Variation of Lake Water Chemistry and Spatial Scale

Licentiate Thesis

by

Elisabet Göransson

Department of Environmental Assessment
Swedish University of Agricultural Sciences
Box 7050 SE-750 07 Uppsala

2001 Rapport 2001:22

Variation of Lake Water Chemistry and Spatial Scale

Elisabet Göransson

Akademisk avhandling som för vinnande av licentiatexamen kommer att offentligen försvaras i Hörsalen, Loftet, SLU, Ultuna, torsdagen den 20 december 2001, kl. 13.15.

Abstract

Lake chemistry generally reflects the ecological, geological, and climatic processes in and around the lake. This thesis addresses the importance of scale, from small (variability within a lake) to large (catchment and ecoregion variability), on selected chemical metrics. In *Paper I,* the importance of ecoregions and catchment characteristics for prediction of pH and total phosphorus (TP) were studied. Predictive models were evaluated both on ecoregional (six ecoregions) and whole-country scales. The models were based on catchment-specific variables alone or using a combination of catchment-specific and chemical variables. We found that grouping lakes by ecoregion worked well when within-ecoregion variance was high. For example, partitioning spatial variance by ecoregions worked better for the ecoregions in the south of Sweden than in the north. However, TP and pH were best predicted by other chemical variables, and many of the "better" predictor variables were known to depend on characteristics of the catchment. For example calcium concentration was a good predictor for pH. Future studies should focus on determining the linkage between catchment-scale variability (e.g. geology and weathering) and in-lake chemistry. *Paper II* addressed the question of how representative a mid-lake sample is of in-lake conditions. This information is important for interpreting the long term changes within lake monitoring programs. The surface water chemistry of 34 Swedish lakes, spatially-stratified across three main ecoregions and according to residence time and mean depth, was studied. This study also included two seasons; namely a period when biological activity is expected to be high (late summer or August) and a period when hydrological (i.e. wind-induced) disturbance is expected to be high (autumn or October). Where a sample was taken within a lake was found to affect the sample's representativity of lake water surface chemistry. However, a mid-lake sample was shown to be more representative than a random sample. We conclude that mid-lake samples can be considered representative of average surface water chemistry. This information is crucial for how we interpret data taken from monitoring programs that are temporally extensive.

Distribution:
Swedish University of Agricultural Sciences
Dep. of Environmental Assessment
Box 7050, 750 07 Uppsala, Sweden.

Uppsala 2001
ISSN 1403-977X.
Rapport 2001:22

Variation of Lake Water Chemistry and Spatial Scale

Elisabet Göransson
Department of Environmental Assessment
Uppsala

ISSN 1403-977X

Licentiate Thesis
Uppsala 2001
Swedish University of Agricultural Sciences

Abstract

Göransson, E., 2001. *Variation of Lake Water Chemistry and Spatial Scale.*
Licentiate Thesis, Uppsala.
ISSN 1403-977X. Rapport 2001:22

Lake chemistry generally reflects the ecological, geological, and climatic processes in and around the lake. This thesis addresses the importance of scale, from small (variability within a lake) to large (catchment and ecoregion variability), on selected chemical metrics. In *Paper I,* the importance of ecoregions and catchment characteristics for prediction of pH and total phosphorus (TP) were studied. Predictive models were evaluated both on ecoregional (six ecoregions) and whole-country scales. The models were based on catchment-specific variables alone or using a combination of catchment-specific and chemical variables. We found that grouping lakes by ecoregion worked well when within-ecoregion variance was high. For example, partitioning spatial variance by ecoregions worked better for the ecoregions in the south of Sweden than in the north. However, TP and pH were best predicted by other chemical variables, and many of the "better" predictor variables were known to depend on characteristics of the catchment. For example calcium concentration was a good predictor for pH. Future studies should focus on determining the linkage between catchment-scale variability (e.g. geology and weathering) and in-lake chemistry. *Paper II* addressed the question of how representative a mid-lake sample is of in-lake conditions. This information is important for interpreting the long term changes within lake monitoring programs. The surface water chemistry of 34 Swedish lakes, spatially-stratified across three main ecoregions and according to residence time and mean depth, was studied. This study also included two seasons; namely a period when biological activity is expected to be high (late summer or August) and a period when hydrological (i.e. wind-induced) disturbance is expected to be high (autumn or October). Where a sample was taken within a lake was found to affect the sample's representativity of lake water surface chemistry. However, a mid-lake sample was shown to be more representative than a random sample. We conclude that mid-lake samples can be considered representative of average surface water chemistry. This information is crucial for how we interpret data taken from monitoring programs that are temporally extensive.

Author's address: Department of Environmental Assessment, Swedish University of Agricultural Sciences, P.O. Box 7050, SE-750 07 Uppsala, Sweden.
E-mail: Elisabet.Goransson@ma.slu.se

"A lake is the landscape's most beautiful and expressive feature.

It is earth's eye; looking into which the beholder measures the

depth of his own nature. The fluviatile trees next the shore are

the slender eyelashes which fringe it, and the wooden hills

and cliffs around are its overhanging brows."

(From "Walden or life in the woods"
by Henry David Thoreau, 1854)

A lake being the earth's eye – reflecting the landscape that surrounds it – is somewhat what this thesis is about.

Contents

Appendix

Papers I and II

The present thesis is based on the following papers which are referred to by their Roman numerals.

I. Göransson, E., and Johnson, R. K., 2001. The importance of ecoregions and catchment characteristics for prediction of pH and total phosphorus in lakes. Manuscript

II. Göransson, E., Johnson, R. K., and Wilander, A., 2001. Representativity of a mid-lake surface water chemistry sample. Manuscript

Background

Integrating pattern and scale is a central theme in ecology (Levin, 1992) and, not surprisingly, ecoregional ecology and landscape ecology are two relatively new areas that focus on the interaction between regional patterns and ecological processes at different scales (Turner, 1998; Ormerod, 2000). The importance of a lake's position in the landscape, and especially underlying geology and land use/cover, has provided insight into differences among aquatic systems (e.g. Magnuson and Kratz, 2000). For example, the lake trophic classification schemes developed by Thienemann (1925) and Naumann (1932) are examples of the importance of interactions between terrestrial and aquatic systems.

The chemistry of a lake generally reflects the ecological, geological, and climatic processes in and around the lake. Moreover, a number of processes influence lake water chemistry at different scales; from small-scale, within-lake variability reflecting in situ processes to large-scale, among-lake variation being primarily determined by catchment geology and hydrology. In particular, geographic position, catchment-scale variability in geology, land use, and vegetation cover have provided insight into differences among lake ecosystems (e.g. Soranno et al., 1999; Kratz and Frost, 2000). More recent studies have shown that spatial organization and connectivity are important descriptors of lake types. For example, lake chemistry and biology have been shown to be correlated with simply the position of a lake in the landscape (e.g. Riera et al., 2000). However, the scale at which we perceive our study objects may influence our interpretation of driving processes and mechanisms. For example, relationships between climate and vegetation that are evident at a broad scale may disappear when finer scales are studied (e.g. Wiens, 1989).

Ecoregions are areas of relatively homogeneous ecological systems, with similar vegetation and climate. It is generally thought that ecoregions with different landscapes have different water quality (e.g. Hughes and Larsen, 1988). For example, studies have assessed how well landscape (ecoregion) level classifications can partition physico-chemical and biological variance of lakes (e.g. Hawkins et al., 2000). If the transition zone between regions is somewhat diffuse, however, boundary effects might not be evident and differences in biotic assemblages between regions may be difficult to distinguish. Hawkins et al. (2000) showed that ecoregions were effective in partitioning variance when distinct differences in geomorphology or strong gradients in physiology and/or climate occurred. Such gradients are found in Sweden; for example, the large-scale regional effect of the *limes Norrlandicus* ecotone (basically a demarcation between deciduous forests in the south and coniferous forests in the north) has been shown to strongly affect different biological communities (e.g. littoral invertebrates, Johnson, 2000). Large-scale patterns in lake water chemistry have also been shown to correlate with ecoregion delineations in Sweden (e.g. Johnson, 1999). Adopting a spatial approach to partition variance, the European Water Framework Directive proposes the use of ecoregions to classify lake typologies (European Commission, 2000).

Although recent focus has been placed on monitoring whole catchments or ecosystems (Omernik, 1994 and references cited therein) in regional monitoring programs, individual lakes often exhibit marked spatial and temporal variability. Though the spatial – vertical variation generally found within a lake is usually accounted for in sampling, much less attention is given to surficial or horizontal spatial variation. Spatial variation and correlation related to horizontal distances between sampling locations is thoroughly studied within terrestrial systems (Webster and Oliver, 1990). A number of lake studies have focused on the spatial (horizontal) variation associated with phyto- and zooplankton communities (e.g. Avois et al., 2000; Lacroix and Lescher-Moutoué, 1995; Pinell-Alloul, 1995; Visman et al., 1994), but much less is known of the horizontal variability associated with lake chemistry. Indeed, spatial patterns in water chemistry are less often reported, and if mentioned, it is usually done in the context of explaining the spatial variation of plankton (e.g. Dickman et al., 1993). In monitoring programs that are temporally extensive (i.e. long-term), the horizontal variation in lake water chemistry is important for interpreting long-term changes. Factors such as lake morphometry (e.g. depth, shoreline development and volume), distance to shoreline, relative importance of surface and groundwater inputs and biological activity may separately or combined affect the spatial variation within a lake. In addition, temporally important factors may also affect water chemistry, such as within- and among-year variability in precipitation, temperature, wind speed and direction.

Besides naturally occurring variability, aquatic ecosystems often show the effects of a number of human-induced perturbations. For example, lakes in Sweden are affected by eutrophication and acidification, and much of this variance is contagiously distributed with lakes in the southern parts of the country being more strongly affected than lakes in the north. According to the 1995 national lake survey, between 3.6% (> 50µg P/l) and 5.5% (> 1.5 mg N/l) of 4113 lakes were classified as very nutrient-rich (Wilander et al., 1998). In the highly populated areas in the south of Sweden with intensive agriculture, 63% of the lakes were classified as very nutrient rich (> 50µg P/l). Furthermore, due to acidification many lakes, in some areas as many as 60% of those sampled, are limed or indirectly affected by liming.

Water chemistry has been monitored in long-term monitoring programs in Sweden since the 1960s. Knowledge of indicator variability, both spatial and temporal, and natural and human-induced, is essential for our interpretation of anthropogenic effects. Much effort has been focused on understanding the importance of scale in explaining variability among lakes. For example, relationships between catchment characteristics and lake water chemistry can be useful for providing information on preindustrial conditions. The European Water Framework Directive (European Commission, 2000) recommends moreover the use of ecoregions and typologies (e.g. lake size) to partition natural variability. Understanding the spatial variability within and among individual lakes is useful for optimizing sampling design. Moreover, this information is needed for improving our awareness of the processes and mechanisms driving indicator metrics (e.g. total phosphorus).

Objectives

In this thesis I have focused on how spatial variability both within and among lakes affects our interpretation of lake monitoring data. In particular, I was interested in assessing the importance of spatial scale, from small (variability within a lake) to large (catchment and ecoregion variability), on selected physico-chemical metrics.

Paper I

The importance of ecoregions and catchment characteristics for prediction of pH and total phosphorus (TP) are studied. Predictive models are evaluated both on ecoregional and whole-country scales. The models are based on catchment-specific variables such as geographic position, catchment land use/cover, and deposition. The hypotheses of the study are that: (i) models based solely on catchment-specific variables can be used to predict pH and TP and (ii) ecoregion based models will have higher predictive power than whole-country models.

Paper II

The horizontal variation of surface water chemistry within lakes is studied. Here we addressed the question if a mid-lake sample can be considered representative of lake chemistry. Our hypotheses are that: (i) where a sample is taken within a lake will affect its representativity and (ii) a mid-lake sample can be considered as representative of average surface water chemistry.

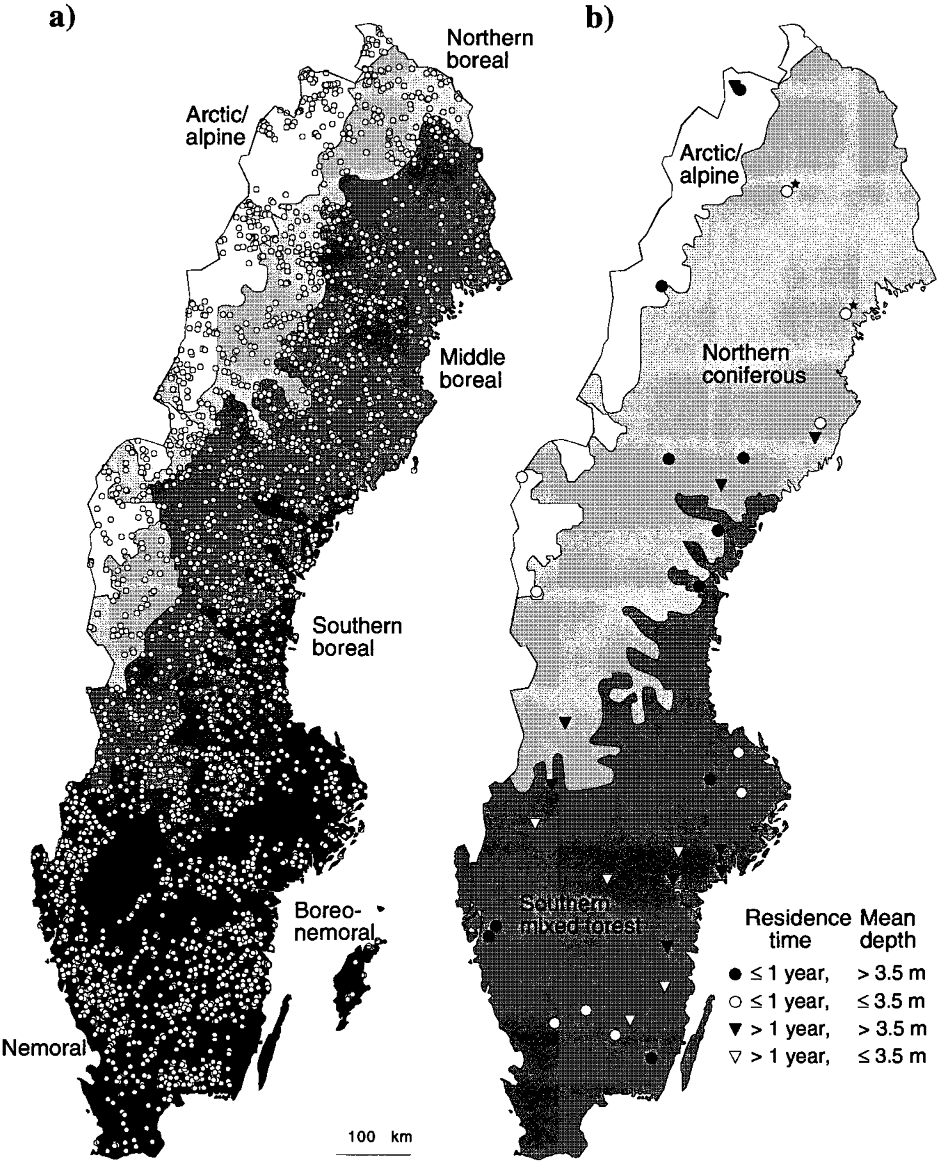

Figure 1. a) Map showing the six major ecoregions of Sweden and the location of the 2765 lakes sampled for physico-chemical and catchment-specific (GCD) variables in the 1995 Swedish national lake survey. **b)** Classification of the 34 lakes according to three ecoregions, water residence time, and mean water depth. The lakes were sampled in August and October 1999. ★ two lakes not sampled in October.

Materials and Methods

Paper I

A Swedish national lake survey was performed in the autumn of 1995. A mid-lake sample was taken in 4113 selected lakes in Sweden (Wilander et al., 1998). The lakes were randomly selected with some predefined criteria to cover different lake types and to obtain a spatial distribution that covered the whole country (summarized in Henriksen et al., 1996). The samples were collected in the autumn during the mixing period to obtain a representative whole-lake sample. The physico-chemical variables analyzed can be grouped in three categories: indicators of acidity, nutrients, and organic compounds. The survey also included variables of geographic position, altitude, catchment area, and lake surface area. Some 2700 lakes were also classified according to land use/cover (see Table 2 in paper I). A total of 2765 lakes were used in this study. For pH modeling, lakes affected by liming were removed resulting in 2046 lakes in the pH models.

All lakes were stratified according to ecoregion (Fig. 1a). Partial least squares regression (PLS) was used to evaluate spatial patterns in lake surface water chemistry and assess the predictive power of models for pH and total phosphorus using descriptive and chemical variables (Esbensen et al., 1996; Geladi and Kowalski, 1986). PLS is a multivariate regression method related to principal component analyses. One advantage with PLS is that it works well with intercorrelated data. PLS was used to calibrate and validate predictive models for pH and total phosphorus. Two types of models were constructed: (i) models using geographic (G); catchment area classification (C), and deposition (D) variables as independent variables (GCD models); and (ii) models using both GCD and chemical variables and independent variables (GCD+CHEM models). Models were developed for the six ecoregions and for the whole country. Root mean square error of prediction (RMSEP) together with r^2 (explained variance) were used to compare the various models.

Paper II

The lakes in this study are a part of the Swedish national lake monitoring program. These lakes are sampled annually for physico-chemical parameters. The total number of reference lakes is about 100, and 34 of these were included in this study. The lakes were selected to cover a wide range of lake types. This was done by classifying the lakes according to: (i) geographic region (3 major ecoregions), (ii) water residence time (2 categories, residence time ≤ 1 year or residence time > 1 year), and (iii) water depth (2 categories, mean depth ≤ 3.5 m or > 3.5 m) (Fig. 1b). Since most reference lakes are situated in the southern part of Sweden that is more densely populated, 5 lakes were selected for each classified group in the southern mixed forest region. In contrast, due to fewer monitoring lakes situated in the north it was difficult to fill all the groups in the arctic/alpine region and the northern coniferous forest region (Table 1).

Ten samples were taken from each lake from quadrates randomly distributed to cover the lake surface area. Each lake was covered with a grid matrix and the quadrates were numbered, starting in the north end of each lake and moving southward in west to east direction. The total number of quadrates per lake was divided by ten to determine the interval between sampling sites. The first sampling quadrate was randomly placed in the first interval. The quadrate closest to the center of the lake, according to the lake surface area and where the maximum depth of the lake was located, was chosen as the mid-lake quadrate. Surficial water samples (0.5 meters' depth) were collected from each quadrate during each season; two sampling seasons, representing high (late summer or August) and low (October) biological activity, were studied. All 34 lakes were sampled in August and 32 in October.

The chemistry variables studied here consisted of; pH, alkalinity (meq/l), total phosphorus (µg/l), NO2+NO3-N (µg/l, hereafter referred to as inorganic nitrogen), conductivity (mS/m 25 °C) and absorbance (420nm) of filtered water (i.e. water color), absorbance (420nm) of unfiltered water, and water temperature (°C).

Large-scale, spatial differences in water chemistry were analyzed by a three-factor unbalanced analyses of variance (ANOVA), with ecoregions (three ecoregions), water residence time (two categories), and mean depth (two categories) as fixed factors. The interaction terms of these three factors were also determined. A Tukey test was carried out to investigate significant differences among the three factors and their interaction terms. A two sample *t*-test was used to determine if the mid-lake sample differed from the nine remaining samples. This was performed on each lake and variable (significant level $\alpha = 0.05$). A jack-knifing technique was used to determine if the distance to the center part of the lake would affect the sample representativity. This was done by sequentially comparing each sample with the remaining nine samples using a two sample *t*-test and Bonferroni correction of α error to a significant level $\alpha = 0.005$.

Table 1. Number of lakes in each classified group.

Residence time	> 1 year		≤ 1 year		
Mean depth	> 3.5 m	≤ 3.5 m	> 3.5 m	≤ 3.5 m	Total
Arctic/alpine region	1	—	2	2	5
Northern coniferous region	3	—	3	3*	9
Southern mixed forest region	5	5	5	5	20
Total	9	5	10	10	34

* Two lakes were not sampled in October

Ecoregions

In Paper I, data was divided into the six major ecoregions of Sweden using the boundaries given by the Nordic Council of Ministers (1984), but modified according to Gustafsson and Ahlén (1996). The ecoregions range from the nemoral region in the south to the arctic/alpine region in the north (Fig. 1a). The nemoral region is characterized by deciduous forests, mean annual temperature greater than 6 °C, and a relatively long growth period (180 - 210 days). In contrast, the arctic/alpine region is characterized by relatively low mean annual temperature (< 2°C) and short growth period (< 100 - 120 days). The number of lakes in each ecoregion were (the numbers within parenthesis are the number after removing the lakes affected by liming): 238 (238) in the arctic/alpine region; 191 (188) in the northern boreal region; 978 (843) in the middle boreal region; 389 (206) in the southern boreal region; 912 (529) in the boreonemoral region; and 57 (42) in the nemoral region.

In Paper II, the ecoregions used correspond closely to those recommended by the European Water Framework Directive (European Commission, 2000). These three regions were originally based on macroinvertebrate distribution studies done in the 1960s (e.g. Illies, 1966). The three ecoregions are composed of the ecoregions above. Hereafter the three regions will be referred to as the arctic/alpine region, northern coniferous region, and southern mixed forest region (Fig. 1b).

Results and Discussion

Paper I

Our hypothesis was that ecoregional based models would have higher predictive power (high r^2) and lower error (RMSEP) than the whole country models. This was based on the principle that ecoregions, as integrators of regional patterns in climate, topography, land use, and natural vegetation, would partition part of the natural variability. This seemed to be true for the southern part of Sweden and especially for the nemoral region. Total phosphorus (TP) and pH models developed for the nemoral region had higher predictive power than the model for the whole country and the other ecoregional models (Fig. 2). For the pH and TP nemoral models, GCD variables explained 90% and 57% of the variation, respectively. The total variation within a region seemed to affect the predictive power. For example, total variation within the arctic/alpine region was low resulting in low predictive power, whereas the total variation within the nemoral region was high resulting in high predictive power. Overall for both pH and TP, the predictive power of the ecoregional models based on GCD variables was slightly better or about the same as the predictive power for the whole-country models. The prediction error was usually about the same size or somewhat lower

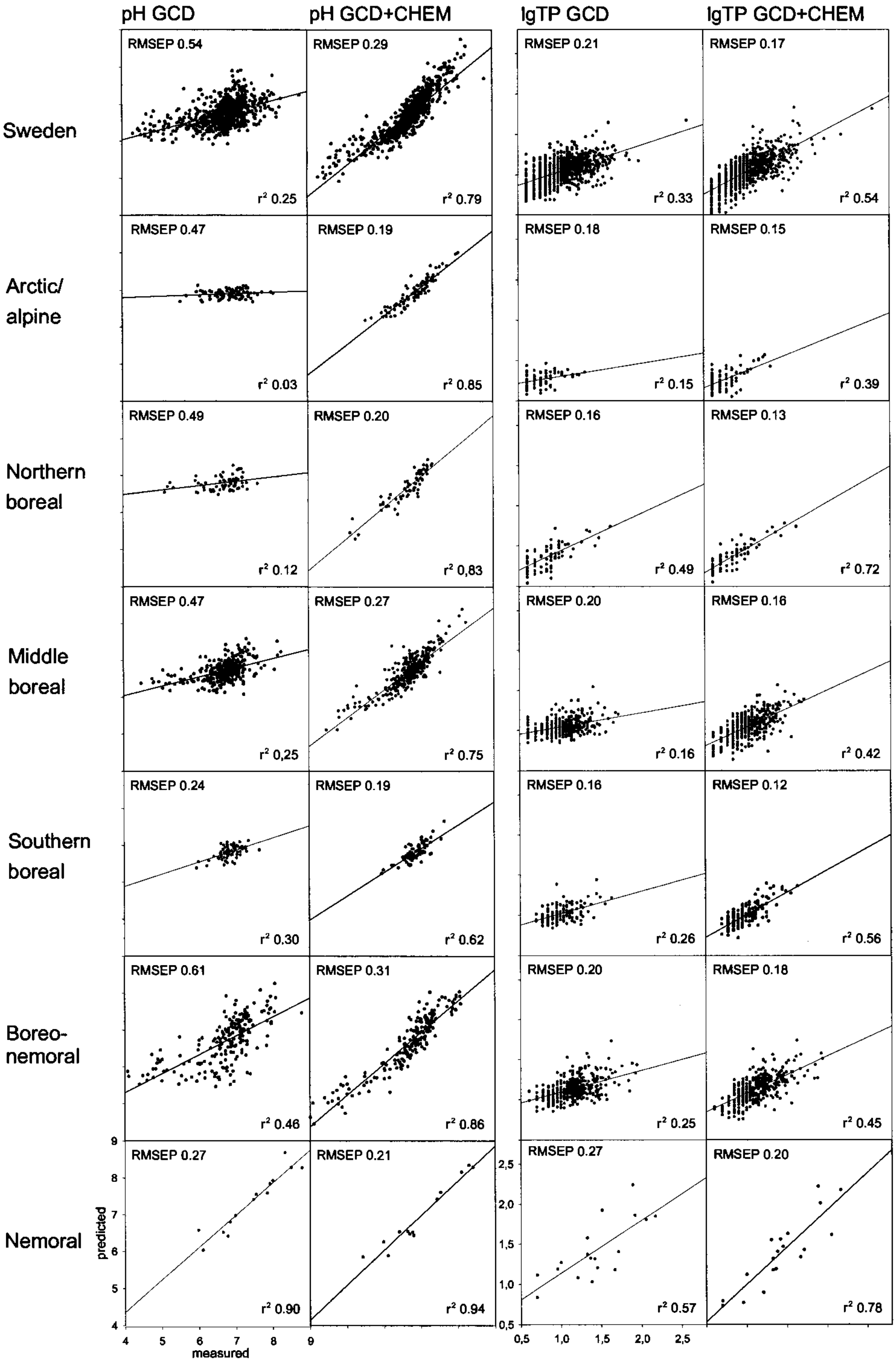

Figure 2. Predicted vs. measured values for the pH and the $\log_{10}$TP models' validation data sets. The explained variance (r^2) and the prediction error values RMSEP are given for the whole-country models (Sweden) as well as the models for the ecoregions.

for the GCD models than the whole-country model. An exception was the TP nemoral model that had a higher prediction error than the whole country model. One reason might be that there were few lakes in the nemoral region compared to the other regions and the whole country. Some of the pH models seemed to be nonlinear across the pH-gradient, with a breakpoint at approximately pH 6. This implies that it might be better to model pH above and below pH 6 instead of across whole interval. Preliminary studies for the middle boreal region supports this conjecture.

Catchment and lake surface area were important descriptive variables for both pH and TP (GCD models). Other variables such as land use (arable, forest) and cover (mire and deciduous forest), and soil type (till) were important for the pH GCD models. Classifications of forest and arable land explained a large part of the variation in pH and TP within the nemoral region. Other important descriptors for the TP GCD models were altitude, and the amount mires and forests within the catchment (especially in the northern parts of the country)

As expected, adding the chemical variables improved the models predictive power considerably. The most obvious example was the arctic/alpine pH model. In this model the explained variance increased from 3% to 85% when the chemical variables were included. The most important predictor for pH was calcium (Ca), whereas the most important for TP were water color (absorbance) and in some cases TOC and the nutrients NH_4-N and total nitrogen (TN). Interestingly, these chemical variables (Ca, water color, TOC, NH_4-N, and TN) all reflect conditions in the catchment.

Paper II

Comparison of the mid-lake sample with the nine remaining samples for all lakes and variables showed that out of 254 (August) and 253 (October) possible combinations only 10 of them differed in August and 9 in October. Jack-knifing showed that a samples location was important for its' representativity of average surficial water chemistry. For example, samples taken in the center of the lakes had few differences in water chemistry indicating that these samples are more representative of average conditions. In contrast, samples taken furthest from the lakes' center often differed markedly from the other samples (Table 2) (e.g. samples taken in the first quadrate). This was especially evident for samples taken close to lake inlets. To further investigate this we also determined if the mid-lake sample could be considered more representative of lake water chemistry than simply randomly selecting a site. The test was performed in the same way as for the mid-lake sample, the only difference was that the sample was randomly selected from the ten samples for each lake. Of the variables tested, 24 differed significantly in August and 23 in October, compared to 10 and 9, respectively, for the mid-lake sample. Although comparison of only one random sample can not be considered a robust measure of sample representativity, these findings indicate nonetheless that a mid-lake sample is more representative of average lake conditions.

Grouping lakes by ecoregion, residence time, and mean depth had a significant effect on a number of variables (except pH, alkalinity, and inorganic nitrogen). Ecoregion seemed to be the most important factor describing lake chemistry. Not surprisingly, temperature in August varied significantly for all regions indicating the large climate gradient that was part of this study. Variables sensitive to biological activity were also influenced by the climate gradient; for example, absorbance of filtered water and total phosphorus also varied with ecoregion. Residence time was also found to be a more important explanatory variable in August than October. This finding might indicate the greater spatial variability expected in late summer (when lakes are relatively hydrologically stable) than autumn (when lakes are turning over). Indeed, several of the variables that varied significantly when the mid-lake samples were compared with the other samples also varied somewhat with season. For example, nutrients seemed to vary more in August than the other variables whereas variables indicative of weathering and hydrology (e.g. alkalinity and conductivity) varied more in October.

Some of the variation within a lake may be attributed to analytical or laboratory-induced variance. For variables exhibiting small, within-lake variability, such as conductivity, this is especially important. We compared the coefficient of variation (CV) of the ten samples in each lake for each chemical variable (CV_{10}) with the CV of the analytical variation (CV_{av}). We found that CV_{av} was greater than CV_{10} in two of the ten differences noted in August and six of the nine differences in October. If $CV_{av}>CV_{10}$ for the variables tested, then we can not conclude that the mid-lake sample is significantly different for those lakes.

Table 2. A 2-sample *t* – test comparing each sample quadrate with the remaining nine by jack-knifing method. Each column represents the sample quadrate that was compared to the other sample quadrates. The values given show the number of significant differences. P-values were adjusted by Bonferroni corrections ($\alpha \leq 0.005$).

Variable	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
					AUGUST					
pH	9	2					1			1
Alkalinity	2		1	1	1		3			
Conductivity	2						1			
Abs. unfiltered								1		1
Abs. filtered	7		2							1
Total phosphorus										
NO2+NO3-N	1				1			1		1
Temperature	1					1			1	1
Sum	**22**	**2**	**3**	**1**	**2**	**1**	**5**	**2**	**1**	**5**
					OCTOBER					
pH	3									2
Alkalinity	3		3	2		1			2	1
Conductivity	2			1		1		1	1	2
Abs. unfiltered	2		1	1				1		1
Abs. filtered	6	1		1						
Total phosphorus	1	1			1			1		
NO2+NO3-N	2	1								3
Temperature	4		1	1						
Sum	**23**	**3**	**5**	**6**	**1**	**2**	**0**	**3**	**3**	**9**

Conclusions and future perspectives

Both large and small spatial scales were found to affect lake water chemistry. In Paper I, we hypothesized that: (i) models based solely on catchment-specific variables can be used to predict pH and total phosphorus and (ii) ecoregion based models will have higher predictive power than whole-country models. Our study showed that pH and total phosphorus were best predicted by other in-lake physico-chemical variables, implying that in-lake chemistry was a better predictor than catchment-scale descriptors. However, many of the “better” predictor variables are known to be dependent on characteristics of the catchment. For example, calcium was shown to be a good predictor of lake pH. Calcium is also strongly correlated with catchment geology and soil chemistry. Hence, pH should be correlated with these catchment-scale variables as well. In our study we did not have access to soil or geological information. Future studies should focus on determining the linkages between catchment-scale variability (e.g. geology and weathering) and in-lake chemistry.

Ecoregion models were not always better than whole-country models in predicting pH and TP. In particular, models developed for the southern parts of the country performed better. In contrast, when the within-ecoregion variance was low (e.g. in the arctic/alpine region) ecoregion models performed poor. Better ecoregion-based models may be developed by partitioning among-lake variance using additional landscape delineations (e.g. geology and highest post-glacial coastline). However, predictive models for some regions may be more difficult to develop. For example, geology in the arctic/alpine region is known to be very variable.

In Paper II, it was hypothesized that: (i) where a sample is taken within a lake will affect it's representativity and (ii) a mid-lake sample can be considered as representative of average surface water chemistry. Where a sample was taken within a lake was shown to be important for how representative it is of lake surface chemistry. In particular, samples taken from quadrates that were situated close to a lake inlet were found to differ. Mid-lake samples, on the other hand, were shown to be representative of average surface water chemistry. However, the finding that water residence time was an important explanatory variable for August water chemistry implies that lakes are more heterogeneous in late summer than autumn. In autumn, many of the lakes were probably circulating which may partly explain the less heterogeneous conditions measured at this time. However, my results indicate that mid-lake samples taken during either August or October can be considered representative of average conditions and should continue to be used in lake monitoring and assessment programs.

References

Avois, C., Legendre, P., Masson, S., and Pinel-Alloul, B., 2000. Is the sampling strategy interfering with the study of spatial variability of zooplankton communities? Canadian Journal of Fisheries and Aquatic Sciences 57, 1940-1956.

Dickman, M., Stewart, K., and Servant-Vildary, M., 1993. Spatial heterogeneity of summer phytoplankton and water chemistry in a large volcanic spring-fed lake in northern Iceland. Arctic and Alpine Research 25, 228-239.

Esbensen, K., Schönkopf, S., and Midtgaard, T., 1996. Multivariate analysis in practice. Wennbergs Trykkeri AS, Trondheim, Norway.

European Commission, 2000. Directive 2000/60/EC of the European Parliament and of the Council – Establishing a framework for Community action in the field of water policy. Brussels, Belgium, 23 October 2000.

Geladi, P., and Kowalski, B. R., 1986. Partial least squares regression: A tutorial. Analytica Chimica Acta 185, 1-17.

Gustafsson L. and Ahlén, I. (eds.), 1996. Geography of plants and animals. Almqvist & Wiksell International, Stockholm.

Hawkins C. P., Norris, R. H., Gerritsen, J., Hughes, R. M., Jackson, S. K., Johnson, R. K., and Stevenson, R. J., 2000. Evaluation of landscape classifications for biological assessment of freshwater ecosystems: synthesis and recommendations. Journal of North American Benthological Society 19, 541–556.

Henriksen, A., Skjelkvåle, B. L., Lien, L., Traaen, T. S., Mannio, J., Forsius, M., Kämäri, J., Mäkinen, I., Berntell, A., Wiederholm, T., Wilander, A., Moiseenko, T., Lozovik, P., Filatov, N., Niiniokja, R., Harriman, R., and Jensen, J. P., 1996. Regional Lake Surveys in Finland - Norway – Sweden – Northern Kola – Russian Karelia – Scotland – Wales 1995. Coordination and design. NIVA Report 40/1996.

Hughes, R. M., and Larsen, D. P., 1988. Ecoregions: an approach to surface water protection. Journal of the Water Pollution Control Federation 60, 486–493.

Illies, J., 1966. Verbreitung der Süßwasserfauna Europas. Verhandlungen der Internationalen Vereinigung für Theoretische und Angewandte Limnologie 16, 287–296.

Johnson, R. K., 1999. Classification of Swedish lakes and rivers using benthic macroinvertebrates (Bottenfauna, Bakgrundsdokument till bedömningsgrunder för sjöar och vattendrag). Rapport, Stockholm, Swedish Environmental Protection Agency. (In Swedish).

Johnson, R. K., 2000. Spatial congruence between ecoregions and littoral macroinvertebrate assemblages. Journal of North American Benthological Society 19, 475–48.

Kratz, T. K., and Frost, T. M., 2000. The ecological organisation of lake districts: general introduction. Freshwater Biology 43, 297 – 299.

Lacroix, G., and Lescher-Moutoué, F., 1995. Spatial patterns of planktonic microcrustaceans in a small shallow lake. Hydrobiologia 300/301, 205-217.

Levin, S. A., 1992. The problem of pattern and scale in ecology. Ecology 73, 1943-1967.

Magnuson J. J., and Kratz T. K., 2000: Lakes in the landscape: Approaches to Regional Limnology, Verhandlungen der Internationalen Vereinigung für Theoretische und Angewandte Limnologie 27 (1), 74 – 87.

Naumann, E., 1932. Grundzüge der regionalen Limnologie. Die Binnengewässer 11, 176 pp.

Nordic Council of Ministers, 1984. Naturgeografisk regionindelning av Norden. Nordiska ministerrådet 1984.

Omernik, J. M., 1994. Ecoregions: A spatial framework for environmental management. - In Simon, T., and Davis, W., (editors). Biological assessment and criteria. Tools for water resource planning and decision making, 49–62. Lewis Publishers, Chelsea, Michigan.

Ormerod, S. J., and Watkinson, A. R., 2000. Large-scale ecology and hydrology: an introductory perspective from the editors of the Journal of Applied Ecology. Journal of Applied Ecology 37, 1–5.

Pinel-Alloul, B., 1995. Spatial heterogeneity as a multiscale characteristic of zooplankton community. Hydrobiologia 300/301, 7-42.

Riera, J. L., Magnuson, J. J., Kratz, T. K., and Webster, K. E., 2000. A geomorphic template for the analysis of lake districts applied to the Northern Highland Lake District, Wisconsin, U.S.A. Freshwater Biology 43, 301–318.

Soranno, P. A., Webster, K. E., Riera, J. L., Kratz, T. K., Baron, J. S., Bukaveckas, P. A., Kling, G. W., White, D. S., Caine, N., Lathrop, R. C. and Leavitt, P. R., 1999. Spatial variation among lakes within landscapes: Ecological organization along lake chains. Ecosystems 2, 395–410.

Thienemann, A., 1925. Die Binnengewässer Mitteleuropas. Eine limnologische Einführung. Die Binnengewässer 1, 255 pp.

Turner, M. G., 1998: Landscape ecology, living in a mosaic. – In: Dodson, S. I., Allen, T. F. H., Carpenter, S. R., Ives, A. R., Jeanne, R. J., Kitchell, J. F., Langston N. E., and Turner, M. G. (eds.): Ecology, 78-122. Oxford University Press, New York.

Visman, V., McQueen, D. J., and Demers, E., 1994. Zooplankton spatial patterns in two lakes with contrasting fish community structure. Hydrobiologia 284, 177-191.

Webster, R., and Oliver, M. A., 1990. Statistical Methods in Soil and Land Resources Survey. Oxford University Press, Oxford.

Wilander, A., Johnson, R.K., Goedkoop, W., and Lundin, L., 1998. Riksinventering 1995. En synoptisk studie av vattenkemi och bottenfauna i svenska sjöar och vattendrag. Rapport 4813, Stockholm, Swedish Environmental Protection Agency. (In Swedish)

Wiens, J. A., 1989. Spatial scaling in ecology. Functional Ecology 3, 385-397.

Acknowledgements – TACK!

Äntligen är den klar! Tack alla ni som engagerat er och inspirerat mig på vägen.

Richard Johnson, min huvudhandledare, tack för allt stöd och engagemang. Du har introducerat mig till limnologins vetenskapliga värld samt många nya engelska ord. Mina biträdande handledare har varit ovärderliga både som kunskapskällor och stöd när det känts motigt. Anders Wilander – du kan allt om vattenkemi och en hel del statistik också! Ewa Bringmark – med dig kan man prata om allt som varierar i rummet, och mycket mycket mer. Jag vill även tacka Anders Grimvall för all hjälp och stöd, främst med de multivariata analyserna.

Bakom alla data finns många betydelsefulla personer. Alla fältprovtagare – tack för att ni visade intresse i mitt projekt genom att samarbeta med mig och ta tio extra prov i varje sjö. Till personalen på institutionens vatten kemi lab.– ni är fenomenala! Tack för att ni tog hand om och analyserade alla extra prov. Jag vill rikta ett särskilt tack till Kjell Östling som organiserade provtagningen (Paper II). Tack Joakim Dahl för att du höll i mitt projekt när jag åkte till New York.

På sluttampen när det var som mest stressigt, när redigering och layout skulle fixas var Mikael Östlund oumbärlig. Tack – figurerna och tabellernas mästare!

Institutionens doktorander är ett härligt gäng – ni gör det roliga ännu lite roligare. Ett gott skratt förlänger livet sägs det – vi blir nog 100 år allihop. Jag vill även tacka Leonard Sandin och Jens Fölster, som var med från början. Lars Rapp, vid det här laget är du väl doktor, – du är en klok man med goda råd.

Till alla på miljöanalys – jag uppskattar att få arbeta här med er - tack för en positiv och trevlig atmosfär.

Det finns personer utanför institutionen som också har betytt en hel del för mig. Erik Karltun – tack för att du introducerade mig till SLU och PLS. Dessutom peppade du mig till att ta doktrorandtjänsten. Ulla Edberg, min mentor, det är alltid lätt att prata med dig - tack för alla goda råd och trevliga middagar.

Till sist vill jag tacka min Tobias – du står mitt hjärta närmast - tack för att du finns!

The importance of ecoregions and catchment characteristics for prediction of pH and total phosphorus in lakes

Elisabet Göransson and Richard K. Johnson

Department of Environmental Assessment, Swedish University of Agricultural Sciences, Box 7050, SE-750 07, Uppsala, Sweden

Abstract

Ecoregions are areas of relatively homogenous ecological systems, with similar vegetation and climate. It is generally thought that ecoregions with different landscapes have different water quality. Here we evaluate the use of landscape characteristics (ecoregion and catchment classification) for predicting surface water pH and total phosphorus (TP). Predictive models were evaluated both on ecoregional and whole-country scales. The models were based on catchment-specific variables alone or a combination of catchment-specific and physico-chemical variables. The hypotheses were that: (i) models based on catchment specific variables could be used to predict pH and TP and (ii) ecoregional based models would have higher predictive power than whole-country models. The lakes used were part of the 1995 national lake survey. Partial least squares regression (PLS) was used for evaluating the models. We found that grouping lakes by ecoregions worked well if the total variation of the data (pH or TP) was large. For example, ecoregion delineations worked better for the ecoregions in the south than in the north. However, in general TP and pH were best predicted by other physico-chemical variables, implying that in-lake chemistry was a better predictor than catchment-specific variables. Indeed, many of the better predictor variables were known to depend on characteristics of the catchment. Our findings also indicate that models would probably be improved by adding catchment and lake characteristics such as soil chemistry and residence time, as well as improving the classification of the included land use/cover parameters.

Introduction

Geographic position and catchment-scale variability in geology and land use and vegetation cover have provided insight into differences among aquatic systems (e.g. Soranno et al., 1999; Kratz and Frost, 2000). Some studies have, for example, assessed how well landscape (ecoregion) level classifications can partition physico-chemical and biological variance of lakes (e.g. Hawkins et al., 2000; Magnuson and Kratz, 2000). Ecoregions are areas of relatively homogeneous ecological systems, with similar vegetation and climate, and one would expect that ecoregions are able to partition within-lake variance in water chemistry. Indeed, it is generally thought that regions with different landscapes have distinctly different water quality (e.g. Hughes and Larsen, 1988). However, if the transition zone between regions is somewhat diffuse, boundary effects may not be evident and differences in biotic assemblages between regions may be difficult to distinguish. This conjecture was recently exemplified in a review by Hawkins et al. (2000); these authors showed that ecoregions were effective in partitioning variance when distinct differences in geomorphology or strong gradients in physiology and/or climate occurred. Such gradients are found in Sweden; namely the large-scale, regional effect of the *limes Norrlandicus* ecotone. For example, Johnson (1999) found clear differences in water chemistry between ecoregions in the northern (arctic/alpine) and southern (nemoral) parts of the country. Moreover, although contemporary paradigms suggest that lake ecosystems can be viewed in a spatial, landscape context (i.e. lakes are nested in and part of the landscape mosaic), recent studies have also shown how connectedness (e.g. lake order) across a landscape can improve predictive power (e.g. Kratz et al. 1997; Sorrano et al. 1999).

If ecoregions are able to partition natural variability within the landscape, then ecoregion-based models might be able to predict lake surface water chemistry. However, variation within- and among lakes is affected by multiple pathways and mechanisms. Indeed, determinants of lake-water chemistry are acting at both on regional, catchment, and local scales. Factors such as climate, soil chemistry, weathering, land use/cover affect the lake water chemistry mainly at a regional and catchment scale, whereas lake morphometry, water residence time, internal processes such as nutrient cycling, and wind affects a lake water chemistry at a local scale. For example, longer residence time generally results in lower total phosphorus (TP) concentrations in lake water compared to the water in the inlets to the lake; simply due to biologically active phosphorus being incorporated into primary production and subsequently sedimentation and deposition into the bottom sediment. On the other hand, TP concentration in a shallow lake may increase due to wind induced resuspension of sediments.

In Sweden, as elsewhere, simply establishing the importance of how patterns in natural variability correspond to lake chemistry are generally not adequate to predict the chemistry of an individual lake. Indeed, many if not most lakes are affected by overlying anthropogenic gradients, mainly eutrophication and acidification. For example, in the southernmost ecoregions of Sweden (nemoral and boreonemoral) lakes are strongly affected by diverse human impacts; two of the most important are land use (agriculture) and airborne pollution (e.g. acidification by S and N compounds).

In this study, the spatial variation of pH and total phosphorus was partitioned using ecoregional classification. pH and TP were chosen as they are good indicators of land use/cover and of acidification and eutrophication, respectively. We analyzed the predictive power (r^2) and prediction error (RMSEP) using two model types. Models based solely on catchment-specific variables such as geographic position, catchment land use and cover classification, and deposition variables (here referred to as GCD models) were compared to models based on both GCD-variables and surface water chemistry (CHEM). Partial least squares regression (PLS) was used to evaluate the predictive power of the two model types and analyses were made for whole-country and ecoregion-based models. Specifically, we were interested in assessing: (i) if ecoregion-based models had a higher predictive power than whole-country models and (ii) if GCD variables can be used to predict pH and TP.

Materials and methods

Sampling

Surficial water samples were taken from 4113 randomly selected lakes as part of the Swedish national lake survey performed in the autumn of 1995 (Wilander et al., 1998). The lakes were randomly selected with predefined criteria to cover different lake types and to obtain a spatial distribution that covered the whole country. For example, five lake surface area classes were used (i.e. 0.04-0.1, 0.1-1, 1-10, 10-100, and >100 km^2) with the proportions of lakes for each lake-surface-area class being 1:1:4:8, lakes >100 km^2 were all included in the survey (summarized in Henriksen et al. 1996). Besides lake-surface-area classes, the lakes were distributed across 24 counties. The variation of selected chemical metrics was also used in selecting lakes to obtain the desired level of statistical power. The water samples were collected from mid-lake, directly in a polyethylene bottle or with a metal-free Ruttner sampler. Autumn sampling, during the mixing period, was selected to obtain a representative whole-lake sample (Wilander et al. 1998). Sampling started in September in the northern part of Sweden and it ended in December in the southern part of the country.

Physico-chemical and environmental or descriptive variables

Water samples were analyzed for a suite of physico-chemical variables that can be grouped in four categories: indicators of acidity, nutrients, natural organic matter, and major constituents (Table 1). All analyses were done by the SWEDAC (Swedish Board for Accreditation and Conformity Assessment) certified laboratory at Department of Environmental Assessment, Swedish University of Agricultural Sciences (Wilander et al. 1998).

The national lake survey also included variables of geographic position (national grid coordinates), altitude (m), catchment area (km^2), and lake surface area (km^2) as well as land use/cover (Table 2). For a subset of the lakes sampled in the national survey (some 2700 of the 4113 lakes), catchment areas were classified according to 10 land use and vegetation cover parameters (Wilander et. al., 1998). For forested areas additional parameters were estimated and expressed as the percentage of the forest area (a-d, Table 2). Information on the areal distributions of land use/cover categories and forest uptake of base cations and nitrogen were taken from the Swedish national forest inventory (Anonymous, 1997) and Swedish national survey of forest soils and vegetation (Anonymous, 1987).

Table 1. The physico-hemical variables used for the GCD+CHEM models.

	Unit
Acidity indicators	
pH	
Alkalinity	meq/l
ANC or acidity	meq/l
Major constituents	
Conductivity	mS/m
Calcium (Ca)	meq/l
Magnesium (Mg)	"
Sodium (Na)	"
Potassium (K)	"
Sulfate (SO_4)	"
Chloride (Cl)	"
Fluoride (F)	mg/l
Nutrients	
Ammonium-nitrogen (NH_4-N)	µg/l
Nitrate-nitrogen (NO_2+NO_3-N)	"
Total nitrogen (TN)	"
Total phosphorus (TP)	"
Silicon (Si)	mg/l
Natural organic matter	
Absorbance, filtered (AbsF)	abs/5cm
Total organic carbon (TOC)	mg/l

Table 2. Catchment-specific (GCD) variables used in model calibration. For forested areas additional parameters were estimated and expressed as the percentage of the forest area (a-d). The four groups (a-d) were independent of each other and classified separately (i.e. apart from forest clear-cut, each group's sum was approximately 100%).

	Unit		Unit
Catchment area classification		**Geographic variables**	
Forest	%	X national grid coordinates	latitude
Pasture-land	"	Y national grid coordinates	longitude
Arable-land	"	Catchment area	km^2
Mire	"	Lake surface area	km^2
Bedrock outcrops	"	Altitude	m
High-mountain	"		
High-mountain coniferous forest	"	**Deposition**	
Populated (urban) area	"	Sulfur deposition	meq/m^{2}year
Fresh water	"	Nitrogen deposition	meq/m^{2}year
Other land	"		
Forest area classifications		**Other**	
(a) Mineral soils	%	Forest uptake of base cations	eq/ha year
(a) Peat land	"	Forest uptake of nitrogen	eq/ha year
(b) Till	"	Mean run off	l/km^2 year
(b) Sorted sediments	"		
(c) Forest clear-cut	"		
(d) Pine	"		
(d) Spruce	"		
(d) Deciduous forest	"		

Ecoregion classification

The lakes were spatially divided among the major ecoregions of Sweden using the boundaries given by the Nordic Council of Ministers (1984), but modified according to Gustafsson and Ahlén (1996) (Fig. 1). The ecoregions range from the nemoral region in the south to the arctic/alpine region in the north. The nemoral region is characterized by deciduous forests, mean annual temperature greater than 6 °C and a relatively long growth period (180 – 210 days). In contrast, the arctic/alpine region in the north is characterized by relatively low mean annual temperatures (< 2 °C) and short growth periods (< 100 – 120 days). The number of lakes in each ecoregion are given in table 3.

Table 3. Number of lakes in each ecoregion and for the whole country. About 700 lakes in the study were affected by liming and therefore removed when modeling pH.

	TP models	**pH models**
Sweden (whole country)	2765	2046
Arctic/alpine	238	238
Northern boreal	191	188
Middle boreal	978	843
Southern boreal	389	206
Boreonemoral	912	529
Nemoral	57	42

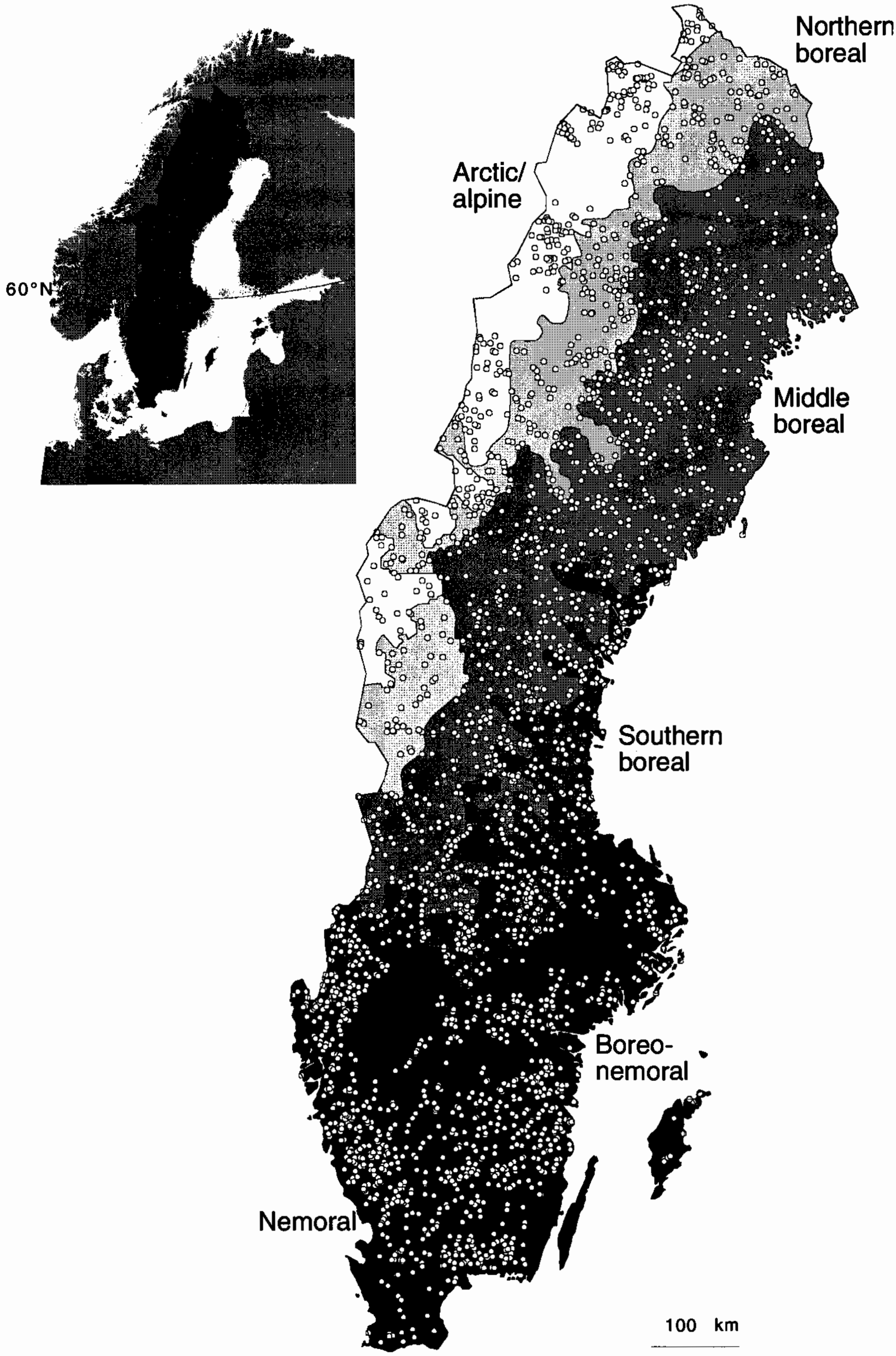

Figure 1. Map of Sweden showing the six major ecoregions and the 2765 lakes sampled for physico-chemical and catchment-specific (GCD) variables in 1995 Swedish national lake survey.

Statistical analyses

Partial least squares regression (PLS) was used to evaluate spatial patterns in lake surface water chemistry and assess the predictive power of models for pH and total phosphorus using descriptive and chemical variables (Esbensen et al., 1996; Geladi and Kowalski, 1986). PLS is a multivariate regression method related to principal component analyses (PCA). One of the advantages of PLS is that it works well with intercorrelated data. In contrast to other commonly used regression methods, such as multiple linear regression and principal component regression, PLS extracts the variation in the X-data (or independent data) that best correlates with the variation in the Y-data (or dependent data) by allowing the Y-data structure to intervene with the decomposition of the X-data. The program UNSCRAMBLER 6.1 was used for the model evaluation (Unscrambler 1996).

PLS was used to calibrate and validate predictive models for pH and total phosphorus. Two types of models were constructed: (i) models using geographic (G), catchment area classification (C), and deposition (D) variables as independent variables (GCD models), and (ii) models using both GCD and chemical variables as independent variables (GCD + CHEM models). The different types of models were evaluated for each ecoregion and for the whole country. Catchment characterization was available for 2765 of the lakes sampled in the 1995 survey. Some 700 of these lakes were influenced by liming and thus were excluded from the data set used for the pH models (i.e. 2046 lakes were used for modeling pH, Table 3). All 2765 lakes were used for the total phosphorus models. The predictive models were evaluated with a randomly selected subset of data. The validation data sets were usually 35 to 40 percent of the total data set for each model (Table 3). For example, the whole-country total-phosphorus model was calibrated using 1765 lake samples and validated with 1000 randomly selected lake samples.

In PLS the variables were standardized (1/standard deviation) and mean centered (mean = 0 and variance = 1), and most variables were $\log_{10}$ or arcsine transformed in order to approximate normally distributed random errors. The catchment area classification variables were arcsine transformed, while the physico-chemical variables (except pH, alkalinity, and ANC) were $\log_{10}$ transformed. Nitrogen and sulfur deposition, mean run off, forest uptake of base cations and nitrogen, and the geographic variables (except X and Y coordinates) were also $\log_{10}$ transformed.

The root mean square error of prediction (RMSEP) was used to compare the GCD and GCD+CHEM models. RMSEP is a direct measurement of the prediction error in the same unit as the dependent variable, and is an estimate of forecasting error. Following Esbensen et al. (1996), RMSEP was used in evaluating the models of the validation data set. The corresponding measure for the calibration data set is RMSEC (root mean square error of calibration).

Regression coefficients and visual observation of plots of x and y loading weights (first two components) were used to determine the importance of the independent variables for each model. Variables with low regression coefficients and/or low variation were excluded from the models. Systematically excluding and including variables was another method used to determine if individual variables were important for the models. Sites that had extreme values for the Y or X variables (i.e. outliers) were removed to improve the models. Although these sites belonged to the population of sites, the "outliers" usually consisted of sites at the edges (extreme low and high values) of the gradient. Removing extreme values was justified in that it is usually more difficult to predict values in the extreme value range since the predictive ability is generally lower at the edges of the population and there are too few sites at the edges to give a sufficient prediction. Accordingly, using this approach, no more than 1 to 4 sites were generally removed from the models.

The number of PLS-components for each of the models was determined by studying the residual variance of Y and X validation sets. Each PLS-component's residual variance was calculated for the Y and X validation data sets. The residual variance was in general inversely related to the number of PLS-components, especially for the X data. For the Y data set, the residual variance usually increased after a certain number of components (i.e. the residual variance curve for the Y validation data displayed a minimum value). The optimal number of components for the models was set to the value of the minimum residual variance for the Y data set. Minimizing the residual variance for the Y validation data also minimized the prediction error.

Results

PLS models and errors

pH GCD models

Stratification according to the six ecoregions reduced the variation of the data for all regions except the boreonemoral and the nemoral regions (Fig. 2). Standard deviations for these regions were 0.84 and 0.88, respectively, compared to 0.63 for the whole country. The lowest standard deviation was recorded for the arctic/alpine region (0.45). The predicted vs. measured plots (Fig. 4) showed that the predictive power (explained variance, r^2) in the regional data sets was higher for the southern regions than in the northern regions. For the nemoral model, 90% of the variance was explained, whereas 3% and 12% of the variance was explained for the arctic/alpine and northern boreal models, respectively. The remaining four models explained > 25% of the variance. Error estimates

(RMSEP) ranged from 0.24 pH units for the southern boreal model to 0.61 for the boreonemoral model. Except for the boreonemoral model, RMSEP was lower for the regions (0.24 – 0.49) than for the whole country (0.54). Some of the models (i.e. whole country, middle boreal, and boreonemoral) seemed to be nonlinear across the pH gradient with a breakpoint at approximately pH 6 (Fig. 4).

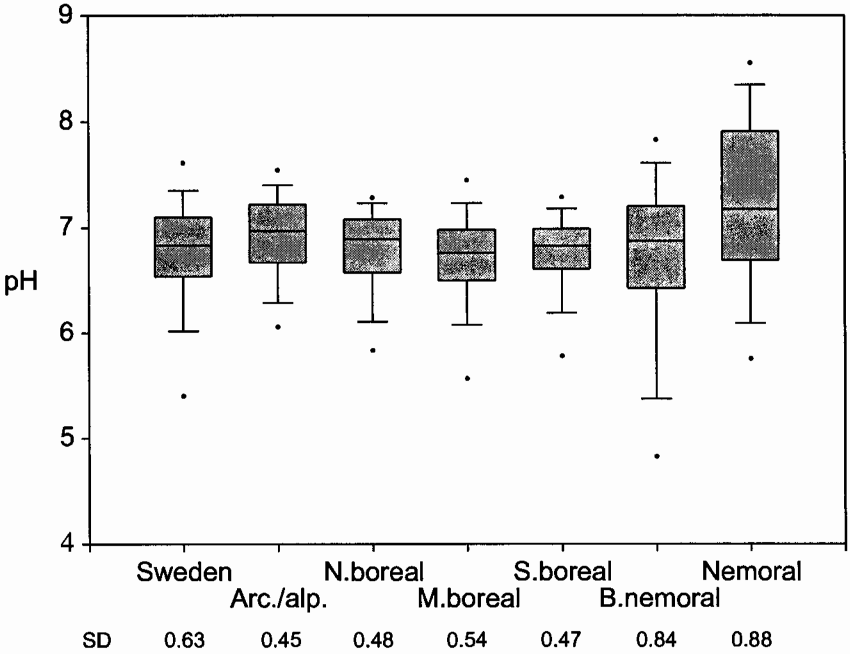

Figure 2. pH variation for the whole country and for each of the ecoregions. The 5th, 10th, 25th, 50th (median), 75th, 90th, and 95th percentiles for the total data set and the data sets for the six ecoregions. The total variation within each data set is given as standard deviation (SD). A contrast test on the mean pH values for each ecoregion (Tukey) was performed with overall error rate 0.05 and individual error rate 0.00321. The test showed that the arctic/alpine region was significantly different from all other ecoregions except the northern boreal and the nemoral region. Whereas the northern boreal was only significantly different from the nemoral region. The three ecoregions: middle boreal, southern boreal, and boreonemoral were significantly different from the arctic/alpine and the nemoral region. Finally, the nemoral region was significantly different from all ecoregions except the arctic/alpine.

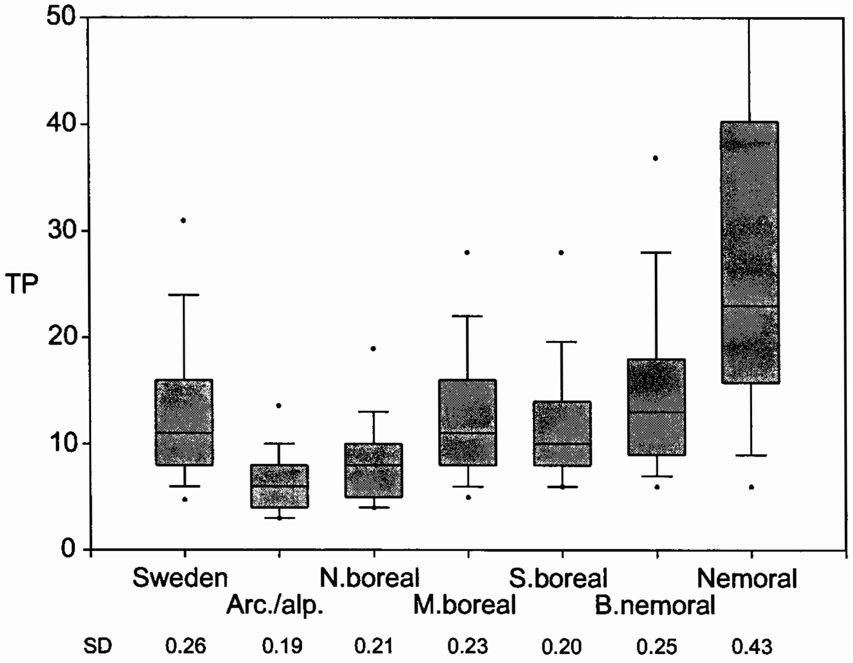

Figure 3. TP variation for the whole country and for each of the six ecoregions. The 5th , 10th, 25th, 50th (median), 75th, 90th, and 95th percentiles for the total data set and the data sets for the six ecoregions. The total variation within each data set is given as standard deviation (SD). A contrast test (Tukey) was performed on mean $\log_{10}$ TP values for each ecoregion, with overall error rate 0.05 and individual error rate 0.00321. The test showed that all the ecoregions except the middle and southern boreal were significantly different.

pH GCD+CHEM models

GCD+CHEM models explained more of variation than the models based on the GCD variables alone. The predictive power increased and the prediction error (RMSEP) decreased when the chemical variables were included in the models. For the arctic/alpine and northern boreal models the predictive power increased from 3% to 85% and 12% to 83%, respectively. Conversely, prediction error for the two models decreased from 0.47 to 0.19 and 0.49 to 0.20 pH units respectively.

TP GCD models

TP concentrations (median values) showed a strong latitudinal gradient, with high TP concentrations in lakes situated in the southern, more agricultural areas, of Sweden (Fig. 3). Median TP concentration in the nemoral region was 23 μg/l, compared to a value of 6 μg/l for the arctic/alpine region. The total variation within the regional (standard deviation, SD, = 0.19 – 0.25) and the whole country (SD = 0.26) data sets were similar, except for the nemoral region (SD = 0.43) (Fig. 3). In contrast to pH, there were no obvious north – south gradients with respect to explained variance and error estimates for the TP-GCD models (Fig. 4). With the exception of the arctic/alpine and middle boreal regions, the amount of variance that could be explained was > 25%. The highest explained variance was recorded for the nemoral region (57%) and the lowest for the arctic/alpine region (15%). The total variation within a region seemed to affect the predictive power. The larger total variation within the nemoral region resulted in a higher explained variance compared to the other regional models. The only model that did not seem to follow this pattern was the northern boreal model. Total variation within the northern boreal region was relatively low (SD = 0.21), nevertheless the model explained 49% of the variation, compared to 33% for the whole country (with SD = 0.26). Error estimates were lower for the regional (RMSEP = 0.16 – 0.20) than for the whole country (0.21) models, except for the nemoral region which had a predicted error estimate of 0.27 $\log_{10}$TP units.

TP GCD+CHEM models

The explained variance increased by 20% to 30% when the chemical variables were included in the models (Fig. 4). The nemoral model had the highest predictive power (78%) and a predictive vs. measured regression line had a slope close to 1, but the nemoral models had the highest error estimate (RMSEP 0.20) of the TP-GCD+CHEM models. One reason that the error estimate was high compared to the other regions, might be that the nemoral model was based on few lakes compared to the other ecoregional models.

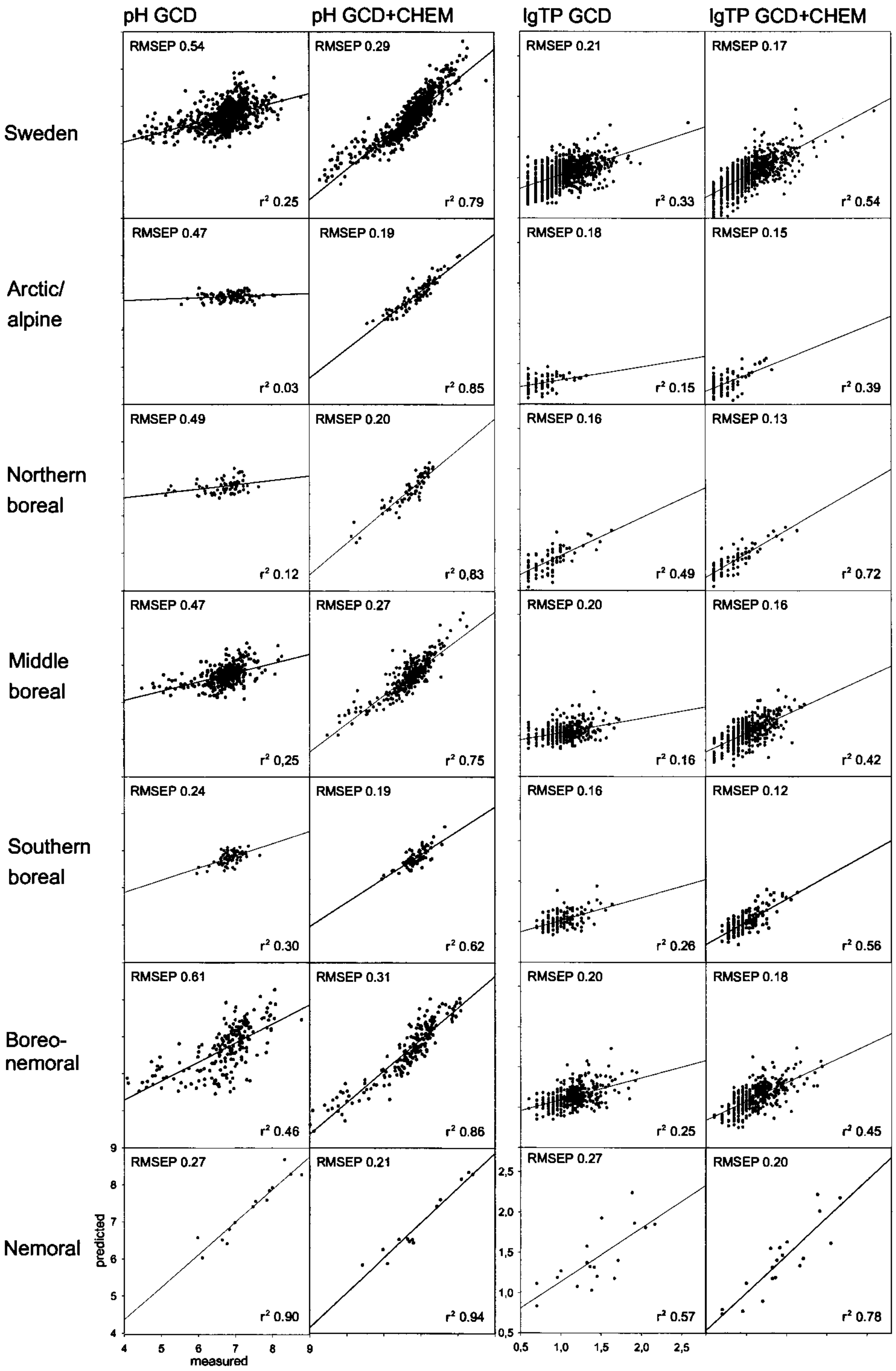

Figure 4. Predicted vs. measured values for the pH and the $\log_{10}$TP models' validation data sets. The explained variance (r^2) and the prediction error values RMSEP are given for the whole-country models (Sweden) as well as the models for the ecoregions.

PLS models and predictor variables

In the pH-GCD models, pH was generally positively correlated with catchment area and lake surface area along the first PLS-component (Table 4). Catchment area was included in all models except the arctic/alpine model where deciduous and spruce forests contributed to most of the variation along the first PLS-component. For the southern regions (boreonemoral and nemoral), where most of the agricultural activity is prevalent, arable land was (not surprisingly) an important variable. Y-coordinates or a longitudinal gradient was included in all ecoregion models except those developed for the arctic/alpine and the nemoral regions. Sulfur deposition that generally affects Sweden across a longitudinal gradient (highest deposition in the west and lowest in the east) was included in the whole country, middle boreal, and boreonemoral models. Mire was negatively correlated with pH for the whole-country, middle boreal region and southern boreal models.

For pH-GCD+CHEM-models, calcium concentration [Ca] was the most important predictor (Table 4). Besides [Ca], pH was positively correlated with lake surface area, alkalinity, ANC, and magnesium concentration along the first PLS-component. Deciduous forests were important descriptors in the arctic/alpine region (high regression coefficient), whereas arable land was more important in the nemoral region.

Total phosphorus was generally positively correlated with arable land and negatively correlated with altitude along the first component of the GCD models (Table 4). Altitude was an important variable along the first component for five of the seven models, whereas arable land was included in four of the seven models. Arable land seemed to be most important for the southern boreal and the nemoral regions. Latitude was also shown to be an important predictor of TP in the nemoral region. TP was also seemingly correlated with longitude (Y-coordinates) for the northern and the middle boreal regions. Longitude, together with catchment area, altitude, freshwater, and sulfur deposition, contributed to most of the variation along the first PLS-component for the boreonemoral model. Besides altitude, catchment area and land cover classified as mire were important predictors for the arctic/alpine region. Forest, high-mountain, high-mountain coniferous forest, and lake surface area represented the largest variation along the first PLS component for the northern boreal region.

In the TP-GCD+CHEM-models, water color (absorbance of filtered water or a proxy for dissolved organic carbon) and TOC, together with the nutrients NH_4-N and TN, were found to be important descriptor variables (Table 4). However, the importance of individual variables varied between the ecoregions. For example, for the northern boreal region, water color, high-mountain, and pH are important predictors, whereas [K], water color, and arable land seem to be the most important predictors for the southern boreal region.

Table 4. Variables with the highest regression coefficients, hence those that contributed most to the variation along the first PLS-component (the first component explained most of the variation for models). The variables listed do not necessarily represent all variables in the models. The optimal numbers of PLS-components for each model are also given.

TP GCD models				**TP GCD+CHEM models**		
	PLS comp	**Variables with highest regression coefficient**	**Variation along PLS component 1**	**PLS** comp	**Variables with highest regression coefficient**	**Variation along PLS component 1**
Sweden (Whole country)	4	Arable-land(0.21), Altitude(-0.19), Y-coord.(0.17), High-mountain(-0.13), Mire(0.13), Mean run off(-0.12), Bedrock outcrops(-0.12), Freshwater(-0.11)	Arable-land vs. altitude	3	AbsF(0.21), TOC(0.18), Y-coord.(0.18), TN(0.16), Arable-land(0.14), K(0.13), NH_4 (0.12)	TOC, AbsF, TN, K, Mg, arable-land, and NH_4 vs. X-coord. and altitude
Arctic/alpine	2	Altitude(-0.34), Catchment area(-0.25), Freshwater(-0.17), Mire(0.17), Nitrogen dep. (-0.12), Mean run off(-0.07), Lake surface area (-0.06), High-mountain coniferous forest(-0.04)	Mire vs. altitude and catchment area	3	AbsF(0.21), Si(0.18), TOC(0.15), TN(0.15), NH_4(0.12), pH(-0.11), Catchment are (-0.10)	TOC, AbsF, TN, and NH_4, vs. altitude and catchment area
Northern boreal	6	Y-coord.(0.59), X-coord.(-0.52), Lake surface area(-0.32), Mean run off(-0.22), High-mountain coniferous forest(-0.20), Spruce(0.13), Sulfur dep.(0.13), High-mountain(-0.13)	Forest vs. high-mountain, high-mountain coniferous forest, and lake surface area	2	AbsF(0.39), High-mountain(-0.25), pH(-0.23), NH_4(0.15), Nitrogen-dep.(0.11), Mean run off(-0.10)	AbsF vs. mountian and pH
Middle boreal	3	Altitude(-0.22), Y-coord.(0.18), Mire(0.13), Lake surface area(-0.12), Sulfur dep.(0.09), Freshwater(-0.08), Forest clear-cut(0.08), Bedrock outcrops(-0.08), Arable-land(0.08)	Y-coord. and arable-land vs. altitude	5	AbsF(0.43), NH_4(0.30), Y-coord. (0.15), NO_2+NO_3(-0.13), K(0.11), Mg(0.11), Catchment area(0.10), pH(-0.10)	AbsF, NH_4, and TN vs. pH and freshwater
Southern boreal	2	Arable-land(0.39), Altitude(-0.21), Freshwater(-0.15), Bedrock outcrops(-0.13), Lake surface area(-0.12), Y-coord.(0.09)	Arable-land vs. altitude	3	K(0.40), AbsF(0.36), Arable-land(0.21), SO_4(-0.18), Mg(0.17), Si(-0.17), Y-coord. (0.16), NH_4(0.15), Ca(-0.14)	K, arable land, Mg, TN, AbsF, and NH_4
Boreo-nemoral	4	Catchment area(0.29), Freshwater(-0.25), Lake surface area(-0.22), Altitude(-0.18), Bedrock outcrops(-0.18), Y-coord.(0.18), Forest(-0.16), Sulfur dep.(-0.13)	Y-coord. and catchment area vs. altitude freshwater, and sulfur dep.	3	AbsF(0.36), TN(0.24), Mg(0.20), K(0.20), Y-coord.(0.15), Alk(0.14), Catchment area(0.13)	TN, Mg, K, and AbsF
Nemoral	2	X-coord.(-0.39), Arable-land(0.22), Forest uptake of base cations(-0.22), Populated area(-0.12), Forest(-0.11)	Arable-land vs. X-coord, forest uptake of base cations, and forest	2	TOC(0.29), Alk(0.20), K(0.19), SO_4(-0.18), pH(0.15), Mg(0.12), TN(0.11), Bedrock outcrops(-0.11)	Alk, K, TOC, Mg, pH, and TN vs. pine and bedrock outcrops

Table 4 continued

pH GCD models				pH GCD+CHEM models		
	PLS comp	Variables with highest regression coefficient	Variation along PLS component 1	PLS comp	Variables with highest regression coefficient	Variation along PLS component 1
Sweden (Whole country)	2	Sulfur dep.(-0.24), Lake surface area (0.21), Forest(-0.19), Catchment area (0.18), Mire(-0.15), Arable-land (0.10), Mean run off(-0.10), Altitude(-0.05)	Mire vs. lake surface area, and catchment area	5	Ca (0.85), Mg (0.26), SO_4 (-0.23), Cl (-0.16), Lake surface area (0.14), AbsF (-0.14)	AbsF vs. Ca, Alk, and ANC
Arctic/alpine	1	Deciduous forest(0.08), Spruce(-0.07), Altitude(-0.04), Forest clear-cut(-0.04), Pine(-0.04), Sorted sediments(0.04), Till (-0.04)	Deciduous forest vs. spruce	3	Ca (0.36), Mg (0.27), Cond. (0.27), Deciduous forest (0.19), NH_4 (-0.15), AbsF (-0.12)	Ca, Cond., and Mg
Northern boreal	2	Catchment area(0.18), Till(-0.10), Bedrock outcrops(-0.09), Lake surface area(0.08), Y-coord.(-0.06), High-mountain(0.05), Sorted sediments(0.04)	Till vs. catchment and lake surface area	5	Ca (0.65), Mg (0.31), TP (–0.26), AbsF (-0.24), Altitude (0.20)	Ca, Cond., Alk, ANC, Mg, Na vs. TP, AbsF, TOC
Middle boreal	2	Mean run off(-0.20), Y-coord.(-0.20), Catchment area(0.20), Sulfur dep.(-0.19), Lake surface area(0.17), Nitrogen dep. (-0.10), Spruce(0.09), Freshwater(0.07)	Sulfur dep. and mean run off vs. catchment and lake surface area	5	Ca (0.62), Cl (-0.24), SO_4 (-0.24), AbsF (-0.23), Na (0.21), Mg (0.21), Lake surface area (0.17)	Ca, Alk, and Mg vs. AbsF, TP, Nitrogen dep.
Southern boreal	2	Lake surface area(0.29), Mire(-0.24), Catchment area(0.16), Y-coord.(0.12), Bedrock outcrops(-0.09), Altitude(-0.05)	Altitude and mire vs. lake surface area, catchment area, and Y-coord.	3	Ca (0.30), Lake surface area (0.26), Alk (0.23), NO_2+NO_3 (–0.20), ANC (0.17), Catchment area (0.13), Cl (-0.12), SO_4 (-0.12),	Ca, Alk, ANC, and Cond. vs. NO_2+NO_3
Boreonemoral	3	Y-coord.(0.32), Lake surface area(0.23), Arable-land(0.14), Catchment area(0.13), Mire(-0.10), Nitrogen(-0.09) and sulfur dep.(-0.08), Pasture-land(0.06)	Sulfur dep., nitrogen dep., and mire vs. Y-coord., lake and catchment area	5	Ca (1.01), SO_4 (-0.24), TOC (-0.23), Cl (-0.23), Mg (0.23)	Ca, Mg, K, Alk, ANC, SO_4, and lake surface area vs. Si and TOC
Nemoral	2	Catchment area(0.33), Forest(-0.29), Arable-land(0.27), X-coord.(-0.27), Peat land(-0.03), Mineral soils (0.03)	Forest and X-coord. vs. arable-land and catchment area	3	Ca (0.32), Alk (0.26), Lake surface area (0.24), Arable-land (0.22), TP (0.21), Deciduous forest (0.19), Mineral soils (-0.13)	Ca, Alk, and arable-land

Discussion

An underlying view of landscape ecology is that the spatial position of an ecosystem within the landscape influences the properties of that ecosystem. Although the mechanisms influencing specific lake ecosystems are diverse and not always clearly understood, there is little doubt that landscape effects are important in determining average physical and chemical conditions. We expected that ecoregion-based models would have higher predictive power (high r^2) and lower prediction error (RMSEP) than whole-country models. This assumption was based on the principle that ecoregions, as integrators of regional patterns in climate, topography, land use and natural vegetation, would partition part of the natural variability. This seems to be true for the southern part of Sweden, and especially for the nemoral region. pH and TP models developed for the nemoral region had higher predictive power than the model for the whole country and the other ecoregional models. For the pH and TP nemoral models, GCD variables explained 90% and 57% of the variation, respectively. The total variation (SD) within a region seemed to affect the predictive power. For example, total variation within the arctic/alpine region was low resulting in low predictive power, whereas the total variation within the nemoral region was high resulting in high predictive power. Overall for both TP and pH, the predictive power of the ecoregional models based on GCD variables was slightly better or about the same as the predictive power for the whole-country models. The prediction error, RMSEP, was usually similar or somewhat lower for the GCD models than the whole-country model. An exception was the TP nemoral model that had a higher RMSEP than the whole-country model. One reason might be that there were few lakes in the nemoral region compared to the other regions and the whole country.

Some of the pH models seemed to be nonlinear across the pH-gradient, with a breakpoint at approximately pH 6. This was especially evident for the middle boreal, boreonemoral and whole-country models, and it implies that it might be better to model pH above and below pH 6 instead of across the whole interval. Preliminary studies for the middle boreal region supports this conjecture. The prediction power increased from 25% to 38% and from 25% to 30% modeling low (4.43 – 5.42) and high (> 5.42) pH values, respectively. The prediction error decreased from 0.47 to 0.21 and 0.35 for models with low and high pH values, respectively.

Both catchment area and lake surface area were shown to be important in explaining the among-lake variance in pH. The only exception was found for the arctic/alpine region where deciduous and spruce forests explained the variation along the first PLS-component. An altitude gradient for deciduous and spruce forest is prevalent within the arctic/alpine region, with spruce at lower and deciduous forest at higher altitudes. However, the predictive power of the arctic/alpine model is low (r^2 = 0.03). Besides area to pH relationships, other variables such as land use (arable, forest) and cover (mire and deciduous forest), and soil type (till) are correlated with the first PLS-components for the various pH-GCD models. For the nemoral model, forest and arable land explained a large part of the variation in pH. For the boreonemoral region, Y-coordinates (longitude) and sulfur and nitrogen deposition were important predictor variables, which can be expected since the deposition of sulfur and nitrogen is greatest in the western parts of the country, and decreases in an easterly direction across the boreonemoral ecoregion.

Besides catchment and lake surface area, altitude is also an important descriptor of TP concentrations. Land use/cover parameters important in the northern part of the country seemed to be mire, forest, and high-mountain coniferous forest. In the southern regions one important predictor seemed to be the amount of arable land. Similar to the pH, arable land and forest explained a large part of the variation in the TP nemoral model. The nemoral region is probably the area in Sweden with most intense agriculture.

Our findings agree with a number of earlier studies regarding the importance of land use and cover on surface water chemistry (Liegel et al., 1991; Nilsson and Hakanson, 1992; Kernan, 1995). Håkanson (1995) was able to explain 56% of the variance in TP in small glacial lakes in Sweden using map information alone. He found that the most important descriptors were the percentage cover of rocks and open land (often cultivated) normalized for the distance from the lake. Our GCD models often explained less than 50% of the TP variation in the regions, and similar to Håkanson (1995) arable land, which is open cultivated land, was one of the most important predictors. The amount of arable land has also been shown to be a significant albeit weak (r^2 values < 0.23) predictor of lake pH (Nilsson and Håkanson 1992). In our pH-GCD models arable land is an important predictor in the southern part of Sweden. Liegel et al. (1991) were able to explain 32% of the variance in lake water pH. These authors showed, for example, that lake pH in the northeastern U.S. was correlated with agricultural land use, the amount of wetlands, open water, and horticulture, as well as deposition and precipitation and runoff. In our study pH-GCD-models, except the arctic/alpine and northern boreal, explained between 25% and 90% of the variation. The best predictors were catchment area and lake surface area, but also variables like mire, sulfur-deposition, arable land, and mean run-off were important predictors for some of the models.

As expected, adding the chemical variables improved the models considerably. The most obvious example was the arctic/alpine pH model; including the chemical variables increased the predictive power from 3% to 85%. The single most important predictor variable for the pH-GCD+CHEM models was [Ca], compared to water color for the TP-GCD+CHEM models. For some TP-GCD+CHEM models, TOC and the nutrients NH_4-N and TN were also important. Interestingly, these chemical variables, [Ca], water color, TOC, NH_4-N, and TN all reflect conditions in the catchment. Calcium concentration is strongly related to the soil chemistry and geology within the catchment, whereas water color, TOC, NH_4-N, and TN are dependent on land use and vegetation cover (e.g. forest agriculture, and mire) and water residence time. Alan et al. (1998) studied the relation between stream chemistry and land use data. They showed that chloride, nutrients (NO_3-N and TP), ANC, and base cations (e.g. [Ca]) were the chemical variables most strongly related to land use/cover. Hence, adding other catchment characteristics like soil chemistry, that partly reflects the variability in geology within a catchment, should improve the predictive ability of our models. Also, improving the classification of the land use variables (e.g. forest- and arable land that are already included in the models), and adding other lake-specific characteristics such as residence time might help to improve the predictive ability of the GCD-models.

There is no doubt that the landscape affects the physical and chemical condition of a lake, though the mechanism(s) behind these effects are more difficult to describe. In this study we can conclude that separating the data set in ecoregions works well if the total variation of the data (pH and TP) within each region is high (e.g. the nemoral region). In general, models developed for the southern parts of the country performed better than those developed for the north. In our study, the correlation between chemical variables within the lake is much stronger than the correlation between chemical variables and GCD variables (Geographic, Catchment classifications, and Deposition variables). This was expected, especially for pH since the correlation between [Ca] and pH is strong. For future studies, however, it is important to take soil chemistry and weathering within the catchment into consideration when modeling chemical variables in lakes.

Acknowledgements

We thank the many people who assisted in the planning, sampling and processing of the large number of samples collected in the 1995 national lake survey; Anders Grimvall for commenting the PLS-models; and Anders Wilander for comments on an earlier draft.

References

Alan, H. T., Stoddard, J. L., and Burch Johnson, C., 1998. The relationship between stream chemistry and watershed land cover data in the mid-atlantic region, U.S. Water, Air and Soil Pollution 105, 377-386.

Anonymous, 1987. Fältinstruktioner för ståndortskarteringen av permanenta provytor vid riksskogtaxeringen 1986. Department of Forest Soils, Swedish University of Agricultural Sciences, Uppsala.

Anonymous, 1997. Instruktioner för fältarbete vid riksskogstaxeringen år 1996. Department of Forest Resource Management and Geomatics, Swedish University of Agricultural Sciences, Umeå.

Esbensen, K., Schönkopf, S., and Midtgaard, T., 1996. Multivariate analysis in practice. Wennbergs Trykkeri AS, Trondheim, Norway.

Geladi, P., and Kowalski, B. R., 1986. Partial least squares regression: A tutorial. Analytica Chimica Acta 185, 1-17.

Gustafsson, L. and Ahlén I. (eds.)., 1996. Geography of plants and animals. Almqvist & Wiksell International, Stockholm.

Hawkins, C.P., Norris R.H., Gerritsen J., Hughes R.M., Jackson S.K., Johnson R.K., and Stevenson R.J., 2000. Evaluation of landscape classifications for biological assessment of freshwater ecosystems: synthesis and recommendations. Journal of North American Benthological Society 19(3), 541–556.

Henriksen, A., Skjelkvåle, B. L., Lien, L., Traaen, T. S., Mannio, J., Forsius, M., Kämäri, J., Mäkinen, I., Berntell, A., Wiederholm, T., Wilander, A., Moiseenko, T., Lozovik, P., Filatov, N., Niiniokja, R., Harriman, R. & Jensen, J. P. 1996. Regional lake surveys in Finland - Norway – Sweden – Northern Kola – Russian Karelia – Scotland – Wales 1995. Coordination and design. NIVA Report 40/1996.

Hughes, R. M., and Larsen, D. P., 1988. Ecoregions: an approach to surface water protection. Journal of the Water Pollution Control Federation 60, 486–493.

Håkanson, L., 1995. Models to predict lake annual mean total phosphorus. Journal of Aquatic Ecosystem Health 4, 25-58.

Johnson, R.K., 1999. Regional representativeness of Swedish reference lakes. Environmental Management 23, 115–124.

Kernan, M., 1995. The use of catchment attributes to predict surface water critical loads: A preliminary analysis. Water, Air and Soil Pollution 85, 2479-2484.

Kratz, T. K., Webster, K. E., Bowser, C. J., Magnuson, J. J., and Benson, B. J., 1997. The influence of landscape position on lakes in northern Wisconsin. Freshwater Biology, 37, 209–217.

Kratz ,T. K., and Frost, T. M., 2000. The ecological organisation of lake districts: general introduction. Freshwater Biology 43, 297–299.

Liegel, L., Cassell, D., Stevens, D., Shaffer, P., and Church, R., 1991. Regional characteristics of land use in northeast and southern blue ridge province: Association with acid rain effects on surface-water chemistry. Environmental management 15, 269-279.

Magnuson, J. J., and Kratz, T. K., 2000. Lakes in the landscape: approaches to regional limnology. Verhandlungen der Internationalen Vereinigung für Theoretische und Angewandte Limnologie 27: 74–87.

Nilsson, Å., and Håkanson, L., 1992. Relationships between drainage area characteristics and lake water quality. Environmental Geology and Water Sciences 19, 75-81.

Nordic Council of Ministers, 1984. Naturgeografisk regionindelning av Norden. Nordiska ministerrådet 1984.

Soranno, P. A., Webster, K. E., Riera, J. L., Kratz, T. K., Baron, J. S., Bukaveckas, P. A., Kling, G. W., White, D. S., Caine, N., Lathrop, R. C., and Leavitt, P. R., 1999. Spatial variation among lakes within landscapes: Ecological organization along lake chains. Ecosystems 2, 395–410.

Unscrambler ver. 6, 1996. Camo AS, Trondheim, Norway.

Wilander, A., Johnson, R.K., Goedkoop, W., and Lundin, L., 1998. Riksinventering 1995. En synoptisk studie av vattenkemi och bottenfauna i svenska sjöar och vattendrag. Rapport 4813, Stockholm, Swedish Environmental Protection Agency. (In Swedish)

Representativity of a mid-lake surface water chemistry sample

Elisabet Göransson, Richard K. Johnson, and Anders Wilander

Department of Environmental Assessment, Swedish University of Agricultural Sciences, Box 7050, SE-750 07 Uppsala, SWEDEN

Abstract

A mid-lake sample is the conventional sampling method used in Swedish lake monitoring programs. Addressing the question of how representative a mid-lake sample is of in-lake conditions is important for interpreting the long term changes of temporally extensive monitoring programs. Our hypotheses were that: (i) where a sample was taken within a lake would affect its representativity and (ii) a mid-lake sample could be considered as representative of average surface water chemistry. The study focused on the surface water chemistry of 34 Swedish lakes that are a part of the Swedish long-term lake monitoring program. To obtain a wide diversity of lake types, the lakes were spatially-stratified across three main ecoregions; arctic/alpine, northern coniferous forest, and southern mixed forest region. In addition, recognizing that mean depth and water residence time also affect in-lake chemistry, lakes were also grouped according to these factors. Lastly, cognizant that sampling season will affect water chemistry, we selected two seasons, representative of high (late summer or August) or low (autumn or October) biological activity. We could conclude that where a sample was taken within a lake would affect the sample's representativity of lake water surface chemistry. A mid-lake sample could be considered more representative than a random sample. Of the eight water chemistry metrics studied and 34 and 32 lakes respectively, only 3.9% (August) or 3.6% (October) of the total number of tests varied significantly. Our study showed that the ideas behind the mid-lake sampling technique are valid. A mid-lake surface sample for water chemistry is a sufficient sampling method and should continue to be used in lake monitoring and assessment programs.

Introduction

Samples can seldom be considered independent, but are often correlated in both space and time. Terrestrial ecologists have long recognized that spatial correlation needs to be taken into account in data interpretation. Analyses of spatial variation within soil sciences are relatively well established. Examples of spatial studies within soil sciences range from infiltration of soil waters and transport of solutes within the unsaturated zone to characterizing spatial variability of microbiological and physico-chemical soil properties (e.g. Berndtsson, 1987; Yasuda et al., 1994; Goovaerts, 1998). Techniques have also been developed using the spatial variation of soil properties (e.g. critical thresholds of heavy metals) for environmental management (Oliver and Webster, 1996). In aquatic ecology, and in particular lake ecology, much less effort has been focused on how spatial variation affects sample independence. Indeed, many lake ecologists would argue that lakes are independent entities, simply because they appear to be "isolated" ecosystems. However, studies by Kratz et al. (1991) show that lakes within a region may behave similarly through time; e.g. as to temporal coherence.

Lake chemistry varies markedly both temporally and spatially (Wetzel, 1983), and sampling designs are often implemented taking into account this variability (e.g. Wetzel and Likens, 1991). Wetzel and Likens (1991) concluded, for example, that a water chemistry sample taken from the water column overlying the central depression of a lake is the conventional sampling method. Although vertical (water column) variation is usually accounted for in sampling, much less attention is given to surficial or horizontal spatial variability. Biologically active variables such as phyto- and zooplankton assemblages and to some extent biologically-sensitive chemical variables (e.g. biologically available phosphorus) may, however, exhibit considerable spatial (horizontal) variability (Wetzel, 1983). Blomqvist (2001) argues that a volume-weighted sampling design (consisting of a composite sample) is best suited for capturing the vertical and horizontal variation found within lake ecosystems.

Addressing the question of how representative a mid-lake sample is of in-lake conditions is important for interpreting the long term changes of temporally extensive monitoring programs. A number of factors may, however, affect the representativity of a mid-lake water sample. Lake morphometry (e.g. depth, shoreline development and volume), distance to shoreline, relative importance of surface and groundwater inputs and biological activity may singly or combined affect the spatial variation within a lake. In addition, temporally important factors may also affect water chemistry, such as within- and among-year variability in precipitation, temperature, wind speed and direction. In particular, the size and shape of a lake may strongly influence spatial variance. For example, small lakes with short residence times may appear more stochastic as they respond more readily to external influences, whereas large lakes may show greater inertia, confounding interpretation of early-warning signals.

Our hypotheses are that: (i) where a sample is taken within a lake will affect its representativity and (ii) a mid-lake sample can be considered as representative of average surface water chemistry. This study focused on the surface water chemistry of 34 Swedish lakes that are a part of the Swedish long-term lake monitoring program (Wilander, 1997). In order to obtain a wide diversity of lake types, the lakes were spatially-stratified across the three main ecoregions of Sweden. In brief, the three ecoregions used here can be characterized as: (i) arctic/alpine, (ii) boreal coniferous forests in the north and (iii) the mixed forest and open landscape in the south. In addition, recognizing that depth and water residence time also affect in-lake chemistry, lakes were also grouped according to these factors. Lastly, cognizant that sampling season will affect water chemistry, we selected two seasons that are representative of high (late summer or August) or low (autumn or October) biological activity.

Material and methods

Study area

The lakes included in this study can be generally described as relatively small (mean surface area = 1.12 km^2, range 0.1 to 6.65 km^2) and mesotrophic (Table 1). Total phosphorus averaged 13.4 µg/l, and lakes ranged from nutrient poor (total phosphorus ≈ 3 µg/l) to nutrient rich (ca 50 µg/l). Both water color and pH varied markedly across the lakes included in this study. Water color ranged from 0.003 absorbance units per 5 cm of filtered water to 0.79 (corresponding to 2 – 400 Hazen units) and pH ranged from 4.6 to 7.5.

Table 1. Mean, minimum, and maximum values for selected physico-chemical variables calculated for the 32 lakes sampled both in August and October 1999.

Variable	Mean	Min	Max
pH	6.51	4.59	7.46
Alkalinity (meq/l)	0.143[a]	-0.037[a]	1.017[a]
Conductivity (mS/m)	4.45	1.41	15.2
Absorbance filtered (420 nm)	0.156	0.003	0.671
Absorbance unfiltered (420 nm)	0.195	0.003	0.792
Temperature (°C)	12.2	5.8	16.4
Total phosphorus (µg/l)	13	3	48
NO_2+NO_3-N (µg/l)	20	2	87
Lake area (km^2)[b]	1.12	0.1	6.65

[a] based only on October samples

[b] based on all 34 lakes

Sample classification

The lakes included in this study are part of the Swedish national lake monitoring program, consisting of some 100 lakes monitored four times per year for physico-chemical variables (Wiederholm and Johnson, 1997). Information on mean depth and water residence time was only available for 77 of the 100 lakes. Thirty-four of these remaining 77 lakes were selected so as to cover a wide spectrum of lake types. This was done by classifying lakes according to: (i) geographic region (3 major ecoregions), (ii) water depth (2 categories, mean depth ≤ 3.5 m or > 3.5 m) and (iii) water residence time (2 categories, residence time ≤ 1 year or residence time > 1 year). Since more lakes are situated in the southern parts of the country, which are more densely populated, 5 lakes were selected for each classified group in the southernmost region (i.e. the southern boreal, boreonemoral and nemoral ecoregions). In contrast, due to fewer monitoring lakes situated in the north, it was difficult to fill all the groups in the arctic/alpine region and the northern coniferous forest regions. Hence, the number of selected lakes in these two ecoregions varied with availability of lakes (Table 2).

Table 2. Number of lakes in each classified group.

Residence time	> 1 year		≤ 1 year		
Mean depth	> 3.5 m	≤ 3.5 m	> 3.5 m	≤ 3.5 m	Total
Arctic/alpine region	1	—	2	2	5
Northern coniferous region	3	—	3	3*	9
Southern mixed forest region	5	5	5	5	20
Total	9	5	10	10	34

* Two lakes were not sampled in October

Ecoregion classification

The ecoregions used here correspond closely to those recommended by the European Water Framework Directive (European Commission, 2000). These regions were originally based on macroinvertebrate distribution studies done in the 1960s (e.g. Illies, 1966). A brief description of the ecoregions is given here and more detailed information can be found in the Nordic Council of Ministers (Anonymous, 1984) and Gustafsson and Ahlén (1996). The three regions used in this study are composed of several ecoregions and hereafter these three regions will be referred to as the arctic/alpine region, northern coniferous region, and southern mixed forest region (Fig. 1). The arctic/alpine region in the north is characterized by relatively low mean annual temperature, ranging from –3 to 2 °C and growing season from < 100 – 120 days. The northern coniferous forest region consists of the northern and middle boreal ecoregions and is characterized by coniferous woodlands, mean annual temperature ranging from 1 to 4 °C, and growing season from 120 to 160 days. The southern mixed forest region is composed of the southern boreal, boreonemoral, and the nemoral ecoregions. The characteristics of the southern mixed forest region are spruce and pine forest in the north, deciduous forest in the south, growing season from 140 – 180 days in the north to 180 – 210 days in the south, and mean annual temperatures are 4 to 5 °C in the north and greater than 6 °C in the south.

Sampling design

Ten samples were taken from each lake from quadrates randomly distributed over the whole lake surface (Fig. 2). Each lake was covered with a grid matrix, with the scale of the grid matrix adjusted so that each lake fit onto one 29.5 x 21 cm sheet of plain white paper. The number of lakes in each grid-size were: 2 lakes with grid-size 25 x 25 m; 1 lake with 40 x 40 m; 12 lakes with 50 x 50 m; 16 lakes with 100 x100 m; and 3 lakes with 200 x 200 m. The quadrates were numbered, starting in the north end of each lake and moving southward in west to east direction. All quadrates with more than two thirds of its area in the lake (visually determined) were numbered. The total number of quadrates per lake was divided by 10 to determine the interval between sampling sites. For example, if 64 quadrates covered a lake, then every 6th quadrate would be sampled. The first sample was randomly placed among the first 6 quadrates and the remaining samples were even dispersed thereafter. If more than 10 quadrates were situated on a lake sampling grid, then quadrates were randomly removed, with however, the constraint that two consecutive samples could not be removed.

One sample quadrate was selected to represent the mid-lake sample. The quadrate closest to the “center” of the lake according to lake surface area and where the maximum depth of the lake was located was chosen as the mid-lake quadrate.

Field sampling

Water column samples were taken from each lake in August and October of 1999. All 34 lakes were sampled in August, but unfortunately due to logistical problems only 32 of the 34 lakes were sampled in October. Visual triangulation or GPS was used to establish position, one surficial (0.5 m depth) water sample was collected with a Ruttner sampler from the middle of each quadrate. For sampling quadrates close to the shoreline, the water sample was to be taken as far from the shore as possible but still within the square. For each sample, the water temperature, depth, occurrence of vegetation, distance to closest vegetation (if any), and occurrence of rocks above the water surface was recorded. The approximate wind speed and direction was also recorded at the time of sampling.

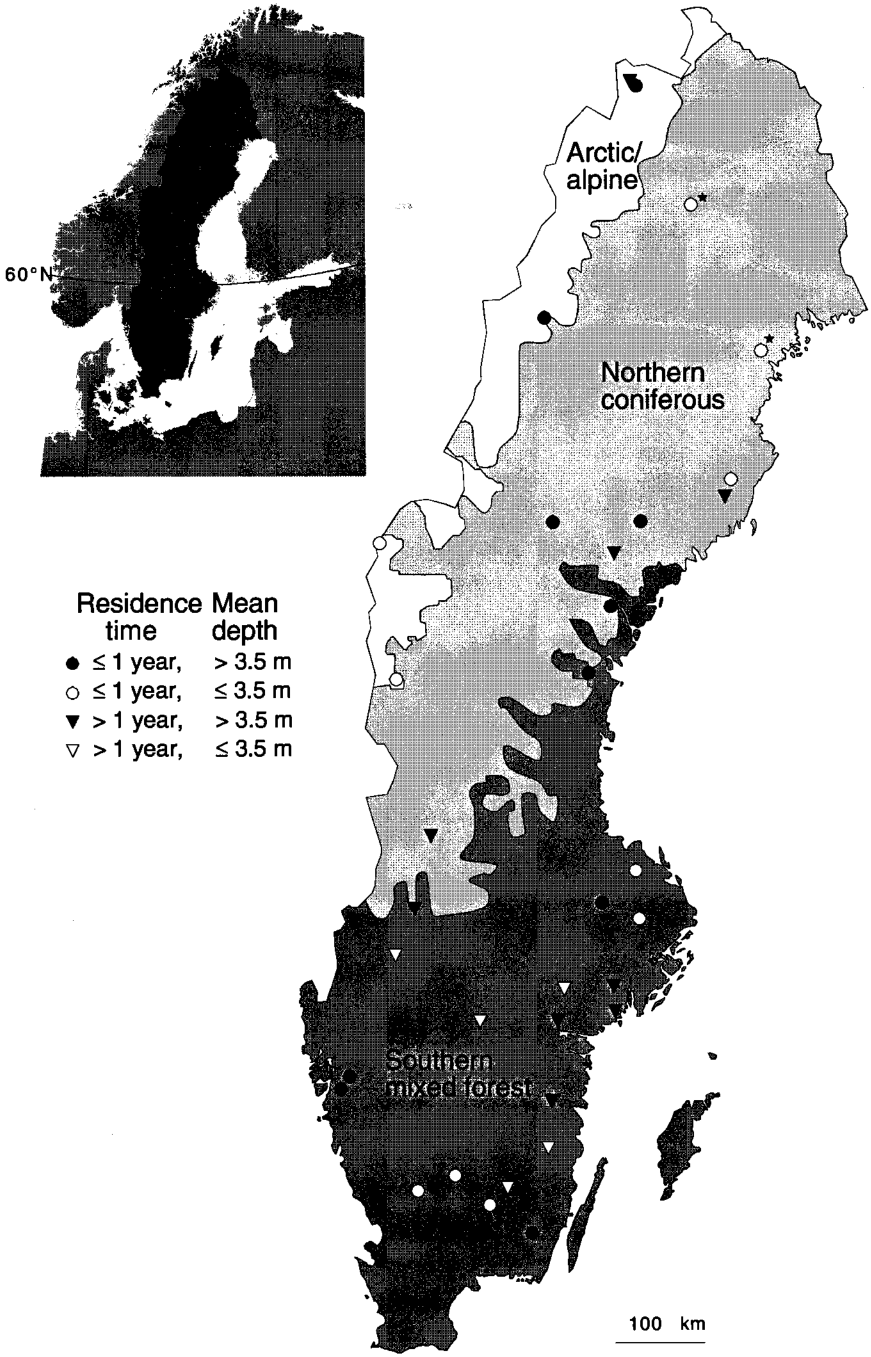

Figure 1. Classification of the 34 lakes according to geographic regions (3 ecoregions), mean water depth, and water residence time. ★ - the two lakes not samples in October.

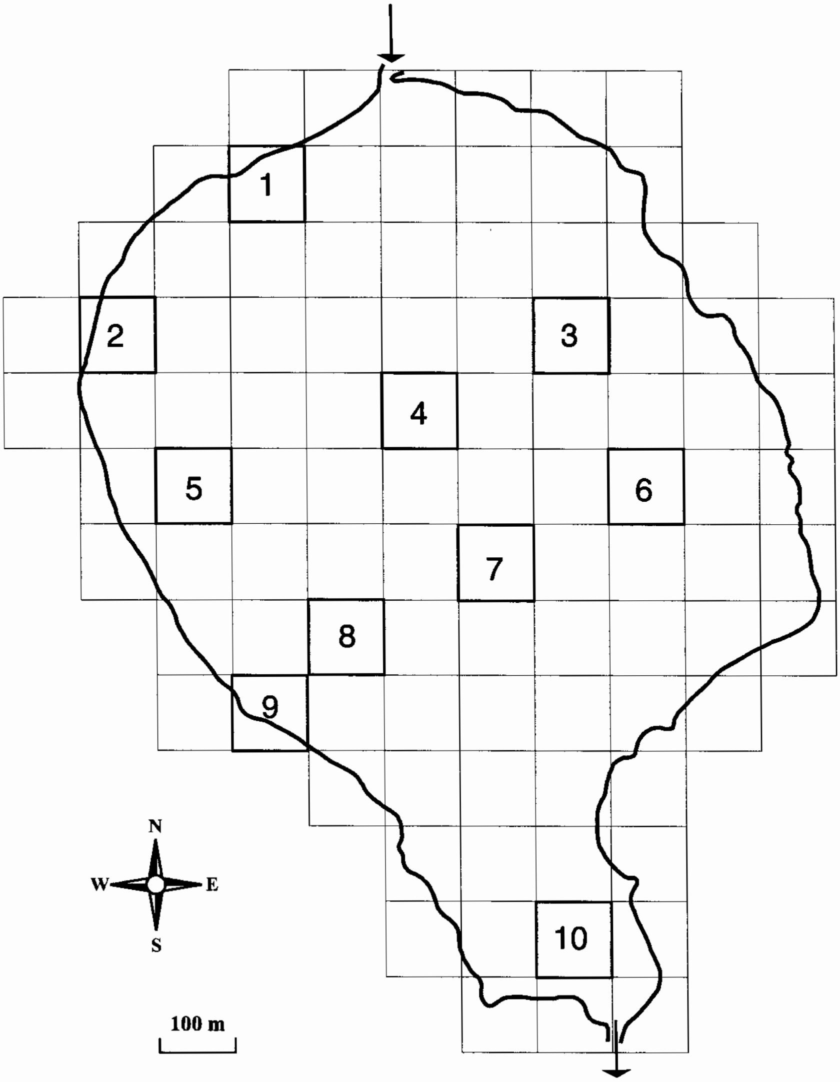

Figure 2. Schematic diagram showing an example of the sampling design. The lake shown is lake Latnjajaure in the arctic/alpine region.

Chemical analyses

Selected water chemistry variables were analyzed to determine the representativity of a single, mid-lake sample. These variables consisted of metrics indicative of acidification (i.e. pH and alkalinity, meq/l) and eutrophication (total phosphorus and nitrite-nitrate nitrogen, NO_2+NO_3-N µg/l, hereafter referred to as inorganic nitrogen). Conductivity (mS/m at 25 °C) and absorbance of filtered (420 nm) water (i.e. water color), were assumed to reflect weathering and leaching of inorganic and organic matter within the catchment. Absorbance (420 nm) of unfiltered water was taken to reflect suspended matter from tributaries and phytoplankton. Alkalinity was analyzed for all lakes sampled in October, but only for 17 lakes sampled in August. Water temperature (°C) was also included in the analyses. All analyses were done by the SWEDAC (Swedish Board for Accreditation and Conformity Assessment) certified laboratory at the Department of environmental assessment, Swedish University of Agricultural Sciences, using national or international standard methods (Wilander et al., 1998)

Statistical analyses

Water chemistry of the two sampling seasons (August and October) was compared using a paired *t*-test to determine if sampling season had an effect. The average value for each lake and parameter was calculated for the two seasons and compared with $\alpha = 0.05$. The paired *t*-test was calculated for the 32 lakes sampled both in August and October.

Large-scale, spatial differences in water chemistry were analyzed by a three-factor unbalanced analysis of variance (ANOVA), with ecoregion (three ecoregions) residence time (two categories) and mean depth (two categories) as fixed factors. The additive effects of these three factors, as crossed factors or interaction terms (i.e. ecoregion x residence time, ecoregion x mean depth, and residence time x mean depth) was also determined. Post-hoc Tukey's tests were carried out to investigate significant differences among "ecoregions", "residence time" and "mean depth". The ANOVAs were run using MINITAB (MINITAB, 1997).

A two-sample *t*-test was used to determine if the mid-lake sample differed from the nine remaining samples. This test was performed on each lake and variable according to the formula below. Significant level was set to $\alpha = 0.05$.

$$t = \frac{\bar{x}_c - \bar{x}_9}{\sqrt{s_9^2\left(\frac{1}{n_c} + \frac{1}{n_9}\right)}} \qquad (1)$$

where $\bar{x}_c$ = mid-lake (i.e. center) sample, $\bar{x}_9$ = mean of the remaining nine samples, s_9^2 = variance of the nine remaining samples, and n_c and n_9 = the sample sizes of the mid-lake (n = 1) and nine remaining samples (n = 9).

A jack-knifing technique was used to determine if the distance to the center part of the lake (or mid-lake sample) would affect sample representativity. This was done by sequentially comparing each sample with the remaining nine samples using a two-sample *t*-test and Bonferroni correction of α error to a significance level $\alpha = 0.005$. As the sample quadrates were similarly placed in all lakes, i.e. the first quadrate was usually situated in the northern part of the lake and the 10th quadrate was located in the southern part of the lake sample number is somewhat indicative of placement on the lake. For example, with this design sample quadrates with numbers 4 to 7 were usually situated in the middle part of the lake.

All variables were $\log_{10}$ transformed (except pH) to approximate normally distributed errors and meet to assumptions of *t*-tests and ANOVA. Alkalinity was $\log_{10}$ x +1 transformed to adjust for negative values. Since alkalinity was measured for only 17 lakes in August, alkalinity was not included in the 3-factor ANOVA analyses for August. Except for alkalinity, the only missing data was temperature for one lake in August and for three lakes in October.

Results

Seasonal effects

The paired *t*-test, comparing the average values for each lake and parameter, showed a seasonal effect for conductivity, total phosphorus, inorganic nitrogen, and temperature. There were no seasonal differences for pH, alkalinity (test based on 15 lakes), unfiltered absorbance, and water color. The average total phosphorus value for all lakes was higher in August (15 µg/l) than October (12 µg/l). On the other hand, the average value for inorganic nitrogen was lower in August (13 µg/l) than October (26 µg/l). Not surprisingly the average temperature is higher in August (16.3 °C) than October (8.0 °C). The difference between average values for conductivity was small (August = 4.26 mS/m and October = 4.51 mS/m) but the test showed that conductivity generally was somewhat higher in October than August.

Spatial classification

A 3-factor ANOVA showed that conductivity, absorbance unfiltered and filtered, total phosphorus, and temperature differed among the three ecoregions, but pH, alkalinity and inorganic nitrogen did not differ (Table 3). These patterns were especially evident for the arctic/alpine region; water chemistry in this region generally differed from the other two regions. Nutrient levels were generally lower in lakes situated in the northern compared to southern regions. Conductivity was on average twice as high in lakes situated in the southern mixed forests region compared to the two other regions. Water color (i.e. absorbance of filtered water) and suspended particulate matter (i.e. absorbance of unfiltered water) also showed marked spatial variation correlated with latitude; higher values were noted in the south compared to the north. Temperature in the arctic/alpine region averaged 10.4 °C in August, compared to 15.2 °C in the northern coniferous region and 18.4 °C in the southern mixed forest region.

In addition to ecoregion effects, other factors such as season and lake morphometry clearly affected lake chemistry. For example, aside from differences in water chemistry that could be explained by ecoregion effects, more differences were noted in August than October that could be explained by lake morphometry. In August, conductivity, absorbance both filtered and unfiltered, and temperature differed with residence time. Also significant interaction terms were noted between ecoregion and water residence time for absorbance of unfiltered water, inorganic nitrogen and temperature. In August, a significant interaction term was also noted between ecoregion and mean depth for temperature. In October, five of the eight variables tested differed with ecoregion, but only one was also affected by residence time (conductivity). A significant interaction between ecoregion and mean depth (temperature) was also noted.

Comparison of a mid-lake and nine remaining samples using a two-sample *t*-test

Water chemistry of the mid-lake sample differed from that of the nine remaining samples in some lakes. However, of a total of 254 (August) or 253 (October) possible combinations of lakes and variables, only 10 of them differed in August and 9 in October (Table 4). Total phosphorus, absorbance of filtered (August) or unfiltered (October) water, and alkalinity differed both in August and October. In addition, inorganic nitrogen and temperature differed in August and conductivity in October.

Comparison of the distance to mid lake and sample representativity by jack-knifing and a two-sample *t*-test

Distance to the center of the lake also seemed to affect sample representativity, with samples close to the center of the lake being more similar to each other than samples further from the center. Water chemistry for the first and last quadrates often differed from the other samples (Table 5). The differences were especially evident for samples taken in the first sampling quadrate. In general, samples close to the center of the lake showed fewer significant differences (i.e. quadrates 4 to 7 showed fewer differences than quadrates 1 to 3 and 8 to 10).

Table 3. A 3-factor unbalanced ANOVA calculated on the mean value for 10 samples in each lake. The three factors were ecoregion, residence time, and mean depth. The crossed effects of the three factors were also included in the analyses. For the significant factors in the ANOVA analyses a pairwise Tukey comparison was made. All variables are $\log_{10}$ transformed (alkalinity $\log_{10}X+1$). Alkalinity was not analyzed for August, since there were only data for 17 of the 34 lakes. Arctic/alpine region (1), northern coniferous region (2), southern mixed forest region (3). Residence time $\leq$ 1 year (short time = ST) or residence time > 1 year (long time = LT). Mean depth $\leq$ 3.5 meters (shallow = sh) or mean depth > 3.5 meters (deep = d).

Variable	**Significant factors** (p<0.05)	**Pairwise comparisons** Significantly different pairs (p<0.05)
	AUGUST	
pH	-	-
Conductivity	Ecoregion	1-3, 2-3
	Residence time	
Absorbance unfiltered	Ecoregion	1-2, 1-3
	Residence time	
	Ecoregion*Residence time	1LT – 1ST, 2ST, 2LT, 3ST, 3LT* 1ST – 3ST
Absorbance filtered	Ecoregion	1–2, 1-3
	Residence time	
Total phosphorus	Ecoregion	1-3
NO_2+NO_3-N	Ecoregion*Residence time	(3ST – 3LT p= 0.08)
Temperature	Ecoregion	1-2, 1-3, 2-3
	Residence time	
	Ecoregion*Residence time	1ST – 1LT, 2ST, 3ST, 3LT 1LT – 2ST, 2LT, 3ST, 3LT 2ST – 3ST, 3LT 2LT – 3LT
	Ecoregion*Mean depth	1d – 2d, 2sh, 3d, 3sh 1sh – 2d, 2sh, 3d, 3sh 2d – 3d, 3sh 2sh – 3d, 3sh
	OCTOBER	
pH	-	-
Alkalinity	-	-
Conductivity	Ecoregion	1-3, 2-3
	Residence time	
Absorbance unfiltered	Ecoregion	1-2, 1-3
Absorbance filtered	Ecoregion	1–2, 1-3
Total phosphorus	Ecoregion	1-2, 1-3
NO_2+NO_3-N	-	-
Temperature	Ecoregion	1-3
	Ecoregion*Mean depth	1sh – 2d, 3d, 3sh

* 1LT – 1ST, 1LT – 2ST, 1LT – 2LT, 1LT – 3ST, 1LT – 3LT

Table 4. A 2-sample *t*-test testing whether there was a significant difference ($\alpha \leq 0.05$) between the mid-lake sample and the remaining nine samples. The number of lakes with a significant difference is listed for each variable for both August and October.

Variable	Mid-lake sample August	Mid-lake sample October
pH		
Alkalinity	1	2
Conductivity		3
Abs. unfiltered		2
Abs. filtered	2	
Total phosphorus	2	2
NO2+NO3-N	4	
Temperature	1	
Sum	**10**	**9**

All variables are $\log_{10}$ transformed (alkalinity $\log_{10}X+1$)

Table 5. A 2-sample *t* – test comparing each sample quadrate with the remaining nine by jack-knifing method. Each column represents the sample quadrate that was compared to the other sample quadrates. The values given show the number of significant differences. P-values were adjusted by Bonferroni corrections ($\alpha \leq 0.005$).

Variable	1	2	3	4	5	6	7	8	9	10
					AUGUST					
pH	9	2					1			1
Alkalinity	2		1	1	1		3			
Conductivity	2						1			
Abs. unfiltered								1		1
Abs. filtered	7		2							1
Total phosphorus										
NO2+NO3-N	1				1			1		1
Temperature	1					1			1	1
Sum	**22**	**2**	**3**	**1**	**2**	**1**	**5**	**2**	**1**	**5**
					OCTOBER					
pH	3									2
Alkalinity	3		3	2		1			2	1
Conductivity	2			1		1		1	1	2
Abs. unfiltered	2		1	1				1		1
Abs. filtered	6	1		1						
Total phosphorus	1	1			1			1		
NO2+NO3-N	2	1								3
Temperature	4		1	1						
Sum	**23**	**3**	**5**	**6**	**1**	**2**	**0**	**3**	**3**	**9**

Lake typologies

Mid-lake chemistry was compared across twelve different lake typologies, unfortunately no lakes were represented in two classes (Table 6). Usually, only one of eight variables was significant for those lakes that had a significantly different mid-lake sample. In August, all classes but one, had at least one lake with a significantly different variables for the mid-lake sample. There were a total of 10 significant differences in 10 different lakes for August data. The distribution of these lakes across the ecoregions was; 3 of the total of 5 lakes had one significant variables in the arctic/alpine region; 4 of the 9 in the northern coniferous region; and 3 of the 20 in the southern mixed forest region. Whithin the arctic/alpine region, inorganic nitrogen was the only variable that differed significantly. Total phosphorus differed in two lakes, both in different ecoregions, but both lakes had similar water residence time (> 1 year) and depth (> 3.5 m). Water color (absorbance of filtered water) also differed in two lakes in two ecoregions. Both of these lakes had water residence times ≤ 1 year and depth ≤ 3.5 m.

For lakes sampled in October, there were a total of nine significant differences distributed in six different lakes. No differences were found for lakes situated in the arctic/alpine region. Three lakes differed in the northern coniferous region and three in mixed forest region. Among the lakes with significant variables there were lakes with both long (> 1 year) and short (≤ 1 year) water residence times. In contrast, a higher number of differences occurred in deep (> 3.5 m, n = 5) than shallow (≤ 3.5 m, n = 1) lakes.

Table 6. Lake types where the mid-lake sample differed significantly for at least one variable. Lakes were sampled in August and October 1999. Classification by ecoregions, water residence time, and mean depth.

Ecoregion	**Arctic/alpine region**				**Northern coniferous region**				**Southern mixed forest region**			
Water residence time (year)	> 1		≤ 1		> 1		≤ 1		> 1		≤ 1	
Mean depth (m)	>3.5	≤3.5	>3.5	≤3.5	>3.5	≤3.5	>3.5	≤3.5	>3.5	≤3.5	>3.5	≤3.5
AUGUST												
Total number of lakes in groups	1	-	2	2	3	-	3	3	5	5	5	5
Number of significant lakes[a]	1	-	1	1	1	-	2	1	1	0	1	1
Significant parameter	NO_2+ NO_3		NO_2+ NO_3	NO_2+ NO_3	TotP		Alk, NO_2+ NO_3	AbsF	TotP		Temp	AbsF
OCTOBER												
Total number of lakes in groups	1	-	2	2	3	-	3	1	5	5	5	5
Number of significant lakes[ab]	-	-	-	-	2	-	1	-	1	-	1	1
Significant parameter					Alk, AbsOF		Kond, Alk		Kond, AbsOF TotP		TotP	Kond

all parameters were $\log_{10}$ transformed (alkalinity $\log_{10}$ X+1)

[a] number of lakes with at least one significant variable

[b] two of the lakes had more than one significant parameter

Discussion

Our findings showed that the mid-lake surface sample can be considered more representative of lake water chemistry than a random sample. Moreover, jack-knifing analysis showed that where a sample is taken within a lake will affect its representativity of average surficial water chemistry. For example, samples collected in the central part of the lake had few differences in water chemistry indicating that these samples were more representative of average conditions. In contrast, samples taken furthest from the center of the lake (e.g. the first quadrate sampled) often differed markedly from the other samples. For almost all lakes (31 of 34), the first quadrate was in the northern part of the lake, and several of these lakes had inlets situated in the northern sections of the lake as well. Hence, the differences often noted in water chemistry samples taken from that quadrate may often be explained by inlet-water effects. The influence of inlet water chemistry on samples taken in the first sampling quadrate was very obvious in lake Stor-Tjulträsket. This is a relatively large lake (5.3 km^2) situated in the northern Sweden with an inlet close to the first sampling quadrate. Water temperature and conductivity in the first quadrate in October were 1.5 °C and 6.81 mS/m, respectively, whereas temperature and conductivity in the other nine samples varied between 7 – 7.5 °C and 4.05 – 4.18 mS/m, respectively. In October, air temperature is going down, which lowers the stream water temperature in the inlet.

To further investigate whether the location in the lake had an effect on water chemistry we also determined if the mid-lake sample could be considered more representative of lake water chemistry than simply randomly selecting a site. In the same way as for the mid-lake sample a two sample *t*-test was done comparing the randomly selected sample with the nine remaining samples for each lake. As for the mid-lake sample, the possible number of combinations of lakes and variables in the *t*-test was 254 (August) and 253 (October). Of the variables tested for the randomly selected sample, 24 differed significantly in August and 23 in October, compared to differences of 10 and 9, respectively, when comparisons were made with a mid-lake sample. Although, comparison of only one random sample can not be considered a robust measure of sample representativity, this indicates, nonetheless that a mid-lake sample is more representative of average lake chemistry. This conjecture is also supported by the findings from the other tests of sample representativity performed here.

Grouping lakes by ecoregion and morphometric variables, such as water residence time and mean depth, had a significant effect on a number of variables, but did not affect pH, alkalinity, or inorganic nitrogen. Ecoregion was the most important descriptor of lake chemistry, for both seasons sampled. Not unexpected, August temperature varied significantly among all regions, indicating the large climatic gradient between the north and south of Sweden. However, as temperature strongly affects the biological production, we would expect to find indirect, ecoregion effects on water chemistry variables that are sensitive to biological activity. Indeed, variables such as absorbance of filtered water and total phosphorus (which include phyto- and zooplankton) varied significantly with ecoregion supporting this conjecture (Table 3).

Residence time was also found to be a more important explanatory variable in August than October. This finding may indicate the greater spatial variability expected in late summer (when lakes are relatively hydrologically stable) than autumn (when lakes are turning over). Indeed, several of the variables that varied significantly when the mid-lake samples were compared with other samples also varied somewhat with season. For example, in August nutrients seemed to vary more significantly than the other variables measured (Table 4). In contrast, in October variables more indicative of weathering and hydrology (e.g. alkalinity and conductivity) varied more. These findings were expected. Water chemistry variables sensitive to in-lake production (biological-active variables) varied more in late summer, whereas variables more dependent on regional-scale variation in catchment geology and run-off varied more in autumn. However, large-scale ecoregion differences were also noted, especially for the sampling in August. For example, in the arctic/alpine region inorganic nitrogen was the only variable of those studied that varied significantly (Table 6). Total phosphorus varied significantly in large (residence time > 1 year) and deep (mean depth > 3.5 m) lakes in the northern coniferous and southern mixed forest regions. In contrast to total phosphorus, absorbance of filtered water (i.e. water color) varied among lakes, but was more variable in small (residence time < 1 year) and shallow (mean depth < 3.5 m) lakes. To further study the influence of lake morphometry on the representativity of the mid-lake sample, we also analyzed the effect of shoreline development (or irregularity) on lake chemistry. Regression of shoreline development versus the *t*-value for each lake from the two-sample *t*-test of the mid-lake sample did not indicate any effect, suggesting that shoreline development was not an important descriptor of the representativeness of the mid-lake sample.

Although selected water chemistry variables were found to differ significantly both within and across lakes, the differences were often small, particularly within lakes. If the natural within-lake variance associated with a variable is small, then significant effects may arise that are not ecologically meaningful. Moreover, variation consists of a combination of natural as well as analytical or laboratory-induced variance. To account for and separate analytical variability from within-lake variability we compared the sample variance with the detection limits of the specific analyses. In other words, comparing the coefficient of variation (CV) of the 10 samples in each lake for each chemical variable (CV_{10}) with the CV of the analytical variation (CV_{av}). Analytical variation was greater than CV_{10} in 2 of the 10 differences noted in August and 6 of the 9 differences noted in October. If $CV_{av} > CV_{10}$, for the variables tested, then we can not conclude that the mid-lake sample is significantly different for lakes. Hence, taking into account analytical variability, the number of "significant" differences should be adjusted downwards. Water color in two lakes in August can therefore not be considered as differing, but significant differences still occurred for alkalinity, inorganic nitrogen, and total phosphorus. Taking into account analytical variability in the October sampling, total phosphorus and alkalinity still vary significantly in one and two lakes, respectively. Since the October samples were most likely taken during turnover, we expected to find a lower number of differences, as the lakes should be more homogenous in October than in August. We tested this by comparing CV_{10} for each lake and season with a paired *t*-test. Our findings, however, did not support the conjecture of higher spatial variability in late summer than autumn. Indeed, weak albeit significant seasonal differences were noted for two variables ($\alpha = 0.05$). CVs for alkalinity and conductivity were higher in October than in August. These findings may be partly explained by the influence of runoff from autumn rains.

Most studies on the horizontal variation within lakes concern biological parameters such as phyto- and zooplankton (e.g. Avois et al., 2000; Lacroix and Lescher-Moutoué, 1995; Pinell-Alloul, 1995; Visman et al., 1994). When the spatial patterns in water chemistry are reported they are usually done in the context of explaining the spatial variation of plankton (e.g. Dickman et al., 1993). Another way of sampling a lake with regard to spatial variation is proposed by Blomqvist (2001). He describes a method for a volume-weighted composite sample both for plankton and chemical parameters. This method compensates both for the horizontal and vertical variation within a lake. The composite sampling method is similar to sampling methods within soil sciences (Falck, 1973). Taking several soil samples and mix them thoroughly is a common way of sampling different horizons within a soil profile. Combining several samples reduces the variation and gives an average value for the different locations sampled. Using a combined sample or not is a matter of purpose and scale of the study. In our study, the variation between the sample quadrates was important, since the aim was to analyze whether the location of a sampling site was important for the representativity.

In summary, we can conclude that where a sample is taken within a lake will affect the sample's representativity of lake surface water chemistry. However, our study also showed that a mid-lake sample can be considered more representative than a random sample. Moreover, of the eight water chemistry metrics studied here, only 3.9% (August, n = 34 lakes) or 3.6% (October, n = 32 lakes) of the total number of tests varied significantly. These values are lower than 5%, or the number expected to occur by chance alone (if a α level of 5% is used). The finding that a mid-lake sample can be considered representative of surface water conditions was not unexpected. Indeed, a number of studies done during the last few decades have often used a mid-lake in sampling (e.g. according to Rainwater and Thatcher, 1960). However, despite its widespread use surprisingly little is known of just how representative a mid-lake sample is of water chemistry. Our study showed that mid-lake samples can be considered representative, and this information is crucial for how we interpret data taken from monitoring programs that are temporally extensive.

Acknowledgements

We would like to thank all who helped us with the field sampling, in particular Kjell Östling who organized the sampling and Inger Sjöstedt and other lab personnel who assisted with the water chemistry analyses; Ulf Olsson who helped us with the statistics; and finally Ewa Bringmark for comments on an earlier draft.

References

Anonymous, 1984. Naturgeografisk regionindelning av Norden. Nordiska ministerrådet 1984.

Avois, C., Legendre, P., Masson, S., and Pinel-Alloul, B., 2000. Is the sampling strategy interfering with the study of spatial variability of zooplankton communities? Canadian Journal of Fisheries and Aquatic Sciences 57, 1940-1956.

Berndtsson, R., 1987. Application of infiltration equations to a catchment with large spatial variability in infiltration. Hydrological Sciences Journal 32, 399–413.

Blomqvist, P., 2001. A proposed sampling method for composite sampling of water chemistry and plankton in small lakes. Environmental and Ecological Statistics 8, 121-134.

Dickman, M., Stewart, K., and Servant-Vildary, M., 1993. Spatial heterogeneity of summer phytoplankton and water chemistry in a large volcanic spring-fed lake in northern Iceland. Arctic and Alpine Research 25, 228-239.

European Commission, 2000. Directive 2000/60/EC of the European Parliament and of the Council - Establishing a framework for Community action in the field of water policy. Brussels, Belgium, 23 October 2000.

Falck, J., 1973. A sampling method for quantitative determination of plant nutrient content of the forest floor. Royal College of Forestry, Dept. of Silviculture, Research Notes, No. 1.

Goovaerts, P., 1998. Geostatical tools for characterizing the spatial variability of microbiological and physico-chemical soil properties. Biology and Fertility of Soils 27, 315–334.

Gustafsson, L., and I. Ahlén (eds.)., 1996. Geography of plants and animals. Almqvist & Wiksell International, Stockholm.

Illies, J., 1966. Verbreitung der Süßwasserfauna Europas. Verhandlungen der Internationalen Vereinigung für Theoretische und Angewandte Limnologie 16, 287–296.

Kratz, T. K., Benson, B. J., Blood, E. R., Cunningham, G. L., and Dahlgren, R. A., 1991. The influence of landscape position on temporal variability in four North American ecosystems. The American Naturalist 138, 355–378.

Lacroix, G., and Lescher-Moutoué, F., 1995. Spatial patterns of planktonic microcrustaceans in a small shallow lake. Hydrobiologia 300/301, 205-217.

MINITAB release 12, 1997. Minitab Inc., USA.

Oliver, M. A., Webster, R., and McGrath, S. P., 1996. Disjunctive kriging for environmental management. Environmetrics 7, 333–358.

Pinel-Alloul, B., 1995. Spatial heterogeneity as a multiscale characteristic of zooplankton community. Hydrobiologia 300/301, 7-42.

Rainwater, F. H., and Thatcher, L. L., 1960. Methods for Collection and Analysis of Water Samples. Geological Survey Water-Supply paper 1454, U. S. Government Printing Office, Washington.

Visman, V., McQueen, D. J., and Demers, E., 1994. Zooplankton spatial patterns in two lakes with contrasting fish community structure. Hydrobiologia 284, 177-191.

Wetzel, R. G. 1983. Limnology, second edition. CBS College Publishing.

Wetzel, R. G., and Likens, G. E., 1991. Limnological analyses, second edition. Springer-Verlag, New York.

Wiederholm, T., and Johnson, R. K., 1997. Monitoring and assessment of lakes and watercourses in Sweden. 317–329 In: Ottens, J. J., Claessen, F. A. M., Stoks, P. G., Timmerman, J. G., and Ward, R. C. (eds.). Monitoring Tailor-made II, Information strategies in Water, Nunspeet, The Netherlands.

Wilander, A., 1997. Referenssjöars vattenkemi under 12 år; tillstånd och trender. Rapport 4652, Stockholm, Swedish Environmental Protection Agency. (In Swedish)

Wilander, A., Johnson, R.K., Goedkoop, W., and Lundin, L., 1998. Riksinventering 1995. En synoptisk studie av vattenkemi och bottenfauna i svenska sjöar och vattendrag. Rapport 4813, Stockholm, Swedish Environmental Protection Agency. (In Swedish)

Yasuda, H., Berndtsson, R., Bahri, A., Persson, H., Gullberg, A., and Jinno, K., 1994. Spatial correlation analysis of two-dimensional solute transport in the unsaturated zone. Stochastic and Statistical Methods in Hydrology and Environmental Engineering 2, 127–138.