UNIVERSITE DU QUÉBEC À MONTRÉAL

MORPHOMETRIC, NETWORK AND LANDSCAPE PREDICTORS FOR CARBON SPECIES IN BOREAL LAKES

MASTERS DEGREE THESIS

IN

BIOLOGICAL SCIENCES

BY

JULIA JAKOBSSON

AUGUST 2018

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

PRÉDICTION DES FORMES DE CARBONE DANS LES LACS BORÉAUX, À PARTIR DE LA MORPHOMÉTRIE, LE RÉSEAU HYDROGRAPHIQUE ET LA STRUCTURE DU PAYSAGE

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAITRISE EN BIOLOGIE

PAR

JULIA JAKOBSSON

AOÛT 2018

ACKNOWLEDGEMENTS

I am very grateful for the support from NSERC's CREATE training program in lake and fluvial ecology (ÉcoLac), the Group for Interuniversity Research in Limnology and aquatic environment (GRIL), the Industrial Research Chair in Carbon Biogeochemistry in Boreal Aquatic systems (CarBBAS), my supervisors Paul del Giorgio and Yves T. Prairie, and everyone who has been involved in the sampling campaigns the last decade for making these two years of research possible.

Thanks to Richard Vogt, Tonya DelSontro and Joan Pere Casas for encouragement and improving my writing, and to Roy Nahas for bouncing GIS ideas with me. Thanks to all colleagues in the research group: Ryan Hutchins, Marie Gerardin, Karelle Desrosiers, Cynthia Soued, Martin Demers, Shoji Tottahill, Patricia Pernic, Felipe Rust, Paula Reis, Ji-Hyeon Kim, Tristy Vick-Majors, Sophie Crevceur, Sara Mercier-Blais, Alex Ducharme, Alice Parkes and Annick St-Pierre; without your invaluable feedback, laughs, chats and shared knowledge, Montreal would have been a much colder place. Also thanks to Sandro, Ariel, Lulu and Amir.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABREVATIONS

C – *Carbon*

CDEM – *Canadian Digital Elevation Model*

CDOM – *Colored Dissolved Organic Matter*

CH_4 – *Methane*

pCH_4 – *Methane partial pressure*

CO_2 – *Carbon Dioxide*

pCO_2 – *Carbon Dioxide partial pressure*

DEM – *Digital Elevation Model*

DIC – *Dissolved Inorganic Carbon*

DOC – *Dissolved Organic Carbon*

GHG – *Greenhouse gas*

GIS – *Geographical Information Systems*

GLM – *General Linear Model*

ISO – *Iterative Self-Organizing*

LC - *Landcover*

ML – *Maximum Likelihood*

NPP – *Net primary production*

PCA – *Principal Component Analysis*

Z mean – *Mean Depth*

Z max – *Maximum Depth*

%Littoral – *Percent Littoral*

RÉSUMÉ SCIENTIFIQUE

Les eaux continentales jouent un rôle important dans le cycle du carbone et l'émission des gaz à effet de serre et la majorité des lacs dans le monde se trouve dans la région boréale. Alors que la technologie et les connaissances pour mesurer les propriétés de l'eau sont établies, les techniques pour les estimer par télédétection sont toujours à développer. Puisque le biome boréal est vaste et souvent inaccessible, il est nécessaire de développer des outils pour estimer les propriétés des lacs sans les échantillonner de façon exhaustive. Dans cette mémoire, nous avons réalisé une carte de la morphologie des lacs et nous avons exploré le rôle de la morphologie dans la variabilité de quatre formes de carbone dans le paysage boréal du Québec, Canada.

Le chapitre 1 vise à définir les régions morphologiques des lacs du Québec en utilisant un ensemble de paramètres morphologiques qui peuvent être quantifiés par télédétection. Ces paramètres incluent : l'aire, le périmètre, la profondeur moyenne/maximale, le volume, la complexité de la rive, la forme du lac, le pourcentage de littoral et le ratio dynamique. Ces paramètres bathymétriques ont été extraits de la forme du bassin, que nous avons estimé à partir de la topographie autour du lac. Nous avons ensuite exploré les relations allométriques entre les paramètres et nous avons identifié trois variables indépendantes : la taille, la forme bathymétrique et la complexité. Nous avons utilisé ces trois variables pour séparer les 1,27 million de lacs du Québec ayant une superficie supérieure à 0.005km^2 entre 10 catégories. Étant donné que le paysage du Québec a subi l'action de la glaciation, certaines régions ayant le même matériel géologique de base ont une structure de surface similaire et ainsi les aspects morphologiques des lacs sont souvent redondants dans ces régions. Les régions morphologiques des lacs s'alignent donc souvent avec la roche mère et les sédiments sous-jacents.

Dans le deuxième chapitre, nous examinons la relation entre neuf paramètres morphologiques, sept caractéristiques du réseau hydrologique et quatre formes de carbone : le méthane (CH_4), le dioxyde de carbone (CO_2), le carbone organique dissous (COD) et le carbone inorganique dissous (CID). Nous avons utilisé une base de données de 300 lacs, échantillonnés au travers Québec, pour examiner l'importance de la relation entre la quantité de carbone et les aspects morphologiques. La meilleure corrélation a été entre la morphologie et la concentration de CH_4, expliquant 60% de la variabilité des concentrations. Quant au CO_2, c'était les variables qui décrivaient l'interface aquatique-terrestre des lacs qui ont expliqué la plupart de la variabilité (30%). Par contre le CID a été presque indépendant de la morphologie des lacs, mis à part d'une faible influence de volume et de la position dans le réseau hydrographique (14%). Les différentes relations entre la morphologie et les formes de carbone sont largement dues aux différentes sources de carbone et leur parcours jusqu'aux lacs.

Les conclusions de ce mémoire mènent à plusieurs perspectives futures. Les sujets abordés dans les deux chapitres élargissent nos connaissances sur la morphologie des lacs boréaux, la distribution de leurs paramètres et leurs impacts sur le cycle de carbone. D'autres études sur le sujet ou sur d'autres processus physiques, chimiques ou écologiques peuvent également utiliser des régions morphologiques des lacs afin d'améliorer le design d'échantillonnage ou la modélisation à grande échelle.

SCIENTIFIC SUMMARY

Inland waters play a significant role in carbon processing and emitting greenhouse gases to the atmosphere. Where the largest abundance of lakes globally is contained in the boreal. While the knowledge and technology in how to physically measure several water chemistry parameters is well developed, the techniques for estimation are still quite rough. Because the boreal biome is vast and largely inaccessible, it is necessary to develop tools to approximate lake properties without the need for extensive sampling. In this thesis we have developed a tool in the form of a lake morphological map and explored what role lake morphology plays in the variation of four carbon species across the boreal landscape in Québec, Canada.

Chapter 1 focuses on establishing the lake morphological regions of Québec through an ensemble of morphometric parameters that can be estimated remotely. These metrics include: area, perimeter, mean/maximum depth, volume, edge complexity, lake shape, percent littoral area and dynamic ratio. The bathymetric parameters of that set are derived from projecting an estimated basin shape from the topography surrounding the lake. We explored the allometric relationships between the lake morphometric parameters where three components were identified to vary relatively independently of each other: size, bathymetric shape and complexity. We found that the 1.27 million lakes of Québec larger than 0.005 km^2 divide into 10 distinct categories based on these three aspects. Because the Québec landscape was formed from glacial processes, many regions with the same geologic base material have similar surface structure and so the morphological features of the lakes are often reoccurring within these areas. The lake morphological regions that thus appear often align with the local bedrock and sediments.

In the second chapter we examine the relation of the nine morphometric parameters and seven additional metrics describing the upstream lake and river network

configuration to four carbon species: methane (CH_4), carbon dioxide (CO_2), dissolved organic carbon (DOC), and dissolved inorganic carbon (DIC). From 300 lakes sampled across the Québec landscape, we were able to see the degrees to which the carbon species concentrations covaried with morphological features. CH_4, was the most influenced by morphological and network configuration metrics where 60% of the variation in concentrations could be predicted. In the case of CO_2, it was the lake metrics related to terrestrial contact area that explained the most (30%), for DOC it was the dynamic ratio and metrics related to network configuration (36%). DIC on the other hand was largely unaffected by lake morphology except for a small influence of volume and network position (14%). These differences in relation to the metrics for each of the carbon species can largely be explained by their sources and methods of delivery to the lakes.

There is much prospect for further studies based on the conclusions from this thesis. The subjects covered by these two chapters expand our knowledge regarding the morphology of boreal lakes, the distribution of their features and the impact it has on carbon cycling. Further studies on this subject and other physical, chemical or ecological processes can use lake morphological regions for more informed sampling design or upscaling exercises.

Keywords: lake morphology, carbon biogeochemistry, boreal, limnology, physical geography, limnogeography

INTRODUCTION

Out of all the biomes on earth, the boreal by far contains the highest abundance of inland waterbodies (Messager et al. 2016; Feng et al. 2016; Verpoorter et al. 2014) making the aquatic networks in boreal regions a substantial part of the landscape as a whole. Inland waters (i.e., lakes, streams, rivers and reservoirs) transport, store and process considerable amounts of carbon. They also emit carbon in various forms at levels capable of offsetting the terrestrial carbon sink (Cole et al. 2007). The boreal zone not only contains a majority of the non-glaciated fresh water on Earth, it also contains a significant amount of carbon in peatlands, soils and vegetation (367.3-1715.8 Pg.) (Bradshaw et al. 2015). It is therefore important to incorporate and quantify the carbon emissions from these dynamic systems into global and regional budgets to get an accurate view of its role in the global carbon cycle, also with respect to climate change (Buffam et al. 2011; Raymond et al. 2013; IPCC 2013; Seekell et al. 2014). The higher northern latitudes are expected to be disproportionately affected by climate change (IPCC, 2007). In fact, northern lakes have already experienced a more rapid rate of surface water warming than lakes in other regions (O'Reilly et al. 2015). Because our understanding of how different landscape features functions (terrestrial and aquatic) is fragmented, it therefore becomes critical to improve on this knowledge, in order to predict potential impacts of our changing environment. However, the features used for quantitively upscaling all kinds of physical, chemical and ecological processes are still quite rough. This hampers our ability to understand, upscale and predict the impacts due to our changing climate on aquatic carbon cycling.

The sheer number of lakes and the remoteness of the boreal landscape in general, represents a major challenge in terms of extrapolation of processes to larger scales. Methods that can to a greater extent utilize data that are obtained remotely are therefore important to develop in order to understand the northern landscapes. As multispectral satellite imagery improves, we are able to see that the abundances of lakes in the boreal are skewed towards smaller sized waterbodies (Cael & Seekell, 2016; Downing et al.

2006). We are also able to utilize more detailed features because of continuously improving spatial and radiometric resolutions. The body of literature showing which processes scale with size is growing (see table 1.1.2) (Post et al. 2000; Bastviken et al. 2004; Kortelainen et al. 2006; Staehr et al. 2012; Kankaala et al. 2013; Schilder et al. 2013; Hayden et al. 2014; Seekell et al. 2014 & 2018b; Rasilo et al. 2015; Hall et al. 2016; Holgerson & Raymond. 2016). However, there are likely other morphological features that can give us insight about the biogeochemical profiles of inland waters.

Metrics such as lake area, perimeter and network configuration are now easily obtained through remote sensing (Quinlan et al. 2003; Verpoorter et al. 2014; Downing 2010; Winslow et al. 2013). Algorithms to estimate lake volume and depth based on the surrounding topography are also available (Heathcote et al. 2015; Sobek et al. 2011; Oliver et al 2016) and it has been shown that lake size is important to carbon processing (Rasilo et al. 2015). However, studying the influence of various facets of lake shape and what it may inform us about the functioning of lakes seems to be overlooked in the literature. Some studies show the effects of connectivity and architecture of the aquatic network on water chemistry, although these studies are usually performed with a single metric for several water chemistry parameters (Sadro et al. 2012; Kratz et al. 1997). In addition, the carbon species often become overlooked in favor of nutrients, such as phosphorous, which correlate well to metrics like lake order (Soranno 1999). Objectives of studies oriented around biogeochemical profiles of lakes can be divided into two categories: i) The need to predict, extrapolate and upscale and ii) explorations of the factors that influence observed concentrations and fluxes. This thesis is aimed to examine some of the morphological features of lakes that may contribute to their carbon dynamics.

The first chapter will focus solely on defining the geomorphological aspects of lakes, from which the results of this geographical classification can be applied broadly for multiple limnological purposes. The only features we are exploring are those that can be obtained or estimated remotely for it to be applicable over large regions. These include: area, perimeter, depth, volume, edge complexity, shape, percent littoral, and dynamic ratio.

The second chapter is of a biogeochemical nature, exploring which of the metrics developed in the first chapter is most relevant for each of the four carbon species in question: dissolved organic carbon (DOC), dissolved inorganic carbon (DIC), carbon dioxide (CO_2) and methane (CH_4). DOC and DIC are important carbon species in aquatic systems, either produced internally or imported from the surrounding landscape. The carbonic gas species, CO_2 and CH_4, are greenhouse gases and regularly produced, consumed and emitted from inland waters

The goal of the study is to focus on drivers or proxies that can be obtained remotely (i.e. derived from remote sensing or geo-databases) such that these approaches can be used as tools for upscaling and prediction of carbon species. While chemical factors, such as pH or phosphorous, may be good predictors of carbon, their solubility and impact on metabolism and other biological processes, limit their estimation to in situ sampling only. Thereby the main issue of quantities and inaccessibility of boreal lakes is not solved. Colored dissolved organic matter (CDOM) and chlorophyll are linked to the carbon cycles and can be derived through remote sensing to a certain degree. However, the accuracy in the derived values are highly inconsistent across regions, time and types of waterbodies (Brezonik et al. 2015). They may therefore not be suitable as predictive factors for the time being.

It is not a new approach to hypothesise that the morphology of lakes or features of the watershed could influence water chemistry (Rasmussen 1989; Noges 2009; Dodds et al. 2010). An increasing number of studies are looking at morphology, landscape factors and configurations of aquatic networks as potential drivers of biogeochemical properties of lakes and streams (Soranno et al. 1999 & 2009; Prepas et al. 2001; Sobek et al. 2003 & 2007; Quinlan et al. 2003; Martin et al. 2006; Schilder et al. 2013;

Winslow et al. 2013; Hall et al. 2015; Fergus et al. 2017; etc.). There have been studies showing that watershed characteristics, geology and climate influence aquatic carbon (Larsen et al. 2011), however, there have been few, if any, study explicitly looking at lake morphology on regional scales or how lake shape may influence lake carbon dynamics. Considering the effects of for example small to large surface areas or volumes on key physical factors in carbon processes such as temperature, water movement and contact areas (see table 1.1.2), it is not inconceivable that other morphological features would also influence the biogeochemical lake environment.

In order to investigate spatial limnological carbon dynamics, there needs to be a large number of sampled lakes covering a wide range of sizes, shapes, carbon concentrations, network configurations, watershed characteristics and regions. Québec is thus well suited as a boreal study region in the context of the overall goal of this work. The landscape was created by the Laurentian ice sheet during the last ice age, it generated a wide range of geomorphological features, including many differently shaped waterbodies (Fulton 1989; Håkansson 2005). All lake types encountered here are therefore likely to be reoccurring throughout the global regions shaped by glacial processes and therefore widely replicable and comparable.

CHAPTER 1

LAKE MORPHOLOGICAL REGIONS ACROSS TEMPERATE, BOREAL AND SUBARCTIC LANDSCAPES

1.1 INTRODUCTION

Lakes are dynamic components of the landscape with significant impact on the global environment (Cole et al. 2007). It is estimated that global inland waters cover about 3.7% of the worlds non-glaciated areas, and the arctic and boreal contain the highest concentration of waterbodies, where the lakes also have the largest areas and perimeters (Verpoorter et al. 2014; Messager et al. 2016). The morphology of lakes is a critical factor in their ecological, biogeochemical and physical functioning (Wetzel 2001; Håkansson 2005).

1.1.1 Lake origins

The occurrence of a lake is a consequence of the surrounding landscape geomorphology. In 1957, Hutchinson compiled a list of all the lake types that originated from different geomorphological processes. Glacial processes were one of 11 major types that was further broken up into more detailed subgroups. The glacial subgroups included: (A) lakes in direct contact with ice, (B) glacial rock basins, (C) morainic and outwash lakes and, (D) drift basins. This categorization has remained fairly intact with only small alterations over time, for example: group D is replaced by periglacial basins (thermokarst) by Cohen (2003). Lakes in the northern landscapes are almost exclusively formed in the Pleistocene glacial period (Hutchinson 1957; Fulton 1989; Wetzel 2001; Håkansson 2012). In Québec in particular the A-types are unlikely to appear since there are no active glaciers. The majority are shield lakes, which represent inundated crevasses in areas where the glacial concourse milled the landscape down to bare bedrock with little to no sediment present. Shield lakes are included under the subgroup *glacial rock basins* along with cirque, valley rock basin, glint and fjord lakes. The other large subgroup of glacial lakes are *Glacial deposit dammed basins* which includes: moraine dammed lakes, kettle lakes and dammed subglacial ice tunnels (Cohen 2003). The resulting landscape from glacial processes alternate somewhat depending on the local geology and the ice movements, and in doing so, create distinct regions characterized by relatively homogenous geomorphological features.

1.1.2 Lake feature effects

While some lake features have been more studied than others, such as surface area and depth, all morphological features contribute in some way to establishing local conditions that influence different ecological and biogeochemical processes, as well as organisms and their adaptations (Hershey et al. 2006; Håkansson 2005, Staehr et al. 2012; Kraemer et al. 2015). Table 1.1.2, summarizes some of the most studied features and their effects in three categories of lake functioning: physics, chemistry and biology. Out of all features, depth has been the most studied. While surface area, depth, volume, shape and land/water interface are the main features, the metrics of them are sub-categorized. For example; depth may include both direct measurements, ratios of mean/max, and surface area/depth. The depth related ratios could be classified as a basin shape measurement, however in this case we would like to highlight shape as referring to the littoral slope or overall hypsographic profile in that category. The cascading effects from each process have not been included, the table is thus showing only what is directly influenced by the feature.

There are therefore multiple ecological, biogeochemical and physical reasons for further understanding the interaction between lake size and shape. The effect of lake size has been widely studied, and we know that the size distribution of lakes varies regionally, and this has strong implications on a number of lake features (Seekell et al. 2013; Downing et al. 2006). More recently, it has been shown that mean and maximum depth is also regionally structured, and that as a result, lake volume per unit surface, and related factors, such as water residence time per unit area, also vary systematically between regions (Heathcote et al. 2015; Cael et al. 2017). Other studies, which are also supporting the notion of the need for additional ways of describing the hydrological setting, have been focusing on aspects of connectivity (Fergus et al. 2017) or geomorphological surface textures (Wolock et al. 2004). Here we are mainly providing tools for in-lake processes.

Table 1.1.2 Lake feature effects in the categories of physics, chemistry and biology. Only the effects studied as being directly influenced by the feature is included. Each of the effects listed have further cascading effects in all categories.

Feature	Factor	Physics	Chemistry	Biology
Surface area	•Area •Shoreline distances	•Wave growth •Heat transfer •Horizontal diffusion •Evaporation	•Gas flux •Photolysis	•Total primary productivity •Planktonic distributions
Depth	•Mean •Max •Ratios	•Light penetration •Stratification •Mixing depth •Sedimentation •Resuspension •Temperature •Ebullition	•Oxidation •Hypoxia •Gas supersaturation •Isotopic profiles	•Number of trophic levels •Primary production •Fish communities •Macrophytes/vegetation •Vertical migration
Volume	•Total volume	•Residence time	•Dilution •Concentration	•Productivity •Number of trophic levels
Shape	•Fetch •Basin shape/slope	•Outgassing •Waves •Mixing/Stratification •Internal seiches •Sedimentation •Sediment transport	•Littoral/Pelagic influence majority •Allochthonous/Autochthonous main carbon sources	•Planktonic distributions •Macrophytes/Vegetation
Land/Water Interface	•Contact area		•Gas supersaturation •pH •Nutrient input/exchange •Mineral availability •Redox potential	•Microbial respiration/activity •Bioturbation •Macrophytes/Vegetation •Community compositions
References	Bengtsson & Herschey 2012; Wetzel 2001; Meybeck 1995; Håkansson 2005, Staehr et al. 2012; Vachon & Prairie 2013; Hayden et al. 2014; Steele et al. 2014; Korteleinen et al. 2006; Kraemer et al. 2015; Holgersson & Raymond 2016; Bastviken et al. 2004; Schilder et al. 2013; Post et al. 2000.			

The patterns in lake shape, on the other hand, have been very sparsely explored. While it is often repeated that lake shape is of a fractal nature, we do not actually know how the roundness of lake shape, the complexity of their shoreline, or the shape of their basins vary as a function to their size. Neither do we know how these additional features relate to each other, and if they also have a spatial structuring that determines a regionality in these properties.

Because lake size and horizontal and vertical shape may not be completely coupled, this will generate a wide range of morphometric configurations, which will in turn have multiple, and often diverging influences on ecological, biogeochemical and physical processes. For example: both a small lake and a large lake could have a deep bowl profile or a shallow plate profile and still have either a complex, convoluted shoreline or a smooth round shape. Where a shallow plate profile is less likely to stratify (Wetzel 2001) and is more inviting for macrophytes to establish in the littoral zone (Rooney & Kalff 2000), a deeper bowl profile is more likely to be stratified and have less relative macrophyte coverage. If one of these shallow plate lakes has a more complex shoreline than a round one of the same area and profile, it will have relatively more contact area with the surrounding landscape and a larger percentage of the total area could be overlying shallow sediments. It is then important to understand the relationships that exist between these dimensions of lake morphometry, and how these relationships vary along gradients of lake size, and across different landscapes.

More interesting, perhaps, is the type of distribution of lake features that emerges from superimposing these different dimensions, and how this may expose regions where lakes are characterized by a common set of morphometric features, which may be quite different from a regional distribution obtained on the basis of lake size or eco-regions alone, and which may generate large-scale, emergent patterns in lake function that are not obvious from considering any single dimension. In addition, determining links between the various aspects of lake size and shape and their distribution across the landscape may provide tools for more effectively upscaling a variety of lake processes at a regional level. To illustrate this latter point, consider a process that becomes enhanced in lakes with a certain combination of morphological features, if the number of those specific lakes out of the whole population can be estimated with a fair accuracy, the effects of that process can be more precisely quantified. If in addition to that it is known where lakes with these features are more commonly occurring, more details on the environment and landscape can be added to the calculations.

Here we explore the allometry of lake shape (i.e. coupled or uncoupled morphological traits), and the regional patterns in lake morphological structure across temperate, boreal and subarctic landscapes in Québec. The basic questions we are looking to answer are;

- Is there any allometry associated to lake morphology?
- How do different metrics of lake shape relate to each other?
- Are there geomorphological regions determined on distinct combinations of lake morphometric features?

We expect to see regional differences in various aspects of lake morphology due to the dynamic behaviour of the Laurentian ice sheet that shaped the Québec landscape. For example, whereas the lakes towards the St Lawrence lowlands on the Canadian shield (south – south-west) lie in deposits from the glacier retreat, the lakes in central Québec will mainly appear in glacial scars with shallow soil depths (Fulton, 1989). These different topographies are likely to result in systematically different patterns in lake morphology.

1.2 METHODS

1.2.1 Spatial Data

The basic spatial data of digital elevation models (DEMs) and lake shapes was obtained from the Canadian governmental open data source Geogratis, which is provided by the department of Natural Resources Canada; Earth Sciences Sector under the Open Government License - Canada. The software used to process the spatial data is ArcMap v10.2-10.4 from ESRI.

The DEMs are the third issue of the Canadian Digital Elevation Model (CDEM) which have a 30m resolution.

The lake shapes are part of the National Hydro Network (GeoBase - NHN) which have been derived from 1:50,000 scale maps or better where source image resolution is about 30m. The distinction of lake vs wetland has thus already been made at the stage of production (GeoBase 2010). A 0.005 km^2 minimum surface area cut-off was decided on based on the average resolution of the source data, so that elements of shape and shape complexity would be adequately represented throughout the whole region. This cut-off reduced the number of waterbodies in the dataset from $2.71*10^6$ to $1.27*10^6$. A total of 3,832 known reservoirs were also removed from the dataset.

1.2.2 Morphometry

In addition to the basic metrics of area and perimeter, derived from the lake shapes, seven other metrics were calculated: Maximum Depth *(Z_{max})*, Volume *(V), Mean Depth (Z_{mean}),* Edge Complexity, Lake Shape, Depth Profile *(Dynamic Ratio)* and Sediment Edge Contact *(Percent Littoral)*.

Maximum Depth and *Volume* was calculated using the method described in *Heathcote et al. 2015,* where the elevation change within a 25-meter buffer around the lake is the key factor:

$$Log10\ (Z_{max}) = 0.35 + Log10\ (Elevation\ change_{25}) * 0.79$$

$$Log10\ (V) = Log10\ (Lake\ Area) * 0.96 + Log10\ (Elevation\ change_{25}) * 0.77$$

Where Z_{max} is Maximum Depth, *V* is Volume and *Elevation change*$_{25}$ is the maximum topographic change within a 25-meter buffer around the lake. Similar methods have been tested using 50 and 100-meter buffers (Sobek et al. 2011; Oliver et al. 2016), however this is most likely an adaptation to the resolution of the available elevation data or for regional differences.

Mean Depth was calculated from the volume and lake surface area:

$$Z_{Mean} = \frac{V}{area_{lake}}$$

Edge Complexity is a perimeter to area ratio, where increasing values expresses an increasing complexity. The square root of area is used to exclude any size dependency since that aspect is already accounted for in the basic metrics.

$$Edge\ Complexity = \frac{perimeter_{lake}}{\sqrt{area_{lake}}}$$

Lake Shape uses a smallest circumscribing circle around the lake and compare the area of which to the actual lake surface area. In effect, it differentiates between oblong (shape approaches 1) and circular lakes (shape approaches 0). This metric is also size independent and unitless.

$$Lake\ shape = 1 - \frac{area_{lake}}{area_{circle}}$$

Depth Profile was accounted for using the Dynamic ratio (Håkansson 1982; Ferland et al. 2012) which is the square root of lake area divided by mean depth:

$$Dynamic\ Ratio = \frac{\sqrt{area_{lake}}}{Z_{mean}}$$

Low dynamic ratios indicate lakes that are bowl-shaped, whereas high values of the ratio denote lakes that are dish-like.

Sediment Edge Contact was included in the metrics to highlight the littoral zones of lakes, which are hotspots for biogeochemical activity. For this we calculated the *Percent Littoral* as the percentage of the lake area where the sediments are above 3 meters in depth as:

$$\%Littoral = 1 - \left(1 - \frac{3}{Z_{Max}}\right)^{q_{bathymetric\ shape}} * 100$$

Where $q_{bathymetric\ shape}$ is a scaling factor that represents an idealized lake bottom profile:

$$q_{bathymetric\ shape} = \frac{Z_{max}}{Z_{mean}} - 1$$

Flowcharts of the metric calculation process can be found in Appendix I.

1.2.3 Statistics

We carried out regression analysis to explore the presence of allometric relationships between lake size and the various metrics of lake morphology. The regressions are simple yet effective explorative tools which also give insight needed for the principal component analyses. All metrics were log10 transformed to maintain a normal distribution and all analysis was conducted with JMP Pro v13 statistical software from SAS Institute Inc.

1.2.3.1 Principal Component Analysis

We used principal component analysis (PCA) to visualize the distribution of lakes based on the ensemble of morphometric variables. The PCA scores on the first three axes for each individual lake were subsequently used as the basis for the spatial analysis as integrative morphometric vectors.

1.2.4 Spatial Analysis

The scores from the first 3 PCA axes for all the lakes were imported to the GIS software, and in order to carry out a surface interpolation of the data, the point-vector format (lake center point coordinates with PCA scores attributed) was transformed to Thiessen polygons and then converted to raster format with a 15m resolution (Figure 1.2.4). Thiessen polygons are created by equally dividing the space between all neighboring point features (Yamada 2016). This procedure results in an equal spatial expression, regardless of actual physical expansion of an individual lake but rather on the lake center point relation to its neighbors. This implies that a small lake and a large lake will be spatially more equally represented in the regional characteristics. This method of interpolation is performed without using algorithms that alter the input values and requires less processing power, as long as the input values are transformed from floating-point to integer data.

A flowchart of the spatial analysis process can be found in Appendix I.

1.2.4.1 Outlier Resolution

When the PCA scores were imported to the GIS software, ArcGIS desktop v10.x (ESRI, 2016), there was a high degree of so-called *salt-and-pepper noise*. This occurs when there is a scatter of extremely high and low values throughout the image. The general approach for removing salt-and-pepper noise is to employ a median filter. In our case the noise was not produced by any error but rather from true regional outliers. The data therefore still needed to be regionally homogenized. Thus, when resampling the initial raster resolution (15 m) to a coarser scale (1.5 km) the cell aggregation technique was driven by median selection. Prior to the resampling the raster values were rounded off through a reclassification using quantile breaks. The three PCA raster layers were then combined into a composite RGB image on which the classification was performed.

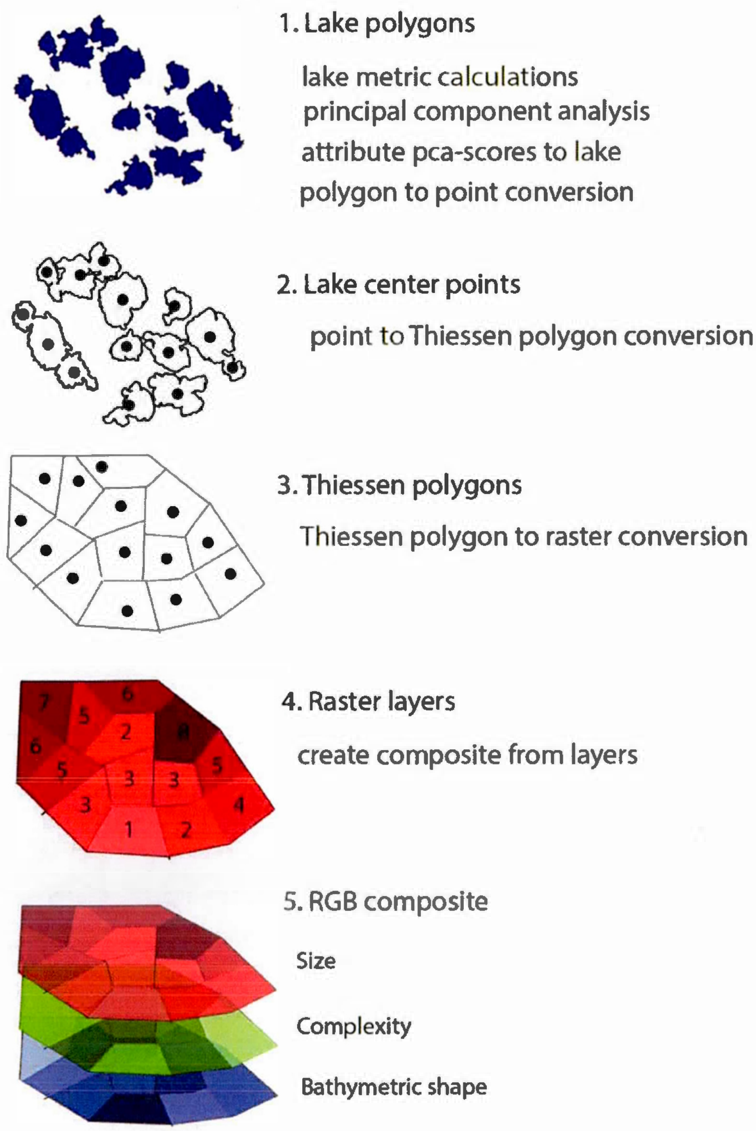


Figure 1.2.4. An illustrated example of the transition from lake polygons to composite image. The 1st step accounts for the calculations made from the lake polygons; the 2nd have had the lake shapes replaced with their center point; the 3rd have had Thiessen polygons outlined; the 4th is a representation of the conversion from polygon to raster format; the 5th is the composite image of the three raster layers which will be the input to the classification.

1.2.4.2 Iterative Self-Organizing (ISO) Cluster and Maximum Likelihood Classification

The ISO Cluster classification procedure is an unsupervised method in the ArcGIS Desktop software (ESRI, 2016), which functions much like k-means clustering. A k-means clustering algorithm works by dividing the data points into groups where the sum of squares within the group is optimized to a minimum (Hartigan et al.1979). Technically that means a number of clusters must first be specified for the algorithm to iteratively find the optimal mean value for each group given the number of specified clusters. In this case the optimal mean value is where the points assigned to the cluster are at the shortest distance compared to the mean of the other clusters. There is however no direct way to determine the optimal number of clusters, apart from specifying the minimal amount of data points to compose an individual cluster. Therefore, the suitable number of clusters that effectively describe a certain set of observations has to be estimated by iteration.

The ISO cluster method thus allows to derive an optimized number of classes from a set of observations and produce a signature file that can then be used as input to the maximum likelihood classification procedure (ESRI, 2016).

Maximum likelihood classification considers the variance and covariance of the input values that characterize an item and makes a Bayesian decision on which class the item likely belongs to. More specifically, Bayes theorem makes estimations of probability based on prior conditions.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Where: A and B are events, P(B) ≠ 0

The ISO Cluster signature file thus becomes a primer for the maximum likelihood that a certain data point belongs to a certain class. In the postprocessing of the ML-classification the raster is aggregated to a 5x5 km resolution to enhance the different regions.

A flowchart of the image analysis process can be found in Appendix I.

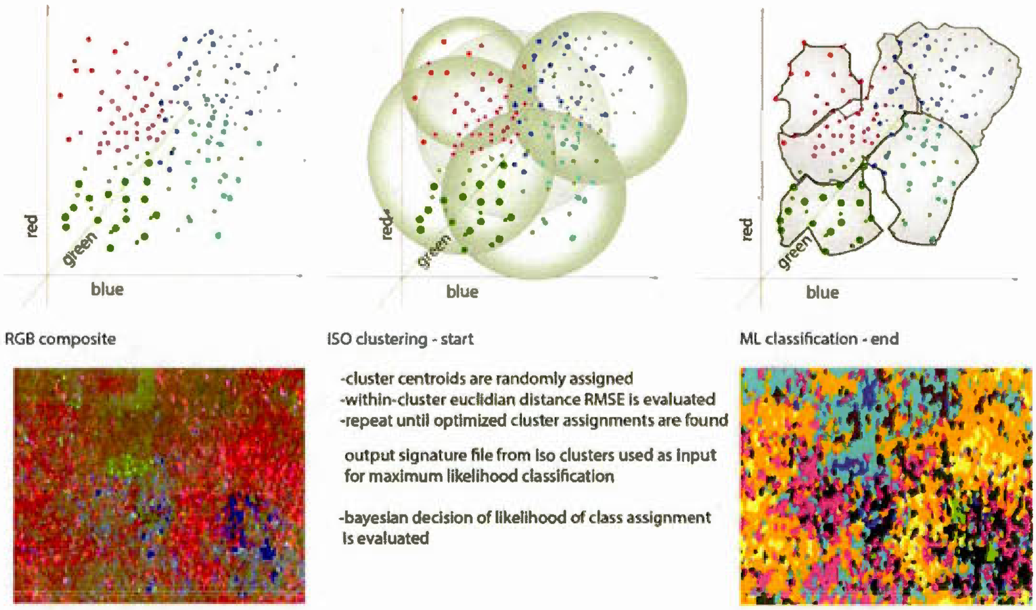


Figure 1.2.4.2 An illustrated example of the transition from an RGB-composite image of 3 PCA axes representing the red, blue and green, to an ISO cluster classification and finally a processed maximum likelihood classification.

1.3 RESULTS

1.3.1 Distribution of lake sizes and morphological features

The upper lake size in our dataset corresponds to Lac Mistassini which is 2,100 km^2, but the lower limit is somewhat arbitrary, since it depends on the cutoff that was chosen for the lake shapes. The lowest cutoff possible, based on the source data resolution would be 0.001 km^2, which would yield 2.71x10^6 lakes, where 48% is in the 0.001 to

0.005 km^2 size class. These ultra-small lakes are at the edge of effective resolution of the source image data. The largest resolution of the satellite imagery used to derive the lake shapes was 30 x 30 m (GeoBase 2010), so in order not bias any regions as having more smaller lakes because the resolution of satellite imagery was higher there, all regions must be held to the same cutoff. We therefore chose the more conservative cutoff of 0.005 km^2 so that lake shapes and elements of shape complexity could be adequately represented. After removing 3,832 catalogued reservoirs from the dataset, so as to not bias the morphometric parameters, the total number of lakes that were finally included in the analysis was 1,270,878. The average lake in this final dataset has a surface area of 0.15 km^2 (15 hectares). The cumulative lake surface area is ~187,000 km^2, which represents 12.3% of the total surface area of Québec covered in this study (1.52 x10^6 km^2).

The lakes extracted covered very large ranges in all of the morphometric parameters (Table 1.3.1 and Figure 1.3.1)

The shortest perimeter in the dataset is 0.25 km, whereas the longest perimeter is 5,711 km and interestingly does not belong to the largest lake in the dataset. The mean perimeter length is 1.73 km, whereas the cumulative lake edge is 2,147,751 km (Table 1.3.1).

The empirical model based on the mean elevation change around the lake yields an average maximum depth of 9.93 m, and the range of maximum lake depths of 0.8 – 144 m (Table 1.3.1). While this model may not adequately approximate depth for extreme situations such as for Lac Mistassini (max. depth 183 m) or for smaller lakes in highly variable topographies, or for kettle lakes, it was calibrated for lakes of glacier origin with surface areas between 0.005 to 1 030 km^2 and 1.5 to 125 m max depth (Heathcote et al. 2015), which represent the majority of lakes in this region and should therefore be generally suitable.

The calculated volumes using the above approach range several orders of magnitude, from $1.13 \cdot 10^{-6}$ to 169 km^3 (Lac Mistassini, actual: 150 km^3) where the mean volume is 2.6 x 10^{-3} km^3. The resulting average mean depth for the lakes in Québec is 5.34 m (Table 1.3.1). The upper range in the distribution of mean depth (up to 143 m) is unrealistic, however, and likely reflects biases in the algorithms used, but these extreme values hardly influence the overall regional lake mean depth.

The edge complexity of the lakes in Québec spans a large range, from 3.6 to 222 (Table 1.3.1), but with a distribution that is clearly skewed towards the less intricate (lower values) part of the spectrum (Figure 1.3.1.f).

The lake shape parameter approaches 0 for lakes that are perfectly round and 1 as lakes become more oblong and elongated. The mean value and overall distribution of the lake shape parameter, with a mean of 0.63 (Table 1.3.1 and Figure 1.3.1.g) thus indicates that lakes in Québec tend to have more oblong shapes.

Table 1.3.1 Summarizing statistics for Québec lakes surface areas, perimeters, depths, volumes, edge complexities, shapes, percent littoral areas and dynamic ratios.

Lake Feature	Min	Max	Mean	Std Dev	Std Err Mean	Sum
Surface Area (km^2)	0.005	1,369	0.15	3.14	0.003	185,762
Perimeter (km)	0.25	5,711	1.73	11.2	0.010	2,147,751
Maximum depth (m)	0.83	143.7	9 93	8.14	0.007	12,295,877
Mean depth (m)	0.21	143.4	5.34	5.72	0.005	
Volume (km^3)	$1.13e^{-6}$	70.1	0 0026	0.15	0.0001	3,206
Edge complexity	3.57	222.4	5.65	2.74	0.002	
Shape (circularity)	0.02	0.998	0.63	0.16	0.0001	
Percent littoral (%)	0.004	100	51.5	33.4	0.03	
Dynamic ratio	4.34	7,115	76.5	94.3	0.085	

That the average lake in Québec is relatively shallow is also reflected in the metric describing the percent littoral habitat of these lakes. The mean percentage of the surface area with depths that are shallower than 3 m was 51.5%, with most lakes ranging between 20-80% littoral (Table 1.3.1 and Figure 1.3.1.h). Likewise, the distribution of values of dynamic ratio, which increases as lakes go from bowl- to plate-shaped, suggest that most lakes in Québec tend to have a flat, plate-like profile, since the values were strongly skewed towards the right, with a mean value of 75.3 (Table 1.3.1 and Figure 1.3.1.i).

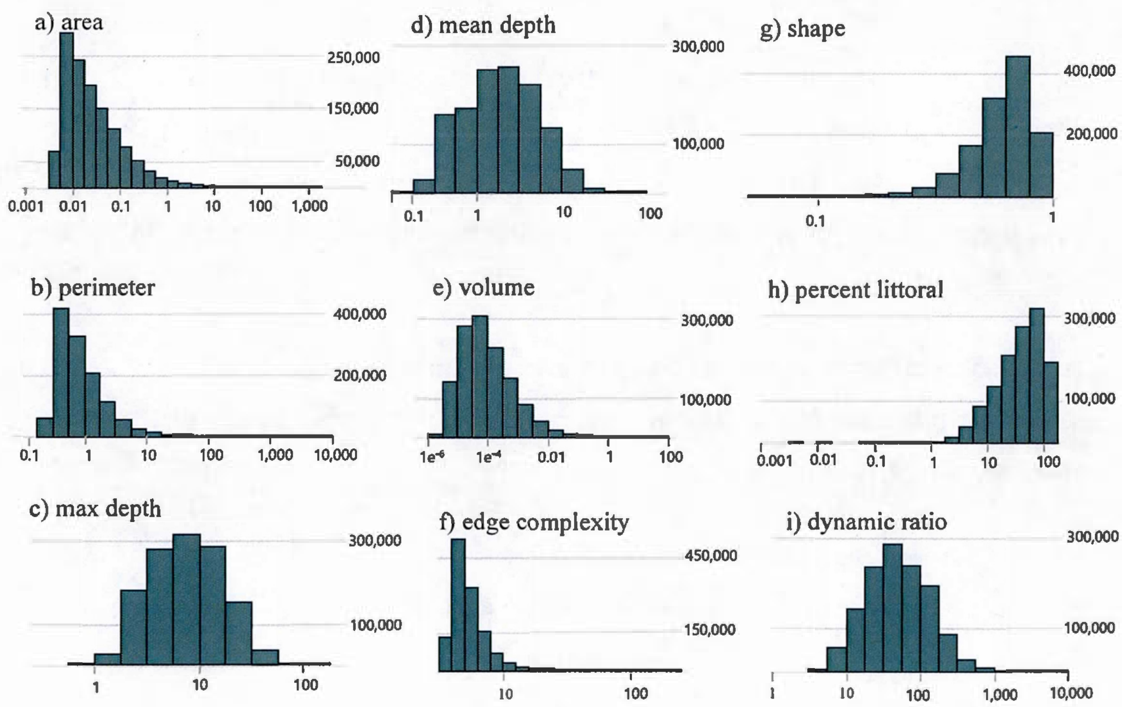


Figure 1.3.1.a-i Distribution of 9 morphometric parameters for lakes with between 0.005-1,370 km^2 surface area in Québec. All distributions are on a log10 scale.

1.3.2 Allometric relationships between lake metrics

We explored the relationships between the various lake metrics and lake area using regression analysis. All the variables were log-transformed to ensure normality, and we further binned the variables into logarithmic-scale categories, in order to expose the underlying patterns that are difficult to visualize due to the large density and scatter of points.

Not surprisingly, perimeter was strongly related to lake area (R^2=0.92) (Figure 1.3.2.1.a) but the log-log relationship was slightly non-linear and better captured by a first order polynomial model.

Likewise, as lake area increases, so does the mean and maximum lake depth (R^2=0.15) (Figure 1.3.2.1.b and c), but in both cases, there was a non-linear trend where depth tends to plateau with increasing lake size. While the range of our data is very large, these trends offer support to the model used to estimate depth, as it in itself does not incorporate lake size.

The mean depth tends to increase less with area than maximum depth and therefore the ratio of the two tends to shift as lakes become larger, with consequences on the shape of the depth profile of lakes, as described below.

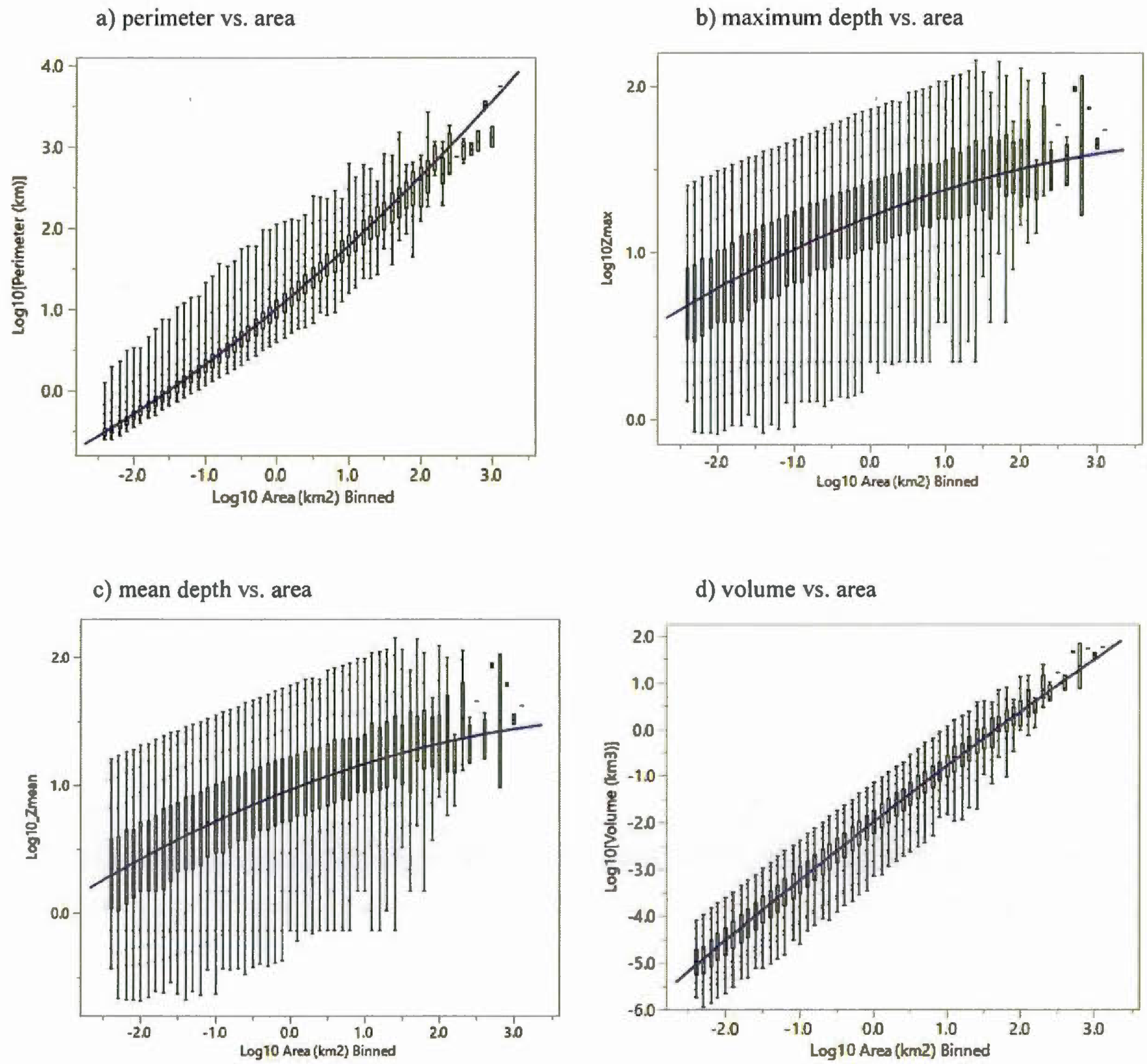
Volume has a strong linear log-log relationship to area (R^2=0.79) (Figure 1.3.2.1.d), which is partly attributable to surface area being a parameter in the volume algorithm. Because depth also increases with lake area, the log-slope of the relationship between volume and area is significantly greater than 1.

The edge complexity metric also showed a wide range for any given lake area, but on average complexity tends to increase with lake size (R^2=0.36) (Figure 1.3.2.1.e). This could partially be related to the fractal nature of shorelines and the spatial resolution of source data (Mandelbrot 1967). It is more likely, however, that it is a landscape scaling factor, such that a lake surface expanding over a large area in the boreal is more likely to follow a variety landscape contours than one with a smaller extent.

As mentioned before, a shape value of 1 is an almost complete diversion from a circle and there is a clear trend for larger lakes to converge to this value of 1 (R^2=0.09) (Figure 1.3.2.1.f), suggesting that in this landscape lakes tend to become proportionally narrower as they increase in size.

On a log-log scale, the percent of lake area overlaying sediments shallower than 3 m depth (% littoral) declined linearly with lake area (R^2=0.17) (Figure 1.3.2.1.g). However, this relationship was rather weak and the range in percent littoral is very large for lakes for any given size of lake (Figure 1.3.2.1.g). Likewise, the dynamic ratio increases with size (R^2=0.10) (Figure 1.3.2.1.h), suggesting that although larger lakes tend to be deeper, as described above, they also tend to become more plate-like as they become larger.

We further explored the relationships that exist between the different aspects of lake shape, and we found that there was structure in lake morphology beyond the above size scaling (Figure 1.3.2.2). For simplicity we divided the metrics into two groups: those that depend on the depth profiles (max depth, percent littoral and dynamic ratio) and those that are associated to shape (edge complexity and shape). The depth related metrics all strongly covaried, with R^2 ranging from -0.75 to -0.96 and 0.70, and the two shape metrics were also strongly coupled to each other (R^2 = 0.69, Figure 1.3.2.2). The relationships between the groups of variables were much weaker, with R^2 ranging from -0.35 to -0.05 (Figure 1.3.2.2), indicating that these different dimensions of lake morphology do not necessarily covary across the landscape, and that any given type of depth profile may be associated to widely varying lake shapes, and vice versa.



	y =	RMSE	R^2
a)	$1.01 + 0.73x + 0.04x^2$	0.31	0.923
b)	$1.22 + 0.18x - 0.02x^2$	0.39	0.154
c)	$0.96 + 0.23x - 0.03x^2$	0.39	0.154
d)	$-1.96 + 1.28x$	0.39	0.786
e)	$0.98 + 0.23x + 0.04x^2$	0.11	0.360
f)	$-0.12 + 0.04x - 0.01x^2$	0.12	0.089
g)	$1.13 - 0.27x$	0.36	0.171
h)	$2.06 + 0.27x + 0.02x^2$	0.39	0.095

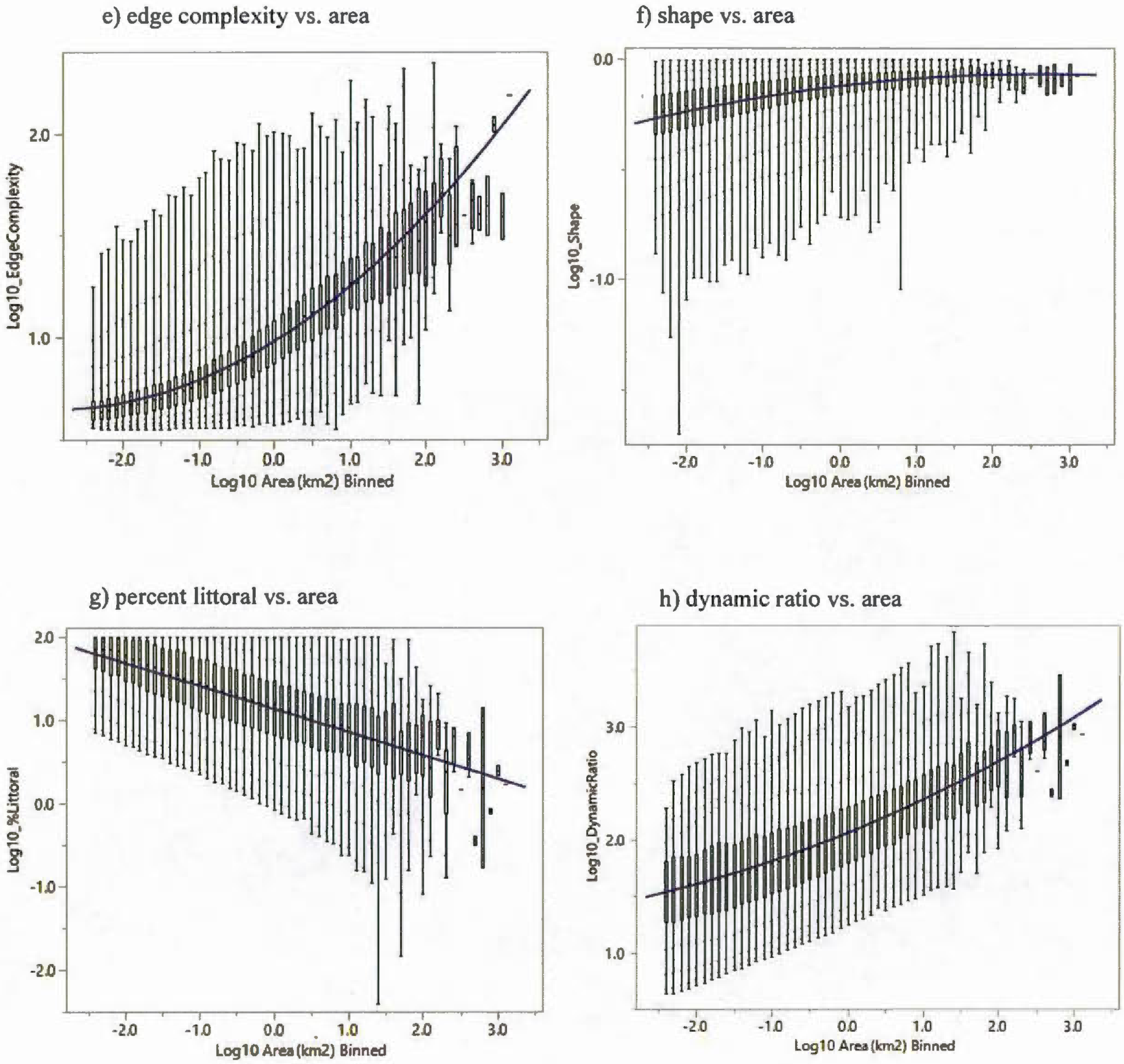


Figure 1.3.2.1 Allometry regressions between all metrics and Area. a) Perimeter (R^2=0.92), b) Maximum Depth (R^2=0.15), c) Mean Depth (R^2=0.15), d) Volume (R^2=0.79), e) Edge Complexity (R^2=0.36), f) Shape (R^2=0.09), g) Percent Littoral (R^2=0.17), h) Dynamic Ratio (R^2=0.10). The Area data have been binned, all parameters are log-transformed and the whiskers of the boxes represent the quantile ranges. The fitted line through the data is based on the mean values within each bin.

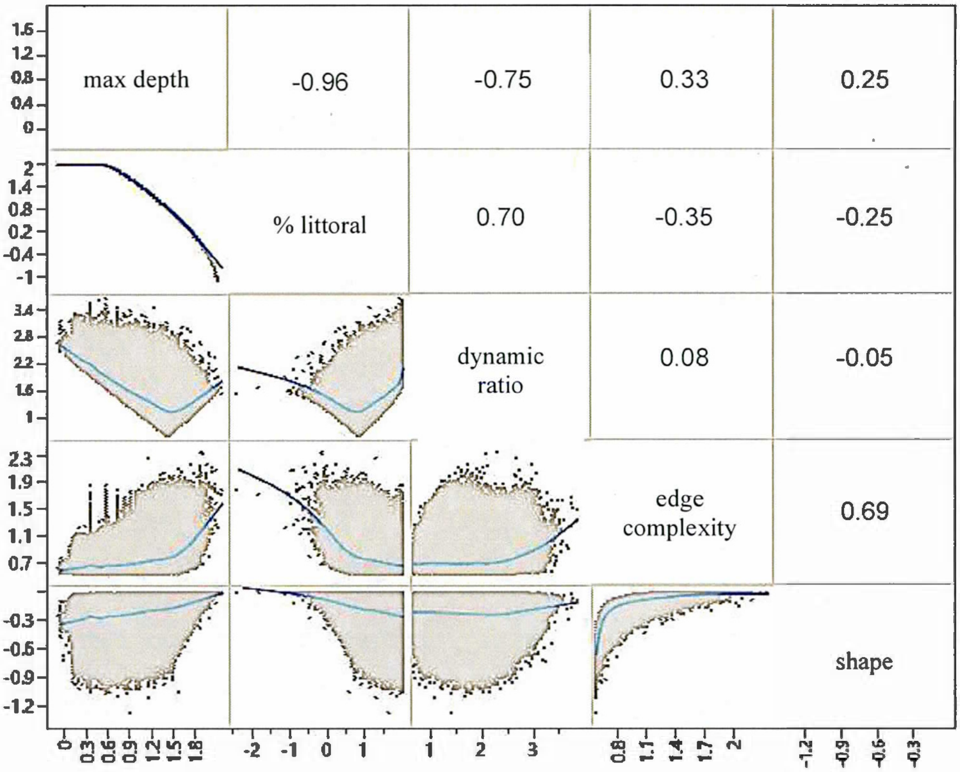


Figure 1.3.2.2 Multivariate correlation scatter-plots and correlations of the 5 metrics: Maximum Depth, Edge Complexity, Shape, Percent Littoral, and Dynamic Ratio.

1.3.3 Principal component analysis of lake metrics across the landscape

We explored the distribution of sites based on the ensemble of morphometric parameters using principal components analysis (PCA), with all input variables log-transformed to account for the large ranges and the often skewed distributions. The resulting eigenvalues of the components quickly declined from 5.2 to 2.5 and 1.0, thus only the first three components were retained as the most informative. These are the three dimensions of variation which is later used as the basis for the classification. The

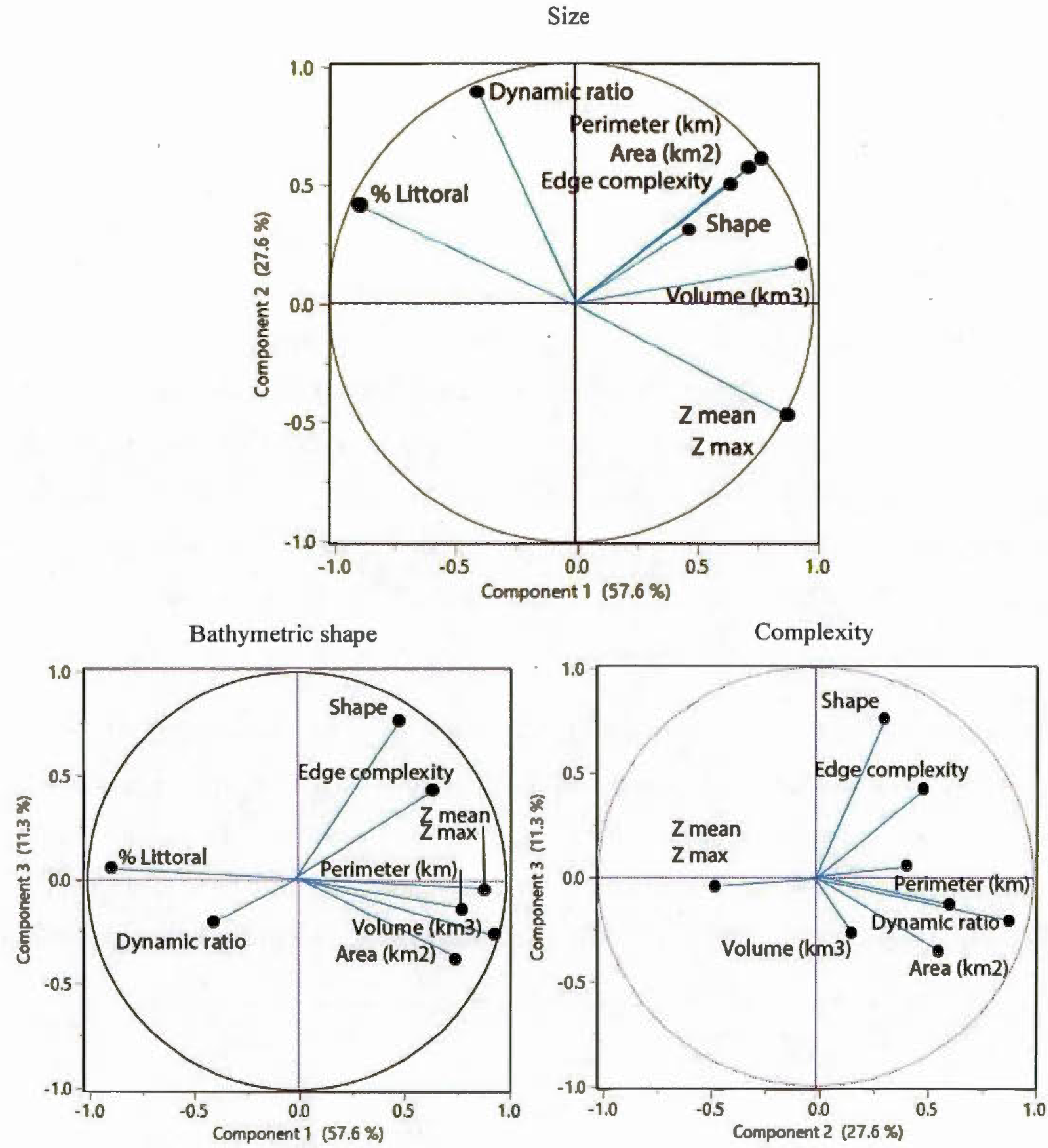


Figure 1.3.3.1 The Principal Component Analysis graphs showing the vectors of the 9 metrics across the 3 first components where the first component (57.6%) is related to size, the second one to bathymetric shape (27.6%) and the third to Complexity (11.3%). The top plot shows the relation between components 1 and 2, the lower plots show components 1-3 and 2-3.

PCA vectors clearly reflected both the importance of lake size, as well as the allometry of lake morphology and the links that exist between specific lake metrics.

The first axis explained up to 57.5% of the composition of the data (Figure 1.3.3.1), and integrated several lake metrics, which also weighed to a lesser extent on axis 2 and 3, which explained 27.8% and 11.3% of the total variance, respectively. Mean and maximum depth, and percent littoral weighed heavily on component 1, with opposing effects which reflects that deeper lakes tend to have proportionately less shallow area and tend to be less flat. Almost orthogonally to depth load area, perimeter and edge complexity increase, suggesting that the larger lakes tend to have more complex shorelines but that this is relatively independent of lake depth. Weakly increasing with the area group is the lake shape metric, suggesting that the larger lakes tend to be more oblong, however the feature is not exclusive to the large lakes.

The volume vector increases in the direction between depth and area, slightly closer to the area, suggesting that both lake area and depth modulate lake volume. In a 90-degree angle to volume and in the same quadrant as percent littoral the dynamic ratio loads most strongly on the 2nd component axis, again reinforcing the trend that lakes with a plate-like profile are more likely to have sediments shallower than 3 m depth and

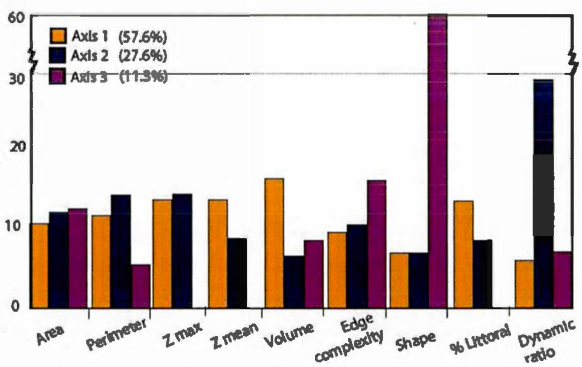


Figure 1.3.3.2 Bar-chart displaying the relative contribution of each metric to the 3 first PCA components. Showing more clearly that the axes can be interpreted as: Size, Bathymetric Shape, and Complexity.

therefore a high proportion of littoral area. Weakly opposing the dynamic ratio is Z max and Z mean, suggesting that the deeper lakes are more likely to have a bowl-shape although there is a certain degree of decoupling in this relationship.

The strongest vectors loading on the 3rd component axis were lake shape and edge complexity. All other vectors are only marginally affecting or not correlated at all, as is shown in figure 1.3.3.2. Overall, the first component therefore seems to comprise

Table 1.3.4.1 Statistics summary of minimum, maximum, mean, and standard deviation for each of the 9 metrics and 10 lake classes.

		Lake class									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Area (km^2)	Min	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
	Max	188	74	480	417	195	879	1,370	545	244	1,220
	Mean	0.35	0.07	0.07	0.23	0.11	0.14	0.20	0.14	0.30	1.30
	Std Dev	5.9	0.8	1.5	3.2	2.2	2.3	4.3	1.8	2.6	18.8
Perimeter (km)	Min	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26
	Max	2,752	521	820	816	3,641	3,018	5,711	785	1,177	1,841
	Mean	3.7	1.2	1.1	2.1	1.5	1.7	2.3	1.9	2.9	5.7
	Std Dev	66.8	6.3	5.4	13.0	9.6	9.0	18.7	5.4	11.3	42.7
Edge complexity	Min	3.6	3.6	3.6	3.6	3.6	3.6	3.6	3.6	3.6	3.6
	Max	222	107	137	148	140	121	154	115	125	113
	Mean	5.9	5.4	5.2	5.7	5.3	5.7	5.9	6.1	6.3	6.9
	Std Dev	6.3	2.2	2.1	3.5	2.3	2.6	3.0	2.9	3.8	5.9
Shape	Min	0.16	0.12	0.09	0.10	0.09	0.02	0.08	0.10	0.10	0.10
	Max	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Mean	0.61	0.63	0.58	0.62	0.61	0.65	0.64	0.67	0.66	0.67
	Std Dev	0.16	0.15	0.16	0.17	0.16	0.16	0.16	0.16	0.16	0.17
Z max (m)	Min	1.3	1.3	0.8	0.9	0.8	0.8	0.8	1.3	1.1	1.4
	Max	83	116	107	99	144	141	100	128	121	120
	Mean	3.4	8.8	4.3	6.2	6.8	12.3	8.5	19.1	11.7	19.6
	Std Dev	4.0	5.0	3.2	5.7	4.5	7.9	6.8	10.7	8.8	12.8
Z mean (m)	Min	0.4	0.4	0.2	0.2	0.2	0.2	0.2	0.4	0.3	0.4
	Max	72	110	99	90	143	140	90	123	115	115
	Mean	1.4	4.4	1.8	3.0	3.2	6.8	4.3	11.8	6.5	12.3
	Std Dev	2.9	3.3	1.9	3.7	2.9	5.6	4.6	8.2	6.3	10.1
Volume (km^3)	Min	$2.0e^{-6}$	$2.0e^{-6}$	$1.2e^{-6}$	$1.5e^{-6}$	$1.1e^{-6}$	$1.7e^{-6}$	$1.9e^{-6}$	$2.3e^{-6}$	$2.0e^{-6}$	$2.2e^{-6}$
	Max	9.9	8.0	13.8	11.4	54.8	55.9	57.7	45.1	15.7	70.1
	Mean	$9.2e^{-3}$	$0.9e^{-3}$	$0.5e^{-3}$	$2.9e^{-3}$	$1.2e^{-3}$	$2.6e^{-3}$	$2.8e^{-3}$	$3.9e^{-3}$	$5.2e^{-3}$	$50.6e^{-3}$
	Std Dev	0.261	0.043	0.042	0.086	0.100	0.117	0.168	0.135	0.102	1.144
% Littoral	Min	0.6	0.2	0.3	0.0	0.0	0.3	0.1	0.1	0.1	0.1
	Max	100	100	100	100	100	100	100	100	100	100
	Mean	92.1	48.5	82.3	71.3	63.1	37.2	57.3	21.8	43.9	25.3
	Std Dev	20.6	26.9	24.0	32.0	29.0	27.9	32.7	22.9	32.7	27.7
Dynamic ratio	Min	4.7	4.4	5.6	4.5	4.4	4.3	4.4	4.4	4.5	4.5
	Max	7,115	1,051	3,592	5,291	4,630	2,287	2,759	1,341	4,260	2,554
	Mean	210.0	48.3	125.0	138.7	86.9	47.7	102.3	30.4	85.5	53.2
	Std Dev	326.9	45.2	110.5	192.7	85.4	53.5	113.9	39.1	112.4	86.5
N		1,784	42,385	127,772	35,398	339,434	332,133	124,448	133,651	53,378	8,151

lake area and the lake metrics that strongly covary with area, the second component integrates the bathymetric shape which is often decoupled from lake area, and the third integrates various aspects of shoreline complexity, which are often decoupled from both lake area and bathymetric profiles.

1.3.4 Determining lake morphological regions across the Québec landscape

The ISO cluster lake classification based on the combination of the three PCA components yielded 10 distinct lake morphological classes, which are patches of landscape where the majority of lakes share a defined combination of morphologic features. Table 1.3.4.1 presents the basic statistics that describe the morphology of the lakes that populate each of these 10 classes. The classes containing the most lakes are 5 and 6, which suggests that the most common and widely distributed lake type in Québec is about 0.11 to 0.14 km^2 and has a maximum depth of 6.8 to 12.3 meters.

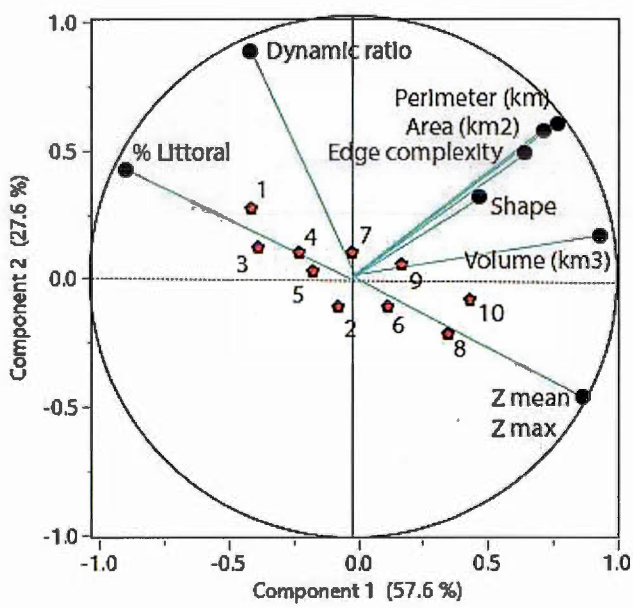


Figure 1.3.4.1 PCA plot showing the distribution of the 10 lake classes along component axis 1 and 2.

Some lake classes or morphological regions differ substantially in average lake area, for example between classes 1 and 10 (Table 1.3.4.1), but other classes share a similar average lake area but differ greatly in average bathymetry (i.e. classes 6 and 8, Table 1.3.4.1), or contour metrics (i.e. classes 2 and 3, Table 1.3.4.1). This interplay between area-related metrics, bathymetry and contour in determining the distribution of lake classes can be clearly seen in a PCA of lakes where we have overlain the vectors for the lake metrics and projected the center points of the 10 lake classes (Figure 1.3.4.1). The lake classes are spread out mostly along a gradient of lake area, complexity and bathymetry in the plot of the first two components, with some classes more influenced by one of the other group of metrics.

The lake morphological regions not only differ in the average lake size, but importantly, they also differ in their size distributions (Figure 1.3.4.2) and their respective lake abundance per lake area relationships (Figure 1.3.4.3). Whereas some

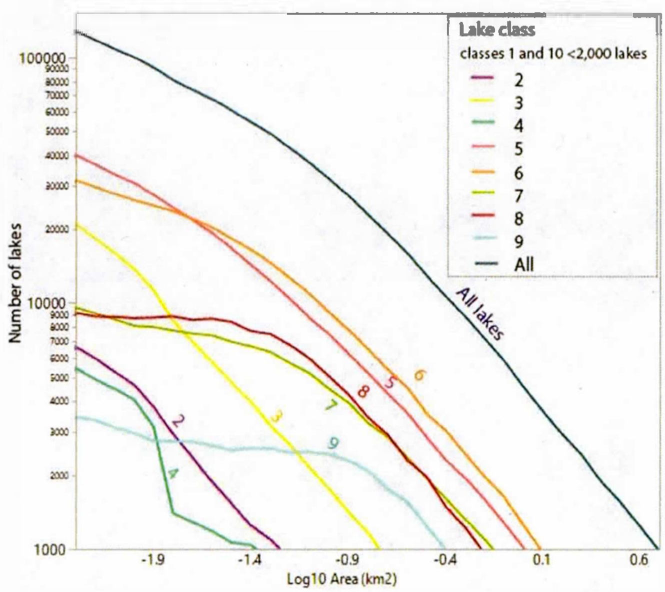


Figure 1.3.4.2. Abundance distribution of lake sizes. Size is one of the defining features for each class. Here it is shown which other lake sizes are distributed within the classes and the differences in slopes of the abundance distribution depending on which class (or all) is regarded.

of the morphological regions clearly follow the expected Pareto distributions (i.e. 5 and 6), others, such as classes 8 and 9, deviate considerably from this pattern. The different intercepts of the individual abundance to area relationships reflect major differences in the average lake density across the lake morphological regions, with some of these regions characterized by an extremely low density of lakes, or perhaps a low density of lakes larger than 0.005 km^2 (i.e. class 1), others by a high density of small lakes (i.e. class 5), and yet others are dominated by a few, very large lakes (i.e. class 10, Figure 1.3.4.3). As a result, the regions populated by the different lake classes also differ greatly in the overall lake surface, ranging from 2% to 28% of the local land area (class 1 and 10, Figure 1.3.4.3).

The lake classes described above identified with the classification procedure that combines the ensemble of lake morphometric parameters, showed a very clear spatial

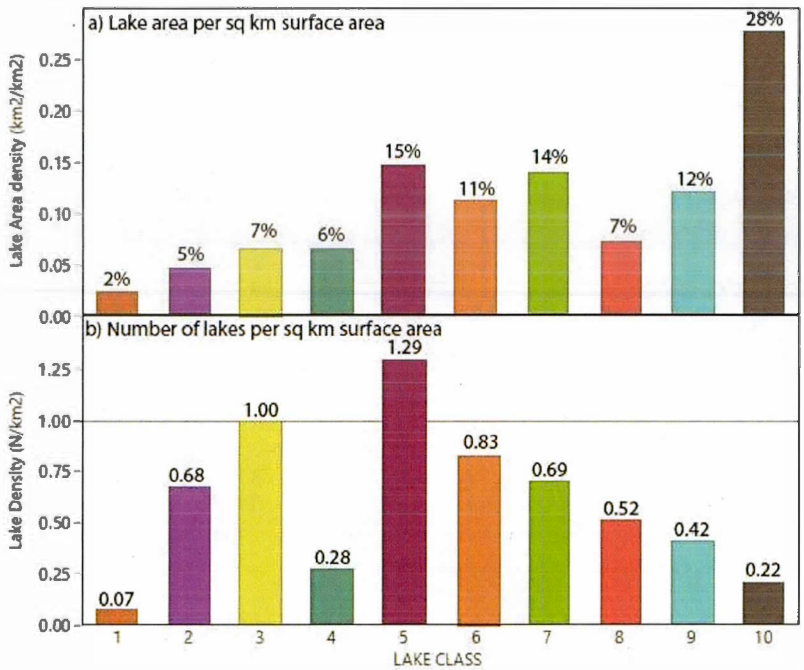


Figure 1.3.4.3.a-b. Bar chart displaying the distribution lakes across the different lake regions. The large class cover more area with fewer lakes, however the other sizes show large differences in composition.

distribution across the Québec landscape (Figure 1.3.4.4) While some classes are patchily distributed essentially across all of Québec, others dominate large portions of the landscape, and yet others are restricted to very specific locations (Table 1.3.4.2).

This generates an extremely heterogenous lake landscape: there are regions of Québec that are dominated by a mosaic of lake classes, such as the Appalachian region in the South, Gaspésie in the Southeast, or some regions in central Abitibi, whereas other regions are overwhelmingly dominated by one or at the most two lake classes, such as the Cote Nord area, the James Bay and some sub-arctic landscapes in the far North, albeit the dominant lake involved not being the same (Figure 1.3.4.4).

Table 1.3.4.2 Summary table of the attributable characteristics of each of the 10 lake region classes.

CLASS	SIZE	BATHY	COMPLEXITY	DISTRIBUTION
1	Very small	Shallow	Low	Small, dispersed patches.
2	Mid	Bowl	Low	Dispersed patches in North and South.
3	Small	Shallow plates	Low	Dispersed patches
4	Small	Shallow plates	Low	Larger dispersed patches
5	Small-mid	Average	Average	Fairly contiguous field, mainly North.
6	Mid	Bowl	Average	Commonly occurring throughout whole area.
7	Small-mid	Plate	High	Dispersed patches in Central area.
8	Large	Deep bowls	Average	Large contiguous area.
9	Large	Plate	High	Dispersed patches in Central area.
10	Very large	Bowl	High	Scattered in South

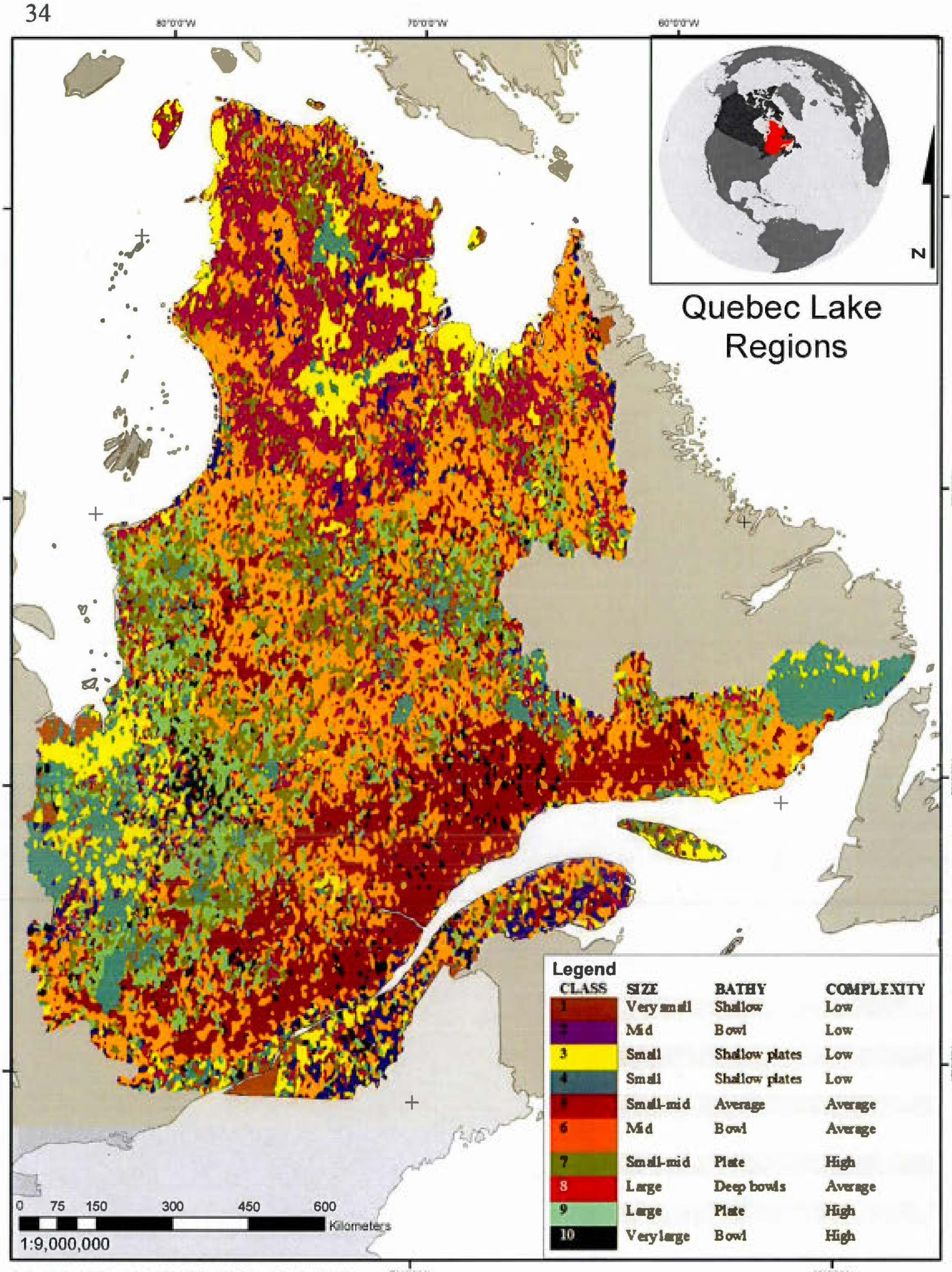


Figure 1.3.4.4 Map of the classification of 10 different lake type regions of Québec. Some classes are distributed across the whole province, while others are very specific to some areas.

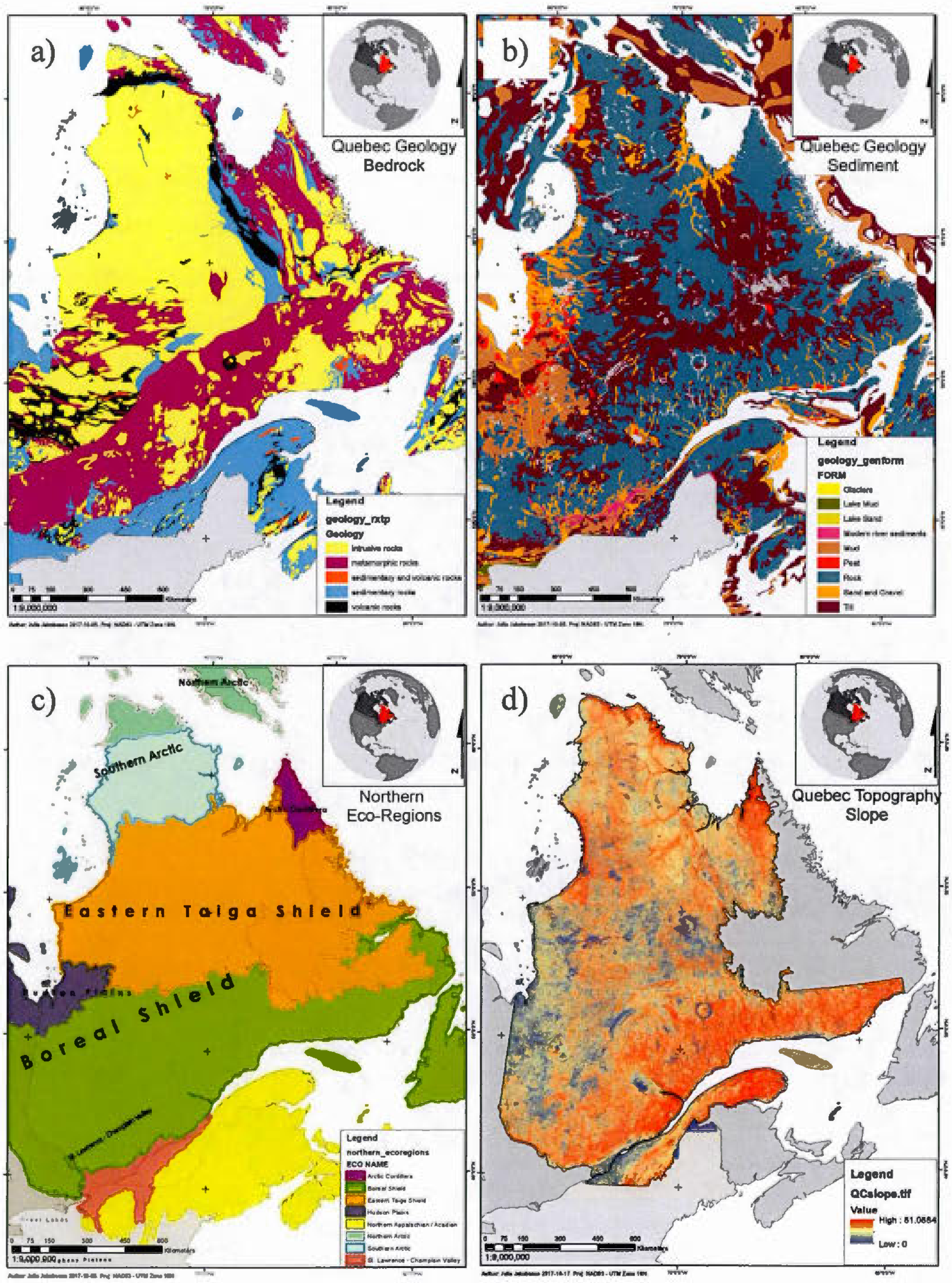


Figure 1.3.5.1.a-d Maps showing the distribution of a) Bedrock, b) Sediment, c) Eco-Regions, and d) Slope in Québec. There is a striking alignment of the 10 lake classes and the bedrock, which was not used as a parameter in the classification. All of these additional layers may be helpful for further specifications of the lake classes depending on purpose.

1.3.5 Links between lake morphological regions, geology and topography

The maps in Figure 1.3.5.1 suggest that the distribution of lake morphological regions shown in 1.3.4.4 aligns very well with a combination of regional geology and topographical structure, suggesting that local lake architecture reflects these broad regional features, which was also found by Noges (2009) for Europe. In contrast, there is almost no correspondence between the established eco-regions and the regional distribution of lake classes.

Most of the lakes (72.0%) are located in areas of glacial deposits, 19.0% is in areas of rock-outcrops and the remaining 9.0% is located in other varieties of glaciofluvial or lacustrine deposits (Figure 1.3.5.2). Table 1.3.5 shows that the underlying bedrock is 62.6% intrusive, 24.9% metamorphic, and the remaining 12.5% is sedimentary or volcanic.

The majority of the lake classes are found in intrusive bedrock with rock and till as the overlying layers, except for classes 1, 8 and 10. These classes have sedimentary or

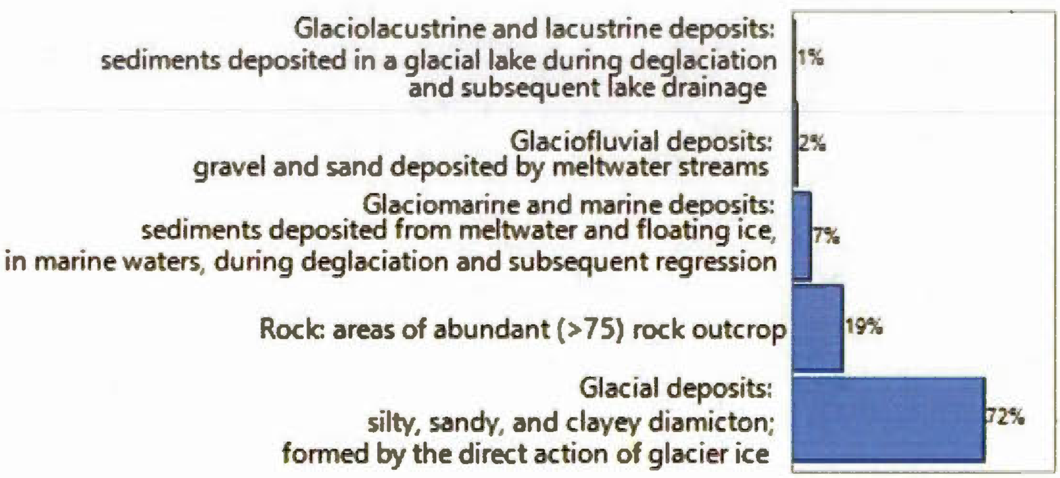


Figure 1.3.5.2 Distribution of surface layer geological compositions for all lakes in Québec.

metamorphic bedrock, where class 1 is most commonly found in peat land. From Table 1.3.5 it thus becomes possible to infer the general lake type for each lake class.

As the majority of class 1 is found in peat and till, it is thus likely to be periglacial ice collapse lakes, i.e. bog (muskeg) lakes. These types of lakes evolve relatively fast (compared to the average occurrence of an ice-age), due to accumulation of slowly decomposing biomass.

The class 2 lakes seem to be a combination of kettle and small shield lakes, depending if they are found in the southern or northern areas.

Class 3 lakes found in rock, till, sand and gravel sediments are likely to be kettle lakes formed by ice collapse in outwash areas. Class 3 and 4 first appears to be very similar, however there is a difference in lakes per km^2, where class 3 is much more densely populated.

Table 1.3.5 Percentage statistics of the geological setting each lake class is located in.

	Lake class										
Sediment	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**All**
Rock (%)	6.7	70.5	39.0	39.0	47.2	62.5	36.6	80.2	47.0	62.0	53.8
Till (%)	19.0	22.7	26.2	44.2	43.2	32.7	53.7	15.2	41.7	28.6	35.5
Sand and Gravel (%)	14.3	4.5	25.0	5.9	7.8	3.9	6.7	3.6	5.6	3.3	7.7
Mud (%)	13.7	2.1	6.1	7.7	1.3	0.9	2.1	0.8	3.7	4.3	2.1
Peat (%)	44.0	0.2	3.7	3.5	0.4	0.2	0.8	0.1	1.8	1.6	0.9
Modern river sediments (%)	2.3	0.1	0.1	0.2	-	-	0.1	-	0.1	0.1	0.1
Lake sand (%)	-	-	-	-	-	-	0.1	-	-	-	-
Bedrock											
Intrusive rocks (%)	30.2	75.9	65.2	52.1	75.4	58.9	63.9	40.0	53.4	33.4	62.6
Metamorphic rocks (%)	7.9	7.0	17.6	35.2	12.3	27.8	24.6	53.7	36.8	53.2	24.9
Sedimentary and volcanic rocks (%)	0.1	0.9	0.1	0.2	0.3	0.3	0.1	0.7	0.4	1.4	0.3
Sedimentary rocks (%)	60.3	10.2	10.8	6.5	7.4	8.3	5.1	4.3	4.6	6.6	7.5
Volcanic rocks (%)	1.7	6.0	6.3	6.0	4.6	4.7	6.2	1.5	4.9	5.4	4.7

Class 4 is mostly found in rock and till with finer particle sediment than sand and gravel. While both class 3 and 4 is probable to include bog land due to the small percentages of peat sediment, class 4 may mostly be shield lakes which are fed fine particle sediment from the further inland waters.

Class 5 is mainly found in the north with only rock and till sediment. Geologically it is similar to class 4 and thus also likely to be shield lakes. Class 5 lakes are more abundant per km^2, with slightly more complex shapes and larger lake sizes than class 4.

Class 6 lakes can be found distributed across landscape in Québec characterized by the absence of sediments. These lakes are of glacier rock basin type where the majority are shield lakes.

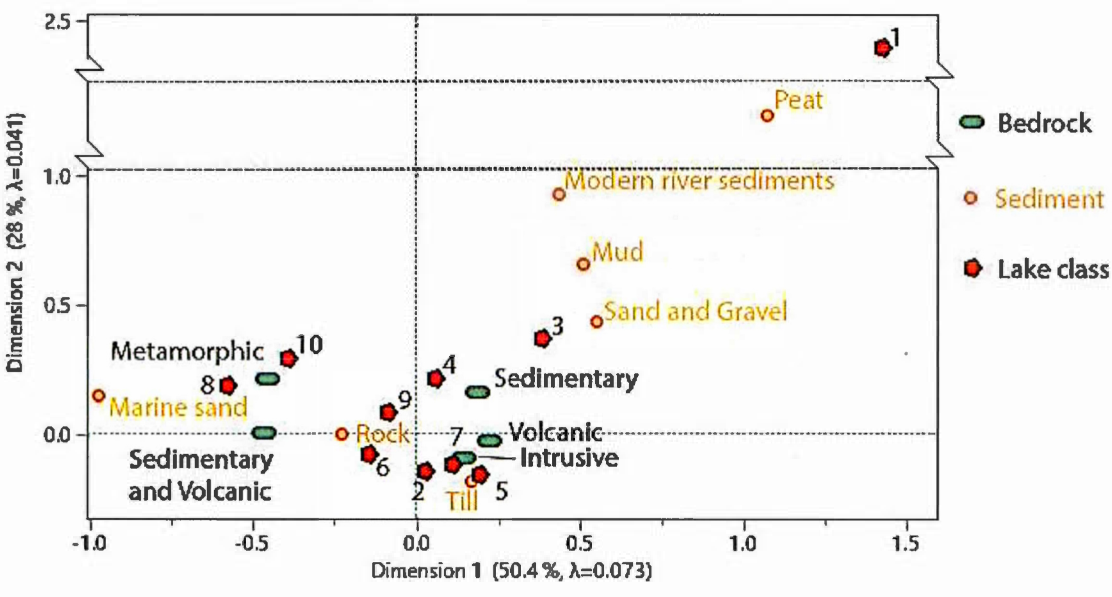


Figure 1.3.5.3 Multiple correspondence plot for geological setting (bedrock and sediment type) for each lake class.

Class 7 lakes are almost always found downstream from class 6 lakes and are primarily found in till sediment. The shapes are complex and likely to be formed by moraine damming or glint-like processes.

Class 8 lakes are large and deep in low-sediment areas, they are therefore most likely to be glacial rock basin types such as shield, valley, fjord or glint.

Class 9 lakes are often found in connection to class 7 lakes, and with equal amounts rock and till sediments, however they are much larger. They are still most likely to have been created through moraine damming.

Class 10 lakes are large with very little sediment and slightly different bedrock. This indicates that they are likely ice-scour lakes (shield lakes). These distinct geological profiles associated to each of the lake classes are summarized in the correspondence analysis presented in Figure 1.3.5.3, which shows how the lake classes align to the geologic setting.

1.4 DISCUSSION

Our results suggest that although there is a vast heterogeneity in lake size, shape and bathymetry across the landscape in Québec, lakes can be grouped into a relatively small set of lake morphological classes, which are characterized by a specific combination of morphological features and a certain distribution within the landscape. We have further shown that these lake morphological classes also have very specific geographic distributions, and that they follow the geomorphology of the landscape. While it is quite "limnocentric" to interpret the landscape from a lake morphological perspective, since in post glacial regions lakes are the result of landscape processes rather than the landscape being a result of lakes, there is a need for studies aimed at exploring lake features at large scales. As we found for the Quebec landscape, these regions may appear simply by overlaying a geological map with the movements of the past glacial flows, however, for landscapes that were formed by different processes, such tools would not be useful. The limnocentric way of reverse-engineering the lakes to the landscape classification may be the most useful path to take if one were to cover regions that were created through a variety of processes.

The lake regions follow the bedrock and sediment geology in a way that confirms that the ice abrasion created different landscape features because of how it moved and what material it was moving over. Regions where it advanced and retreated repeatedly (South) show a different composition of lake types than where the presence was more constant (Central/North). This also accounts for the topographical profile of the landscape, where elevations and slopes differ between regions, which have a large influence on the transportation of material within the watershed. Thus, the lake regions may also be useful for studying watershed-scale processes. If the underlying sediment, or lack of, is taken into account the lake classes could be even further specified. This can have large implications on understanding the continued evolution of the lakes and landscape in Québec.

These results are part of a new way of regarding lake features in the context of large-scale limnology (Cheruvelil et al. 2017). The implications this may have, both in the interpretation of past studies and design of new ones, is of great importance. Rather than relying upon a sole metric such as surface area or divisions of eco-regions, which we and others have shown can have great variance in relation to other features (Wolock et al. 2004), we provide a classification that can aim directly at lake features that will be relevant. Because the lake biological, biogeochemical and physical and environment is to a large degree determined by its morphology (Wetzel 2001, Staehr et al. 2012) and by knowing that there are regions with specific types of lakes future sampling campaigns can be designed to more effectively encompass the range of dominant lake classes. Revisiting previous studies can also be done more wisely. For example, if there are findings that seem to only apply to certain regions, such as CDOM and chlorophyll remote sensing (Brezonik et al. 2015), there may be tools in lake morphological classifications to explain some of the variation.

All of the metrics show some degree of scaling with lake size, but the degree of allometric coupling varies greatly, such that for a given lake size some of the metrics may vary in some cases by several orders of magnitude. The result is that there is a wide range of combinations but always with a certain framework imposed by lake size. This is also why we can draw just as much information from the cases that are absent, as the ones that are present. For example: small lakes can range from being very shallow and very deep but there are hardly any large lakes that are very shallow. As a consequence, in areas where lakes tend to be shallow, they also tend to be small. While this may not be a surprising result it does provide some validation to the depth calculation algorithm that we used (Heathcote et al. 2015), which does not incorporate any form of lake size. This however becomes a problem when it comes to the very small lakes in relatively steep terrain, where the algorithm clearly overestimates the maximum depths. To remedy the possible bias of the analysis because of those artefacts, lakes with unrealistically low values of the dynamic ratio were removed from the set. We deemed values in the lower 90-95 percentiles of the Dynamic Ratio as unrealistically low and excluded these from all further analyses. Ultimately these biases may be accounted for by fine-tuning the depth algorithm and adjusting it for different regions (Oliver et al 2016). The Heathcote et al. (2015) model we use here was mainly calibrated for lakes found in lake classes 6 and 8. Even though those classes are dominant across the Quebec landscape, it would be worthwhile to test it for the other lake classes and expand on the dataset to perhaps develop class-specific algorithms.

As lake size increases we also see a trend for shoreline complexity to increase and the lake shapes to depart from the round forms. We can interpret this as part of what has been termed the fractal nature of lake complexity (Mandelbrot 1967), and also a sign of a shift from one type of origin process to another, i.e. kettle lakes, or shield lakes.

When it comes to the use of metrics to describe shape complexity all methods are in one way or other variations on the area-perimeter ratio. In this regard, although we chose to use edge complexity as one of our shape metrics, we could also have used fractal dimension (Cael & Seekell 2016; Steele & Heffernan 2014). The relationship between the two is not perfectly linear due to the algorithm for fractal dimension being one step further removed from any size dependence, but they are nevertheless not entirely decoupled as can be seen in figure 1.4.2.1.

$$FractalDimension = 2 * \frac{log(perimeter)}{log(area)}$$

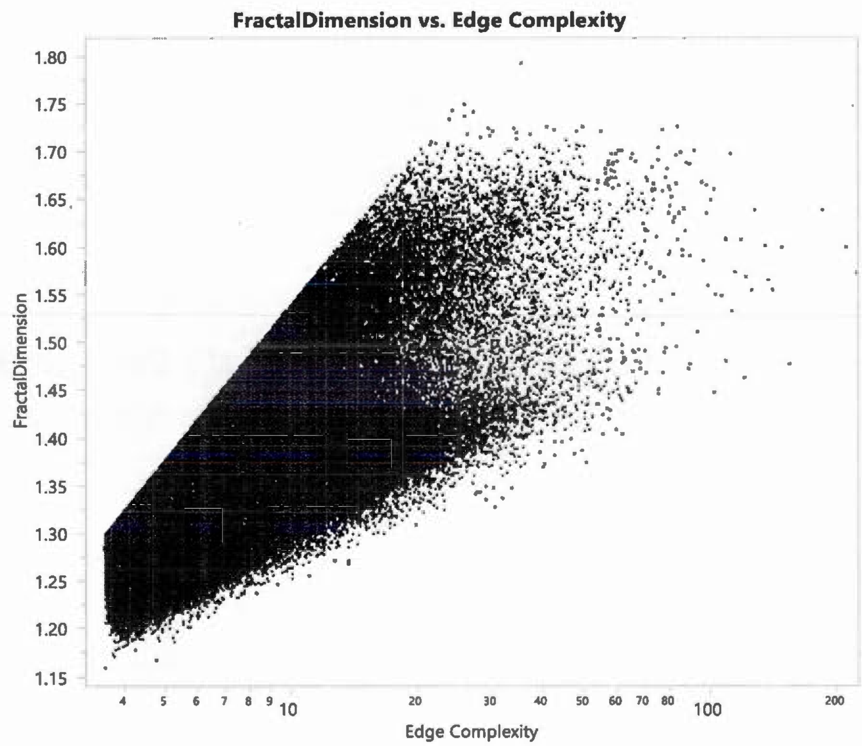


Figure 1.4.2.1 The relationship between the metrics Edge Complexity and Fractal Dimension. The two metrics display a sharp cut-off and then spreads with increasing variation with growing complexity.

The relatively stringent surface area constraint we imposed on our data (0.005 km^2) was based on the use of metrics related to shape. The resolution of the source images may be sufficient to detect smaller waterbodies, but insufficient to adequately represent their morphometries. By using this relatively large cut-off we may thus be under-representing regions where lakes smaller than 0.005 km^2 dominate, but on the other, we are ensuring that our regional analysis of lake shape metrics is robust.

We recognize that the density of lakes and other open water surfaces plays a large role in the functioning of the landscape, however we do not include it until after the division has been made and we still only focus on lakes larger than 0.005 km^2. This is because of our "limnocentric" perspective, where the emphasis was on individual lake morphological features as focal classifiers. In this regard we can clearly see (figure 1.3.4.2 & 1.3.4.4) that each lake class has a distinct landscape distribution, despite the fact that we used this somewhat arbitrary surface area cutoff.

Unsupervised classifications such as ISO clustering and Maximum Likelihood on average do not perform as well as the newer generation of supervised, machine learning classifiers, such as a Random Forest classifier (Cutler et al. 2007, Cheruvelil et al 2017). However, applying a more advanced classification algorithm could in our study have been somewhat redundant. While some regions clearly contain specific combinations of lake classes rather than one or two dominant classes, it is not certain that a more complex algorithm would have been more effective at defining these regions rather than just recognizing the constituent parts.

A validation of the classification is made when the class-values are assigned to each lake metrics and statistically analyzed. Since all lakes that fall in a certain region do not necessarily share the same features as the majority, there is a certain proportion of outliers within each class. A small part of that error is attributable to the generalization part of the post classification process that introduces a slight spatial offset in some smaller, irregular patches. This error is however negligible on a regional scale since it is only a minor shift.

In figure 1.3.4.2 we can also see that the size versus abundance relationships vary greatly between the lake classes that we identified, and therefore between landscapes within Québec where these classes are dominant. However, when all these classes are integrated, these differences are smoothed out and the overall size to abundance relationships for the entire region of Québec follows roughly a power law distribution, which agrees with what was found by Cael & Seekell (2016) for Sweden, although the slope for the ensemble of lakes in Québec starts to deviate from this Pareto distribution in lakes below 0.40 km^2. This supports the conclusion that using a Pareto distribution to estimate lake size abundances may not be the most suitable method, because it may yield biased estimates particularly in the small lake category. These deviations from Pareto distribution were much more obvious for some of the lake classes, which did not conform at all to a power law distribution (i.e. Classes 7, 8 and 9). This implies that extrapolation of a single power law relationship to the ensemble of lakes in Québec, or to the entire landscape, would yield extremely biased and unrepresentative estimates of lake size distributions in some portions of the landscape, and this may explain in part the large discrepancies that exist worldwide in lake accounting. Our results highlight the importance of identifying dominant lake morphological classes, which are characterized not only by a certain average lake architecture, but also associated to a certain size distribution, and further establishing the large-scale spatial distribution of these classes across the landscape.

1.5 CONCLUSIONS

We have identified a core set of ten lake classes, based on a combination of morphometric features, which together capture the main axis of variation of lake properties across the entire landscape in Québec. We have shown that these lake classes have a particular spatial distribution across the landscape in Québec, tied to topographical and geologic features. This in turn results in distinct geomorphological regions for lakes across Québec, and determines what we could refer to as limno-regions, which are characterized by different combinations of these core classes of lakes: Some limno-regions are dominated by a single, sometimes two lake classes, other limno-regions are composed of a mosaic of multiple lake classes. The ten different lake classes that we identified were primarily distinguished on the basis of their average lake size and on their respective size distributions, but they were further differentiated on the basis of their average bathymetric profiles and shape complexity. Had we used lake size alone to establish lake classes within the Québec landscape, we would have arrived at a much more constrained, and one-dimensional set of lake classes.

In addition to classifying and analyzing the lake features based on their PCA scores we have provided the basic statistics (area, perimeter, depth, volume, edge complexity, lake shape, percent littoral and dynamic ratio) of 1,270,878 lakes in Québec larger than 0.005 km^2, excluding reservoirs. With this data we have also explored the relationships between metrics to identify their allometric relationships with lake size, and between each other. While there are large variations for all metrics in relation to area, there are underlying structures through the mean values. The defining characteristics for the division of lake classes (size, bathymetric profile and complexity of shape) show that these different features are roughly allometrically independent.

Finally, we have shown that these lake classes relate strongly to regional features geology, topography, and lake density. Depending on if the lake classification will be used to study physical, chemical or ecological processes these relations might have different importance and implications.

CHAPTER 2

CARBON SPECIES IN RELATION TO LAKE MORPHOLOGICAL AND AQUATIC NETWORK METRICS IN BOREAL QUÉBEC

2.1 INTRODUCTION

Inland waters play a significant role in carbon processing and as sites of greenhouse gas emission to the atmosphere (Cole et al. 2007 Duarte et al. 2005; Tranvik et al. 2009; Bastviken et al. 2011). However, accurately quantifying these emissions for entire regions, where physical sampling of every waterbody is impossible, remains a major challenge. Estimation methods are rough and there is a need to further understand which features could potentially be utilized for more accurate upscaling. In order to be useful these features need to be able to be obtained remotely. Here we have collected a suite of morphological (*see chapter 1*) and network metrics to investigate whether or not they can add predictive value to four different forms of carbon. The water chemistry data has been collected in 7 different regions of Québec during the last decade and has been published in previous studies.

2.1.1 BOREAL INLAND WATERS

Boreal aquatic inland waters play an important role in carbon emission, processing and storage (Algesten et al. 2004; Cole et al. 2007; Bastviken et al. 2011). While it is estimated that global inland waters cover about 3.7% of the worlds non-glaciated areas, the arctic and boreal contain the highest concentration of waterbodies, where the lakes also have the largest areas and perimeters (Verpoorter et al. 2014; Messager et al. 2016). In chapter I we explored the morphological characteristics of Québec's lakes, where we found that at least 12.3% of the surface area consisted of lakes larger than 0.005 km^2 (reservoirs excluded).

2.1.2 CARBON SPECIES

The carbon species we are interested in looking at are CH_4, CO_2, DOC and DIC. CH_4 and CO_2 are powerful GHG's when released to the atmosphere, while DOC and DIC are the main inputs of carbon to lakes which fuel a number of reactions and processes along the whole aquatic continuum (Tranvik, 2009). The fundamental processes underlying the variation of carbon species in lakes, such as sources, chemical properties and major pathways are relatively well known, but how lake morphology may modulate these processes and the dynamics of these C species has hardly been studied. However, to accurately explore additional effects of morphology the known parameters must be accounted for.

DIC becomes a prominent form of carbon in lakes of higher northern latitudes and boreal forests in carbonate terrain. The increasing concentrations of DIC, even within carbonate rock regions with little organic material, can be partially attributed to the retrograde solubility of carbonates. Retrograde solubility is a counterintuitive effect meaning that solubility increases as temperature decreases, due to the water potential to hold more dissolved CO_2 thus lowering pH (James & Jones, 2015). From the surface the type of forest can affect soil pH, through for example, the litter it deposits on the forest floor (Osman 2013). Once in the open water, transported from groundwater, most of the DIC that is CO_2 is quickly released to the atmosphere (Öquist, 2009). This implies that part of the carbon dioxide emitted from lakes originate directly from CO_2-rich ground and surface water (Striegl & Michmerhuizen, 1998) or is released through photochemical processes (Koehler et al., 2014) but most of it emerges from respiration within the lake; metabolized from terrestrially derived organic carbon (del Giorgio et al., 1999; Jansson et al., 2000; Battin et al., 2008). These two species might therefore relate to metrics describing relative terrestrial contact, exposure to sunlight and placement in the aquatic network.

DOC becomes more apparent in the noncarbonate boreal forests (Stets et al, 2009; Tranvik, 2009). It is involved in a number of in-lake processes, such as being mineralized or incorporated to heterotrophic metabolism (Battin et al 2008), undergoing photochemical degradation (Molot & Dillon, 1997; Vachon et al 2016) or sequestered into lake sediments (Wachenfeldt & Tranvik, 2008), with the remainder being further transported (Cole et al 2007). Metrics related to water movement and sunlight exposure might thus be relevant to DOC prediction.

In contrast to the other carbon species, most CH_4 in lakes is synthesized by anaerobic archaea (Zehnder et al, 1979), where the majority of it is produced in anoxic sediments, clearly driven by temperature (Crill et al, 1988; Rudd & Hamilton, 1978). Littoral and shallow areas of lakes usually emit more CH_4 since there is less time for the methane to react with the oxygen in the water column before reaching the surface (Yavitt et al. 1990, Kortelainen et al. 2006) and if vegetated, plant stems may transport the gas directly into the atmosphere (Bergström et al. 2007). CH_4 emissions are therefore strongly predictive from features such as lake area, water depth, and anoxic lake volume fraction (Bastviken et al. 2004).

We can expect the carbon species to relate to the metrics more or less strongly but also in different ways. Where conditions are favorable for increased DIC, it is likely to be the opposite for the production of CH_4. The gaseous species are also likely to be more reactive and thus more predictable from morphological features, whereas the dissolved ones are less so. Thus, this chapter explores if introducing further metrics to the prediction or understanding of carbon species processes will have different results and if so, to what degree.

2.1.3 METRICS

There is an extensive literature on the climatic and environmental drivers of lake C species. This literature suggests that each C species has its own set of environmental and climatic drivers, some of them shared, some of them specific for a given C species or location (Kortelainen 1993; Prepas et al. 2001; Sobek et al. 2003; Bade et al 2004; Larsen et al. 2011; Sadro et al. 2012; Lapierre et al. 2012 & 2015; Wik et al. 2014 & 2016; Pinho et al. 2016). Among these, lake area has been identified as a feature that influences multiple aspects of lake functioning, including GHG dynamics and C processing (Håkansson 2005; Kankaala et al. 2013; Vachon et al. 2013; Rasilo et al. 2015; Holgerson et al. 2016). There has been little work done beyond assessing the role of size on lake biogeochemical functioning, however there are some (Rasmussen et al. 1989; Soranno et al. 1999; Schilder et al. 2013; Steele et al. 2014; Fergus et al. 2017; Mosquera et al. 2017). Despite the scarcity of studies there is reason to hypothesize that other aspects of lake morphology might also influence C dynamics in lakes. For example, a large portion of the pCO_2 that is found in lakes originate from terrestrial soil carbon which is transferred to lakes via groundwater (Sobek et al. 2005; Lapierre et al. 2013), lakes with a higher degree of contact surfaces with the terrestrial soils and lake sediments should perhaps therefore show higher average pCO_2 (Kortelainen et al. 2006). Likewise, most of the pCH_4 that is found in lake waters originates from anoxic lake sediments (Rudd & Hamilton 1978). Thus, lakes with a high ratio of sediment surface area to water volume, as well as shallower lakes where CH_4 oxidation in transit is minimized, should have higher average pCH_4.

It might be expected that not only the shape but the position of lakes within the aquatic network should influence C dynamics. For example, much of the DOC that is found in lakes is of terrestrial origin (Sobek et al. 2007) and the pathways of delivering this C to lakes thus have a profound influence on the amount and quality of DOC that is found there (Jonsson et al. 2001). Lakes that have tighter hydrologic connections to their watersheds, that receive multiple riverine sources and that are positioned upstream within the aquatic network should be expected to have higher average DOC concentrations.

There is indeed an increasing number of studies that explore network configuration and the positioning of lakes within the continuum as having a key role in the biogeochemical functioning (Kratz et al. 1997; Soranno et al. 1999; Noges 2009; Sadro et al. 2012; Fergus et al. 2017). When applied, these studies usually focus on nutrient dynamics (Martin et al. 2006; Zhang et al. 2012). The influence of lake morphology and network configurations on C dynamics in lakes have to date seldom been assessed, and furthermore, no study to date has looked at the possible interaction between the two. What complicates matters further is that each of the major C species in lakes; DOC, DIC, CO_2 and CH_4, likely have different responses to morphological and network variables. Such that identifying their influence on a given C species does not necessarily allow interference on the behavior of the other species; they all need to be independently assessed if we are to have an overall perspective of how these variables influence lake C dynamics. Furthermore, these potential influences are likely subtle and are only detectable across a wide range in the ambient concentrations of the four C species. This can only be generated through a cross-regional perspective that maximizes the environmental gradients and landscape features. In this study we have explicitly addressed the issues above and have explored the potential influence of lake morphology and network configuration across >200 lakes that span the entire size, shape and configuration gradient, and which are located in varying regions throughout the boreal biome of Québec.

2.2 METHODS

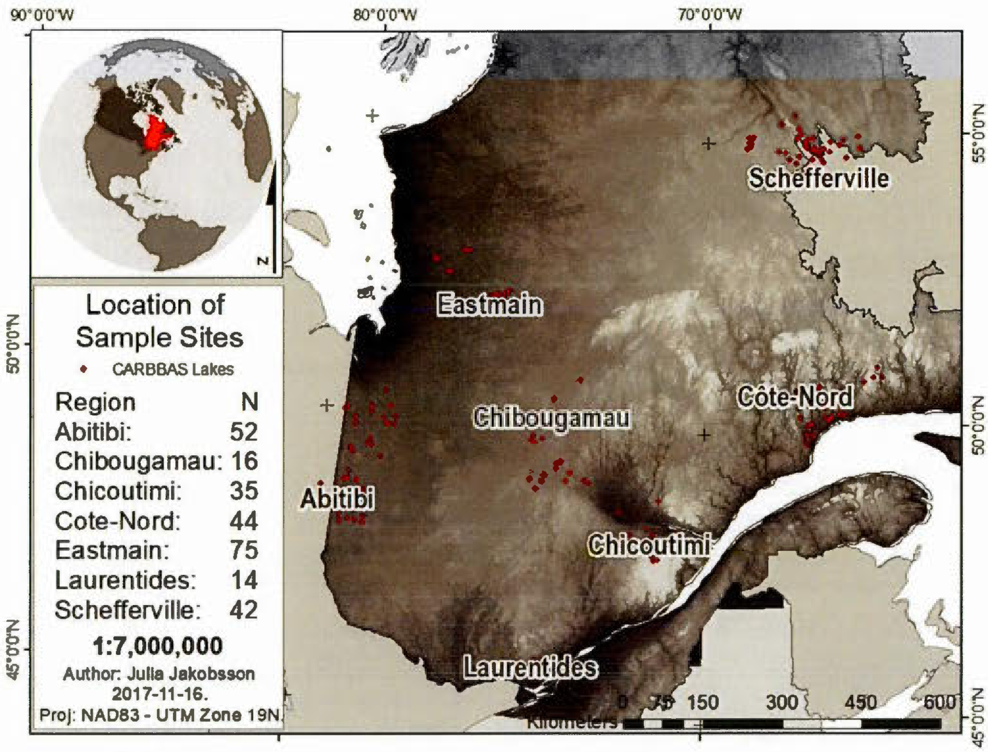


Figure 2.2.1 Overview map of the sampling locations in Québec, Canada. Seven regions were sampled between 45° and 56° North.

2.2.1 Sampling

Sampling and analysis of DIC, DOC, pCO_2 and pCH_4 is described in detail in the papers by Lapierre & Del Giorgio 2012 and Rasilo et al. 2015. The timespan of the sampling campaigns was: 2007-2009 Eastmain region, 2010; Abitibi, Baie James & Laurentides, 2011; Chibougamau & Chicoutimi, 2012; Schefferville, and since 2013-2015; Cote Nord (Figure 2.2.1). Some sites have been sampled more than once, the values have in those cases been averaged. The sites were accessed by boat or in remote areas by hydroplane. Water samples for analysis was taken at 0.5-m depth over the deepest part of the lake or an estimated deepest part if bathymetric maps were not available. The values and ranges of the carbon species for each region can be seen in Table 2.2.1.

The aquatic surface water partial pressure of CO_2 (pCO_2) (in μatm) was in the earlier years measured in field with a MiniModule equilibrating module and an EGM-4 gas analyser. The partial pressure of CH_4 (pCH_4) (in μatm) was measured using the headspace method with triplicate 60ml polypropylene syringes, 30ml lake water, 30ml ambient air headspace, 1-minute shaking with the sample injected to a 30ml glass vial, stored cold and dark, then analyzed in a gas chromatograph in the lab. In later years, this method has been replaced with a headspace method for both pCO_2 and pCH_4 where a 1L glass bottle, fitted with two vents (50/50 water-headspace), is vigorously shaken for 3min after which the headspace is extracted by pumping more of the same lake water into the bottle, pushing the headspace gases into a dark, vented gas container and then analyzed on an LGR Ultraportable Greenhouse Gas Analyzer.

Table 2.2.1. Overview of the carbon species (CH_4, CO_2, DOC and DIC) minimum, maximum, mean and standard deviation values for all regions.

	pCH_4 (ppm)				pCO_2 (ppm)			
	min	max	mean	std	min	max	mean	std
Abitibi	9.7	3612.0	279.5	589.8	89	2881	754.0	570.2
Chibougamau	13.7	128.5	35.0	29.9	403	863	582.4	140.4
Chicoutimi	21.3	965.0	285.3	254.1	432	1972	978.0	342.3
Cote-Nord	14.4	1829.0	361.9	444.5	424	1550	832.3	291.3
Eastmain	11.6	245.8	52.9	47.5	338	2400	806.9	352.3
Laurentides	74.9	1593.0	610.2	466.0	162	787	428.7	164.4
Schefferville	17.1	1378.0	187.0	330.4	329	1392	542.0	198.9
	DOC (mg/L)				DIC (mg/L)			
	min	max	mean	std	min	max	mean	std
Abitibi	4.37	22.25	11.62	4.35	0.434	13.53	4.077	3.82
Chibougamau	4.03	13.24	7.94	2.68	0.896	8.20	3.342	2.40
Chicoutimi	3.91	14.28	7.85	2.14	0.923	7.92	2.419	1.81
Cote-Nord	3.16	13.11	8.54	2.72	0.431	2.34	0.975	0.43
Eastmain	4.96	19.40	8.84	3.20	0.671	2.23	1.439	0.58
Laurentides	6.33	15.49	8.19	2.58	0.734	5.54	2.596	1.79
Schefferville	1.41	7.68	3.80	1.69	0.282	14.42	4.241	4.41

DOC and DIC were measured in lake water samples filtered through 0.45μm PES cartridges and later analyzed on a OI 1010 TOC analyzer after sodium persulfate digestion.

2.2.2 Metrics & Data

The metrics were divided into two groups. The first one being Lake Morphology, where the different features are exhaustively described in chapter 1. The parameters tested under the Lake Morphology are: Area, Perimeter, Volume, Maximum Depth, Mean Depth, Edge Complexity, Shape, % Littoral Area, and Dynamic Ratio. The other group of metrics are Network Configuration, which include: Number of Inlets to the lake, Upstream Length, Number of Upstream Lakes, Lakes/Inlet, Lakes/Upstream Length, and Upstream Length/Inlet. In addition to these metrics, the variable Latitude, was included in the analysis. This is to account for the non-morphological parameters affecting the carbon species, as a proxy for temperature, precipitation, soil carbon, vegetation, land use and other regional-scale drivers that tend to vary along latitudinal gradients in this region (see table 2.3.1.1).

Some of the Network Configuration data was extracted using tools for geometric networks in ArcGIS 10.3, which included Number of Inlets, Upstream Length, and Number of Upstream Lakes. Number of Inlets accounts for the total number of rivers or streams entering the lake in question. Upstream Length is the total length of all rivers, streams and waterbodies upstream of the lake. Number of Upstream Lakes counts the total number of lakes in the upstream network from the focal lake.

In addition, three ratios were taken from those metrics: Lakes/Inlet, Lakes/Upstream Length, and Upstream Length/Inlet. By taking ratios of the simple metrics, we aim at characterizing the structure of the network and potential differences in processing the water masses and carbon entering lake have been through. The Lakes/Inlet metric is a ratio of the total number of upstream lakes and total number of inlets to the lake. This metric is used to assess if the lake in question is mostly fed by streams with or without lakes, where a lower value indicates mostly stream and a higher value indicates mostly lake. The Lakes/Upstream Length is a ratio of total number of upstream lakes and total length of watercourses upstream of the lake. Lakes/Upstream length therefore informs us if the upstream network has tightly or distantly connected lakes. Upstream Length/Inlet is a ratio of the total upstream watercourse length and the total number of inlets. This metric thus infers a network position of the lake, where it is either closer or further from the headwaters.

To incorporate a hydrological component, we calculated Water Retention Time (WRT) for each of the lakes. WRT was calculated using Lake Volume (m^3), catchment Runoff (mm m^{-2} yr^{-1}) and Catchment Area (m^2). While this is a rough estimate method, it can be applied using data obtained remotely.

$$WRT = \frac{Volume_{Lake}}{Flow}$$

$$Flow = Runoff * Catchment\ Area$$

The origin data for the geometric river network was obtained from Canadas Natural Resources Department on the open data portal GEOGRATIS. This is also where spatial data for the proxy latitude variable was collected. Annual mean temperature (°C yr^{-1}), annual mean precipitation (mm yr^{-1}), annual mean runoff (mm m^{-2} yr^{-1}), landcover and geology was used for the substitute latitude and the WRT calculation. The composition of Geology, Sediment and Landcover was converted into relative percentages of the total watershed area. The watersheds (n=56), by which those metrics were delineated, was also collected from the same portal. The watersheds are derived from the HydroBASINS product produced for the World Wildlife Fund US (Lehner et al. 2013). It is divided into 9 scales, from continental (level 1 = $15.9*10^6$ km^2) to headwater stream orders (level 9 0.1-5470 km^2), we used mostly level 7 (1.4 – 120 km^2) for this data. In addition to using latitude as a proxy in the analysis, lake area was also included.

In the case of DOC, area was replaced with dynamic ratio. This is to account for the strong influence of those morphological features on the carbon species.

2.2.3 Statistics

To identify which morphometric or network parameters could have any effect on the variation in partial pressures and concentrations an elastic net generalized regression was used (Zou & Hastie 2005). The two groups of parameters, Lake Morphology and Network Configuration, were tested separately to more clearly see the effects of each in the analysis. The elastic net is in a category of penalizing regression models which is suitable for datasets with a large number of parameters. The penalties are dealt by shrinking uninteresting variables to zero in order to avoid overfitting. In that class of penalty methods, the elastic net has advantages over similar models such as lasso in how it treats parameters that may be highly correlated, where it can for example: either include or exclude parameters as a group. The elastic net regression is also known for its ability to handle instances where there may be more predictors than sample points. While this is not the case for our data (278 observations for 9+7 predictor variables) it is a strong fundamental feature of the model.

The validation method used for the predictor selection is minimum AIC (Akaike 1979). By employing AIC validation criteria in the model, the potential outcomes are continuously compared against each other, to select the one of highest quality. AIC can thus include every predictor that contributes predictive power, even if the individual variable may not be statistically significant.

Outliers for each carbon species were identified and excluded from the analysis through a multivariate k-nearest neighbor method. This method entails locating and excluding values which are on a Euclidian distance far enough from any neighboring data points to be considered an outlier. The values had been log10 transformed before the outlier analysis. All parameters were also log10 transformed to account for the distribution of the data.

The metrics that were found to contribute to the variation of each carbon species by the elastic net model are then fed into a decision tree partitioning model. With the partitioning model we are able to determine the ranks and partial contribution of each parameter. The decision tree works by recursively branching predictor variables until the best fit model for the data points are found (SAS Institute Inc., 2017). This results in a tree-like structure of rules, cataloging values into increasingly precise groups based on their commonalities.

2.3 RESULTS

Lake area has previously been shown to be an important determinant of various aspects of lake functioning (Bastviken et al. 2004; Håkansson et al. 2005; Kortelainen et al. 2006; Post et al 2000; Vachon et al. 2013; Rasilo et al. 2015; Hall et al. 2016; Holgerson et al. 2016). The four main carbon species showed various degrees of covariation with lake area across the boreal biome (Figure 2.3.1). The strongest relationship was with pCH_4, which declined steeply as a function of lake area (R^2=0.31) (Fig 2.3.1.a), as did pCO_2, although this relationship was weaker (R^2=0.08) (Fig 2.3.1.b). DOC did not show any relationship with area (R^2=0.00) (Fig. 2.3.1.c), whereas DIC had no significant relationship with area (R^2=0.04) (Fig 2.3.1.d). Because various aspects of lake morphology tend to covary with lake area, as shown in chapter 1, we included lake area in all of the analysis (exchanged for dynamic ratio in the case of DOC), such that we could extract the potential influence of other morphological variables, once those have been accounted for. As we show below, lake area was a significant predictor for pCH_4, pCO_2 and DOC in the multiple regression models, but was not significant in the case of DIC.

2.3.1 Latitude as a proxy for environmental variables

To account the more well-known effects to each species of carbon, latitude was included as a proxy in all subsequent analyses. While the elastic net regression is able to manage multiple parameters, the potential links to the various aspects of lake

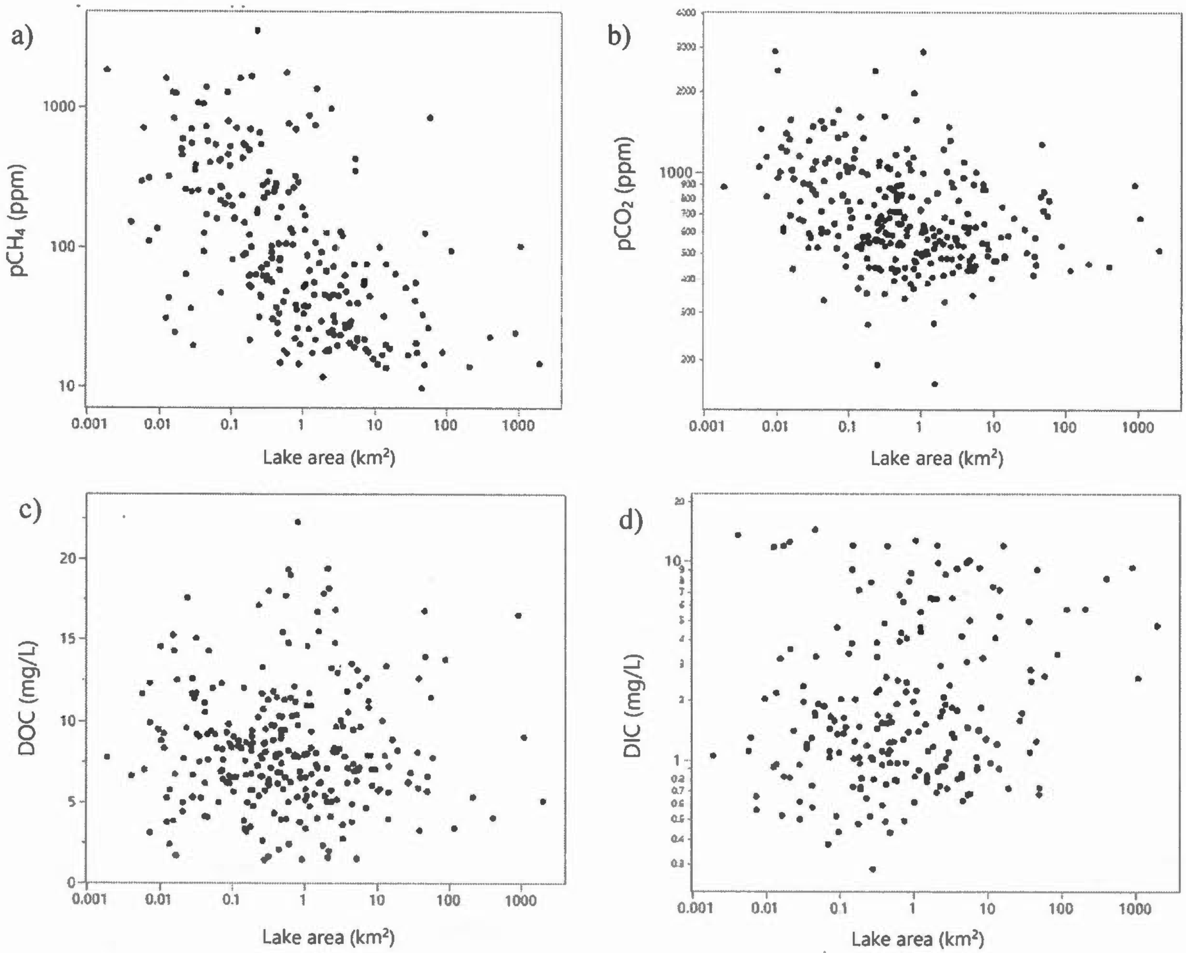


Figure 2.3.1 Overview of the carbon species; CH_4, CO_2, DOC and DIC in relation to lake surface area. a) CH_4: R^2=0.31, RMSE=0.48. b) CO_2: R^2=0.08, RMSE=0.19. c) DOC: R^2=0.00, RMSE=0.22. d) DIC: R^2=0.04, RMSE=0.39.

morphometry and network position might become obscured, since carbon species may be linked to a whole suite of environmental and climatic factors. In Québec many of these factors tend to covary with geographic position, and in particular with latitude. In table 2.3.1.1 we show the environmental and climatic features that are significantly correlated to latitude (i.e. mean annual temperature, soil carbon, NPP, evapotranspiration, precipitation, runoff, and landcover types). Figure 2.3.1.1 shows the predicted latitude based on these features, showing a near perfect correlation. Latitude can therefore be used latitude as an integrative variable that may account for these additional environmental and climatic drivers, in order to aid in exposing relationships to lake morphology and network configuration once these additional variables have been accounted for.

2.3.1.2 The influence of lake morphology on the four carbon species

The elastic net analysis showed that the carbon species are related to morphological features, albeit differently. CH_4 showed the strongest link to lake metrics, while DIC showed the least connection to lake morphometry or network position.

Once latitude was accounted for in the analysis, lake pCH_4 strongly declined as a function of lake area, and further declined as a function of increasing depth, and edge complexity. Other morphometric features also influenced pCH_4, albeit weakly, including: Volume, Shape, % Littoral Area and Dynamic Ratio. The model nevertheless included all of the parameters in the predictive model, highlighting the fact that lake pCH_4 is strongly influenced by lake morphology. (Figure and Table 2.3.1.2).

As with pCH_4, lake pCO_2 showed a strong relationship to latitude. Edge complexity emerged with a strong positive relationship once the negative relationship to area was accounted for. Less influential predictors were mean depth, shape, and dynamic ratio (Table 2.3.1.3). Metrics that were excluded from the predictive model were: perimeter, maximum depth, volume, and % littoral. This suggests that the strongest morphological influences on pCO_2 are related to the degree of contact area with the surrounding watershed (Table 2.3.1.3).

The strongest morphometric predictor of lake DOC concentration, after the proxied environmental and climatic effects were accounted for, was dynamic ratio. Volume and % littoral were weakly related to DOC but were nonetheless included in the elastic net predictive model (Table 2.3.1.4). This would suggest that the strongest influence on DOC concentrations are related to the degree of contact between water and lake sediments, i.e. suspension and sedimentation processes.

Lake DIC concentrations were completely unrelated to latitude, suggesting that the factors related to DIC at the regional scale do not follow a latitudinal pattern but rather local geology. There was an interplay between lake area and volume in determining lake DIC concentrations: area having a positive influence while volume has a negative. This indicates that DIC concentrations are generally unaffected by lake morphology, apart from a potential dilutive effect reflected in lake volume (Table 2.3.1.5).

2.3.1.3 The influence of network configuration on the four carbon species

We analyzed the potential influence of network configuration metrics on the four carbon species, and included in each case both latitude and either lake area or dynamic ratio. As previously described, this is to account for the overarching effect of those variables on the carbon species, in order to more clearly perceive the potential effects of the network metrics. As in the case of the morphometric parameters there was a gradient of influence of the network metrics on the four carbon species and the selected metrics differed between the four.

Lake pCH_4 were significantly linked to WRT once latitude and area had been accounted for. WRT was the strongest predictor but number of inlets, total upstream lakes, and total upstream length were all included in the elastic net predictive model (Table 2.3.1.2). This would suggest that pCH_4 is sensitive to local lake conditions that are influenced by both hydrologic connectivity to the watershed as well as effects of network position.

There was barely any relationship between network configuration and lake pCO_2 when latitude and area were accounted for. WRT and the ratio of lakes/upstream length had a weak negative relationship, this suggests that pCO_2 is not strongly affected by the aquatic network configuration, however there may be a slight effect of the lake order in the network (Table 2.3.1.3).

In the case of DOC concentration, we included latitude and dynamic ratio to account for regional and lake effects. The weakly influencing but nonetheless included metrics to the elastic net model are: WRT, number of inlets, and the ratio of upstream length to number of inlets (Table 2.3.1.4). These metrics are reflecting the hydrologic connectivity with the watershed, suggesting that when regarding the network configuration, DOC is mostly affected by water movement within the watershed and also of in lake degradation.

DIC concentrations were inversely related to both WRT and the ratio of upstream length to number of inlets (Table 2.3.1.5). This suggests that lake DIC concentrations are also linked to hydrologic connectivity within the network.

2.3.1.4 Integrative Predictive Models

Following the elastic net procedure, we combined the morphological and network configuration metrics that emerged as significant for each carbon species into a General Linear Regression (GLR) model that further selected significant parameters through a stepwise minimum AIC selection for each carbon species. In all cases we included latitude to account for regional environmental effects. These morphological/network configuration models performed best in predicting lake pCH_4 and moderately for DOC and pCO_2, whereas only a very small fraction of the among lake variability in DIC was explained by any combination of these variables.

The best integrated predictive model for lake pCH_4 explained 60% of the variation and retained the following variables (Figure 2.3.1.2):

Log pCH_4(ppm) = 3.55 - 0.0592 (latitude) + 1.46 (Log-perimeter) + 0.243 (Log-N inlets) + 0.137 (Log-total upstream length) -1.068 (Log-area) -0.0684 (Log-mean depth) -1.48 (Log-Edge Complexity) -0.169 (Log-% littoral) -0.241 (Log-WRT) -0.001 (Log-Total Upstream Lakes)

The best integrated predictive model for lake pCO_2 explained 30% of the variation and retained the following variables (Figure 2.3.1.3):

Log pCO_2 (ppm) = 3.45 + 0.364 (Log-Edge Complexity) + 0.0217 (Log-WRT) -0.0194 (latitude) - 0.127 (Log-area) - 0.0371 (Log-mean depth) - 0.461 (Log-Shape) - 0.0468 (Log-Lakes/upstream length)

The best integrated predictive model for lake DOC concentrations explained 36% of the variation and retained the following variables (Figure 2.3.1.4):

Log DOC concentration (mg/L) = 2.75 + 0.0441 (Log-volume) + 0.196 (Log-%littoral) + 0.0812 (Log-Dynamic Ratio) + 0.0183 (Log-upstream length/inlet) - 0.0447 (latitude) - 0.0408 (Log-WRT) - 0.0604 (Log-N inlets)

The best integrated predictive model for lake DIC concentrations explained 14% of the variation and retained the following variables (Figure 2.3.1.5):

Log DIC concentration (mg/L) = 0.293 + 0.539 (Log-area) - 0.0016 (latitude) - 0.260 (Log-perimeter) - 0.188 (Log-volume) - 0.0404 (Log-WRT) - 0.0428 (Log-upstream length/inlet)

2.3.1.5 Partitioned variables and parameter ranks

The parameters of both morphological and network metrics which were found to contribute to the variation of the C species in the GLR were then analyzed in a decision tree partitioning model. For all species, latitude was ranked first or second in importance, however, it varied at which latitude the division was established and what metrics had more impact on the variation above or below that degree.

Lake area is the first divider for pCH_4 at 0.28 km^2. Following that it is split into further branches of latitude groups. Some of those groups tangent more towards network configuration metrics while the other branches into lake morphological features such as %littoral, WRT, edge complexity and mean depth. However, eventually all combinations of parameters appear, highlighting the sensitivity of pCH_4 to changes in the environment (Table 2.3.1.2).

pCO_2 is first divided by lake area where the very small lakes (<0.09 km^2) are on their own branch while the others are divided into groups of latitude from which the morphological and network metrics begin showing effect: >52.6°: Shape, 52.6 - 48.2°: mean depth, edge complexity and lakes/upstream length (Table 2.3.1.3). This indicates that lake morphology and network configuration have different effects on pCO_2 depending on what type of landscape it is found in.

DOC is first divided by latitude at 54.3°. The lakes in the far north respond mainly to dynamic ratio while the lakes below 54.3° branch out in further groups of volume, dynamic ratio, WRT and upstream length/inlet, depending on which latitude it is on. This suggests that lake DOC concentrations are foremostly affected by the surrounding landscape while the bathymetric shape and stream network configuration regulate the expression of that within each landscape group (Table 2.3.1.4).

The differences in DIC concentrations are first divided by lake area (0.60 km^2). For lakes smaller than that WRT have the most influence and lakes larger than that it depends on the latitude before branching out into smaller groups of WRT, area and volume (Table 2.3.1.5).

Table 2.3.1.1 The parameters that are significantly correlated to Latitude and therefore able to be accounted for through the inclusion of Latitude in the following models. LC stands for landcover.

Parameter	Estimate	Std Error	Prob>|t|
Temp mean	-0.518	0.050	<.0001
Soil carbon	0.038	0.011	0.0004
NPP	-4.465	0.935	<.0001
Evapotranspiration	-0.045	0.018	0.0126
Precipitation (mm/yr)	-0.001	0.001	0.0235
Runoff (mm/yr)	0.001	0.000	<.0001
LC Developed %	15.088	4.567	0.0011
LC Bryoids %	9.179	2.730	0.0009
LC Wetland - Treed %	-4.581	1.806	0.0118
LC Annual Cropland %	6.001	2.780	0.0318
LC Perennial Cropland & Pasture %	-8.677	3.918	0.0277
LC Mixedwood Dense %	-2.824	1.308	0.0318

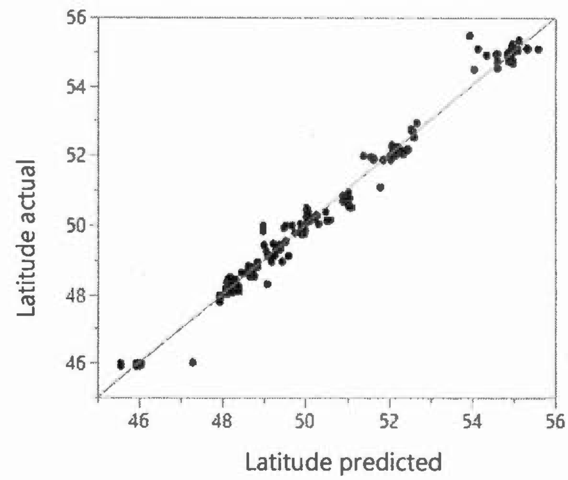


Figure 2.3.1.1 Plot of Actual vs. Predicted values of Latitude from the parameters in the dataset. The line going through is the 1:1 reference.

Table 2.3.1.2 Summary of how much each metric contributes to the variation of pCH_4 from the elastic net analysis and the results of a decision tree partitioning model.

	Elastic Net Estimate	Elastic Net Std Error	Partitioning Rank	Partioning Portion
pCH_4 Morphology				
Latitude	-0.081	0.011	2	0.30
Log10[Lake area (km2)]	-1.514	0.267	1	0.51
Log10[Lake perimeter (km)]	1.597	0.470	10	0.00
Log10[Lake volume (km3)]	0.414	0.297	-	-
Log10[max depth (m)]	0.462	0.458	-	-
Log10[mean depth (m)]	-1.250	0.403	9	0.00
Log10[Edge Complexity]	-1.260	0.507	8	0.00
Log10[Shape]	-0.340	0.410	-	-
Log10[% littoral area (3m depth)]	-0.302	0.171	3	0.10
Log10[Dynamic Ratio]	-0.063	0.404	-	-
pCH_4 Network				
Latitude	-0.066	0.012		
Log10[Lake area (km2)]	-0.295	0.079		
Log10[Water Residence Time (yr)]	-0.209	0.056	5	0.02
Log10[N inlets]	0.175	0.112	11	0.00
Log10[Total Upstream Lakes)]	0.010	0.072	7	0.01
Log10[total upstream length]	0.131	0.080	4	0.05
Log10[Lakes/inlet]	0	0	-	-
Log10[Lakes/upstream length]	0	0	-	-
Log10[upstream length/inlet]	0	0	-	-

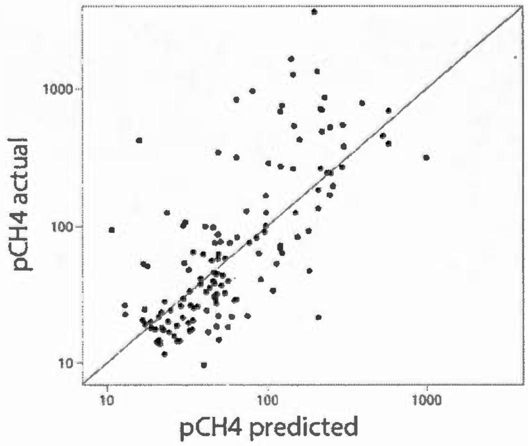


Figure 2.3.1.2 Plot of Predicted vs. Actual of pCH_4 (ppm) from a minimum AIC selection model combining both Morphology and Network metrics. RSq=0.60, RMSE=0.35. The diagonal line is the 1:1 reference.

Table 2.3.1.3 Summary of how much each metric contributes to the variation of pCO_2 from the elastic net analysis and the results of a decision tree partitioning model.

	Elastic Net Estimate	Elastic Net Std Error	Partitioning Rank	Partioning Portion
pCO_2 Morphology				
Latitude	-27.696	7.119	2	0.36
Log10[Lake area (km2)]	-121.306	68.573	1	0.40
Log10[Lake perimeter (km)]	0	0	-	-
Log10[Lake volume (km3)]	0	0	-	-
Log10[max depth (m)]	0	0	-	-
Log10[mean depth (m)]	-161.354	155.556	4	0.10
Log10[Edge Complexity]	494.160	113.276	3	0.11
Log10[Shape]	-306.358	242.444	6	0.01
Log10[% littoral area (3m depth)]	0	0	-	-
Log10[Dynamic Ratio]	-83.336	157.435	-	-
pCO_2 Network				
Latitude	-18.508	10.002		
Log10[Lake area (km2)]	-85.969	43.032		
Log10[Water Residence Time (yr)]	-0.287	40.414	7	0.00
Log10[N inlets]	0	0	-	-
Log10Total Upstream Lakes)]	0	0	-	-
Log10[total upstream length]	0	0	-	-
Log10[Lakes/inlet]	0	0	-	-
Log10[Lakes/upstream length]	-2.646	53.633	5	0.03
Log10[upstream length/inlet]	0	0	-	-

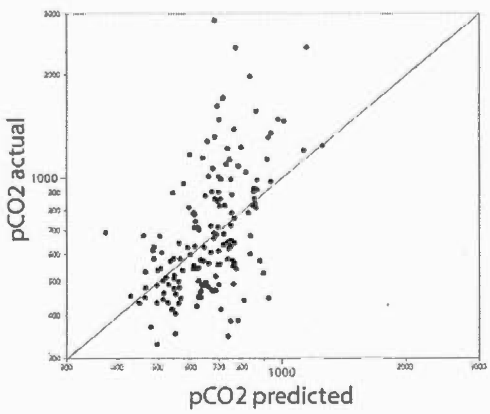


Figure 2.3.1.3 Plot of Predicted vs. Actual of pCO_2 (ppm) from a minimum AIC selection model combining both Morphology and Network metrics. RSq=0.30, RMSE=0.15. The diagonal line is the 1:1 reference.

Table 2.3.1.4 Summary of how much each metric contributes to the variation of DOC from the elastic net analysis and the results of a decision tree partitioning model.

	Elastic Net Estimate	Elastic Net Std Error	Partitioning Rank	Partioning Portion
DOC Morphology				
Latitude	-0.049	0.005	1	0.68
Log10[Lake area (km2)]	0	0	-	-
Log10[Lake perimeter (km)]	0	0	-	-
Log10[Lake volume (km3)]	-0.016	0.014	6	0.03
Log10[max depth (m)]	0	0	-	-
Log10[mean depth (m)]	0	0	-	-
Log10[Edge Complexity]	0	0	-	-
Log10[Shape]	0	0	-	-
Log10[% littoral area (3m depth)]	0.059	0.066	2	0.11
Log10[Dynamic Ratio]	0.112	0.030	3	0.10
DOC Network				
Latitude	-0.060	0.007		
Log10[Dynamic Ratio]	0.134	0.036		
Log10[Water Residence Time (yr)]	-0.026	0.021	4	0.05
Log10[N inlets]	-0.035	0.038	7	0.01
Log10[Total Upstream Lakes)]	0	0	-	-
Log10[total upstream length]	0	0	-	-
Log10[Lakes/inlet]	0	0	-	-
Log10[Lakes/upstream length]	0	0	-	-
Log10[upstream length/inlet]	0.013	0.015	5	0.04

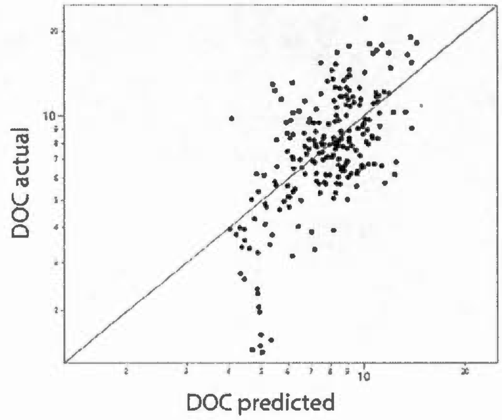


Figure 2.3.1.4 Plot of Predicted vs. Actual of DOC (mg/L) from a minimum AIC selection model combining both Morphology and Network metrics. RSq=0.33, RMSE=0.14. The diagonal line is the 1:1 reference.

Table 2.3.1.5 Summary of how much each metric contributes to the variation of DIC from the elastic net analysis and the results of a decision tree partitioning model.

	Elastic Net Estimate	Elastic Net Std Error	Partitioning Rank	Partioning Portion
DIC Morphology				
Latitude	0	0	1	0.31
Log10[Lake area (km2)]	0.384	0.110	2	0.28
Log10[Lake perimeter (km)]	-0.167	0.150	6	0.00
Log10[Lake volume (km3)]	-0.153	0.058	5	0.08
Log10[max depth (m)]	0	0	-	-
Log10[mean depth (m)]	0	0	-	-
Log10[Edge Complexity]	0	0	-	-
Log10[Shape]	0	0	-	-
Log10[% littoral area (3m depth)]	0	0	-	-
Log10[Dynamic Ratio]	0	0	-	-
DIC Network				
Latitude	-0.023	0.014		
Log10[Lake area (km2)]	0.279	0.048		
Log10[Water Residence Time (yr)]	-0.184	0.055	3	0.21
Log10[N inlets]	0	0	-	-
Log10[Total Upstream Lakes)]	0	0	-	-
Log10[total upstream length]	0	0	-	-
Log10[Lakes/inlet]	0	0	-	-
Log10[Lakes/upstream length]	0	0	-	-
Log10[upstream length/inlet]	-0.133	0.053	4	0.12

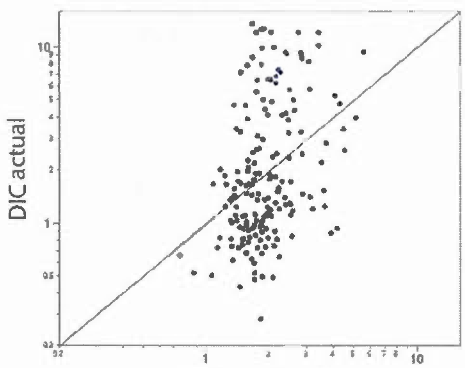


Figure 2.3.1.5 Plot of Predicted vs. Actual of DIC (mg/L) from a minimum AIC selection model combining both Morphology and Network metrics. RSq=0.14, RMSE=0.34. The diagonal line is the 1:1 reference.

2.4 DISCUSSION

Our results show that some C species, notably lake CH_4 and CO_2, are strongly influenced not only by lake area, but also by lake shape, and also by position in the network. Lake morphological and network features significantly may improve the predictions of lake pCH_4, pCO_2, DOC and DIC based on area or environmental parameters alone. No single morphometric or network variable, however, applied across all carbon species, and here we have identified which morphological features contribute to the variation of each one. This means that there is great potential for large scale biogeochemical predictions with morphological features derived from remote sensing data beyond lake surface area. There is still a need for in-situ sampling in regards to obtaining a wider range of data points to calibrate models and algorithms, although relying on sampling alone for the boreal or global estimates of carbon cycling in the landscape is an impossibility.

One of the challenges in this exercise is to identify the influential morphological features in the context of multiple other environmental drivers. Here we have used latitude as an integrative proxy for a large set of environmental, landscape and climatic variables that are known to influence the various C species (Sobek et al. 2007; Larsen et al. 2011), and we have shown that key variables, such as landcover, terrestrial NPP, soil composition and major climatic variables strongly covary with latitude. Three of the C species; CO_2, CH_4 and DOC, had highly significant relationships with latitude, suggesting that this integrative variable captured at least a large portion of the extant environmental drivers. DIC had only a weak relationship with latitude, which reflects the fact that the main environmental drivers of this C species do not necessarily vary along a latitudinal gradient. Our use of latitude as a proxy thus has its limitations due to it being specific for Québec. Since latitude itself is only a coordinate, what is affecting the environment and carbon species are the differences in climate, which in combination with the local geology and anthropogenic activities affect the biogeochemical and biological processes. As such, for the models that we present here to be applied in other regions will require recalibration to the climate and geological conditions of that region. For example, in China using longitude as a proxy will likely prove more effective than using latitude (Liu et al. 2010). It has been suggested that fertilizer use, population density and net primary production could provide a relevant base of division as such in a global context (Seekell et al. 2018a)

The four C species showed different relationships to lake morphology and network configuration metrics both in the strength of the relationships and the metrics that were most strongly linked. Three of the four C species (CO_2, CH_4 and DIC) showed some degree of connection to lake area. Both pCO_2 and pCH_4 declined as a function of lake surface area and size in general, a pattern that has been shown before (Rasilo et al. 2015; Lapierre et al. 2012). This has been related to the extent of connection with high gas production in littoral areas (DelSontro et al. 2017), and also to increased gas exchange and degassing in larger lakes (Bastviken et al. 2004). Interestingly, there was a weak but significant positive relationship between DIC and lake size scaling (Figure 2.3.1. and Table 2.3.1.5), which does not have an obvious explanation, especially since a portion of the DIC is composed of CO_2, which itself has a negative relationship with lake size scaling.

DIC showed the least signs of differences in concentration due to lake morphology or network configuration. A significant fraction of the total DIC pool is composed of carbonates and bicarbonates, which originate from mineral weathering in the bedrock and sediments, and undergo relatively little change within the aquatic network. We did detect a small effect of lake volume, network configuration and WRT, which perhaps reflect some dilution of DIC-rich groundwater in deeper, long residence time lakes.

In contrast, lake pCH_4, appeared to be the C species that was most strongly related to lake morphology, since it linked to virtually every metric we explored. Most of the CH_4 that is found in lakes is locally produced in lake sediments (Rudd & Hamilton 1978), and it is not surprising that aspects of lake shape would influence CH_4 dynamics in lakes. pCH_4 declined as a function of mean depth, a pattern to be expected since that implies that less volumes of water are in contact with the littoral. Surprisingly, pCH_4 also declined as a function of %littoral area and of the dynamic ratio, a counterintuitive result because both are metrics that increase as the relative importance of shallow versus open water pelagic regions decline within lakes. pCH_4 further decreases as a function of declining circularity and increasing edge complexity, again a rather counterintuitive result given that both metrics are also related to the extent of littoral versus open water areas. The littoral areas of lakes are otherwise considered hotspots for pCH_4 because of macrophytes, ebullition from anoxic sediments and decreased oxidation through the water column. This view was challenged by a recent study that in mesocosm experiments still found 90% of methane emissions from a temperate lake originated from oxic lake surface water (Donis et al. 2017). While they did not attempt to in detail explain the "paradoxical process", DelSontro et al. (2017) did, using a physical model where lateral inputs proved to be more influential than previously considered. Therefore, it is likely that the negative relation to the Edge Complexity metric, which was chosen over the Percent Littoral metric, captures more influencing conditions for pCH_4. Since Edge Complexity increases with lake size (see Chapter 1 Figure 1.3.2.1.e), it generally captures lakes that likely originated from moraine damming of a valley or deep scraping into the bedrock by glacial action. This in turn would imply deeper mean depths, less shallow, fine grained sediments and larger volumes of water. However, to accurately upscale pCH_4, the lateral gradients have to be taken into account, in which case a metric like percent littoral may or may not prove to be more useful.

These results highlight the complexity of the sources and regulation of pCH_4 in lakes.

The partial pressures of CO_2 also appeared to be linked to lake morphology, albeit less strongly than pCH_4, once latitudinal and lake-size effects were accounted for. Originating from a multitude of sources, pCO_2 is less likely to show the same pattern across all metrics in all sampled locations. There was therefore few metrics that could significantly explain much of the variation, while there were more parameters that could explain a little.

pCO_2 tended to decline with increasing mean depth, a pattern that is consistent with a major role of lake sediments fueling water column CO_2 supersaturation (Kortelainen et al. 2006). pCO_2 also tended to be higher in oblong lakes (high shape values) and also in lakes with high edge complexity, both of which imply increased terrestrial contact area, i.e. perimeter, per unit of lake area or volume, which highlights the importance of lateral inputs of or for CO_2 originating from terrestrial sources (Sobek et al. 2005; Lapierre et al. 2013). Taken together, the relationships with these morphological metrics reflect the multiple sources of lacustrine pCO_2, both internal and external, reflected in the relationship to lake morphology.

The Network Configuration metrics did not explain much of the variation for either pCO_2 or pCH_4. WRT did explain some of the variation for pCH_4 in a negative relationship, meaning that longer residence times leads to less pCH_4. This is more likely explained by larger lakes having longer residence times than slow flowing water having a negative impact. The weak positive response to the metrics: Number of Inlets, Total Upstream Lakes, and Total Upstream Length, hints at some slight influence of the lake position in the network, although not significantly. The ratio of Lakes per Upstream Length had the same slight but not significant, negative relationship to pCO_2, which could suggest that in some instances the Network Configuration has an impact.

Interpreting dissolved gas partial pressures does in general bring uncertainties regarding the data. What we are aiming to determine, separately from daily or weekly fluctuations is the general patterns associated with the morphological and network features. Morphology and network configurations are in relation to dissolved gases relatively static. Where a low gas concentration in a sample could be due to odd dispersal patterns within the lake which was not captured by a single point sample, or to low production or high outgassing to the atmosphere on that particular day, to name a few. To obtain very exact data one would have to be continuously sampling the lakes from several points, during all seasons. As this is not a likely scenario to obtain data from hundreds of lakes across the boreal, a certain degree of variation in the results can thus be expected. The metrics as well are only proxies for the processes involved, and the actual relation can only be speculated about. In table 1.1.2 in chapter 1, a number of processes associated with each morphological feature is listed. Thus, the relations we present here apply only to the summer period and only to explore the C species larger patterns related to metrics that can be obtained remotely.

DOC was the only species that did not show a clear relationship with lake size but in contrast was strongly positively linked to the dynamic ratio, such that flat, shallower lakes tend to have higher DOC concentrations regardless of their size, once latitudinal effects related to the surrounding watershed soil and vegetation features have been accounted for. This relationship with dynamic ratio may reflect on the balance between sedimentation and suspension in either decreasing or supplying the water column of lakes with DOC. It is also likely that the plate shaped lakes have increased littoral macrophyte growth which have been shown to release large amounts of DOC to the surrounding waters (Demarty and Prairie, 2009). In addition, lakes with flatter profiles are more likely situated in terrain with low slopes which have previously been found to affect the amount and quality of DOC delivered to the surface waters by way of flow paths (Mullholland 2003; Rasmussen et al. 1989; McGlynn & McDonnell 2003).

In chapter 1 we identified a set of generic lake classes based on the ensemble of their morphological features. We further showed that there are regions where these lake types are more prevalent, lake morphological features are thus not randomly distributed across the landscape. In this chapter we show that the concentrations of various C species are linked to some of these lake morphological features. The different lake types should impose a baseline level of C concentrations in the lakes of these regions, solely based on lake morphology and network configurations, separate from the other regional and environmental drivers. Our results suggest that these baseline patterns should be more marked for pCH_4, which showed the strongest links to lake morphological features and the aquatic network structure and secondarily so to pCO_2 and DOC. The link was almost non-existent for DIC, which is overwhelmingly driven by regional geological factors and only a small effect of lake morphology and network configuration. These morphology-driven patterns in C species may account for some of the variation in the concentrations that has not been explained by any of the environmental and climatic factors that have been explored to date, and may further help explain some distinct regional patterns in C species that have been observed across the boreal biome (Lapierre et al. 2015).

2.5 CONCLUSIONS

This study has highlighted that there is definitely an influence of lake morphology and network configurations on carbon.

pCH_4 is strongly influenced by lake morphology and the aquatic network configuration. It would therefore be highly useful to include several of the lake morphological and some of the network metrics in predictive lake pCH_4 models.

The strongest influence to variations in pCO_2 is related to contact area. It is not strongly affected by the aquatic network configuration, however there may be a slight effect of the lake order in the network. The use of the edge complexity metric in addition to lake area improves prediction of pCO_2.

The strongest influence on DOC concentrations, apart from the proxy for environmental factors (latitude), is related to the dynamic ratio. At the network-scale the variation is mostly affected by water movement and watershed drainage. To improve DOC predictions, it is therefore useful to combine the dynamic ratio metric with land cover information.

DIC concentrations are generally unaffected by lake morphology, apart from a dilutive effect from volume and it is somewhat affected by placement in the watershed. Volume and proximity to the source may be the most effective predictors for DIC.

REFERENCES

Akaike, H., 1979. A Bayesian extension of the minimum AIC procedure of autoregressive model fitting. Biometrika, 66(2), pp.237-242.

Algesten, G., Sobek, S., Bergström, A.-K., Ågren, A., Tranvik, L. J. and Jansson, M. (2004), Role of lakes for organic carbon cycling in the boreal zone. Global Change Biology, 10: 141–147. doi:10.1111/j.1365-2486.2003.00721.x

Aufdenkampe, A. K., Mayorga, E., Raymond, P. A., Melack, J. M., Doney, S. C., Alin, S. R., Aalto, R. E. and Yoo, K. (2011), Riverine coupling of biogeochemical cycles between land, oceans, and atmosphere. Frontiers in Ecology and the Environment, 9: 53–60. doi:10.1890/100014

Bade, D.L., Carpenter, S.R., Cole, J.J., Hanson, P.C. and Hesslein, R.H., 2004. Controls of δ13C-DIC in lakes: Geochemistry, lake metabolism, and morphometry. Limnology and Oceanography, 49(4), pp.1160-1172.

Barros, N., Cole, J.J., Tranvik, L.J., Prairie, Y.T., Bastviken, D., Huszar, V.L., Del Giorgio, P. and Roland, F., 2011. Carbon emission from hydroelectric reservoirs linked to reservoir age and latitude. Nature Geoscience, 4(9), pp.593-596.

Bastviken, D., J. Cole, M. Pace, and L. Tranvik (2004), Methane emissions from lakes: Dependence of lake characteristics, two regional assessments, and a global estimate, Global Biogeochem. Cycles, 18, GB4009, doi:10.1029/2004GB002238.

Bastviken, D., L. J. Tranvik, et al. (2011), Freshwater Methane Emissions Offset the Continental Carbon Sink. Science 331(6013).

Battin T., Luyssaert S., Kaplan L., Aufdenkampe A., Richter A., & Tranvik L., (2009), The Boundless Carbon Cycle, Nature Geoscience 2, 598 - 600 (2009) doi:10.1038/ngeo618

Bengtsson, L. and Herschy, R.W., 2012. Encyclopedia of lakes and reservoirs. Monographiae Biologicae.

Bergström, I., Mäkelä, S., Kankaala, P., Kortelainen, P., Methane efflux from littoral vegetation stands of southern boreal lakes: An upscaled regional estimate, Atmospheric Environment, Volume 41, Issue 2, January 2007, Pages 339-351, ISSN 1352-2310, http://dx.doi.org/10.1016/j.atmosenv.2006.08.014

Bradshaw, C.J. and Warkentin, I.G., 2015. Global estimates of boreal forest carbon stocks and flux. Global and Planetary Change, 128, pp.24-30.
Brezonik, P.L., Olmanson, L.G., Finlay, J.C. and Bauer, M.E., 2015. Factors affecting the measurement of CDOM by remote sensing of optically complex inland waters. Remote Sensing of Environment, 157, pp.199-215.

Cael, B.B. and Seekell, D.A., 2016. The size-distribution of Earth's lakes. Scientific reports, 6, p.29633.

Cael, B.B., Heathcote, A.J. and Seekell, D.A., 2017. The volume and mean depth of Earth's lakes. Geophysical Research Letters, 44(1), pp.209-218.

Cheruvelil, K.S., Yuan, S., Webster, K.E., Tan, P.N., Lapierre, J.F., Collins, S.M., Fergus, C., Scott, C.E., Henry, E.N., Soranno, P.A. and Filstrup, C.T., 2017. Creating multithemed ecological regions for macroscale ecology: Testing a flexible, repeatable, and accessible clustering method. Ecology and evolution, 7(9), pp.3046-3058.

Cohen, A.S, 2003. Paleolimnology: The History and Evolution of Lake Systems (pp.21-32). Oxford University Press New York. ISBN: 9780195133530

Cole, J., Y. Prairie, et al. (2007). "Plumbing the Global Carbon Cycle: Integrating Inland Waters into the Terrestrial Carbon Budget." Ecosystems 10(1): 172-185.

Crill, P.M., Bartlett, K.B., Harriss, R.C., Gorham, E., Verry, E.S., Sebacher, D.I., Madzar, L. and Sanner, W., 1988. Methane flux from Minnesota peatlands. Global Biogeochemical Cycles, 2(4), pp.371-384.

Cutler, D.R., Edwards Jr, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J. and Lawler, J.J., 2007. Random forests for classification in ecology. Ecology, 88(11), pp.2783-2792.

Del Giorgio, P.A., Cole, J.J., Caraco, N.F. and Peters, R.H., 1999. Linking planktonic biomass and metabolism to net gas fluxes in northern temperate lakes. Ecology, 80(4), pp.1422-1431.

DelSontro, T., del Giorgio, P.A. and Prairie, Y.T., 2017. No Longer a Paradox: The Interaction Between Physical Transport and Biological Processes Explains the Spatial Distribution of Surface Water Methane Within and Across Lakes. Ecosystems, pp.1-15.

Demarty, M. and Prairie, Y.T., 2009. In situ dissolved organic carbon (DOC) release by submerged macrophyte–epiphyte communities in southern Québec lakes. Canadian journal of fisheries and aquatic sciences, 66(9), pp.1522-1531.

Dodds, W.K. and Whiles, M.R., 2010. Freshwater Ecology: Concepts and Environmental Applications of Limnology. Academic Press.

Dodds, W. K. (2002). Freshwater ecology: concepts and environmental applications. Academic press.

Donis, D., Flury, S., Spangenberg, J.E., Vachon, D. and McGinnis, D.F., 2017. Full-scale evaluation of methane production under oxic conditions in a mesotrophic lake., Nature Communications 8: 1661.

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, J.M. and Middelburg, J.J., 2006. The global abundance and size distribution of lakes, ponds, and impoundments. Limnology and Oceanography, 51(5), pp.2388-2397.

Duarte, C. M. and Y. T. Prairie (2005). "Prevalence of heterotrophy and atmospheric CO2 emissions from aquatic ecosystems." Ecosystems 8(7): 862-870.

Feng, M., J. O. Sexton, S. Channan, and J. R. Townshend. 2016. A global, high-resolution (30-m) inland water body dataset for 2000: first results of a topographic–spectral classification algorithm. International Journal of Digital Earth 9: 113–133.

Fergus, C.E., Lapierre, J.F., Oliver, S.K., Skaff, N.K., Cheruvelil, K.S., Webster, K., Scott, C. and Soranno, P., 2017. The freshwater landscape: lake, wetland, and stream abundance and connectivity at macroscales. Ecosphere, 8(8).

Ferland, M.E., 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 Biogeochemical Cycles, 26(4).

Finlay, K., Leavitt, P.R., Wissel, B. and Prairie, Y.T., 2009. Regulation of spatial and temporal variability of carbon flux in six hard-water lakes of the northern Great Plains. Limnology and Oceanography, 54(6), p.2553.

Fulton, R J (ed.), 1989. Quaternary Geology of Canada and Greenland Geological Survey of Canada, Geology of Canada Series no. 1, 1989, ; 839 pages (5 sheets), doi:10.4095/127905

Hall, E. K., Schoolmaster, D., Amado, A. M., Stets, E., Lennon, J. T., Domaine, L., & Cotner, J. B. (2016). Scaling relationships among drivers of aquatic respiration from the smallest to the largest freshwater ecosystems.Inland Waters, 6(1), 1-10.

Hartigan, J.A. and Wong, M.A., 1979. Algorithm AS 136: A k-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics), 28(1), pp.100-108.

Hayden, B., Harrod, C. and Kahilainen, K.K., 2014. Lake morphometry and resource polymorphism determine niche segregation between cool-and cold-water-adapted fish. Ecology, 95(2), pp.538-552.

Heathcote, A.J., del Giorgio, P.A. and Prairie, Y.T., 2015. Predicting bathymetric features of lakes from the topography of their surrounding landscape. Canadian Journal of Fisheries and Aquatic Sciences, 72(5), pp.643-650.

Hershey, A.E., Beaty, S., Fortino, K., Keyse, M., Mou, P.P., O'BRIEN, W.J., Ulseth, A.J., Gettel, G.A., Lienesch, P.W., Luecke, C. and McDonald, M.E., 2006. Effect of landscape factors on fish distribution in arctic Alaskan lakes. Freshwater Biology, 51(1), pp.39-55.

Holgerson, M.A. and Raymond, P.A., 2016. Large contribution to inland water CO2 and CH4 emissions from very small ponds. Nature Geoscience.

Humborg, C., Mörth, C., Sundbom, M., Borg, H., Blenckner, T., Giesler, R. and Ittekkot, V., 2010. CO2 supersaturation along the aquatic conduit in Swedish watersheds as constrained by terrestrial respiration, aquatic respiration and weathering. Global Change Biology, 16(7), pp.1966-1978.

Hutchinson, G. E., 1957. A Treatise on Limnology. London: Wiley. Geography, Physics and Chemistry, Vol. 2.

Håkanson, L., 1982. Lake bottom dynamics and morphometry: the dynamic ratio. Water Resources Research, 18(5), pp.1444-1450.

Håkanson, L. and Karlsson, B., 1984. On the relationship between regional geomorphology and lake morphometry. A Swedish example. Geografiska Annaler. Series A. Physical Geography, pp.103-119.

Håkanson, L., 2005. The importance of lake morphometry for the structure and function of lakes. International Review of Hydrobiology, 90(4), pp.433-461.

Håkanson, L., 2012. Origin of Lakes and Their Physical Characteristics. In Encyclopedia of Lakes and Reservoirs (pp. 585-591). Springer Netherlands.

IPCC, 2007: Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate

Change [Core Writing Team, Pachauri, R.K and Reisinger, A. (eds.)]. IPCC, Geneva, Switzerland, 104 pp

James, N.P. and Jones, B., 2015., Origin of Carbonate Rocks., John Wiley & Sons.

Jansson, M., Bergström, A.K., Blomqvist, P. and Drakare, S., 2000. Allochthonous organic carbon and phytoplankton/bacterioplankton production relationships in lakes. Ecology, 81(11), pp.3250-3255.

Jonsson, A., Meili, M., Bergström, A.K. and Jansson, M., 2001. Whole-lake mineralization of allochthonous and autochthonous organic carbon in a large humic lake (Örträsket, N. Sweden). Limnology and Oceanography, 46(7), pp.1691-1700.

Kankaala, P., Huotari, J., Tulonen, T. and Ojala, A., 2013. Lake-size dependent physical forcing drives carbon dioxide and methane effluxes from lakes in a boreal landscape. Limnology and Oceanography, 58(6), pp.1915-1930.

Kaplan, L. A. & Newbold, J. D (eds Findlay, S. E. G. & Sinsabaugh, R. L.). 2003. Interactivity of Dissolved Organic Matter (Aquatic ecosystems, Academic Press, Massachusetts.

Koehler, B., Landelius, T., Weyhenmeyer, G.A., Machida, N. and Tranvik, L.J., 2014. Sunlight-induced carbon dioxide emissions from inland waters. Global Biogeochemical Cycles, 28(7), pp.696-711.

Kortelainen, P., 1993. Content of total organic carbon in Finnish lakes and its relationship to catchment characteristics. Canadian Journal of Fisheries and Aquatic Sciences, 50(7), pp.1477-1483.

Kortelainen, P., Rantakari, M., Huttunen, J.T., Mattsson, T., Alm, J., Juutinen, S., Larmola, T., Silvola, J. and Martikainen, P.J., 2006. Sediment respiration and lake trophic state are important predictors of large CO2 evasion from small boreal lakes. Global Change Biology, 12(8), pp.1554-1567.

Kraemer, B.M., Anneville, O., Chandra, S., Dix, M., Kuusisto, E., Livingstone, D.M., Rimmer, A., Schladow, S.G., Silow, E., Sitoki, L.M. and Tamatamah, R., 2015. Morphometry and average temperature affect lake stratification responses to climate change. Geophysical Research Letters, 42(12), pp.4981-4988.

Kratz, T., Webster, K., Bowser, C., Maguson, J. and Benson, B., 1997. The influence of landscape position on lakes in northern Wisconsin. Freshwater Biology, 37(1), pp.209-217.

Larsen, S., Andersen, T. and Hessen, D.O., 2011. Predicting organic carbon in lakes from climate drivers and catchment properties. Global Biogeochemical Cycles, 25(3).

Lapierre, J.F. and Giorgio, P.A., 2012. Geographical and environmental drivers of regional differences in the lake pCO_2 versus DOC relationship across northern landscapes. Journal of Geophysical Research: Biogeosciences, 117(G3).

Lapierre, J.F., Guillemette, F., Berggren, M. and Del Giorgio, P.A., 2013. Increases in terrestrially derived carbon stimulate organic carbon processing and CO_2 emissions in boreal aquatic ecosystems. Nature communications, 4, p.2972.

Lapierre, J.F., Seekell, D.A. and Giorgio, P.A., 2015. Climate and landscape influence on indicators of lake carbon cycling through spatial patterns in dissolved organic carbon. Global change biology, 21(12), pp.4425-4435.

Larsen, S., Andersen, T. and Hessen, D.O., 2011. Predicting organic carbon in lakes from climate drivers and catchment properties. Global Biogeochemical Cycles, 25(3).

Lauerwald, R., Laruelle, G.G., Hartmann, J., Ciais, P. and Regnier, P.A., 2015. Spatial patterns in CO_2 evasion from the global river network. Global Biogeochemical Cycles, 29(5), pp.534-554.

Lehner, B., Grill G. (2013): Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. Hydrological Processes, 27(15): 2171–2186. Data is available at www.hydrosheds.org.

Liu, W., Zhang, Q. and Liu, G., 2010. Lake eutrophication associated with geographic location, lake morphology and climate in China. Hydrobiologia, 644(1), pp.289-299.

Mandelbrot, B., 1967. How long is the coast of Britain? Statistical self-similarity and fractional dimension. science, 156(3775), pp.636-638.

Marotta, H., Duarte, C.M., Sobek, S. and Enrich-Prast, A., 2009. Large CO_2 disequilibria in tropical lakes. Global biogeochemical cycles, 23(4).

Martin Sherry L. , Soranno Patricia A. , (2006), Lake landscape position: Relationships to hydrologic connectivity and landscape features, Limnology and Oceanography, 51, doi: 10.4319/lo.2006.51.2.0801.

McGlynn, B.L. and McDonnell, J.J., 2003. Role of discrete landscape units in controlling catchment dissolved organic carbon dynamics. Water Resources Research, 39(4).

Messager, M.L., Lehner, B., Grill, G., Nedeva, I. and Schmitt, O., 2016. Estimating the volume and age of water stored in global lakes using a geo-statistical approach. Nature communications, 7, p.13603.

Meybeck, M., 1995. Physics and chemistry of lakes. Springer Berlin Heidelberg.

Molot, L.A. and Dillon, P.J., 1997. Photolytic regulation of dissolved organic carbon in northern lakes. Global Biogeochemical Cycles, 11(3), pp.357-365.

Mosquera, P.V., Hampel, H., Vázquez, R.F., Alonso, M. and Catalan, J., 2017. Abundance and morphometry changes across the high mountain lake-size gradient in the tropical Andes of Southern Ecuador. Water Resources Research.

Mulholland, P.J., 2003. Large-scale patterns in dissolved organic carbon concentration, flux, and sources. Aquatic ecosystems: interactivity of dissolved organic matter, pp.139-159.

Noges, T., 2009. Relationships between morphometry, geographic location and water quality parameters of European lakes. Hydrobiologia, 633(1), pp.33-43.

Oliver, S.K., Soranno, P.A., Fergus, C.E., Wagner, T., Winslow, L.A., Scott, C.E., Webster, K.E., Downing, J.A. and Stanley, E.H., 2016. Prediction of lake depth across a 17-state region in the United States. Inland Waters, 6(3), pp.314-324.

O'Reilly, C. M., R. J. Rowley, P. Schneider, J. D. Lenters, P. B. Mcintyre, B. M. Kraemer, G. A. Weyhenmeyer, D. Straile, B. Dong, R. Adrian, M. Allan, O. Anneville, L. Arvola, J. Austin, J. Bailey, J. Baron, J. Brookes, E. de Eyto, M. Dokulil, D. Hamilton, K. Havens, A. Hetherington, S. Higgins, S. Hook, L. Izmest'eva, K. Joehnk, K. Kangur, P. Kasprzak, M. Kumagai, E. Kuusisto, G. Leshkevich, D. Livingstone, S. MacIntyre, L. May, J. Melack, D. Mueller-Navarra, M. Naumenko, P. Noges, T. Noges, R. North, P. Plisnier, A. Rigosi, A. Rimmer, M. Rogora, L. Rudstam, J. Rusak, N. Salmaso, N. Samal, D. Schindler, S. Schladow, M. Schmid, S. Scmidt, E. Silow, M. Soylu, K. Teubner, P. Verburg, A. Voutilainen, A. Watkinson, C. Williamson, and G. Zhang. 2015. Rapid and highly variable warming of lake surface waters around the globe. Geophysical Research Letters **42**: 10773–10781.

Osman, K.T., 2013., Forest Soils: Properties and Management., Springer Science and Business Media.

Pinho, L., Duarte, C.M., Marotta, H. and Enrich-Prast, A., 2016. Temperature dependence of the relationship between pCO2 and dissolved organic carbon in lakes. Biogeosciences, 13(3), pp.865-871.

Post, D.M., Pace, M.L. and Hairston, N.G., 2000. Ecosystem size determines food-chain length in lakes. Nature, 405(6790), pp.1047-1049.

Prepas E, D Planas, J J Gibson, D H Vitt, T D Prowse, W P Dinsmore, L A Halsey, P M McEachern, S Paquet, G J Scrimgeour, W M Tonn, C A Paszkowski, K Wolfstein, (2001) "Landscape variables influencing nutrients and phytoplankton communities in boreal plain lakes of northern alberta: a comparison of wetland and upland-dominated catchments", Canadian Journal of Fisheries and Aquatic Sciences, 2001, 58:1286-1299, 10.1139/f01-081

Quinlan, R., Paterson, A. M., Hall, R. I., Dillon, P. J., Wilkinson, A. N., Cumming, B. F., Douglas, M. S. V. and Smol, J. P. (2003), A landscape approach to examining spatial patterns of limnological variables and long-term environmental change in a southern Canadian lake district. Freshwater Biology, 48: 1676–1697. doi:10.1046/j.1365-2427.2003.01105.x

Rasilo, T., Prairie, Y.T. and Giorgio, P.A., 2015. Large-scale patterns in summer diffusive CH4 fluxes across boreal lakes, and contribution to diffusive C emissions. Global change biology, 21(3), pp.1124-1139.

Rasmussen Joseph B. , Godbout Lyse , Schallenberg Marc , (1989), The humic content of lake water and its relationship to watershed and lake morphometry, Limnology and Oceanography, 34, doi: 10.4319/lo.1989.34.7.1336.

Raymond, P.A., Hartmann, J., Lauerwald, R., Sobek, S., McDonald, C., Hoover, M., Butman, D., Striegl, R., Mayorga, E., Humborg, C. and Kortelainen, P., 2013. Global carbon dioxide emissions from inland waters. Nature, 503(7476), pp.355-359.

Rooney, N. and Kalff, J., 2000. Inter-annual variation in submerged macrophyte community biomass and distribution: the influence of temperature and lake morphometry. Aquatic Botany, 68(4), pp.321-335.

Rudd, J.W. and Hamilton, R.D., 1978. Methane cycling in a eutrophic shield lake and its effects on whole lake metabolism. Limnol. Oceanogr, 23(2), pp.337-348.

Sadro, S., Nelson, C.E. and Melack, J.M., 2012. The influence of landscape position and catchment characteristics on aquatic biogeochemistry in high-elevation lake-chains. Ecosystems, 15(3), pp.363-386.

SAS Institute Inc. 2017. JMP® 13. Predictive and Specialized Modeling, Second Edition. Cary, NC: SAS Institute Inc.

Sawakuchi, H.O., Bastviken, D., Sawakuchi, A.O., Krusche, A.V., Ballester, M.V. and Richey, J.E., 2014. Methane emissions from Amazonian Rivers and their contribution to the global methane budget. Global change biology, 20(9), pp.2829-2840.

Schilder, J., D. Bastviken, M. van Hardenbroek, P. Kankaala, P. Rinta, T. Stötter, and O. Heiri (2013), Spatial heterogeneity and lake morphology affect diffusive greenhouse gas emission estimates of lakes, Geophys. Res. Lett., 40, 5752–5756, doi:10.1002/2013GL057669.

Seekell, D. A.,M. L. Pace, L. J. Tranvik, and C. Verpoorter (2013), A fractal-based approach to lake size-distributions, Geophys. Res. Lett., 40, 517–521, doi:10.1002/grl.50139.

Seekell, D. A., J. A. Carr, C. Gudasz, and J. Karlsson (2014), Upscaling carbon dioxide emissions from lakes, Geophys. Res. Lett., 41, 7555–7559, doi:10.1002/2014GL061824.

Seekell, D.A., Lapierre, J.F. and Cheruvelil, K.S., 2018a. A geography of lake carbon cycling. Limnology and Oceanography Letters, 3(3), pp.49-56.

Seekell, D.A., Byström, P. and Karlsson, J., 2018b. Lake morphometry moderates the relationship between water color and fish biomass in small boreal lakes. Limnology and Oceanography.

Sobek, S., Algesten, G., Bergström, A.-K., Jansson, M. and Tranvik, L. J. (2003), The catchment and climate regulation of pCO_2 in boreal lakes. Global Change Biology, 9: 630–641. doi:10.1046/j.1365-2486.2003.00619.x

Sobek, S., Tranvik, L.J. and Cole, J.J., 2005. Temperature independence of carbon dioxide supersaturation in global lakes. Global Biogeochemical Cycles, 19(2).

Sobek Sebastian , Tranvik Lars J. , Prairie Yves T. , Kortelainen Pirkko , Cole Jonathan J. , (2007), Patterns and regulation of dissolved organic carbon: An analysis of 7,500 widely distributed lakes, Limnology and Oceanography,52, doi: 10.4319/lo.2007.52.3.1208.

Sobek, S., Nisell, J. and Fölster, J., 2011. Predicting the depth and volume of lakes from map-derived parameters. Inland Waters, 1(3), pp.177-184.

Soranno, P., Webster, K., Riera, J. et al. Ecosystems (1999) Spatial variation among lakes within landscapes: ecological organization along lake chains 2: 395. doi:10.1007/s100219900089

Soranno Patricia A., Webster Katherine E., Cheruvelil Kendra S. and Bremigan, Mary T. (2009) "The lake landscape-context framework: linking aquatic connections, terrestrial features and human effects at multiple spatial scales", Verh. Internat. Verein. Limnol. 2009, vol. 30, Part 5, p. 695–700

Staehr, P.A., Baastrup-Spohr, L., Sand-Jensen, K. and Stedmon, C., 2012. Lake metabolism scales with lake morphometry and catchment conditions. Aquatic Sciences, 74(1), pp.155-169.

Steele, M.K. and Heffernan, J.B., 2014. Morphological characteristics of urban water bodies: mechanisms of change and implications for ecosystem function. Ecological Applications, 24(5), pp.1070-1084.

Stets, E.G., Striegl, R.G., Aiken, G.R., Rosenberry, D.O. and Winter, T.C., 2009. Hydrologic support of carbon dioxide flux revealed by whole-lake carbon budgets. Journal of geophysical research: Biogeosciences, 114(G1).

Striegl, R.G. and Michmerhuizen, C.M., 1998. Hydrologic influence on methane and carbon dioxide dynamics at two north-central Minnesota lakes. Limnology and Oceanography, 43(7), pp.1519-1529.

Teodoru, C.R., Del Giorgio, P.A., Prairie, Y.T. and Camire, M., 2009. Patterns in pCO2 in boreal streams and rivers of northern Québec, Canada. Global Biogeochemical Cycles, 23(2).

Tranvik Lars J., Downing John A., Cotner James B., Loiselle Steven A., Striegl Robert G., Ballatore Thomas J., Dillon Peter, Finlay Kerri, Fortino Kenneth, Knoll Lesley B., Kortelainen Pirkko L., Kutser Tiit, Larsen Soren., Laurion Isabelle, Leech Dina M., McCallister S. Leigh, McKnight Diane M., Melack John M., Overholt Erin, Porter Jason A.,Prairie Yves, Renwick William H., Roland Fabio, Sherman Bradford S., Schindler David W., Sobek Sebastian,Tremblay Alain, Vanni Michael J., Verschoor Antonie M., von Wachenfeldt Eddie, Weyhenmeyer Gesa A., (2009),Lakes and reservoirs as regulators of carbon cycling and climate, Limnology and Oceanography, 54, doi: 10.4319/lo.2009.54.6_part_2.2298.

Turner, M.G., Gardner, R.H. and O'neill, R.V., 2001. Landscape ecology in theory and practice (Vol. 401). New York: Springer.

Vachon, D. and Prairie, Y.T., 2013. The ecosystem size and shape dependence of gas transfer velocity versus wind speed relationships in lakes. Canadian Journal of Fisheries and Aquatic Sciences, 70(12), pp.1757-1764.

Verpoorter, C., T. Kutser, D. A. Seekell, and L. J. Tranvik (2014), A global inventory of lakes based on high-resolution satellite imagery, Geophys. Res. Lett., 41, 6396–6402, doi:10.1002/2014GL060641.

von Wachenfeldt, E. and Tranvik, L.J., 2008. Sedimentation in boreal lakes—the role of flocculation of allochthonous dissolved organic matter in the water column. Ecosystems, 11(5), pp.803-814.

Wetzel, R.G., 2001. Limnology: lake and river ecosystems. Gulf Professional Publishing.

Wiens, J.A., 2002. Riverine landscapes: taking landscape ecology into the water. Freshwater biology, 47(4), pp.501-515.

Wik, M., Thornton, B.F., Bastviken, D., MacIntyre, S., Varner, R.K. and Crill, P.M., 2014. Energy input is primary controller of methane bubbling in subarctic lakes. Geophysical Research Letters, 41(2), pp.555-560.

Wik, M., Varner, R.K., Anthony, K.W., MacIntyre, S. and Bastviken, D., 2016. Climate-sensitive northern lakes and ponds are critical components of methane release. Nature Geoscience, 9(2), pp.99-105.

Winslow, L. A., Read, J. S., Hanson, P. C. and Stanley, E. H. (2014), Lake shoreline in the contiguous United States: quantity, distribution and sensitivity to observation resolution. Freshw Biol, 59: 213–223. doi:10.1111/fwb.12258

Wolock, D.M., Winter, T.C. and McMahon, G., 2004. Delineation and evaluation of hydrologic-landscape regions in the United States using geographic information system tools and multivariate statistical analyses. Environmental management, 34(1), pp.S71-S88.

Yamada, I., 2016. Thiessen Polygons. The International Encyclopedia of Geography. John Wiley & Sons, Ltd.

Yavitt, J.B., Lang, G.E. and Sexstone, A.J., 1990. Methane fluxes in wetland and forest soils, beaver ponds, and low-order streams of a temperate forest ecosystem. Journal of Geophysical Research: Atmospheres, 95(D13), pp.22463-22474.

Zehnder, A.J. and Brock, T.D., 1979. Methane formation and methane oxidation by methanogenic bacteria. Journal of Bacteriology, 137(1), pp.420-432.

Zhang, T., Soranno, P.A., Cheruvelil, K.S., Kramer, D.B., Bremigan, M.T. and Ligmann-Zielinska, A., 2012. Evaluating the effects of upstream lakes and wetlands on lake phosphorus concentrations using a spatially-explicit model. Landscape ecology, 27(7), pp.1015-1030.

Zou, H. and Hastie, T., 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), pp.301-320.

Data & Software References

ESRI 2016. ArcGIS Desktop: Release 10. Redlands, CA, USA: Environmental Systems Research Institute.
www.esri.com

ESRI 2017. World Imagery (Map Server) Image Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community
https://services.arcgisonline.com/ArcGIS/rest/services/World_Imagery/MapServer

GEOGRATIS
Contains information licensed under the Open Government License – Canada.
http://geogratis.gc.ca/

GeoBase®, 2010. National hydro Network, Product specifications/Distribution Profile – Edition 1.1. Department of Natural Resources Canada, Earth Sciences Sector, Geomatics Canada, Centre for Topographic Information. http://geogratis.gc.ca/

HydroBASINS
http://www.hydrosheds.org

JMP 2017. JMP Pro : Release 13. Cary, NC, USA. SAS Institute Inc.
www.jmp.com

APPENDIX

I. GIS FLOWCHARTS

Metrics calculation

a)

b)

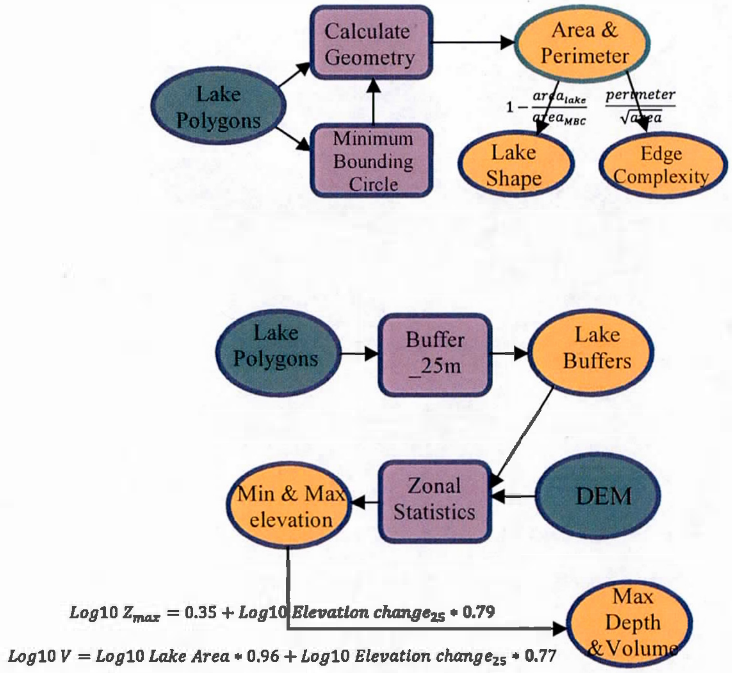


Figure A1.1 Flowchart for calculation of metrics from lake polygons. Where a) shows the steps for area, perimeter, lake shape and edge complexity. b) shows the steps for maximum depth and volume. Blue circles are pre-existing spatial data, purple squares are processing tools and red circles are resulting data.

Classification Data Processing

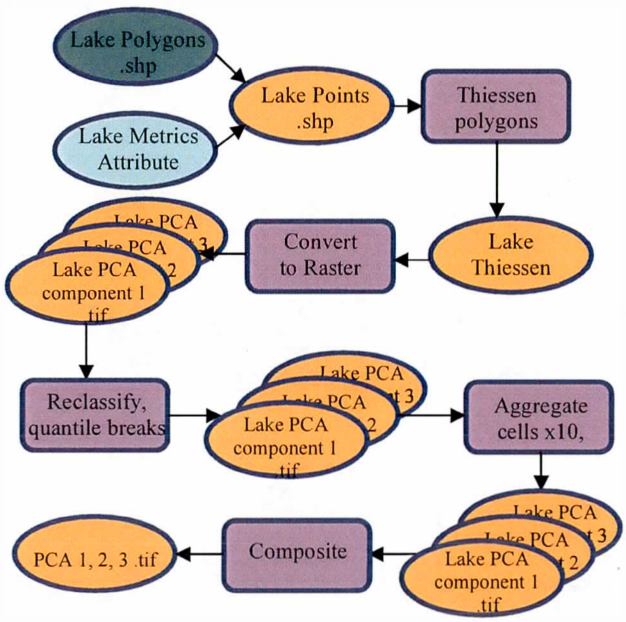


Figure A1.2 Flowchart for producing the data to be used for image classification. Beginning with the lake data shapefile and resulting in the three PCA axis layers. Blue circles are pre-existing spatial data, purple squares are processing tools and red circles are resulting data.

Image Classification

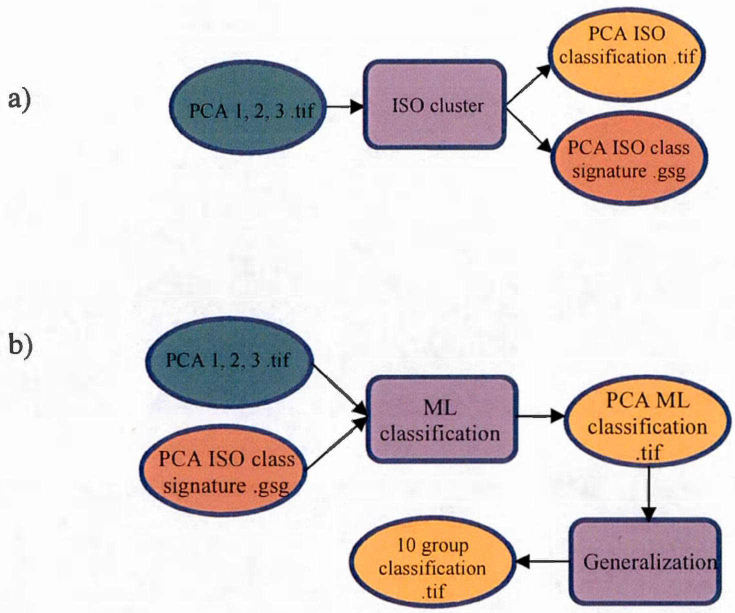


Figure A1.3 Flowchart for the image classification. Where a) produces an ISO classification signature file from the combined PCA layers and b) shows the ML classification which produces the final lake region classification. Blue circles are pre-existing spatial data, purple squares are processing tools and red circles are resulting data.

II. Example images of lake classes

Figure A2.1 Lake class 1. (ESRI 2017)

Figure A2.2 Lake class 2. (ESRI 2017)

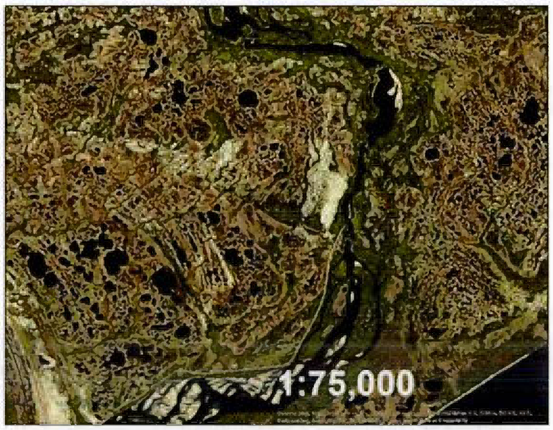


Figure A2.3 Lake class 3. (ESRI 2017)

Figure A2.4 Lake class 4. (ESRI 2017)

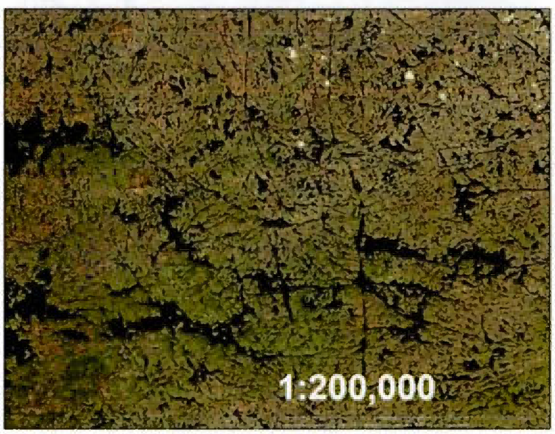


Figure A2.5 Lake class 5. (ESRI 2017)

Figure A2.6 Lake class 6. (ESRI 2017)

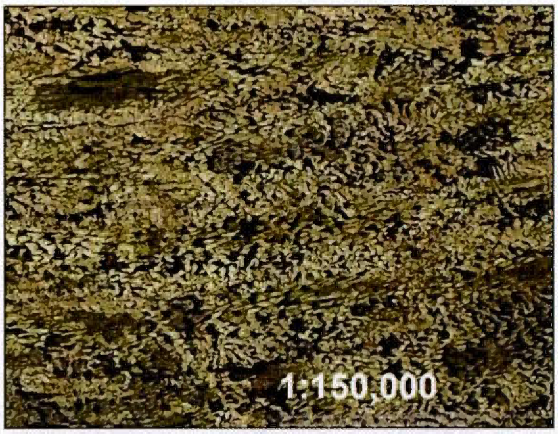


Figure A2.7 Lake class 7. (ESRI 2017)

Figure A2.8 Lake class 8. (ESRI 2017)

Figure A2.9 Lake class 9. (ESRI 2017)

Figure A2.10 Lake class 10. (ESRI 2017)