m<sup>3</sup>. Lake elevation ranged from 100 to 861 m, and the surrounding topography varied between 31 and 931 m above sea

level. The maximum change in elevation between the surrounding topography and the lake surface ranged from 2 to 281 m.

### Bathymetric and geographical data collection

Bathymetric data from Chicoutimi and Eastern Townships were collected from repeated transects using a transom-mounted SONAR coupled to a global position system (GPS) collecting depth and location information every 10 s while travelling at a speed of approximately 4 km·h<sup>-1</sup>. Data from Eastmain were collected as part of a previous study using a sub-bottom profiler capable of distinguishing the sediment–water interface (208 kHz) coupled to a differential GPS (see Ferland et al. 2012 for details). Existing bathymetric data collected by government agencies were compiled and digitized using the GIS software Quantum GIS (QGIS Development Team 2014) from publically available maps via the Ministère des Ressources Naturelles (MRN) and the Société des Établissements de Plein Air du Québec (SEPAQ). Data from the Laurentians were collected from the online database maintained by the Conseil Régional de l'Environnement des Laurentides (R. Carignan, unpublished; [www.crelaurentides.org](http://www.crelaurentides.org)).

Lake areas were derived from polygons contained in the National Hydro Network or digitized directly from Landsat 7 orthorectified imagery. Topographical data were extracted from the Canadian Digital Elevation Data (CDED) database at 1:50 000 scale. Hydrologic, satellite, and elevation data are all publically available via Canada's GeoBase network ([www.geobase.ca](http://www.geobase.ca)).

### Extracting topographical data from surrounding area

Lake size in our dataset spanned seven orders of magnitude (Table 1), which is consistent with recent global estimates of the lake size distribution (Downing et al. 2006). Given this large range in lake size, we assumed that lakes would be related to the surrounding landscape at a distance proportional to their size. Whereas a static buffer is constrained to trying to find a single optimal distance for all lake sizes, a buffer scaled to lake area assumes that small lakes will be related to a narrower band of topography than large lakes. This prevents the incorporation of unrelated topography in small lakes or the exclusion of relevant topography in large lakes. We developed a method to scale the terrestrial buffer zone to the diameter of a circle with an area equal to each lake, hereinafter referred to as the equivalent diameter, using the following equation:

$$D = 2 \cdot \sqrt{\frac{A}{\pi}}$$

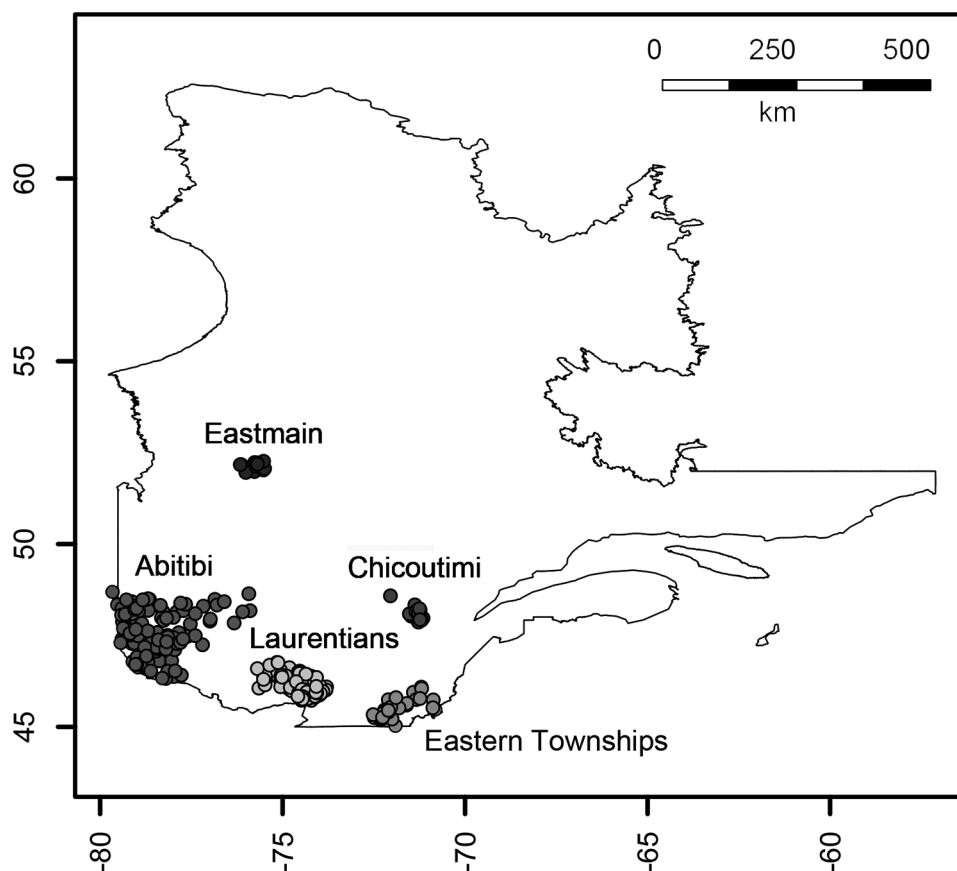
where  $D$  is the equivalent diameter and  $A$  is the lake area. Buffers of 25%, 50%, 75%, and 100% of the equivalent diameter were used to evaluate the optimal distance.

Within each buffer range, topography was summarized as the minimum, maximum, and mean elevation of slope, as well as the mean elevation change, which was defined as the difference between the mean elevation within the buffer and the elevation of the lake surface (similar in concept to mean lake depth). The buffer size used for each parameter was denoted as a subscript following the name (e.g., "elevation change<sub>25</sub>" is the elevation change in the buffer at 25% of the equivalent diameter).

### Statistical analyses

All topographic variables were considered as possible predictors for mean lake depth and lake volume. All variables were normalized by  $\log_{10}$  transformation to prevent heteroscedasticity in variance. Because of the inherent bias introduced by back-transforming estimates from regressions on log-transformed variables (Bird and Prairie 1985), back-transformation of modeled values was done using a correction for a bias towards underestimation following eq. 10 from Ferguson (1986):

**Fig. 1.** Map of the province of Quebec (Canada) showing 433 lakes selected for comparison of lake bathymetry and surrounding topography. Lakes are shaded by region and coordinates are in latitude and longitude (world geodetic system (WGS) 1984).



**Table 1.** Mean and range (in parentheses) of bathymetry and adjacent topography for lakes in this study.

| Region            | Mean elevation (m) | Elevation change (m) | Mean depth (m) | Max. depth (m) | Lake area (km <sup>2</sup> ) | Lake volume (km <sup>3</sup> )          |
|-------------------|--------------------|----------------------|----------------|----------------|------------------------------|---|
| Abitibi           | 326<br>(204–456)   | 12<br>(0–44)         | 6<br>(1–17)    | 19<br>(2–104)  | 14.4<br>(0.01–919.9)         | 0.10<br>(5.95×10 <sup>-6</sup> – 5.52)  |
| Chicoutimi        | 558<br>(155–868)   | 15<br>(4–54)         | 5<br>(1–17)    | 20<br>(5–102)  | 68.4<br>(0.04–1029.7)        | 0.79<br>(6.11×10 <sup>-5</sup> – 11.64) |
| Eastmain          | 292<br>(262–311)   | 6<br>(–1–16)         | 3<br>(1–5)     | 13<br>(2–26)   | 3.0<br>(0.04–24.7)           | 0.06<br>(5.13×10 <sup>-5</sup> – 0.62)  |
| Eastern Townships | 292<br>(173–497)   | 15<br>(3–65)         | 7<br>(1–39)    | 21<br>(2–86)   | 3.7<br>(0.05–32.0)           | 0.05<br>(7.01×10 <sup>-5</sup> – 0.74)  |
| Laurentians       | 338<br>(190–537)   | 13<br>(0–61)         | 7<br>(1–50)    | 19<br>(2–125)  | 1.0<br>(0.005–22.1)          | 0.01<br>(0.50×10 <sup>-6</sup> – 0.387) |
| All lakes         | 336<br>(155–868)   | 13<br>(–1–65)        | 6<br>(1–50)    | 19<br>(2–125)  | 8.0<br>(0.005–1029.7)        | 0.07<br>(0.50×10 <sup>-6</sup> – 11.64) |

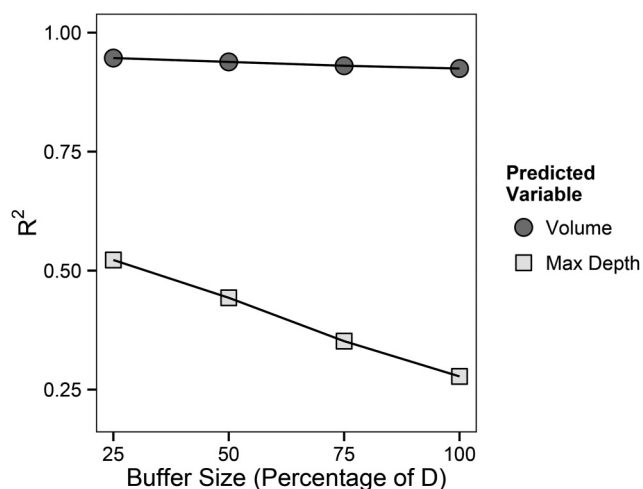
$$\hat{Y}_{corr} = 10^{\log_{10} \hat{Y}} \cdot \exp(2.65 \cdot s^2)$$

where  $\hat{Y}_{corr}$  is the bias-corrected prediction,  $\log_{10} \hat{Y}$  is the uncorrected predicted value,  $s^2$  is the residual variance from the model, and 2.65 is a constant.

Predictive models for lake volume and maximum depth were developed using multiple linear regression with all of the topographic and lake size variables described above. Predictability of mean depth was evaluated by dividing predicted lake volume by observed lake area. Variables that did not explain a significant amount of variation ( $P > 0.05$ ) were removed from the model. The unique and shared explanatory power of each variable was assessed from the Pearson’s partial correlation coefficient, which

partitions the unique and shared variance explained between multiple explanatory variables. Predictive power was assessed based on the residual standard error (RSE). Model performance was further characterized using the Akaike information criterion (AIC) (Akaike 1974) by comparing the full model (all variables), the model with all significant explanatory variables, and the model with only the single best explanatory variable. Because adding additional explanatory parameters will always increase the fit of a model, AIC balances this increasing complexity with principles of parsimony by adding a penalty for each explanatory variable. The best model was considered as the simplest model (least parameters) within two AIC units of the model with the lowest AIC score (Burnham and Anderson 2004). Finally, after

Fig. 2.  $R^2$  for the final regression model versus buffer size (as a percentage of the equivalent diameter,  $D$ ) for lake volume and maximum depth.



selecting the best model, separate regressions were produced for each region to determine the sensitivity of each model to local landscape differences. All statistical analyses were done using the statistical software R (R Core Team 2013), and the “ppcor” package was used to calculate partial correlation coefficients (Kim 2012).

## Results

### Lake volume and mean depth

Lake area ( $R^2 = 0.91$ ,  $F_{[1,431]} = 4435$ ,  $P < 0.001$ ) and elevation change<sub>25</sub> ( $R^2 = 0.63$ ,  $F_{[1,431]} = 731$ ,  $P < 0.001$ ) were the best predictors of lake volume from the multiple regression. Elevation change within the larger buffer sizes was also well correlated with lake volume, but the smallest buffer (25%) performed the best (Fig. 2). No other variables explained a significant portion of variation in lake volume. Both variables were significant predictors of lake volume in all regions except Eastmain (Figs. 3A, 3B). The best model for lake volume ( $V$ ) was

$$\log_{10} V = \log_{10} \text{lake area} \cdot 0.96 + \log_{10} \text{elevation change}_{25} \cdot 0.77$$

The predicted versus observed lake volume are shown in Fig. 4A. This model explained 95% of the variation in lake volume ( $R^2 = 0.95$ ,  $F_{[2,430]} = 3830$ ,  $P < 0.001$ ). Lake area and elevation change<sub>25</sub> uniquely explained 32% and 4% of the variation, respectively. The RSE for the model was  $0.222 \log_{10} \text{m}^3$ . Additionally, when predicted lake volume was converted to mean depth by dividing by observed lake area, the model explained 51% of the variation in mean depth (Fig. 4B;  $R^2 = 0.51$ ,  $F_{[1,431]} = 455$ ,  $P < 0.001$ ) with an RSE of  $0.159 \log_{10} \text{m}$ .

### Maximum depth

Elevation change<sub>25</sub> was the single best predictor of maximum depth ( $R^2 = 0.52$ ,  $F_{[1,431]} = 473$ ,  $P < 0.001$ ) for all lakes and was also significant for each individual region (Fig. 3C). As above, elevation change in the smallest buffer (25 m) explained the most variability in maximum depth (Fig. 2). No other variable explained a significant amount of variation in maximum depth in the multiple regression, and the model with only elevation change<sub>25</sub> had the lowest AIC value (Table 2). The equation for the best model for maximum depth ( $Z_{\text{max}}$ ) was

$$\log_{10} Z_{\text{max}} = 0.35 + \log_{10} \text{elevation change}_{25} \cdot 0.79$$

The RSE for the model was  $0.245 \log_{10} \text{m}$ . Predicted maximum depth versus observed maximum depth is shown in Fig. 4C.

### Bathymetric prediction example

We have provided here a brief example of how this methodology could be used to predict the amount of water across large regions where few bathymetric data are available. Lake depth and volume were modeled in two adjacent regions of Quebec, Chicoutimi and Chibougamau, both having high densities of lakes but differing in topography (Fig. 5). Both regions are dominated by boreal coniferous forest and have 8% (Chicoutimi) to 12% (Chibougamau) of their total area covered by freshwater lakes. Chicoutimi and Chibougamau are at similar mean elevation (430 and 414 m, respectively), but elevation change surrounding lakes is approximately two times greater in Chicoutimi (4.1 and 2.0 m, respectively). The Chibougamau region covers 76 400 km<sup>2</sup> and Chicoutimi covers 24 200 km<sup>2</sup>. Lac Mistassini, within the Chibougamau region, was excluded because of its size and extreme depth (mean depth = 75 m, maximum depth = 183 m; Downing and Duarte 2009). Lac Mistassini is the largest natural lake in Quebec, and there were no analogous systems present in our model.

Lake polygons for both regions were extracted from Canada's National Hydro Network. Large rivers were removed and multi-chain lakes were divided based on visual inspection following basic rules of length-to-width polygon ratios (see Emmerton et al. 2007). Elevation was taken from 1:50 000 scale digital elevation model with 10 m × 10 m pixels. Lakes whose buffers were less than 30 m<sup>2</sup> were removed to ensure that each prediction had at least three data points. The final dataset included 55 391 lakes from Chibougamau and 13 403 lakes from Chicoutimi.

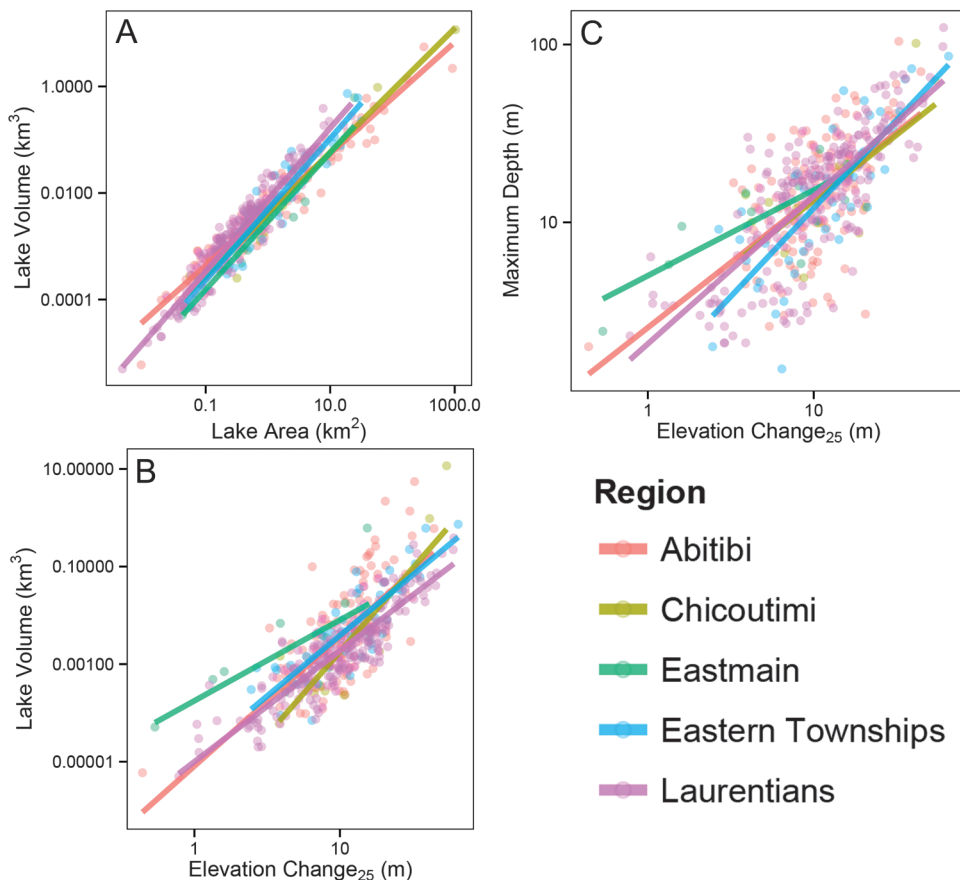
Using the proposed regression models for bathymetry developed above, we estimated the distribution of fresh water over the landscape as predicted lake volume/lake area (or mean depth) and maximum depth (Figs. 6A, 6B) and the difference in areal (land + lakes) volume of water for each region. Mean depth predictions were weighted by dividing lake area by the total area of lakes to produce the mean distribution of water depth over the region. The mean area-weighted mean depth for Chicoutimi was 10.1 m (range: 1.0–38.5 m), approximately twice that of Chibougamau (area-weighted mean: 4.6 m, range: 0.9–18.4 m), reflecting the influence of the topography and the resulting difference in elevation change between the two regions. Similarly, the estimated mean maximum depth for Chicoutimi was 11.1 m (range: 3.9–160.7 m), compared with 6.4 m for Chibougamau (range: 3.9–89.3 m). The total predicted volume of water in freshwater lakes for Chicoutimi and Chibougamau was 25.7 and 42.2 km<sup>3</sup>, respectively, which indicated that lakes accounted for a mean volume of 1.1 and 0.6 m<sup>3</sup>·m<sup>-2</sup> in each region. Based on published mean surface water runoff rates of 526 and 562 mm·year<sup>-1</sup> for Chicoutimi and Chibougamau, respectively (Fekete et al. 1999), we can calculate a lower limit for the time it would take to flush all the water from the lakes by dividing the total volume of lake water by the product of the mean runoff and the total area of the region. This yields a regional lake water retention time of 0.50 (Chicoutimi) and 1.01 (Chibougamau) years for the two regions.

### Discussion

The objective of this study was to produce a model for predicting basic bathymetric parameters (lake volume and depth) that would require only minimal knowledge of geographic features (i.e., lake area, topography) and still perform as well or better than previously proposed methods. We accomplished this by evaluating a range of descriptive topographic variables in relation to a newly compiled dataset of lake bathymetry for five distinct regions across the province of Quebec.

We found strong correlations with relatively low predictive error for all bathymetric variables and did so by only using a simple

**Fig. 3.** Relationships between bathymetry and topography by region: (A) lake volume versus lake area, (B) lake volume versus elevation change<sub>25</sub>, and (C) maximum depth versus elevation change<sub>25</sub>. Lines represent simple linear regression fits for each of the five regions. For the coloured version of this figure, refer to the Web site at <http://www.nrcresearchpress.com/doi/full/10.1139/cjfas-2014-0392>.



buffer rather than having knowledge of catchment area (cf. Håkanson and Peters 1995; Hollister et al. 2011). In comparison with the simplest models predicting lake volume from lake area (Håkanson and Karlsson 1984), our model using elevation change<sub>25</sub> and lake area reduced the RSE by 23% and improved the variance explained by 4%. Although this difference in the coefficient of determination may seem small, Prairie (1996) showed that the predictive power of regression models increases exponentially at high  $R^2$  ( $R^2 \geq 0.65$ ) based on the degree of categorical resolution. Using this criterion (eq. 5; Prairie 1996), we find that an  $R^2$  increase from 0.91 to 0.95 (as seen in our model) accounts for a 34% increase in predictive power.

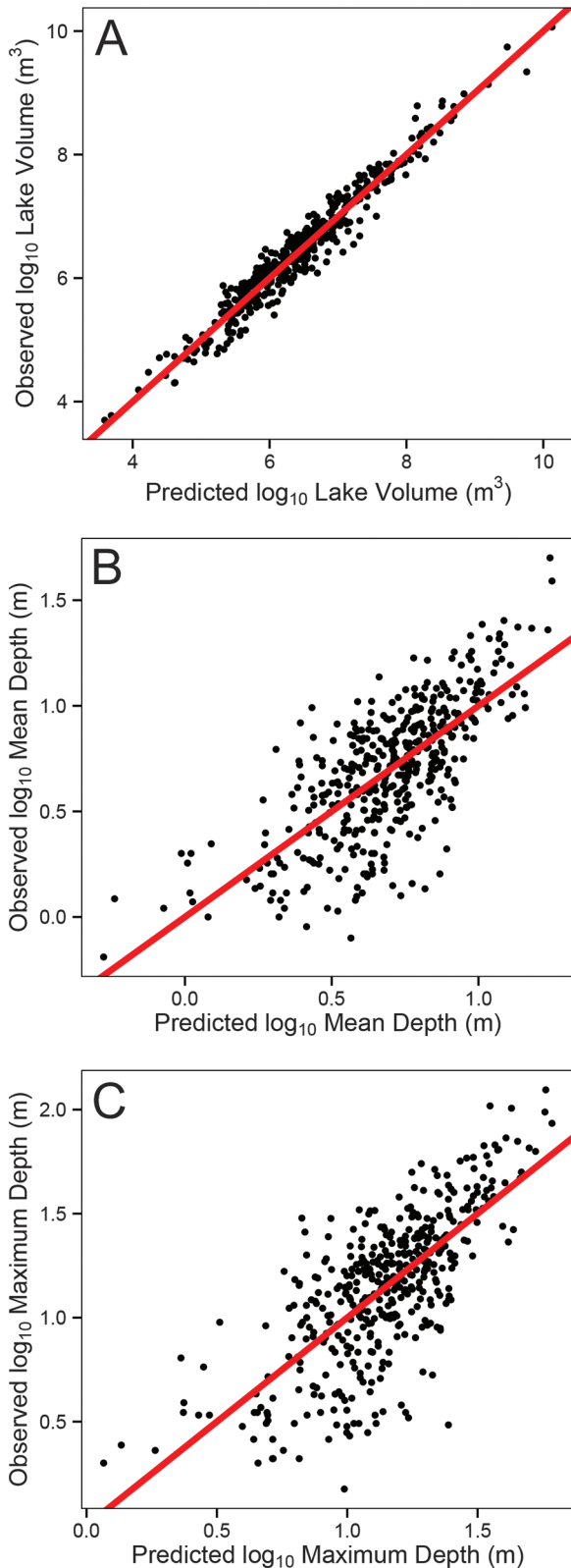
Whereas other studies, including those discussed previously, have found significant correlations between topography and bathymetry, to our knowledge, no other model has performed as well as the one we propose here in terms of predicting lake volume ( $R^2 = 0.95$ ) or lake depth ( $R^2 > 0.5$  for mean and maximum depth) without requiring knowledge of the watershed delineation. This may be explained by two novel approaches used in our study. Firstly, we proposed a new descriptor of topography, mean elevation change, which is calculated similarly to mean lake depth and is scaled to the surface elevation of each lake. This parameter was a better descriptor than slope (maximum or mean), and no other topographic parameter explained a significant amount of variation in our models after accounting for elevation change. Secondly, we did not use a static buffer around the lake from which to draw the topographic parameter. Instead, we suspected that more information might be gleaned from a larger buffer around larger lakes and vice versa. To our knowledge, this description of mean elevation change and the practice of scaling the

buffer to lake area have not been previously incorporated in any prior model.

An alternative explanation for the increased variance explained by our models in comparison with previous work (e.g., Sobek et al. 2011) is differences in sample size ( $n = 433$  versus  $n = 6943$ ). If increasing sample size also introduces additional variance around the relationship between topography and bathymetry, this alone may explain a decrease in the model  $R^2$ . To account for this, we applied the models of Sobek et al. (2011), using lake area and a maximum slope from a static 50 m buffer, to our smaller dataset and compared the  $R^2$  and AIC values with our model. We found that applying the Sobek et al. (2011) model to our 433 lake dataset resulted in a similar performance to that reported in the larger Swedish dataset ( $R^2 < 0.4$  for lake depth (mean and maximum),  $R^2 = 0.93$  for lake volume), and our model still performed best both in terms of variance explained and AIC (Table 2). Though there was a small increase in the  $R^2$  that may be attributed to differing sample sizes, this change is not large enough to account for the improved accuracy gained by using our newly developed method.

Despite the improvements to the prediction of bathymetry shown in the models we have proposed, nearly half of the variability in lake depth remains unexplained. This indicates that other unmeasured environmental factors may be important predictors of lake depth, in at least some systems. These factors could include hydrologic variables that impact lake water balance (and thus lake surface level; Almendinger 1990) or differential sediment accumulation rates (Ferland et al. 2012). Even in relatively well-studied regions such as Quebec, however, such data are limited to a relatively small number of lakes. Consequently, whereas we can only speculate on the relative importance of such factors,

**Fig. 4.** Observed versus predicted values for linear regression models shown with 1:1 line. (A) Lake volume predicted as a function of lake area and elevation change<sub>25</sub>, (B) mean depth calculated from predicted volume divided by lake area, (C) maximum depth predicted as a function of elevation change<sub>25</sub>.



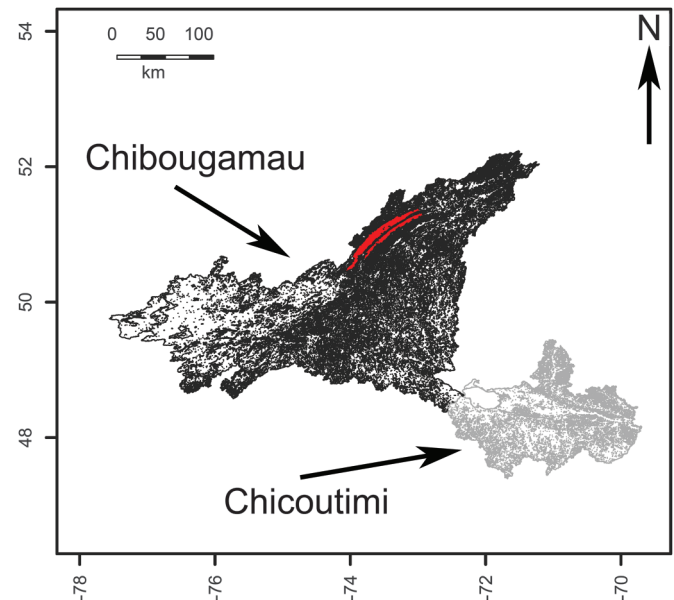
**Table 2.** Akaike information criterion (AIC) and adjusted  $R^2$  for the different predictive models tested for lake volume and maximum depth.

| Model <sup>a,b</sup>  | AIC   | $\Delta$ AIC | $R^2_{adj}$ |
|---|-------|--------------|-------------|
| <b>Lake volume</b>  |       |              |             |
| lake area + mean elevation + max. elevation + min. elevation + elevation change + mean slope + max. slope | -62.4 | 7.5          | 0.95        |
| elevation change + lake area  | -69.9 | 0.0          | 0.95        |
| lake area   | 149.2 | 219.1        | 0.91        |
| ln lake area + max slope <sub>50 m</sub> (Sobek et al. 2011)  | 64.70 | 134.61       | 0.93        |
| <b>Maximum depth</b>  |       |              |             |
| lake area + mean elevation + max. elevation + min. elevation + elevation change + mean slope + max. slope | 16.2  | 0.7          | 0.53        |
| elevation change  | 15.4  | 0.0          | 0.52        |
| ln lake area + max slope <sub>50 m</sub> (Sobek et al. 2011)  | 124.3 | 108.9        | 0.39        |

<sup>a</sup>All variables were  $\log_{10}$ -transformed unless otherwise noted.

<sup>b</sup>All topographic parameters are from the 25% buffer unless otherwise noted.

**Fig. 5.** Map showing the maximum extent and lake density of the Chibougamau and Chicoutimi regions of Quebec (Canada). Lac Mistassini, which was excluded from the analysis, is shown in red. (For the coloured version of this figure, refer to the Web site at <http://www.nrcresearchpress.com/doi/full/10.1139/cjfas-2014-0392>.) Coordinates are in latitude and longitude (WGS 1984).

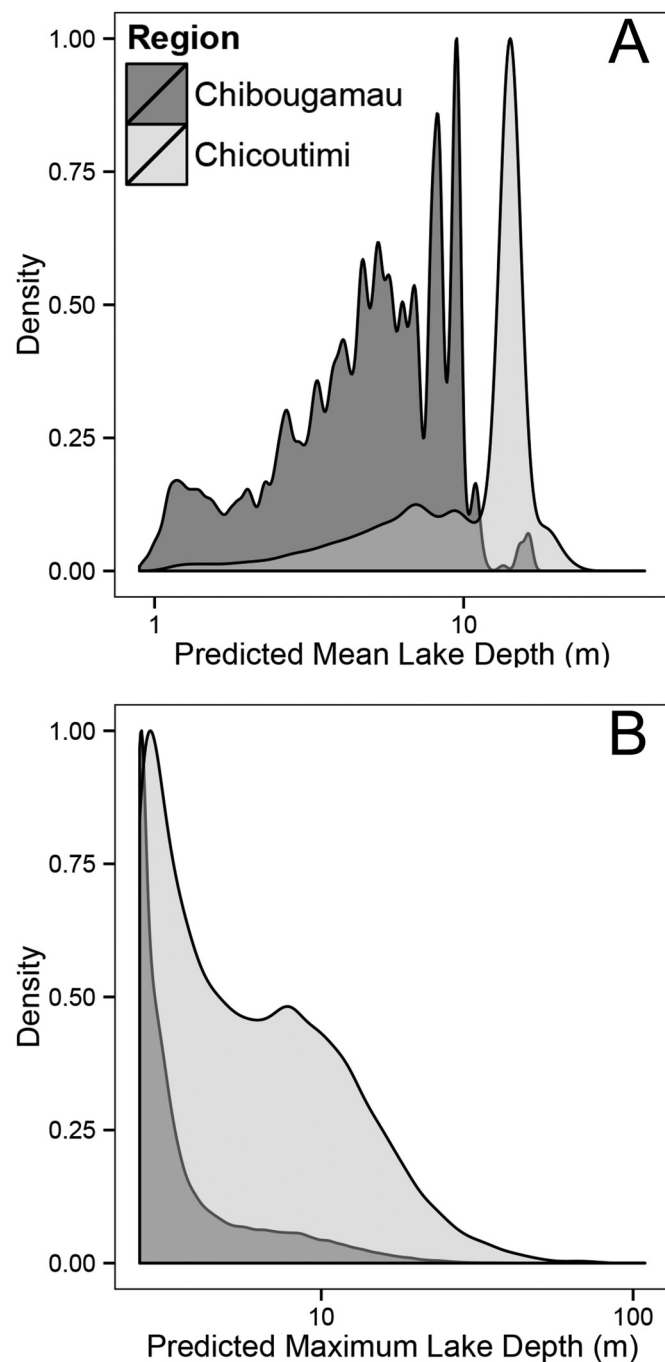


incorporating their additional complexity is at odds with our objective of providing a simple and widely applicable model for upscaling lake bathymetry.

#### Model limitations and application to other regions

Though our model was effective in the estimation of lake bathymetric parameters in Quebec, some caution must be used when applying it to other regions. At the base of our hypothesis is the assumption that similar processes formed the lakes and their landscapes. This assumption is valid in most of the Quebec boreal forest, where the terrain and lakes were shaped during the last glaciation, but may not hold for areas that have not been recently glaciated or are more characteristic of other forms of lake ontogeny (e.g., thermokarst and graben lakes). In any case, globally, glacial lakes are the most abundant in terms of number and areal

**Fig. 6.** Density plot for predicted distribution of (A) mean depth (weighted by lake area) and (B) maximum depth for the regions of Chicoutimi and Chibougamau. Maximum density was scaled to one for each region.



coverage (74 and 50%, respectively; Meybeck 1995), and so these relationships should hold for most systems.

Our model performed poorly when predicting lake volume and depth for a very large lake (surface area  $>1000$  km<sup>2</sup>), such as Lac Mistassini in our example (surface water area = 1900 km<sup>2</sup>). This is likely due to a combination of under-representation in the calibration dataset (extrapolating outside the model space) and unique ontogenetic events that led to the formation of these extremely large bodies of water. In comparison with smaller lakes, which number in the hundreds of millions (Downing et al. 2006; Verpoorter et al. 2014), there are relatively few lakes larger than

1000 km<sup>2</sup> (global estimates range from 121 to 225 lakes; Downing et al. 2006; Verpoorter et al. 2014). Whereas it may be possible to adapt our model to predict lake volume for these largest lakes, the bathymetry of these systems is often already well-described because of their regional importance to ecological and industrial interests (Herdendorf 1982). For this reason we have chosen to focus on most precisely predicting bathymetry for lakes with a surface area less than 1000 km<sup>2</sup>, which represent the major gap in knowledge in terms of worldwide depth and volumetric distributions.

It should be noted that even within Quebec, our study found some inter-regional variation for both predictive models (Fig. 3). Eastmain, however, was the only region where this variation led to a nonsignificant relationship. This region had the shallowest lakes and the smallest elevation change (Table 1) and fits well within the overall Quebec model; however, it also has the smallest range in both the explanatory (topography, lake area) and response (lake volume, maximum depth) variables. Given the fit of the 11 Eastmain lakes to the overall Quebec model, this small range is the likely explanation for the lack of a significant regression; however, this could indicate difficulties in developing models in similarly homogenous glacial landscapes such as the northern prairies of the USA and Canada.

We have provided equations for our model based on lakes from Quebec used in this study, but we do not propose that these same regression coefficients must apply in all regions. Instead, we have identified key topographic parameters that correlate to lake shape (e.g., elevation change<sub>25</sub>) and recommend this approach as a means of developing similar regression models for regions where some lake bathymetry data are available or alternatively as a more general rule when trying to upscale to similar (in topography and landscape formation) regions where no bathymetric data are available.

#### Model application

Our worked example demonstrates the practicality of using our model to estimate important bathymetric features at very large scales. These data could be used to better inform regional and continental predictions of ecosystem processes. For example, Bastviken et al. (2004) demonstrated that methane ebullition was only significant in depths of less than 4 m, and Ferland et al. (2012) showed that the lake dynamic ratio ( $\sqrt{\text{lake area} / \bar{Z}}$ ) was one of the best predictors of lake organic carbon storage. Further, lake mean depth has long been considered one of best predictors of nutrient loading potential (Vollenweider 1975) as well as trophic state (Canfield and Bachmann 1981). We demonstrated how these models could be used to estimate difference in mean areal water coverage (as cubic metres of water per square metre total land surface) between regions as estimates of the density of water and therefore the connectivity between the terrestrial and lentic aquatic ecosystems. This variable is likely correlated to water residence time and could function as a useful proxy to develop estimates of regional water retention time when coupled with other modeled and measured parameters such as mean surface water runoff. The above examples demonstrate how lake depth and volume are critical to the foundation of some of the most basic limnological parameters of interest and that developing models to reliably predict bathymetry is crucial to improving our understanding of the importance of inland waters at regional and global scales.

#### Acknowledgements

We thank Adrienne Huchet and Terhi Rasilo for assistance in the field and with GIS analyses, Marie-Eve Ferland and Dominic Vachon for contributing bathymetric data, and the Société des établissements de plein air du Québec (SEPAQ) for providing access to lakes within protected reserves. This project was part of the large-scale research program of the Industrial Research Chair

in Carbon Biogeochemistry in Boreal Aquatic Systems (CarBBAS), co-funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) and Hydro-Québec to PdG, and also supported by NSERC funding to YTP.

## References

- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **19**: 716–723. doi:10.1109/TAC.1974.1100705.
- Almendinger, J.E. 1990. Groundwater control of closed-basin lake levels under steady-state conditions. *J. Hydrol.* **112**: 293–318. doi:10.1016/0022-1694(90)90020-X.
- Bastviken, D., Cole, J., Pace, M., and Tranvik, L. 2004. Methane emissions from lakes: Dependence of lake characteristics, two regional assessments, and a global estimate. *Global Biogeochem. Cycles*, **18**: GB4009.
- Bird, D.F., and Prairie, Y.T. 1985. Practical guidelines for the use of zooplankton length-weight regression equations. *J. Plankton Res.* **7**: 955–960. doi:10.1093/plankt/7.6.955.
- Burnham, K.P., and Anderson, D.R. 2004. Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociol. Methods Res.* **33**: 261–304. doi:10.1177/0049124104268644.
- Canfield, D.E., Jr., and Bachmann, R.W. 1981. Prediction of total phosphorus concentrations, chlorophyll *a*, and Secchi depths in natural and artificial lakes. *Can. J. Fish. Aquat. Sci.* **38**(4): 414–423. doi:10.1139/f81-058.
- Cole, J.J., Prairie, Y.T., Caraco, N.F., McDowell, W.H., Tranvik, L.J., Striegl, R.G., Duarte, C.M., Kortelainen, P., Downing, J.A., Middelburg, J.J., and Melack, J. 2007. Plumbing the global carbon cycle: integrating inland waters into the terrestrial carbon budget. *Ecosystems*, **10**: 172–185. doi:10.1007/s10021-006-9013-8.
- Downing, J.A. 2010. Emerging global role of small lakes and ponds: little things mean a lot. *Limnetica*, **29**: 9–24.
- Downing, J.A., and Duarte, C.M. 2009. Abundance and size distribution of lakes, ponds and impoundments. *In* Lake ecosystem ecology: a global perspective. Edited by G.E. Likens. Academic Press, San Diego, Calif. pp. 220–229.
- Downing, J.A., Prairie, Y.T., Cole, J.J., Duarte, C.M., Tranvik, L.J., Striegl, R.G., McDowell, W.H., Kortelainen, P., Caraco, N.F., Melack, M.J., and Middelburg, J.J. 2006. The global abundance and size distribution of lakes, ponds, and impoundments. *Limnol. Oceanogr.* **51**: 2388–2397. doi:10.4319/lo.2006.51.5.2388.
- Emmerton, C.A., Lesack, L.F.W., and Marsh, P. 2007. Lake abundance, potential water storage, and habitat distribution in the Mackenzie River Delta, western Canadian Arctic. *Water Resour. Res.* **43**: W05419. doi:10.1029/2006WR005139.
- Fekete, B.M., Voeroversmarty, J., and Grabs, W. 1999. Global composite runoff fields based on observed river discharge and simulated water balances [online]. Complex Systems Research Center, University of New Hampshire. Available from [www.grdc.sr.unh.edu](http://www.grdc.sr.unh.edu).
- Ferguson, R.I. 1986. River loads underestimated by rating curves. *Water Resour. Res.* **22**: 74–76. doi:10.1029/WR022i001p00074.
- Ferland, M.-E., del Giorgio, P.A., Teodoru, C.R., and Prairie, Y.T. 2012. Long-term C accumulation and total C stocks in boreal lakes in northern Québec. *Global Biogeochem. Cycles*, **26**: GB0E04. doi:10.1029/2011GB004241.
- Ferland, M.-E., Prairie, Y.T., Teodoru, C., and del Giorgio, P.A. 2014. Linking organic carbon sedimentation, burial efficiency, and long-term accumulation in boreal lakes. *J. Geophys. Res. Biogeosci.* **119**: 836–847. doi:10.1002/2013JG002345.
- Gorham, E., and Boyce, F.M. 1989. Influence of lake surface area and depth upon thermal stratification and the depth of the summer thermocline. *J. Gt. Lakes Res.* **15**: 233–245. doi:10.1016/S0380-1330(89)71479-9.
- Gudasz, C., Bastviken, D., Steger, K., Premke, K., Sobek, S., and Tranvik, L.J. 2010. Temperature-controlled organic carbon mineralization in lake sediments. *Nature*, **466**: 478–481. doi:10.1038/nature09186. PMID:20651689.
- Håkanson, L., and Karlsson, B. 1984. On the relationship between regional geomorphology and lake morphometry. A Swedish example. *Geogr. Ann. Ser. A Phys. Geogr.* **66**: 103–119. doi:10.2307/520942.
- Håkanson, L., and Peters, R.H. 1995. Predictive limnology: methods for predictive modelling. SPB Academic Publishing, Amsterdam.
- Herdendorf, C.E. 1982. Large lakes of the world. *J. Gt. Lakes Res.* **8**: 379–412. doi:10.1016/S0380-1330(82)71982-3.
- Hollister, J.W., Milstead, W.B., and Urrutia, M.A. 2011. Predicting maximum lake depth from surrounding topography. *PLoS ONE*, **6**: e25764. doi:10.1371/journal.pone.0025764. PMID:21984945.
- Hutchinson, G.E. 1957. A treatise on limnology. Vol. 1. Geography, physics and chemistry. Wiley, New York.
- Kim, S. 2012. ppcor: Partial and semi-partial (part) correlation, version 1.0 [online]. Available from [cran.r-project.org/package=ppcor](http://cran.r-project.org/package=ppcor).
- Lehner, B., and Döll, P. 2004. Development and validation of a global database of lakes, reservoirs and wetlands. *J. Hydrol.* **296**: 1–22. doi:10.1016/j.jhydrol.2004.03.028.
- McDonald, C.P., Rover, J.A., Stets, E.G., and Striegl, R.G. 2012. The regional abundance and size distribution of lakes and reservoirs in the United States and implications for estimates of global lake extent. *Limnol. Oceanogr.* **57**: 597–606. doi:10.4319/lo.2012.57.2.0597.
- Meybeck, M. 1995. Global distribution of lakes. *In* Physics and chemistry of lakes. Edited by D.M. Imboden and J.R. Gat. Springer-Verlag, Berlin. pp. 1–35.
- Nürnberg, G.K. 1995. Quantifying anoxia in lakes. *Limnol. Oceanogr.* **40**: 1100–1111. doi:10.4319/lo.1995.40.6.1100.
- Prairie, Y.T. 1996. Evaluating the predictive power of regression models. *Can. J. Fish. Aquat. Sci.* **53**(3): 490–492. doi:10.1139/f95-204.
- QGIS Development Team. 2014. QGIS Geographic Information System, Version 2.0.1 [online]. Open Source Geospatial Foundation Project. Available from [qgis.osgeo.org](http://qgis.osgeo.org).
- R Core Team. 2013. R: a language and environment for statistical computing [online]. R Foundation for Statistical Computing, Vienna. Available from [www.r-project.org](http://www.r-project.org).
- Raymond, P.A., Hartmann, J., Lauerwald, R., Sobek, S., McDonald, C., Hoover, M., Butman, D., Striegl, R., Mayorga, E., Humborg, C., Kortelainen, P., Dürr, H., Meybeck, M., Ciais, P., and Guth, P. 2013. Global carbon dioxide emissions from inland waters. *Nature*, **503**: 355–359. doi:10.1038/nature12760. PMID:24256802.
- Sobek, S., Nisell, J., and Fölster, J. 2011. Predicting the volume and depth of lakes from map-derived parameters. *Inland Waters*, **1**: 177–184. doi:10.5268/IW-1.3.426.
- Verpoorter, C., Kutser, T., Seekell, D.A., and Tranvik, L.J. 2014. A global inventory of lakes based on high-resolution satellite imagery. *Geophys. Res. Lett.* **41**: 6396–6402. doi:10.1002/2014GL060641.
- Vollenweider, R.A. 1975. Input-output models with special reference to the phosphorus loading concept in limnology. *Schweiz. Z. Hydrol.* **37**: 53–84. doi:10.1007/BF02505178.