Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Department of Aquatic Sciences and Assessment

Lake Biology as a Function of Catchment Characteristics and Water Quality Parameters – Focus on Phytoplankton

Jenny Näslund

Degree project • 30 credits
Uppsala 2019

Lake Biology as a Function of Catchment Characteristics and Water Quality Parameters – Focus on Phytoplankton

Koppling mellan sjöars biologi och egenskaper i avrinningsområdet- fokus på växtplankton

Jenny Näslund

Supervisor:	Faruk Djodjic, Swedish University of Agricultural Sciences, Department of Aquatic Sciences and Assessment
Assistant supervisor:	Stina Drakare, Swedish University of Agricultural Sciences, Department of Aquatic Sciences and Assessment
Examiner:	Jens Fölster, Swedish University of Agricultural Sciences, Department of Aquatic Sciences and Assessment

Credits:	30 credits
Level:	Second cycle, A2E
Course title:	Master thesis in Biology
Course code:	EX0895
Course coordinating department:	Department of Aquatic Sciences and Assessment
Place of publication:	Uppsala
Year of publication:	2019
Cover picture:	Jenny Näslund
Online publication:	https://stud.epsilon.slu.se
Keywords:	phytoplankton, total phosphorus, catchment, soil properties, land use, agriculture, water framework directive, ecological status

Swedish University of Agricultural Sciences
Faculty of Natural Resources and Agricultural Sciences
Department of Aquatic Sciences and Assessment

Abstract

According to the EU Water Framework Directive, phytoplankton should be included for assessment of the ecological status of lakes. Phytoplankton respond rapidly to environmental changes and are a particularly good indicator of nutrient loads. In this large-scale study, it was examined whether the properties in the catchment area can be used to explain the variation of phytoplankton and total phosphorus (TP) in lakes. A large number of variables were investigated through statistical analysis, in particular if the already established linear relationship between phytoplankton and TP in lakes can be improved.

The study used measured values for total biovolume phytoplankton (tot bio), Plankton Trophic Index (PTI) and TP for 487 lakes (represented by 523 waterbodies) in south of Sweden. The lakes associated catchment properties were calculated and analysed through various Geographical Information System (GIS) tools. Each catchment was described regarding land use, soil properties (texture and chemistry), soil distribution, climate and lake properties. In total 59 variables produced with GIS were evaluated from available map data and national soil surveys together with 34 lake variables. All variables were used in Principal Components Analysis (PCA) whereas the 59 catchment variables together with some lake variables were used for other analysis. For each dependent variable (tot bio, PTI and TP) several statistical models were created, and important catchment variables were identified using Partial Least Squares (PLS) analysis. Important variables identified in PLS were then included in multiple regressions.

Result shows that the share of agricultural land in the catchment area is positively correlated with phytoplankton (tot bio and PTI) and TP. For phytoplankton models without TP as explanatory variable, a few catchment variables could explain variation of PTI up to 48 % while tot bio could be explained to a lower extent (33 %). The degree of explanation and variables included differed depending on selected statistical model. While TP alone was the strongest explaining variable for both tot bio (66 %) and PTI (56 %). However, TP together with the share of agricultural land significantly improved the explanation of PTI to 65 %. For the lake TP, catchment properties could statistically explain 55 % of the variation in the TP concentration. In summary, TP was shown to correlate positively with specific soil properties of both non-agricultural and agricultural areas of studied catchments. Higher TP concentration could also be expected in lakes with larger share of agricultural land and urban area and smaller water body area. The results also show that catchment properties derived from continuous map data had a higher explanation of the studied lakes' tot bio, PTI and TP compared to result from field sample point data collected in national soil surveys.

The relationship between catchment properties and water quality is important to understand and catchment properties can help to describe the lake phytoplankton and phosphorus levels, which then should be taken into account when developing assessment criteria for lakes.

Keywords: phytoplankton, total phosphorus, catchment, soil properties, land use, agriculture, water framework directive, ecological status

Sammanfattning

Enligt EU:s vattendirektiv ska växtplankton inkluderas vid bedömningar av sjöars ekologiska status. Växtplankton reagerar snabbt på miljöförändringar och ger särskilt bra indikation på näringsbelastning. I denna storskaliga studie undersöktes om egenskaper i avrinningsområdet kan användas för att förklara sjöars växtplankton och totalfosfor (TP). Genom statistisk analys av ett stort antal variabler testades också om förklaringsgraden av det redan kända linjära sambandet mellan växtplankton och TP i sjöar kan förbättras.

I studien användes uppmätta värden av växtplanktons totala biovolym (tot bio), planktontrofiskt index (PTI) och TP för 487 sjöar (sammantaget 523 vattenförekomster) i södra Sverige. Sjöarnas tillhörande egenskaper i avrinningsområdet togs fram och analyserades genom olika Geografisk Informations System (GIS) verktyg. Varje avrinningsområde beskrevs utifrån dess markanvändning, markegenskaper (textur och kemi), jordartsfördelning, klimat och sjöegenskaper. Sammanlagt utvärderades 59 variabler framtagna med GIS från befintliga kartdata och nationella markundersökningar tillsammans med 34 sjövariabler. Alla variabler användes vid principalkomponentanalys (PCA) och för övrig statistisk analys användes de 59 avrinningsområdes variablerna tillsammans med några sjövariabler. Flera statistiska modeller skapades för varje beroende variabel (tot bio, PTI och TP) där viktiga avrinningsområdes variabler identifierades med hjälp av Partial Least Squares (PLS) analys. De viktigaste variablerna från PLS analysen användes sedan i multipla regressioner för att ta fram statistiskt signifikanta förklarande variabler för variationen hos växtplankton och TP.

Andel jordbruksmark i avrinningsområdet visade ett positivt samband till växtplankton (tot bio and PTI) och TP. Ett fåtal avrinningsområdes variabler kunde tillsammans utan TP förklara variationen av PTI upp till 48 % medan tot bio kunde förklaras till en lägre grad (33 %). Beroende på vald statistisk modell, skiljde sig förklaringsgrad och de ingående variablerna delvis åt. Det bör nämnas att det linjära sambandet mellan TP och växtplankton var mycket starkt och förklarade variationen hos tot bio till 66 % och PTI till 56 %. TP tillsammans med andelen jordbruksmark i avrinningsområdet gav dock en signifikant ökad förklaringsgrad av PTI (65 %). För sjöarnas TP kunde egenskaperna i avrinningsområdet statistiskt förklara 55 % av variationerna i halterna. Sammanfattningsvis visade TP positivt samband till vissa markegenskaper på så väl jordbruksmark som övrig mark för de studerade sjöarnas avrinningsområden. Höga halter av TP förväntas även i sjöar med hög andel jordbruksmark och tätort. Sjöar med mindre yta på vattenförekomsten förväntas också ha högre TP. Studiens resultat visar även att egenskaper i avrinningsområdet framtagna från kartdata hade en högre förklaringsgrad för de studerade sjöarnas tot bio, PTI och TP jämfört med resultat från provpunktsdata som samlades inom de nationella markundersökningarna.

Slutsatsen från denna studie är att sambandet mellan land och vatten är viktigt att förstå för att kunna beskriva en sjös fosfornivåer och växtplankton, vilket bör tas i beaktan vid utveckling av bedömningsgrunder för sjöar.

Nyckelord: växtplankton, totalfosfor, avrinningsområde, markegenskaper, markanvändning, jordbruk, vattendirektivet, ekologisk status

Populärvetenskaplig Sammanfattning

Övergödning av sjöar och vattendrag är idag ett globalt miljöproblem som orsakas av en ökad tillförsel av näringsämnen, framförallt kväve och fosfor, vilket påverkar såväl ekosystem som vattenkvalitén. En ökad näringstillförsel leder ofta till ökad tillväxt av växtplankton eftersom de kan ta upp näringsämnen direkt från vattnet. Växtplanktons snabba respons på näringsbelastning gör de till en bra indikator. För att skapa en långsiktig och hållbar förvaltning av sötvatten i EU infördes vattendirektivet med syfte att skydda och hindra försämring av Europas vattenresurser. Målet är att alla vattenförekomster, sjöar och vattendrag, ska uppnå minst god vattenstatus till 2021. Vid bedömningen av sjöars ekologiska status ingår fyra biologiska kvalitetsfaktorer där växtplankton används som indikator på övergödning. I Sverige behöver mer än 7200 sjöar (de över 50 ha) förvaltas och övervakas enligt vattendirektivet vilket är en resurskrävande process både tidsmässigt och ekonomiskt. Därför är förbättringar av befintliga bedömningsgrunder och metoder för klassificering av sjöars status viktiga.

I denna studie undersöktes om egenskaper i sjöns avrinningsområde från olika typer av kartor kan användas för att förklara och uppskatta växtplanktonförekomst, nivåer och sammansättning, samt halter av totalfosfor. Fördelen med att använda kartdata är att det är billigare, när man inte behöver åka ut till varje sjö för att ta prover. Fosfor har länge varit känt som begränsande näringsämne för tillväxt i sjöar och uppmätt halt av totalfosfor används för att uppskatta mängden växtplankton. I studien ingick ett stort antal sjöar (487 stycken) i södra Sverige med uppmätta värden av totalfosfor och två växtplanktonparametrar: planktontrofiskt index (PTI) och växtplanktons totala biovolym. PTI baseras på indikatorarter för växtplankton, med indikatorvärden som visar om de är känsliga eller toleranta för näringsbelastning. Befintliga kartdata och markundersökningar användes för att beskriva varje avrinningsområde utifrån markanvändning, markegenskaper (textur och kemiska sammansättning), jordartsfördelning och klimat. Hänsyn togs även till ett antal sjöegenskaper som exempelvis sjöns area och position.

Studiens resultat visar att andelen jordbruksmark i avrinningsområdet har positivt samband till sjöns växtplankton och totalfosfor. Högre halter av totalfosfor, en större totalvolym av växtplankton och arter toleranta för näringsbelastning förväntas därmed i sjöar med mer jordbruksmark. Ett urval av egenskaperna i avrinningsområdet visades tillsammans kunna förklara variationen av PTI upp till 48 % medan den totala volymen av växtplankton kunde förklaras till en lägre grad (33 %). Sjöns uppmätta totalfosforhalt förklarar dock starkast variationen i totala volymen av växtplankton (66 %) och PTI (56 %). Däremot kan sjöns totala fosfor tillsammans med andelen jordbruksmark öka förklaringsgraden av PTI till 65 %. För de studerade sjöarna kunde variationen av totalfosforhalterna förklaras till 55 % av egenskaper i avrinningsområdet.

Slutsatsen är att både växtplankton och totalfosfor i sjöar till stor utsträckning kan uppskattas av avrinningsområdets egenskaper från befintliga kartdata, vilket bör beaktas vid utveckling av bedömningsgrunder för sjöar.

Table of Contents

1 Introduction

1.1 Eutrophication

Eutrophication is considered as a global problem for both freshwater and marine systems causing negative impact on and consequences for ecosystems and water quality (Smith 2003; Smith & Schindler 2009). Eutrophication occurs when a waterbody gets an increased supply of minerals and nutrients compared to its natural state, where the potential of eutrophication depends on the available nutrients (Novotny & Olem 1994). The effects of increase in loads of nutrients, nitrogen (N) and phosphorus (P), lead to a whole set of effects on the ecosystem. Firstly, it results in an increased primary production, especially of phytoplankton (Novotny & Olem 1994; Brönmark & Hansson 2005; Smith & Schindler 2009). Secondly, the increase of primary production results in reduced water transparency and increased sedimentation (Brönmark & Hansson 2005; Smith & Schindler 2009). Thirdly, the consequent degradation of dead organic material by bacteria can cause oxygen deficiency in the bottom sediment and bottom water (Brönmark & Hansson 2005; Smith & Schindler 2009). Effects of eutrophication can be seen in different parts of a lake ecosystem, with altered species composition, biomass and shift in the trophic levels (Correll 1998; Smith & Schindler 2009).

Phosphorus has been established as a common limiting nutrient for freshwater system after Schindler (1974) did a whole lake experiment involving different nutrients including P. In that experiment, carbon (C) and N where added in one half of a lake while C, N and P were added in the other half, resulting in an algae bloom only where P also was added (Schindler 1974). Phosphorus can be transported to a waterbody either as bound to soil particles (particulate P) or as dissolved (orthophosphate), where the dissolved form is essential and directly available for primary producers and their growth (Correll 1998).

For phytoplankton growth, the ratio of N and P in the water is important to assess the limited nutrient (Brönmark & Hansson 2005). For instance, an increased atmospheric N-deposition influences the ratio, where phytoplankton in lakes with low N-deposition regions are considered to be N-limited and the opposite is true in regions with high N-deposition, in these lakes P-limitation instead becomes an issue (Elser *et al.* 2009). Phytoplankton in northern Sweden lakes are considered N-limited due to low atmospheric N-deposition and low input of dissolved inorganic N from the catchment, while lakes in the southern part with higher N-deposition and input are P-limited (Bergström *et al.* 2008). Phytoplankton have fast turnover rate and absorb dissolved nutrients directly from the water column (Brönmark & Hansson 2005). The available nutrients are essential for building the cells and their functions where P constitute an important component in the genetic information (DNA and RNA), energy system (ATP) and as phosphorus-lipids in the cell membrane (Brönmark & Hansson 2005). The phytoplankton's properties thus make them a good indicator for rapid response on environmental impacts in general and nutrients loads in particular (Swedish EPA 2010).

1.2 Water Framework Directive

To avoid long-term deterioration of freshwater systems, the Water Framework Directive (WFD) was adopted in year 2000 to protect freshwater resources in European Union and to establish long-term sustainable management (Directive 2000/60/EC). To assess the ecosystem's function and structure, each waterbody's ecological status is assessed with the main goal of achieving at least good ecological status. The waterbodies are assessed during a six-years management cycle where the next cycle ends 2021. The good ecological status defines as:

> "The value of the biological quality elements for the surface water body type show low levels of distortion resulting from human activity, but deviate only slightly from those normally associated with the surface water body type under undisturbed conditions." (Directive 2000/60/EC, Annex V)

Four biological quality elements responding to eutrophication are assessed within the WFD for the ecological status in lakes. Phytoplankton is one of the four elements where biomass, as well as taxonomic composition and abundance are considered (Directive 2000/60/EC). As phytoplankton respond immediately to nutrient supply, they reflect the current eutrophication in the pelagic zone (Lyche-Solheim *et al.* 2013). The other three elements, macrophytes, benthic invertebrates and fish, have longer generation time and thus reflect nutrient response over longer time, i.e. years (Lyche-Solheim *et al.* 2013). Macrophytes and benthic invertebrates mainly respond to changes in littoral zone whereas fish respond to, and therefore indicate, changes across the lake ecosystem (Lyche-Solheim *et al.* 2013). The ecological status is assessed as Ecological Quality Ratio (EQR) for each element parameter and shows the relationship between the observed value in the lake and type-specific reference conditions (Directive 2000/60/EC), which enables a comparison of the status between waterbodies.

The WFD was incorporated 2004 in the Swedish Environmental Code (1998:808) for national legislation, where approach and included assessment criteria for phytoplankton parameters are described in the Swedish Handbook for Swedish ecological classification (Swedish EPA 2010; Swedish Agency for Marine and Water Management 2018). The five year research program WATERS has since then suggested improvements and harmonisation of many methods for the Swedish assessment criteria for ecological status (Lindegarth *et al.* 2016). The existing regulation HVMFS 2013:19 for ecological classification were revised (HVMFS 2018:17) and came into force 1 January 2019. This update shows four phytoplankton parameters that needs to be considered for assessment, namely total biomass, chlorophyll-*a* (chl-*a*), Plankton Trophic Index (PTI) and number of taxa. The three first phytoplankton parameters are indicators of the response to nutrient pressure, while number of taxa is an indicator of the pressure from acidification (Lindegarth *et al.* 2016).

1.2.1 Phytoplankton as an Environmental Indicator

Phytoplankton parameters response to nutrient pressure

Measuring chl-*a* works as an indicator to get a cheap and broad overview of a waterbody’s total phytoplankton biomass (Swedish EPA 2010). Using the pigment, chl-*a*, the algae get energy by absorbing sunlight (Brönmark & Hansson 2005). However, the amount of pigment is dependent on species groups (Brönmark & Hansson 2005). There is a strong positive relationship between chl-*a* and total phosphorus (TP) (Phillips *et al.* 2008; Carvalho *et al.* 2012; Lyche-Solheim *et al.* 2013; Lindegarth *et al.* 2016). A relationship is also seen between chl-*a* and total nitrogen, but it explains less of the variation (Phillips *et al.* 2008). In addition, other variables are related to chl-*a* such as lake depth and alkalinity (Phillips *et al.* 2008). The total biomass of phytoplankton is used as an indicator for nutrient pressure and is measured as biovolume (assuming a similar density of phytoplankton and water) (Swedish EPA 2010). The total biomass thus reflects the lakes primary production, and depends on the species represented (Swedish EPA 2010). Both chl-*a* and total biomass can be predicted using TP, where TP explains 80 % of the variation in chl-*a* and 65 % of the variation in total biomass (Lindegarth *et al.* 2016).

The Swedish Trophic Plankton Index (TPI) was developed by Willén (2007) and is based on indicator species for nutrient pressure. In short, species tolerant to high TP level are assigned a score value from 3 to 1, whereas the sensitive species are given score values from -3 to -1 (Willén 2007; Swedish EPA 2010). The TPI value thus indicates in which part of the oligotrophic and eutrophic scale (-3 to 3) the lake is located base on species composition. One disadvantage of TPI is that most species with score value are in the two ends of the scale, i.e. species being either very sensitive or very tolerant for nutrient pressure (Willén 2007; Lindegarth *et al.* 2016). Consequently, there is the lack of species in the middle of the TPI pressure score, which makes the index less robust in the middle of the nutrient pressure scale. The European developed Plankton Trophic Index (PTI) (Phillips *et al.* 2012), functions in a similar way as TPI, but has the advantage that species are represented with score values throughout the P gradient. The PTI has also been evaluated for Swedish lakes within the WATERS project (Lindegarth *et al.* 2016), with the suggestion to replace TPI with PTI when assessing the species composition in Swedish lakes, which is now regulated in HVMFS 2013:19. Phillips *et al.*, (2012) showed in a study based on a large European lake dataset for PTI that the index is significantly correlated to TP. The WATERS project, concluded that PTI and TP had an significant linear relationship for 361 Swedish lakes, $R^2 = 0.59$ (Lindegarth *et al.* 2016). Based on the relationship between TP and PTI, the index is suggested as strong and sensitive for eutrophication (Carvalho *et al.* 2012; Lyche-Solheim *et al.* 2013). However, PTI is also significantly influenced by other lake water properties independent of TP, such as alkalinity, and factors affecting the lake such as precipitation, temperature and lake surface area (Phillips *et al.* 2012), highlighting the need of different reference values for different types of lakes.

In response to nutrient pressure in lakes, phytoplankton total biomass and PTI should be weighted together and form the basis for the phytoplankton ecological status classification, and in case chl-*a* is sampled it should be included (HVMFS 2013:19; Swedish Agency for Marine and Water Management 2018).

Phytoplankton response to acidic pressure and other pressures

Two particularly important factors determining major phytoplankton groups are gradient along nutrient condition (oligotrophic-eutrophic) and pH gradient (acidic-alkaline) (Brönmark & Hansson 2005). For instance, at pH lower than 5 to 6 the diversity and biomass decrease and dominating species groups shift (Brönmark & Hansson 2005). The number of phytoplankton taxa and pH is thus positively correlated, i.e. the number of taxa increase with pH, however, the relationship only applies under pH 7 (Swedish Agency for Marine and Water Management 2018). The phytoplankton’s response to acidification is not in the focus of this study.

Additionally, several other factors besides nutrient pressure and pH influence the lake phytoplankton community composition, such as alkalinity, latitude and water colour, where some phytoplankton groups are also influenced by lake surface area and depth (Maileht *et al.* 2013). A decline of phytoplankton species richness and biomass is identified for Swedish lakes toward higher latitude (Weyhenmeyer *et al.* 2013). Further, variation in phytoplankton richness of different species groups was in a Danish lake study besides water chemistry, also explained to a small extent by climate, lake morphology and land use (agriculture) in the catchment (Özkan *et al.* 2013).

1.3 Losses of Phosphorus from Land to Surrounding Water

Since phosphorus is considered one of the most important factors for lake phytoplankton growth, reduction of the external load from the catchment is required for eutrophication recovery (Lyche-Solheim *et al.* 2013). The lake productivity depends both on the internal produced material (autochthonous) and the large proportion of material (such as humic substances and nutrients) transported from the catchment area (allochthonous) (Novotny & Olem 1994; Brönmark & Hansson 2005). Understanding of factors influencing P concentration in water is important, where variability of TP concentrations (R^2= 0.71) in Swedish lakes was best predicted by the relationship to natural suspended matter, absorbance and altitude (Huser & Fölster 2013). Lakes at low altitude have generally higher concentrations of N and P throughout Europe (Nõges 2009). Internal fluxes of nutrients by mineralization of organic matter and internal loading with phosphate release from sediment at low redox potential (low oxygen) to the water column (Brönmark & Hansson 2005) are also important sources. Internal loading occurs especially in summer in shallow eutrophic lakes even decades after the reduction of external P supply have taken place (Søndergaard *et al.* 2013). Geographical gradients of both lake morphology and water chemistry are seen throughout Europe, where northern lakes have lower concentration of nutrients, alkalinity and pH but higher content of organic matter (Nõges 2009). The levels of water chemistry parameters (nutrients, alkalinity, pH and organic matter) are generally higher in larger lakes with larger catchment areas (Nõges 2009).

1.3.1 Diffuse- and Point Sources

According to the calculations of the annual nutrient load to the Baltic Sea for year 2014, the major sources of nutrients (N and P) are from agriculture (35 % N and 40 % P) and forestry (36 % N and 30 % P) (Ejhed *et al.* 2016). The nitrogen load is considerably higher than the P load, and the diffuse leakage from agricultural land is the largest anthropogenic source of P followed by wastewater treatment plants for the point effluents (Ejhed *et al.* 2016). Although P transport from land is lower than for N, it is still of concern given that freshwater primary producers mainly are P-limited (Brönmark & Hansson 2005; Smith & Schindler 2009). In Sweden, the most intense and largest area of agriculture is located in southern part of the country (Johnsson *et al.* 2016). Urban areas also have a substantial impact as point sources, although the improved treatment methods from the 1970s have reduced the nutrient emissions from wastewater treatment plants, especially for P (Swedish EPA 2018). Phosphorus emissions from Swedish wastewater treatment plants in 2016 amounted to 237 tonnes with 96 % purification efficiency (Statistics Sweden 2018). Additionally, nutrient emissions from individual sewers, stormwater, and industry are also important to consider for urban impact on water (Swedish EPA 2018).

The biological quality factors govern the assessment of the ecological status of lakes whereas the physical-chemical factors have a supporting function (Swedish EPA 2010). For lakes, TP is assessed due to its role as limiting nutrient. For the estimation of TP reference values in the running waters, the impact of agriculture is considered when there is more than 10 % agricultural land in the catchment (HVMFS 2013:19; Fölster *et al.* 2018). Considering that nutrient retention may occur in lakes and thereby influence the concentration of P, the 10 % agricultural criterion is not applied when assessing the TP reference value for lakes (Fölster *et al.* 2018).

1.3.2 Soil Properties Affecting Phosphorus Losses

Soil is a complex system and understanding the effects of soil properties on nutrient losses is important to prevent potential leaching and negative effects to surrounding water. Generally, factors such as soil texture, structure and permeability, but also soil nutrient content and soil particle-size distribution, are important to understand leaching (Eriksson *et al.* 2014).

Nitrate is a very mobile compound of nitrogen due to low adsorption capacity to soil particles (Eriksson *et al.* 2014) and nitrogen losses are therefore higher in intense agricultural land with low clay content (course texture) soils having high permeability (Eriksson *et al.* 2014; Johnsson *et al.* 2016). On the small catchment scale, highest losses of nitrogen are seen for sandy soil with high precipitation (Kyllmar *et al.* 2014). Leaching of dissolved and particular bound P, depends highly on soil aggregates and pore structure (Eriksson *et al.* 2014). For instance, clay rich soil with strong aggregate structure will rapidly transport P through macropores or as surface runoff during precipitation even if the fine soil can adsorb P more than permeable sandy soil (Eriksson *et al.* 2014; Ejhed *et al.* 2016).

Soil texture governs to a high degree the prevailing transport pathways between fields and water recipients (Eriksson *et al.* 2014; Johnsson *et al.* 2016). Another important factor for P losses is the soil P content in agricultural soils. In a study investigating different soil types with different soil P content Djodjic *et al.*, (2004) concluded that subsoil properties such as water transport pathways and P sorption capacity are important to consider when assessing the potential for P losses. Soil types with high clay content have a generally larger tendency to contribute to P losses (Djodjic *et al.* 2004; Johnsson *et al.* 2016), but this is highly site specific and strongly influenced by the soil properties including the aggregation (Eriksson *et al.* 2014). The soils composition of fine particles (clay), organic matter and mineralogy are in turn all affecting the availability of P in the soil and its potential for leaching though the profile (Eriksson *et al.* 2014, 2016). Soils ability to adsorb and release P is also pH-dependent, where P at low pH is primarily bound to aluminium- and iron oxides, while at higher pH to calcium (Eriksson *et al.* 2014, 2016). On the catchment scale, around 80 % of the terrestrial P losses transported to aquatic system originates from a small area, around 20 % of the catchment area closest to the lake or river, a phenomenon known as the 80:20 rule (Sharpley *et al.* 2009).

On arable land, the slope of the fields is considered an important factor for P losses where increasing slope contributes to higher losses (Johnsson *et al.* 2016). Another factor also affecting P losses is the crop and its soil cover, where lay (grassland) prevent highest losses of P (Johnsson *et al.* 2016). In Swedish agriculture lay is also the dominating crop (Johnsson *et al.* 2016). Areas with high erosion risk are important to detect as important sources of both suspended sediment and P loads (Schoumans *et al.* 2014; Djodjic & Markensten 2018). Soil vulnerability to erosion (soil erodibility), vegetative soil cover and slope length are some influencing factors for soil erosion risk (Djodjic & Markensten 2018). To prevent P impact on water from diffuse sources such as agricultural areas, there are many approaches and strategies to apply at different scales, such as reducing risk of runoff and soil erosion, appropriate fertilizer and manure applications and establishing buffer zones connected to the water for nutrient trapping and water flow delay (Schoumans *et al.* 2014).

1.4 Aims and Study Questions

Ecological status for lakes should be evaluated, monitored and managed according to the WFD for each water management cycle (Directive 2000/60/EC). Sweden has around 95 000 lakes over 1 hectare (Holmgren 2018) whereas WFD demands only that waterbodies over 50 hectare need to be evaluated and reported (Directive 2000/60/EC, Annex II). For Sweden this represents 7223 waterbodies (Drakare

2014). Monitoring of such a large number of waterbodies is a resource-demanding process regarding both the economy and time. Improvement of methods to assess lake status and upscaling of existing measurements is therefore important.

This study aims to increase the understanding of the connection between lake phytoplankton and catchment properties and could thus be used to improve assessment methods for phytoplankton in lakes. It is also intended to improve the estimation of phytoplankton metrics in lakes lacking measurements of TP. The study will therefore investigate potential correlations between phytoplankton and catchment properties and explore if variables from the catchment can be used to describe lake phytoplankton. Of particular interest is if catchment variables can improve the already established correlation between phytoplankton and TP in lakes, as well as, to what extent the catchment properties can explain phytoplankton and TP itself. The lake phytoplankton will be represented by three indicators: total biovolume phytoplankton (tot bio), chl-*a* and PTI. The driving land variables include basic geographical information such as categories of land use, soil distribution, soil texture, soil chemistry and climate variables in the catchments of the studied lakes. The catchment variables will be explained with different categories or more detailed sub-variables. In addition, the lake was also described by basic geographic information of size of lake and catchment as well as location. To investigate the correlation between catchment characteristics and lake phytoplankton the study aims to answer following questions:

Questions regarding phytoplankton:

1. Which catchment variables are important to explain the variation of each phytoplankton indicator in Swedish lakes?
2. To what extent can the variation of each phytoplankton indicator be explained by the variables detected in question 1?
3. Can the linear regression between total phosphorus and phytoplankton indicator be further improved by including important variables found in question 1?

Questions regarding total phosphorus:

4. Which catchment variables are important to explain the variation of total phosphorus concentrations in Swedish lakes?
5. How much of the variation in total phosphorus concentrations in lakes can selected catchment variables detected in question 4 explain?

2 Material

In this study, a dataset was built up by collection and compiling various sources covering lake water quality and associated catchment characteristics. The main components were lake phytoplankton and abiotic factors, physical and chemical, together with lake surrounding properties. *Figure 1* shows the overall structure of the used input data.

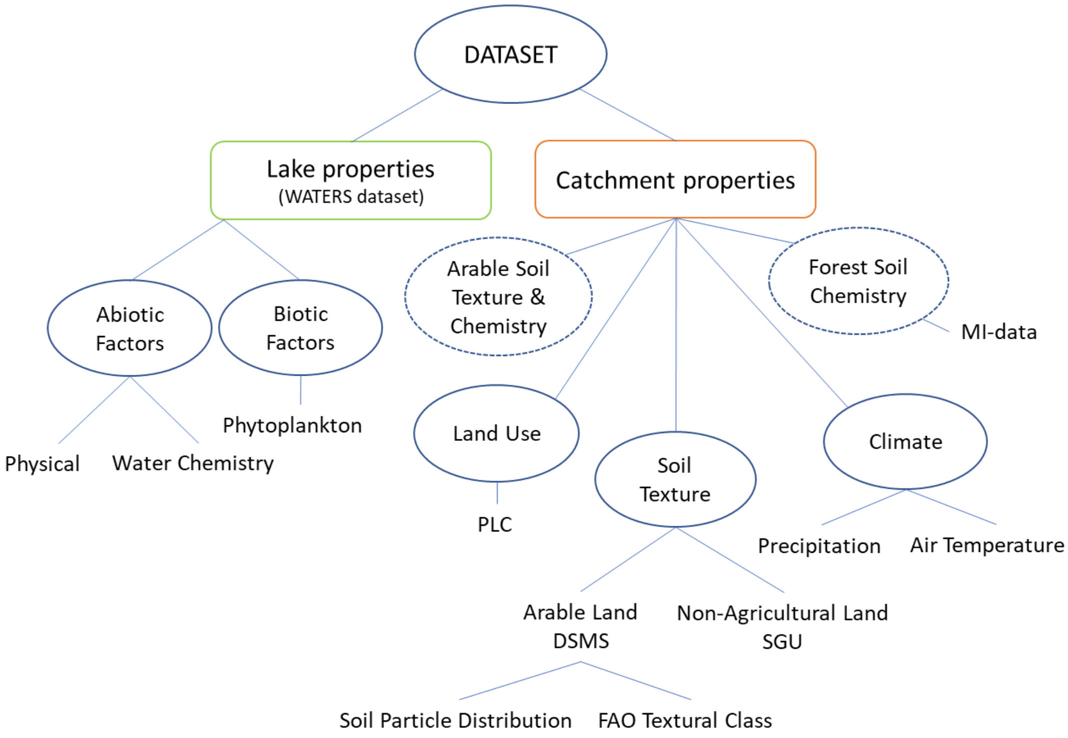

Figure 1. Overview of the dataset variables with lake and catchment properties. In total, 93 variables are represented in the dataset where 34 are from the WATERS dataset and 59 variables from the catchment dataset. Dashed line show material from the national soil surveys while the other catchment datasets represent material from other map sources. Abbreviations illustrate the source and name used for the variable group: PLC-Pollution Load Compilation, DSMS-Digital Arable Soil Map of Sweden, SGU-Geological Survey of Sweden and MI-Markinventering.

2.1 Lake Properties

2.1.1 WATERS Dataset

The lake data used in this study was collected by the WATERS project section that worked with phytoplankton metrics for Swedish lakes (Lindegarth *et al.* 2016). The WATERS phytoplankton dataset was originally selected from the national data base hosted at the Swedish University of Agriculture Sciences (SLU) (Miljödata-MVM 2019). The WATERS criteria for the selection from the national data base was that the selected lakes had at least one reported phytoplankton data occasion between July and August, and years 2000 to 2012 (Lindegarth *et al.* 2016). Water chemistry was added to the dataset if available. All selected lakes in the current study are thus part of the Swedish Monitoring Programs, with standardised collection techniques for sampling and certified laboratories for analyses according to requirements from the Swedish Agency for Marine and Water Management. Usually, one lake represents one waterbody, where lake and associated catchment can be identified with individual ID. The larger lakes are however divided into several waterbodies and are therefore represented several times in the dataset. This means that larger lakes may be represented by different biological and chemistry values collected at individual sampling sites.

The WATERS project made a compiled version from data derived from the national data host, where a single mean value was calculated for 2000-2012 data for each variable and sampling site/lake, when more than one value was available (Lindegarth *et al.* 2016). The geographical information about lake depth, location and altitude (masl) in the WATERS dataset had been derived from Swedish Meteorological and Hydrological Institute (SMHI 2019). The WATERS dataset covers 806 lake sites from the whole of Sweden. For the current study, only data representing areas with enough coverage in the land information was selected, see study boundaries in Section 3.1. In total, 523 sampling locations were used in this study representing 487 lakes. Eight lakes with several waterbodies in each lake are covered by together 44 sampling locations (waterbodies) in the dataset. *Table 1* show all lake variables used in in this study, including phytoplankton, water chemistry, location and physical properties. Variables representing the phytoplankton are total biovolume (biomass), chl-*a*, PTI and number of taxa. Water chemistry variables include pH, water colour (filtered water, absorbance measured at 420 nm), alkalinity, conductivity, metals and nutrients.

Table 1. *Lake variables, mean value July-August 2000-2012, derived from the WATERS dataset and used in this study. Min, max and median for the variable and number of waterbodies (N) included. Note that all variables are not available for all waterbodies, in total 523 waterbody´s are represented. Transformation type used for each variable are also given.*

Variable description	Variable name	Unit	Transf.	N	Min	Max	Median
Lake Mean Depth	L. Mean Depth	m	log(x)	397	0.5	38.7	4.7
Lake Max Depth	L. Max Depth	m	log(x)	449	0.9	120	14
Lake Area	Lake Area	km^2	log(x)	523	0.0086	5550	1.7
Water Body Area	Water Body Area	km^2	log(x)	523	0.0086	3086	1.7
Meter Above Sea Level	Masl	m	No	523	0.1	379.1	75
Latitude	Latitude	-	No	523	55.49	61.17	58.98
Longitude	Longitude	-	No	523	11.23	18.93	15.26
Secchi Depth	Secchi Depth	m	log(x)	320	0.3	15.5	2.2
Lake pH	Lake-pH	-	No	301	4.8	9.8	7.2
Ammonium Nitrogen	NH_4-N	µg/l	log(x)	265	0.5	253.3	14.0
Nitrate and Nitrite	NO_2+NO_3-N	µg/l	log(x)	285	1.0	724.8	10.2
Organic Nitrogen	Org-N	µg/l	log(x)	165	58.0	1220.0	434.5
Total Nitrogen	TN	µg/l	log(x)	297	216.2	2825.0	535.0
Phosphate	PO_4-P	µg/l	log(x)	245	0.96	145.25	2.88
Total Phosphorus	TP	µg/l	log(x)	303	1.0	332.5	16.0
Conductivity	Cond	mS/m	log(x)	301	1.56	53.93	7.42
Calcium	Ca	mekv/l	log(x)	216	0.023	3.039	0.328
Magnesium	Mg	mekv/l	log(x)	216	0.02	1.04	0.12
Sodium	Na	mekv/l	log(x)	213	0.05	1.28	0.22
Potassium	K	mekv/l	log(x)	214	0.004	0.153	0.029
Alkalinity	Alk	mekv/l	log(x+1)	298	-0.040	2.500	0.240
Sulphate	SO4	mekv/l	log(x)	213	0.019	2.105	0.124
Chlorine	Cl	mekv/l	log(x)	216	0.024	1.365	0.204
Fluorine	F	mg/l	log(x)	157	0.019	0.745	0.150
Total Organic Carbon	TOC	mg/l	log(x)	291	0.8	33.0	10.5
Silicon	Si	mg/l	log(x)	180	0.1	8.3	1.0
Absorbance	Abs	420nm/ 5cm	log(x)	263	0.001	0.746	0.106
Turbidity	Turbidity	FNU	log(x)	209	0.25	59.50	1.70
Iron	Fe	µg/l	log(x)	110	8.1	5440.0	191.6
Manganese	Mn	µg/l	log(x)	110	1.2	730.0	49.6
Total Biovolume Phytoplankton	Tot Bio	mm^3/l	log(x)	523	0.05	115.78	1.61
Chlorophyll-*a*	Chl-*a*	µg/l	log(x)	321	0.50	231.95	8.05
Plankton Trophic Index	PTI	-	No	523	-0.933	1.859	0.088
Number of Taxa	Taxa	count	No	523	4.5	95.0	45.3

2.1.2 Lake Catchments

The digitized geographical information about the lake catchments comes from the Department of Aquatic Sciences and Assessment geographical catchment-database at SLU (Miljödata-MVM 2019). The database provided corresponding lake catchment based on the WATERS dataset catchment ID. The lake catchments were delineated in a shapefile containing polygons, with one polygon for each lake catchment. The catchment polygons were thereafter used in the study to extract the different catchment properties, see Section 3.2.

2.2 Catchment Characteristics

For the description of the catchment characteristics, different variables regarding land use, soil properties (texture, content, chemistry) but also climate factors were investigated. The material used comes from two different source types, either as continuous map of studied variables, or in form of point data from national soil survey's sampling points. Overview of the different catchment properties is given in *Figure 1* (Section 2). Variables derived from the same source type are grouped as a map variable or as a soil survey variable. All catchment variables included in this study are divided by the sources and can be found in *Table 2*. Note that some variables may have similar names but represent different content, which mainly depends on different definitions used in the original data source. Globally there are many soil type classification systems used. In this study both the Swedish system (Eriksson *et al.* 2014) and the international classification system FAO/USDA (USDA 2019) were used, where the last mentioned system describing soil texture (composition of sand, silt and clay) was used for arable land.

2.2.1 Land Use - PLC

In order to illustrate the land use distribution for the lake catchments, the information from the developed and compiled geographical map within Pollution Load Compilation 6 (PLC 6) project was used (Widén-Nilsson *et al.* 2016). The PLC 6 Land Use map were originally developed for the calculation of sources of N and P loads to Swedish seas for year 2014, and used for reporting to HELCOME about PLC 6 (Ejhed *et al.* 2016). The PLC 6 shapefile used in this study contain a detailed land use distribution map as polygons and covers southern Sweden up to and including Dalarna county. The land use map contains eleven land use categories: urban area, forest, open land, mountain, water, sea, mire, arable land, clear cutting, wetland and pasture. The PLC 6 Land Use map was produced based on digital maps from different authorities (Widén-Nilsson *et al.* 2016). The main source of the land use map was the GSD-roadmap from Swedish Mapping Cadastral and Registration Authority (Widén-Nilsson *et al.* 2016). The Swedish Board of Agriculture assisted with data to improve distribution of arable land, whereas urban areas were derived from the Statistics Sweden data (Widén-Nilsson *et al.* 2016). Finally, the Swedish Forest Agency provided information regarding areas that have been clear cut.

2.2.2 Soil Texture Distribution

For Soil Texture Distribution, two data sources were used. Firstly, for non-agricultural land, a geographical map from the Geological Survey of Sweden (SGU) was used. A combination of different SGU maps with the best available data developed by SGU and used in Djodjic & Markensten (2018), was also used in this study. The combined SGU soil map contains a raster with spatial resolution 25 x 25 m and covers the southern half of Sweden. The map includes 14 categories ranging from sand, clay, mountain to till soil. The soil definitions are based on the Swedish classification system were for instance soils with more than 15 % clay content defines as clay soil (Eriksson *et al.* 2014). Variables from this dataset are named with SGU as the first letters followed by the variable name.

Secondly, for the detailed description of the topsoil texture properties of the arable land, the Digital Arable Soil Map of Sweden (DSMS) (Söderström & Piikki 2016) was used. This map was developed for the national mapping of arable land in Sweden by SLU in cooperation with SGU (Söderström & Piikki 2016). The DSMS map is built on a concept where reference soil analysis (soil sampling) was connected to available remote sensor data for model prediction by multivariate adaptive regression splines. The DSMS map was produced based on around 15 000 reference soil samples from 2011-2012, combined with gamma radiation data and digital elevation model derived from airborne radiometric scanning and airborne light detection and ranging (LIDAR), respectively. These input data together with quaternary geological map build the two primary DSMS digital layers of clay- and sand content in arable land topsoil. The DSMS clay layer had a 0-80 percentual continuous range and DSMS sand 0-100 percentual range. A silt content map was derived from the primary maps (DSMS clay and DSMS sand) with a 0-99 continuous range. A DSMS map was also developed for twelve texture classes according to FAO/USDA system. For the organic layer, DSMS organic, show just if organic soil is present or not in the area.

In this study the five DSMS maps: clay, sand, silt, FAO and organic were used for describing the arable topsoil texture. All DSMS layers are raster with a spatial resolution of 50 x 50 m and cover arable land up to and including Gävleborg county. The estimated uncertainties when comparing predicted and measured values differ between regions and DSMS layers, with an overall 6 % mean absolute error (R^2=

0.76) for DSMS clay and 11 % (R^2= 0.57) for DSMS sand (Söderström & Piikki 2016). The predicted model for organic are highly uncertain according to Söderström & Piikki (2016).

2.2.3 Arable Topsoil Texture and Chemistry

Underlying material for the extraction of the chemical properties in arable land consists of the data from national soil survey of arable land where almost 12 600 topsoil samples were collected and analysed during 2011-2012 (Paulsson *et al.* 2015). The soil survey covers Sweden's arable land excluding the four northern counties. Sampling locations are distributed as a grid, with an average density of one sample per 200 hectares, with higher sample density in regions with large portion of agricultural land (Paulsson *et al.* 2015). The samples describe the topsoil, down to 20 cm, content according to Swedish particle size scale of clay (< 2 µm), silt (2-60 µm), sand (0.2-2 mm) and gravel (> 2 mm). Soil analyses also included the organic material content, pH, aluminium, and iron and plant available nutrients phosphorus, magnesium, calcium and potassium extracted by ammonium lactate (Paulsson *et al.* 2015). The soils' Phosphorus Sorption Capacity (PSC), was also calculated as a sum of aluminium and iron on molar basis and used as an indicator of available P-binding sites. The potential P leaching risk, estimated as molar ratio between the content of plant available P and PSC, called Degree of Phosphorus Saturation (DPS) was also calculated and included in the analyses. These sampling points with associated analysed variables were used in this study to describe the chemical properties of the arable topsoil for each catchment area, for further specification of the variables see *Table 2*.

2.2.4 Forest Soil Chemistry- MI

The forest soil chemistry data used in this study originate from the Swedish Environmental Protection Agency monitoring stations covering Sweden as a grid. The propose of the inventory program is to provide a basis for both nationwide estimation of the forest soil status and to follow changes in forest soils. Hence, the sampling sites are reinvestigated with regular intervals (Nilsson *et al.* 2015). The Land Inventory Database (Markinventerings databas) host the sample data (Stendahl 2019). Variables derived from this database are named MI as first letters, referring to the Swedish term Markinventering. The samples included in this study where investigated between 2003-2012 to approximately cover the same period as both lake variables and soil survey on arable land. They represent different humus form types in the upper 0-30 cm layer and include pH (extracted with deionized water) as well as carbon and nitrogen content in percent weight (Nilsson *et al.* 2015; Stendahl

2019). All samples are collected in the soils' fine fractions, ≤ 2 mm (Nilsson *et al.* 2015; Stendahl 2019). The samples are mainly collected in forest and mire, with few samples from pasture, rock and impediment. In the inventory program, forest soil samples are included only from the productive forest producing on average at least one cubic meter wood per hectare and year (Nilsson *et al.* 2015).

2.2.5 Climate: Temperature and Precipitation

Material included for climate factors was developed by SMHI (Johansson 2000). Annual mean value of air temperature (°C) and precipitation (mm) from period 1961 to 1990 was used. The original data comes from the metrological station network, covering Sweden, where geostatic interpolation was used to extrapolate measured values to two separate raster files with spatial resolution of 4000 x 4000 m over whole Sweden (Ejhed *et al.* 2016).

Table 2. *Catchment variables used in the study, derived from different map sources and soil surveys shown in the first column. Abbreviations illustrate the source and names used for the variable group: PLC-Pollution Load Compilation 6, DSMS-Digital Arable Soil Map of Sweden, SGU-Geological Survey of Sweden, and MI-Markinventering. Min, max and median for the variable and amount of catchments (N) represented. Note that soil survey variables are not available for all catchments. Transformation type used for each variable are also given.*

	Variable description	Variable name	Unit	Transf.	N	Min	Max	Median
	Catchment Area	Catch. Area	km^2	log(x)	523	0.09	46839.7	48.4
Land Use- PLC	PLC6-2 Urban Area	Urban Area	%	log(x+1)	523	0	85.4	0.2
	PLC6-3 Forest	Forest	%	No	523	0	94.3	63.4
	PLC6-4 Open Land	Open Land	%	log(x+1)	523	0	39.0	4.5
	PLC6-5 Mountain	Mountain	%	log(x+1)	523	0	2.9	0
	PLC6-6 Water	Water	%	log(x+1)	523	0	40.6	11.1
	PLC6-7 Sea	Sea	%	log(x+1)	523	0	0.00001	0
	PLC6-8 Mire	Mire	%	log(x+1)	523	0	63.9	5.2
	PLC6-9 Arable Land	Arable Land	%	log(x+1)	523	0	54.5	5.1
	PLC6-10 Clear Cutting	Clear Cutting	%	log(x+1)	523	0	15.8	2.7
	PLC6-11Wetland and other unknown	Wetland+other	%	log(x+1)	523	0	3.9	0.1
	PLC6-12 Pasture	Pasture	%	log(x+1)	523	0	27.3	1.0
	PLC6-9+11+12 Agriculture	Agriculture	%	log(x+1)	523	0	64.3	7.4
Digital Arable Soil Map of Sweden- DSMS	DSMS-Clay Content	DSMS-Clay Cont	% (mean)	No	523	0	46.0	15.9
	DSMS-Silt Content	DSMS-Silt Cont	% (mean)	No	523	0	75.0	40.0
	DSMS-Sand Content	DSMS-Sand Cont	% (mean)	No	523	0	84.5	32.8
	DSMS-Organic Content	DSMS-Organic Cont	%	log(x+1)	523	0	4.5	0.2
	DSMS FAO 1 Sand	FAO-Sa	%	log(x+1)	523	0	4.5	0
	DSMS FAO 2 Loamy Sand	FAO-LS	%	log(x+1)	523	0	17.8	0
	DSMS FAO 3 Sandy Loam	FAO-SaL	%	log(x+1)	523	0	33.0	0.4
	DSMS FAO 4 Loam	FAO-L	%	log(x+1)	523	0	25.6	0.8
	DSMS FAO 5 Silt Loam	FAO-SL	%	log(x+1)	523	0	62.9	0.1
	DSMS FAO 6 Silt	FAO-S	%	log(x+1)	523	0	0.2	0
	DSMS FAO 7 Sand Clay Loam	FAO-SaCL	%	log(x+1)	523	0	1.3	0
	DSMS FAO 8 Clay Loam	FAO-CL	%	log(x+1)	523	0	19.4	0
	DSMS FAO 9 Silt Clay Loam	FAO-SCL	%	log(x+1)	523	0	15.7	0

	Variable description	Variable name	Unit	Transf.	N	Min	Max	Median
	DSMS FAO 10 Sand Clay	FAO-SaC	%	log(x+1)	523	0	0.03	0
	DSMS FAO 11 Silt Clay	FAO-SC	%	log(x+1)	523	0	20.7	0
	DSMS FAO 12 Clay	FAO-C	%	log(x+1)	523	0	15.3	0
Soil Texture Distribution- SGU	SGU 13 Organic Soil	SGU-Organic Soil	%	log(x+1)	523	0	62.4	7.6
	SGU 14 Clay	SGU-Clay	%	log(x+1)	523	0	41.6	0.4
	SGU 15 Silt	SGU-Silt	%	log(x+1)	523	0	26.0	0.2
	SGU 16 Sand	SGU-Sand	%	log(x+1)	523	0	83.8	0.5
	SGU 17 Gravel	SGU-Gravel	%	log(x+1)	523	0	14.7	0
	SGU 18 Cobbles to Boulders	SGU-Cobbles to Boulders	%	log(x+1)	523	0	1.3	0
	SGU 19 Fluvio-Glacial Sediment. Sand-Block	SGU-Fluvio-Glacial Sed.	%	log(x+1)	523	0	81.1	1.3
	SGU 20 Clay Till	SGU-Clay Till	%	log(x+1)	523	0	26.3	0
	SGU 21 Till (Moraine)	SGU-Till	%	No	523	0	89.1	36.0
	SGU 22 Thin Soil Layer	SGU-Thin Soil Layer	%	log(x+1)	523	0	74.9	4.3
	SGU 23 Rock	SGU-Rock	%	log(x+1)	523	0	86.0	6.7
	SGU 24 Artificial Fill	SGU-A.Fill	%	log(x+1)	523	0	8.3	0
	SGU 25 Other	SGU-Other	%	log(x+1)	523	0	3.9	0
	SGU 26 Water	SGU-Water	%	log(x+1)	523	0	36.7	10.8
[1]Arable Topsoil Texture & Chemistry	Arable Land pH	Arable-pH	-	No	314	4.80	7.74	6.07
	Organic Material Content	OM Cont	%	log(x)	314	1.5	63.3	5.6
	Clay Content	Clay Cont	%	No	314	0	50.0	18.9
	Silt Content	Silt Cont	%	No	314	13.0	100.0	46.6
	Sand Content	Sand Cont	%	No	314	0	82.0	32.0
	Gravel Content >2mm	Gravel Cont	%	log(x)	314	0.02	15.80	1.38
	Phosphorus. Ammonium lactate extracted	P_AL	mg/kg soil	log(x)	314	10.0	220.0	55.7
	Potassium. Ammonium lactate extracted	K_AL	mg/kg soil	log(x)	314	25.0	336.3	109.0
	Magnesium. Ammonium lactate extracted	Mg_AL	mg/kg soil	log(x)	314	5.0	600.0	131.9
	Calcium. Ammonium lactate extracted	Ca_AL	mg/kg soil	log(x)	314	50.0	8814.9	1424.6
	Aluminium. Ammonium lactate extracted	Al_AL	mg/kg soil	log(x)	314	99.0	1900.0	424.6
	Iron. Ammonium lactate extracted	Fe_AL	mg/kg soil	log(x)	314	110.0	2400.0	490.8

	Variable description	Variable name	Unit	Transf.	N	Min	Max	Median
	Degree of Phosphorus Saturation. Ammonium lactate extracted	DPS	%	log(x)	314	1.0	44.8	8.5
	Phosphorus Sorption Capacity	PSC	mmol	log(x)	314	8.0	91.0	24.8
[2]MI- data	MI-Carbon Content	MI-C. Cont	% weight	No	291	4.13	48.80	32.48
	MI-Nitrogen Content	MI-N. Cont	% weight	No	291	0.20	2.14	1.15
	MI-pH deionized H_2O extracted	MI-pH	-	No	291	3.31	5.95	3.96
[3]Climate	Air Temperature	Temp	℃	No	523	2.2	7.6	5.5
	Precipitation	Precip	mm	log(x)	523	558	1134	688

1. Number of soil samples in catchments extent from min 1 to max 2241, median 7. Number of samples constitutes to calculated mean value for each catchment and variable.

2. Number of soil samples in catchments extent from min 1 to max 917, median 10. Number of samples constitutes to calculated mean value for each catchment and variable.

3. Mean value for each catchment with data from 1961-1990 annual mean.

3 Methods

3.1 Study Boundaries

This study has been geographically limited to include lakes with phytoplankton data and associated catchment south of 62°0´0´´N. So, this study thereby includes 487 lakes (523 sampling sites) situated in the southern part of Sweden. Reasons behind this limitation are mainly based on the most limited available map source material. Thus, in order to be able to access and extract the same material for all catchments and thereby get as homogeneous and as representative material as possible, the study focus on the lakes in the southern part of Sweden. For instance, both DSMS map and soil survey on arable land cover only the southern part of Sweden. Note that larger lakes with several waterbodies will be represented by the whole lake catchment, meaning that they have the same catchment characteristics but different lake biology and chemistry data.

3.2 Geographic Information System Analyses

To analyse the different geographical information, ArcGIS 10 (ESRI Inc 2018) was used with SWEREF 99 TM as the coordinate system. The catchment polygon shapefile was the layer on which the extraction of all other data was based on.

Since the smaller polygon catchment areas can be overlapped by larger catchment areas from other lakes the toolbox *Spatial Analyst Supplemental Tools* (ESRI 2017) was used to handle the overlapping polygons. Specifically, *Tabulate Area 2* and *Zonal Statistics 2* were the two tools used to calculate distribution of variables within each catchment. Using these tools, each data source is thus run separately in ArcMap, where the catchment polygon layer defined the zone for which the distribution of a given variable was calculated. Generally, *Tabulate Area 2* tool was used for raster sources with categorical values whereas the *Zonal Statistics 2* tool was used for data variables with continuous values. The results of these calculations were tables containing different statistical values, most often the mean value for a given variable per catchment of interest. The tables produced with ArcMap were then processed and analysed in Excel software (Microsoft Office 2016). The following part describes how the different catchment properties were linked and extracted to the catchment polygons in ArcMap.

To make it easier to implement *Tabulate Area 2* tool, the PLC 6 Land Use feature layer (polygon) was transformed to raster with 25 x 25 m grid size. SGU-Soil Texture Distribution and DSMS FAO were also processed by *Tabulate Area 2* tool. Optimal processing cell size were set to the layer's own cell size (PLC 6 Land Use and SGU-Soil Texture Distribution 25 x 25 m and DSMS FAO 50 x 50 m). The output resulted in tables with calculated area, where proportion was calculated for each category present in the catchment. Agricultural land was calculated by adding arable land, pasture and wetlands on arable land and other agriculture land together. As mentioned earlier, DSMS FAO map (categories 1-12) was used on arable land.

The other arable land DSMS layers (content of clay, silt and sand particles), consisted of continues values (percent). All statistics for these layers were calculated by *Zonal Statistics 2*. A comparison of mean and median in the statistical output was made to identify possible substantial differences. Finally, it was checked that no major deviations exist when these three DSMS percentage values were summarised for each catchment, as the sum of them should amount to 100 %. For DSMS organic, a manually percentual value were calculated based on the DSMS organic area and total catchment area.

In the same way as described above, the *Zonal Statistics 2* tool was also used for the climate variables, temperature and precipitation. Due to the large spatial resolution (4000 x 4000 m) in these climate layers, smaller catchment areas needed to be treated differently. By using *Feature to Point* tool for these small catchments statistical mean value was calculated by *Extract Multi Values to Points* tool. When doing so, the point located in the middle of the small catchment was chosen as representative and mean values of temperature and precipitation were extracted for that point. All catchments were then described by a mean value for temperature and precipitation from 1961-1990 annual mean period.

The point layer with 12 600 arable topsoil samples were linked to the catchment polygon by *Spatial Join* tool, with join criteria that one sample point could belong to several polygons, to account for the overlapping polygons. Sample points not identified in a catchment area were excluded from the study. The number of sample points in each catchment and the mean of these points' associated variables were thereafter calculated in Excel. The soil samples from this data set cover only arable land and were not available in all catchments. Similarly, as the arable soil samples, the MI soil samples dataset was processed in exactly the same way. The MI data were derived only for catchment with available soil sample point(s) present in the catchment.

Finally, all properties produced in ArcMap were compiled in Excel for each catchment and prepared for furtherer statistical analysis.

3.3 Statistical Analysis

3.3.1 Data Processing

All statistical analysis were done with JMP Pro 14 (SAS Institute Inc 2018). The Shapiro-Wilk Test ($\alpha= 0.05$) was used to test normality of all variables in the dataset. To handle non-normal distributed data, logarithmic transformation is usually suggested for continuous data (Zar 1984) and was used for many variables in the dataset. For the individual variable's transformation see *Table 1* and *Table 2* in Section 2. In some cases, untransformed data was used when the transformation did not improved normality according to Shapiro-Wilk Test. A scatterplot matrix was used to visualize data, discover outliers but also as an indication of how variables in the dataset correlate and which variables could be important for the lake biology. The variables SGU water and PLC water, represent the same information (water area), and the latter was then used with an assumption of higher precision of the PLC data. PLC sea and mountain variables were excluded for further analysis since they were present in just a few catchments. Linear regression was used to identify correlation between TP and phytoplankton tot bio as well as PTI and chl-*a* in the lakes. A Principal Components Analysis (PCA) was thereafter performed to show correlation between all variables in the dataset, including both lake and catchment variables.

3.3.2 PLS- Partial Least Squares Analysis

The Partial Least Squares (PLS) analysis method was used to identify the catchment properties explaining the variation of the lake phytoplankton and concentration of TP. The result of PLS gives an indication of important variables in the catchment for further processing. Various models were built and used in PLS for the three dependent variables (Y-variables) in the lakes: tot bio, PTI and TP, with catchment properties as the independent variables (X-variables). For each dependent lake variable, three different models were produced and used in PLS. The included groups of independent catchment variables for the three models are given in *Table 3*: the first four variable groups are data from continuous map sources while the two last variable groups (Arable Soil Texture & Chemistry and Forest Soil Chemistry-MI) represent national soil survey data in the catchment. The remaining two variable groups addressed lake location and area where in total five variables are derived from the WATERS dataset, hereafter also referred to as map variables. The last sixth variable (catchment area) in this group was derived from catchment shape file. Note that PLS analysis were based on available data just for the catchments sharing common variables selected in each model, meaning that catchment or waterbodies with missing values for one chosen variable were excluded from the analysis. Maximal available data for dependent variables tot bio and PTI were all (523) waterbodies and for TP 303 waterbodies.

Model 1 included soil survey data on arable land and forest (MI-data) together with all map variables (first six variable groups in *Table 3*) in the dataset. For model 2, only map data is included as the independent variables. The difference between model 1 and 2 was that model 1 represented fewer catchments but included all available catchment variables compared to model 2, which included all catchments. Model 3 included the same map variables as model 2 but only catchments containing less than 10 % of agricultural land.

Table 3. *Overview of the groups of catchment variables included in each model (M1-M3) for PLS analysis with dependent variable being: total biovolume phytoplankton, Plankton Trophic Index and total phosphorus. Grey cells show variable groups included in each model. The first four variable groups represent map data and the last two soil survey data. Five variables related to lake location and lake area are derived from the WATERS dataset were the catchment area are derived from catchment shape file, and these variables will also be referred as map variables. For detailed description of each variable included see Table 1 and Table 2. Number of catchments with available data showed by maximal number of catchments, meaning that all catchments are not represented in the model. PLC-Pollution Load Compilation, DSMS-Digital Arable Soil Map of Sweden, SGU-Geological Survey of Sweden, and MI-Markinventering.*

	No. Variables	Max No. Catchments	M1	M2	M3*
Land Use - PLC	10	523	▨	▨	▨
DSMS	16	523	▨	▨	▨
Soil Texture Distribution - SGU	13	523	▨	▨	▨
Climate	2	523	▨	▨	▨
Lake Location	3	523	▨	▨	▨
Area (Lake & Catchment)	3	523	▨	▨	▨
Arable Soil Texture & Chemistry	14	314	▨		
Forest Soil Chemistry - MI	3	291	▨		
Total Variables			64	47	47

*M3 including catchment with <10 % agriculture.

The PLS analysis options were set to centering and scaling with NIPALS (Nonlinear Iterative Partial Least Squares) as method specifications. The validation method used was Leave-One-Out, meaning that the PLS cross-validate by leaving out one row (catchment) at a time when running the model to find the best dimension for building the PLS-model. The factor search range (dimensions) was set to default, maximum 15 factors. The first factors should in a good PLS model explain most of the variation in X and Y (Cox & Gaudard 2013). Variable Importance for the Projection (VIP) generated from the PLS analysis works as an indicator for how much each variable contributes to explain the model (Cox & Gaudard 2013). Generally, variables with small VIP value, less than 0.8, are consider as not important while variables with VIP over 1.0 are considered important (Cox & Gaudard 2013). For this reason, the criteria VIP >1.0 generated by the PLS analysis, for each model and dependent lake variable, was used as an indicator for selection of the important independent variables for further processing.

3.3.3 Multiple Regression Analysis

Catchment variables generated by the PLS analysis with VIP >1.0 were included in stepwise multiple regression for each model 1-3 and each dependent lake variable. Before the stepwise analysis was done, the variables representing almost the same information were evaluated and one of the variables was excluded. For example, when both arable land and agricultural land (the sum of arable land, pasture and other agricultural land) were estimated in PLS as important variable only agricultural land was kept to avoid high correlation of independent variables. For all stepwise regression analysis, the default option was used, with minimum BIC (Bayesian Information Criterion) as stopping rule and forward direction. Using these settings, JMP produces the best stepwise regression based on minimum BIC with output displayed in step history. So, for every lake variable (tot bio, PTI and TP), three stepwise regression were made using the catchment variables from the PLS analysis with VIP >1.0. Thereafter, for the phytoplankton dependent variables (tot bio and PTI), the same procedure was repeated in exactly same way but with TP now also included as independent variable in the stepwise regression.

Individual assessment was also made of variables included for each final multiple regression to limit the number of variables without substantially reducing the proportion of variation explained. For this, the output of the stepwise regression, step history's R^2 value, was used. The criteria for the inclusion of a certain variable in the multiple regression was the following: when addition of a given variable at a time suggested by stepwise regression contributed to at least 2 % increase of step history R^2, the variable was included in the multiple regression. If the variable contributed less than 2 %, the variable was excluded, and no further variables were added for the multiple regression. In case the variable passed this criteria, but was in the multiple regression not statistically significant ($p > 0.05$), the variable was also excluded. Additionally, if the multiple regression selected by criteria was the same as the minimum BIC selected regression, then just one regression was the output for the model i.e. the one chosen by the stepwise regression (minimum BIC).

All produced multiple regressions for each model and each dependent lake variable can be seen in associated Appendix. For the TP models, there are thus two multiple regression possible as output for each model based on the variables with VIP >1, namely one selected by the JMP (minimum BIC) and another selected by criteria described above. Additionally, for each PTI and tot bio model, four multiple regression were possible since TP was also included as independent variable: two regressions selected by JMP (minimum BIC), with and without TP as independent variable, and two selected by criteria, with and without TP as independent variable. The flow chart over the statistical analysis can be seen in *Figure 2*.

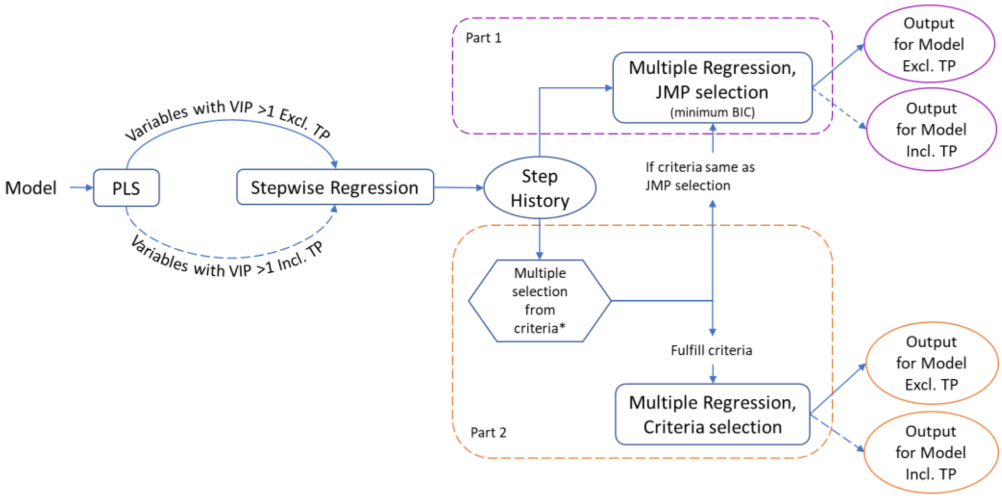
Model
PLS
Variables with VIP >1 Excl. TP
Variables with VIP >1 Incl. TP
Stepwise Regression
Step History
Part 1
Multiple Regression, JMP selection (minimum BIC)
Output for Model Excl. TP
Output for Model Incl. TP
If criteria same as JMP selection
Multiple selection from criteria*
Fulfill criteria
Part 2
Multiple Regression, Criteria selection
Output for Model Excl. TP
Output for Model Incl. TP

Figure 2. Flow chart over statistical analysis with Partial Least Squares (PLS) and stepwise regression with associated output. For each dependent variable: total biovolume phytoplankton (tot bio), Plankton Trophic Index (PTI) and total phosphorus (TP) three models were produced. All models 1-3, underwent the PLS analysis. Variables with VIP >1 from PLS analysis were thereafter processed in the stepwise regression for each model separately resulting in a multiple regression (solid arrow). Same analysis was repeated to include TP as independent variable (dashed arrow) for tot bio and PTI. Stepwise regression process is further divided in part 1 performing the multiple regression based on all variables selected by JMP, minimum BIC (Bayesian Information Criterion), and part 2, where further reduction of variables was done based on following criteria*: Step history's R^2 value were used to assess whether addition of variable one at a time contributed to ≥ 2 % in R^2. If variable addition contributes < 2 %, it was excluded, no more variables were included. Also, in cases when variable contributed to ≥ 2 % but was not statistically significant ($p > 0.05$) in the following multiple regression, the variable was excluded. Additionally, if the criteria were the same as JMP selection, just one regression was the output i.e. the one selected by the stepwise regression. For TP, two multiple regression were possible as output for a model. In total, four multiple regression are possible as output for a model from PTI and tot bio.

4 Results

4.1 Description of the Dataset

A total of 487 lakes located in southern Sweden were used in the analyses. Eight of these lakes are large lakes and represented 44 waterbodies, making total number of included waterbodies to 523. The lakes are distributed in the southern part of Sweden, up to just above latitude 61°N, see *Figure 3*. The dataset contains lakes with various properties sampled in July-August between 2000-2012 (mean value), see *Table 1* in Section 2.1.1. The lake area ranges from smallest 0.0086 km^2 to largest 5550 km^2 with median of 1.7 km^2, while the largest water body area is 3086 km^2. The total biovolume of phytoplankton varies between 0.05-116 mm^3/l (median 1.6 mm^3/l) whereas PTI-value range from -0.9 to 1.9 (median 0.09), with data available for all 523 waterbodies. Chlorophyll-*a* ranges between 0.5-232 µg/l, with a median 8 µg/l (N= 321). Total phosphorus was available for 303 waterbodies with range 1.0- 332.5 µg/l, and median 16.0 µg/l.

The associated catchments had a median area of 48 km^2 but range from the smallest 0.09 km^2 to the largest 46840 km^2, see *Table 2* in Section 2.2. The share of forest (0 to 94.3 %) and agricultural land (0 to 64.3 %) also varied considerably. In the entire dataset of all 523 waterbodies, 310 of those waterbodies were characterised by less than 10 % agricultural land in the catchment, which was used in the sub-analysis, model 3. In the whole dataset the share of urban area was 0.2 % (mean) but varies between 0 to 85.4 % in the catchments.

The comparison of mean and median from the GIS statistical output for the DSMS layer (content of clay, silt and sand particles) showed small differences and therefore the mean values are presented here in the results.

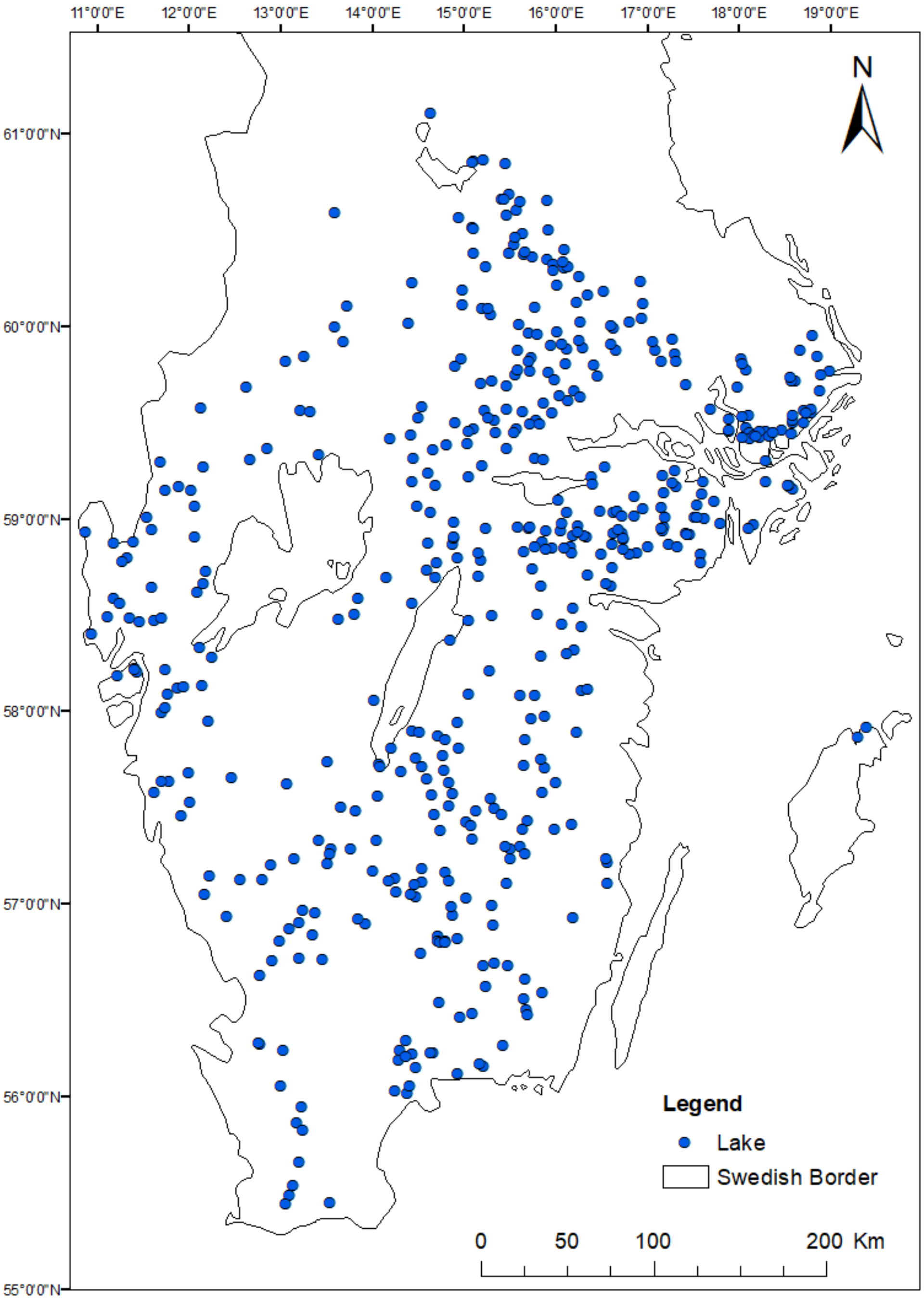

Figure 3. Location of the 487 investigated lakes in southern Sweden. Eight lakes are larger and consists together of 44 waterbodies (not shown in the map) for example lake Värnen, Vättern and Mälaren. A total of 523 waterbodies were included in the dataset.

4.2 PCA Analysis

The PCA analysis shows the correlation between all lake and catchment properties in the dataset, see *Figure 4*. The first PCA component explains 23.8 % and the second 10.9 % of the variation. Variables correlated positively to each other are close to each other, while a negative correlation is indicated by being at the opposite sides of the centre. Variable strength is shown by the distance from the centre, with higher strength farther away from the middle. Generally, the PCA indicates that lake phytoplankton indicators (chl-*a*, tot bio and PTI) are positively correlated with each other and lake TP. These variables also have positive correlation with certain catchment properties, for instance agriculture, and negative correlations with variables such as forest and secchi depth. Number of taxa is located close to the middle with lower strength and weaker correlation to the other variables in the PCA analysis.

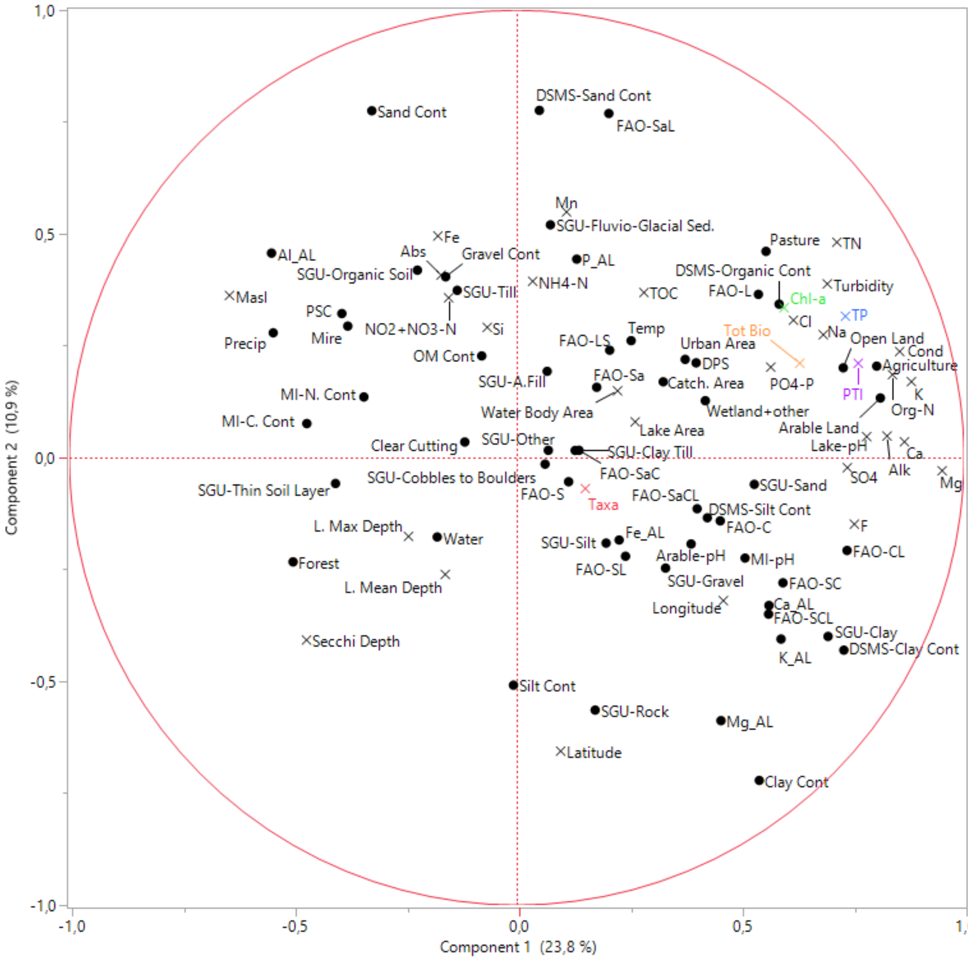

Figure 4. PCA analysis for all 93 variables included in the dataset: 34 lake variables (cross) and 59 catchment variables (filled circles). Coloured variables show the variables in focus of the study. For detailed description of abbreviations see *Table 2* for the catchment variables and see *Table 1* for the lake variables.

4.3 Linear Regression Analysis

Total phosphorus of the lake water was highly correlated with chl-*a* (R^2= 0.78, p < 0.001), tot bio (R^2= 0.66 p, < 0.001) and PTI (R^2= 0.56, p < 0.001) with 303 observations, see *Figure 5*A-C. The correlation between chl-*a* and tot bio was also strong (R^2= 0.75, N= 321, p < 0.001, *Figure 5*D) therefore only tot bio was investigated further.

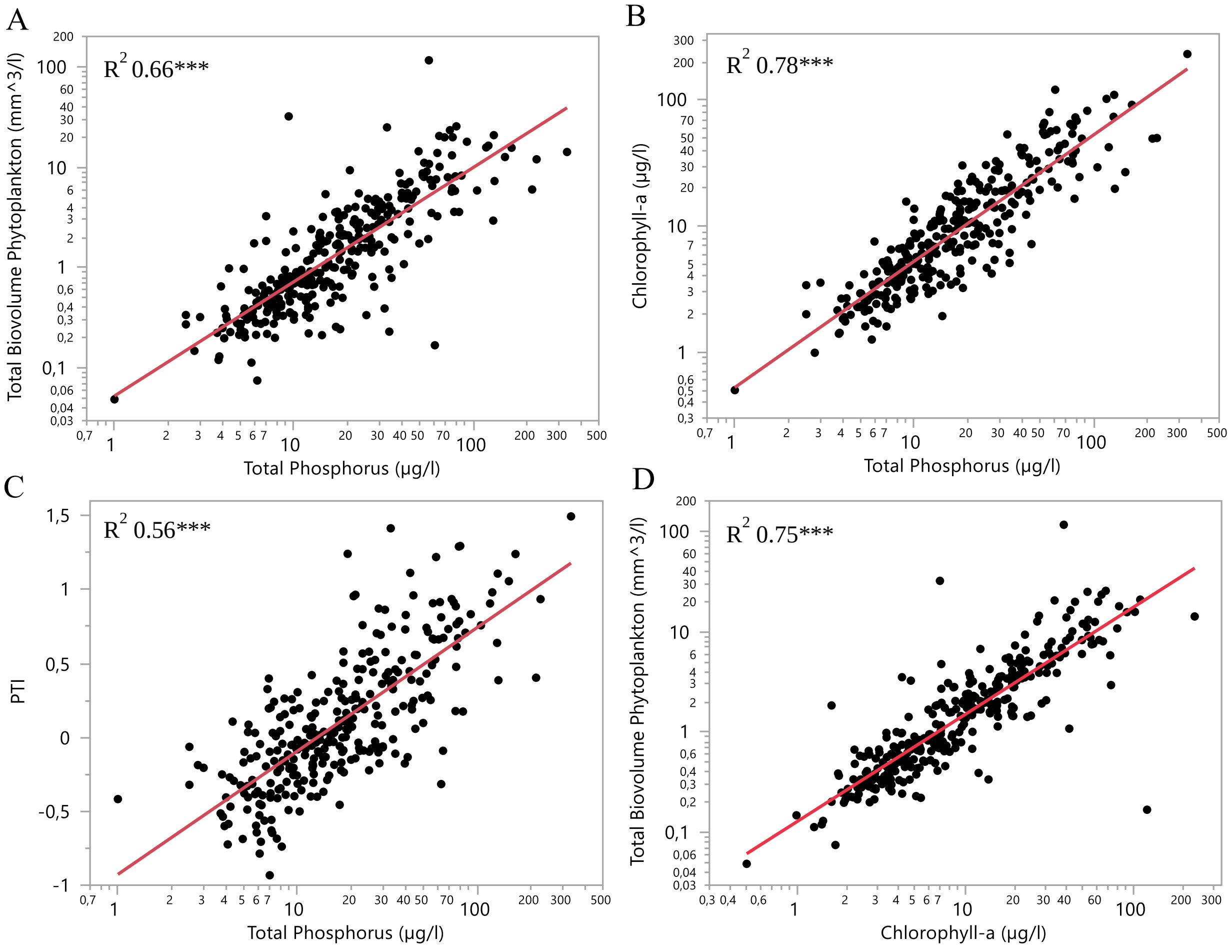

Figure 5. Linear regression for total phosphorus A) total biovolume phytoplankton B) chlorophyll-*a* C) PTI (Plankton Throphic Index) for 303 observations, and D) correlation of total biovolume phytoplankton and chlorophyll-*a* for 321 observations. Note the logarithmic scales. * ($p < 0.05$) ** ($p < 0.01$) *** ($p < 0.001$)

4.4 Total Biovolume of Phytoplankton

The three PLS models produced for tot bio explain the variation of Y and X variables to different extent, see *Table 4*. Model 2, including all 523 catchments, explain the highest variation of Y (43.7 %) and X (50.4 %). The number of catchments included in the models differs due to differences in data availability and includes 250 catchments in model 1 and 310 catchments in model 3, for catchment having less than 10 % of agricultural land. The number of factors explaining the variation differ also between the models.

Table 4. *PLS analysis for total biovolume phytoplankton as Y and catchment variables as X. Included number of X variables, selected catchments and numbers of factors explaining the variation for X and Y in the model.*

Model	No. of X Variables Incl.	Catchments	No. of Factors	Variation Explained for Cumulative X (%)	Variation Explained for Cumulative Y (%)	No. of VIP > 1
M1	64	250	2	32.9	30.6	26
M2	47	523	6	50.4	43.7	19
M3*	47	310	3	33.8	39.1	21

*M3 represent catchment with <10 % agriculture.

Overall, there is a small difference regarding the selection of most important variables (VIP >1) included in each individual model, as seen in *Figure 6*, *7* and *8*. Regardless of model, the share of agricultural land was shown to be important. Map variables shown to be important in all models for tot bio were different land use categories, soil composition, lake location and climate (precipitation).

Important catchment variables with VIP >1 included in model 1 indicate that both map sources and soil survey data were important for the lake tot bio (*Figure 6*). The share of agricultural land use has the highest VIP value, whereas variables from the soil surveys data were in the lower VIP range with MI-pH (forest soil survey pH) at the top (*Figure 6*). Model 2, including only map variables, shows that variables such as land use categories, clay content, FAO-textural classes with finer soils (clay loam and silty clay loam), lake location (longitude and altitude (masl)) and precipitation are important for the model (*Figure 7*). Variables with highest VIP value for model 2 are agricultural land use and forest. The result for model 3 which only include catchments with less than 10 % agricultural land, illustrated with VIP values in *Figure 8*, show that most important variables for lake tot bio are still the share of agricultural land but also lake location along the longitudinal axis.

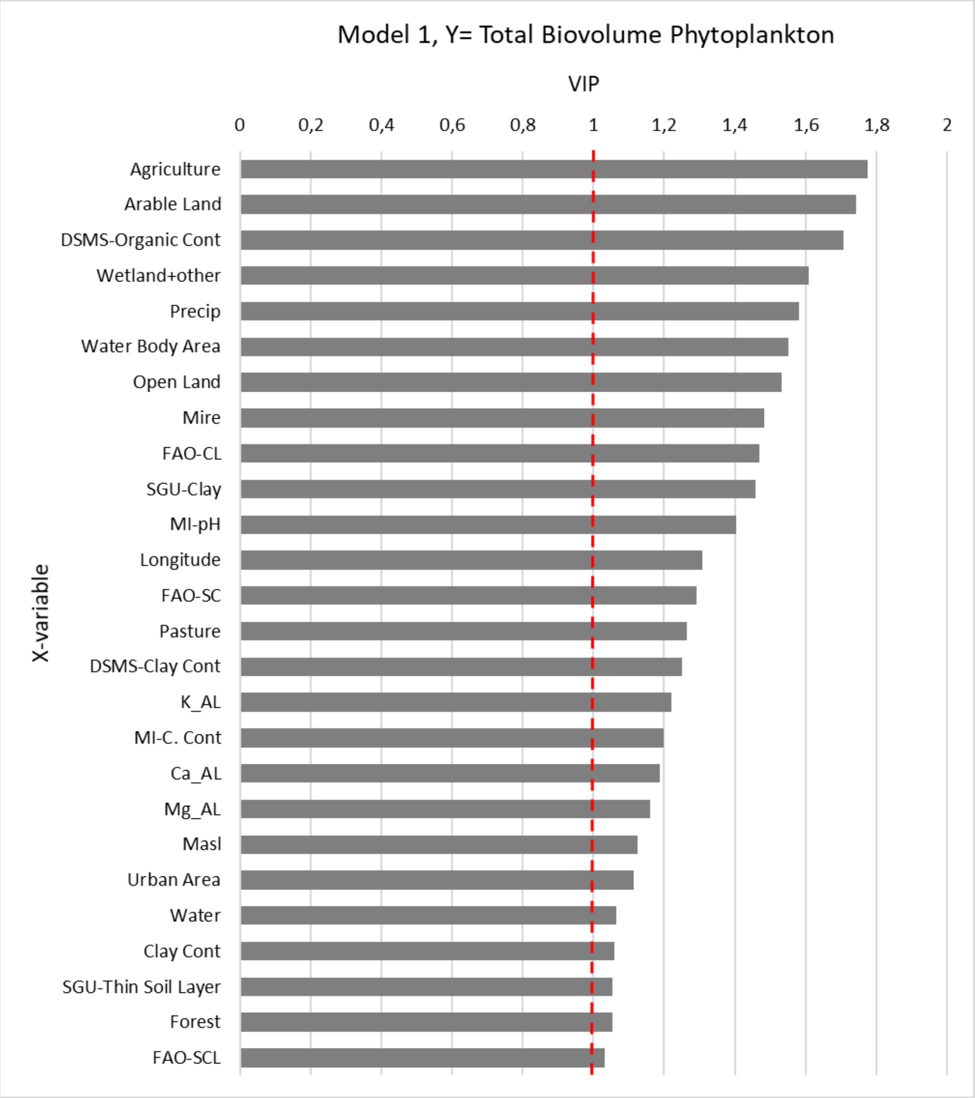

Figure 6. Model 1, catchment variables with VIP >1 from PLS for total biovolume phytoplankton with 250 catchments included and two factors explained. Red dashed line indicates VIP =1.

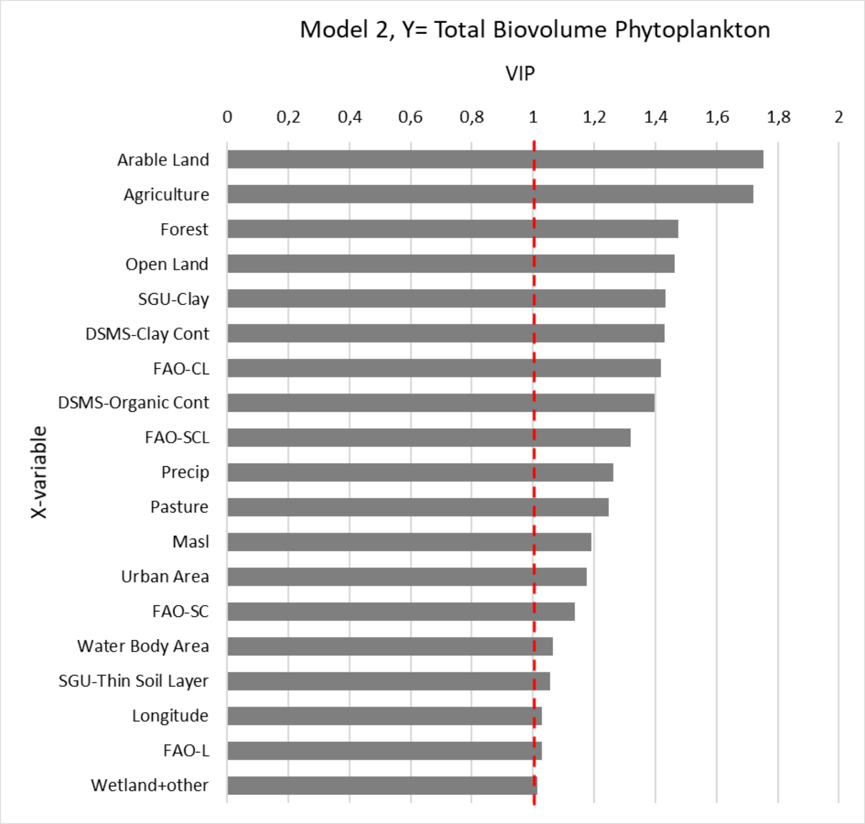

Figure 7. Model 2, catchment variables with VIP >1 from PLS for total biovolume phytoplankton with 523 catchments included and six factors explained. Red dashed line indicates VIP =1.

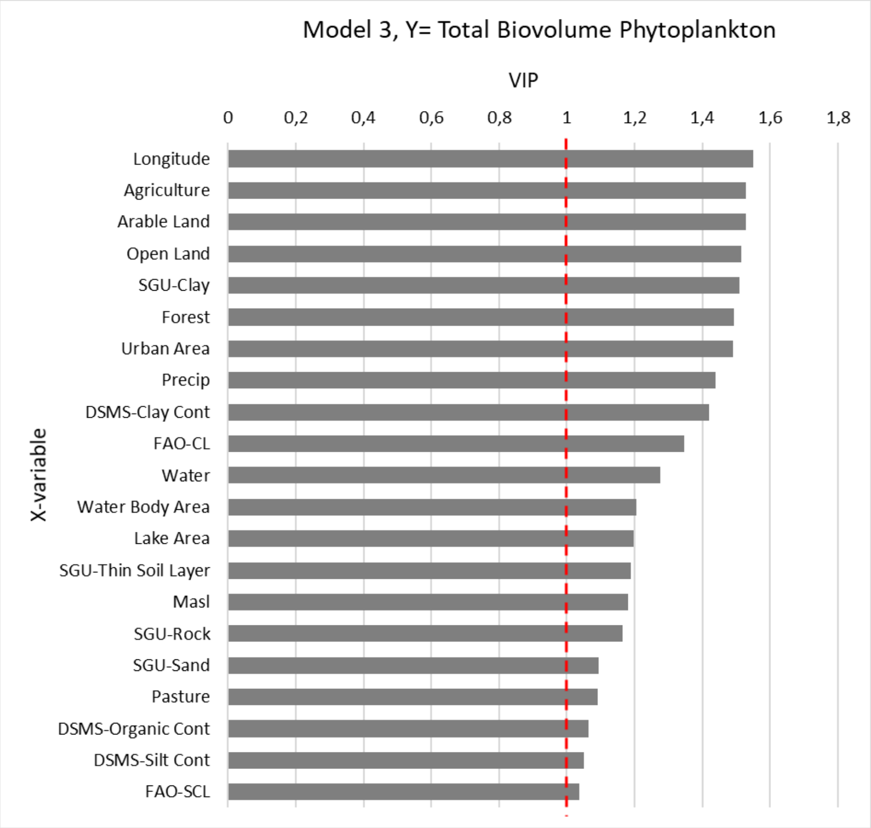

Figure 8. Model 3, catchment variables with VIP >1 from PLS for total biovolume phytoplankton, 310 catchments included with less than 10 % agricultural land and three factors explained. Red dashed line indicates VIP =1.

Results of multiple regressions for tot bio produced from variables important in PLS analysis (VIP >1) for each model are shown in *Table 5*. Depending on the model, catchment variables explain the Y-variable with 29-33 %, without adding TP. Common variables for model 1 and 2 are agricultural land, urban areas and water body area. The share of agricultural land has according to the effect test the highest explanation for these two multiple regressions (*Table 10*, Appendix 1). For model 3 without TP in the regression the share of forest land is the strongest explanatory variable important for the lake tot bio and the share of agricultural land are not indicated as important in this regression (*Table 10*, Appendix 1). A common explanatory variable for tot bio in model 1 and model 3 (excl. TP) is longitude.

When TP was added as independent variable, it was the most important variable regardless of model and explained up to 66 % of tot bio in the dataset (model 2, *Table 5*). Note that for model 1, soil forest survey data (MI-pH) was in combination with TP significantly explain the variation of the lake tot bio. The stepwise regression output with step history can be seen in *Table 11* (Appendix 1) for selection by JMP (minimum BIC) and by criteria. Multiple regression selection based on selected criteria described in Section 3.3.3 showed that the number of variables could be reduced with a small decrease in step history R^2, for all multiple regressions see *Table 10* in Appendix 1.

Table 5. *Multiple regression with R^2 and significance level for the total biovolume phytoplankton (tot bio). Input variables were derived from PLS with VIP >1 for each model, without and with total phosphorus (TP) as variable. Observations show the number of catchments used in multiple regressions and the numbers in parentheses show observations used in stepwise regressions. Multiple regressions selected by JMP (minimum BIC) are shown in bold text and by criteria as Italic text.*

Model	Excl/Incl. TP	R^2	Observ.	Multiple Regression
M1	Excl. TP	**0.32*** **	**523 (250)**	**logTot Bio=-1.142+0.0548*Longitude-0.117*logWater Body Area+0.269*log(Urban Area+1)+0.566*log(Agriculture+1)**
	Incl. TP	**0.65*** **	**185 (161)**	**logTot Bio=-2.133+0.215*MI-pH+1.136*logTP**
M2	Excl. TP	*0.33*** *	*523 (523)*	*logTot Bio=-0.353-0.115*logWater Body Area+0.260*log(Urban Area+1)+0.500*log(Agriculture+1)+0.250*log(SGU-Clay+1)*
	Incl. TP	**0.66*** **	**303 (303)**	**logTot Bio=-1.277+1.137*logTP**
M3[1]	Excl. TP	*0.29*** *	*310 (310)*	*logTot Bio=0.605+0.0766*Longitude+0.00941*DSMS-Clay Cont-0.0179*Forest-0.631*log(Water+1)*
	Incl. TP	*0.64*** *	*199 (199)*	*logTot Bio=-1.294+1.142*logTP*

1. M3 represent catchment with <10 % agriculture. * (p < 0.05) ** (p < 0.01) *** (p < 0.001)

4.5 Plankton Trophic Index (PTI)

Results from PLS analysis for PTI show that model 2, based on map variables only, explained the highest amount of variation for PTI with 52.5 % (Y) and for the catchment variables 26.1 % (X), see *Table 6*. The other two models were weaker. Note that the number of factors included in the model as well as the number of included catchments differ between the models.

Table 6. *PLS analysis for Plankton Tropic Index (PTI) as Y and catchment variables as X. Included total number of X variables, selected catchments and numbers of factors explaining the variation for X and Y in the model.*

Model	No. of X Variables Incl.	Catchments	No. of Factors	Variation Explained for Cumulative X (%)	Variation Explained for Cumulative Y (%)	No. of VIP > 1
M1	64	250	1	22.2	30.8	27
M2	47	523	2	26.1	52.5	18
M3*	47	310	1	17.0	44.6	19

* M3 represent catchment with <10 % agriculture.

Generally, based on the PLS analysis, catchment variables important for explaining PTI in each model (VIP > 1) are mainly the same regardless of the model. Variables related to the share of agricultural land had the highest VIP-value (*Figure 9*, *10* and *11*). Result from model 1 indicates that map data had the higher VIP values compared to the soil survey data, which were found in the middle/lower half of the range of important variables (*Figure 9*). The Degree of P Saturation (DPS) from the arable soil survey reached the highest VIP value among the variables derived from the soil survey data. Important variables in model 2 (*Figure 10*) and model 3 (*Figure 11*) are relatively similar and mostly differ in the order of appearance. The main difference between the models used to describe PTI is that size of the area related to the lake (water body area and entire lake area) appears as an important variable in model 3 but not in model 2 or model 1.

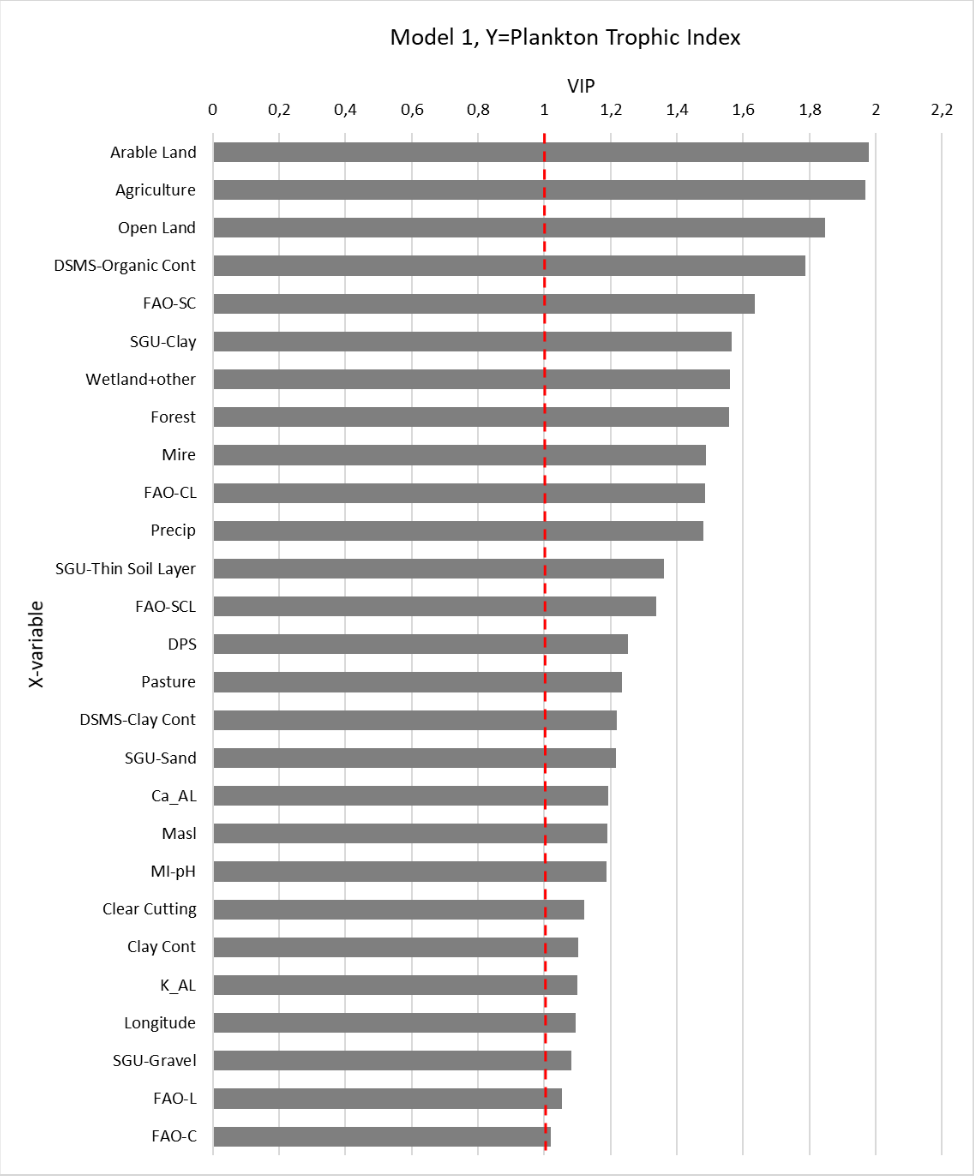

Figure 9. Model 1, catchment variables with VIP > 1 from PLS for Plankton Trophic Index (PTI) with 250 catchments included and one factor explained. Red dashed line indicated VIP =1.

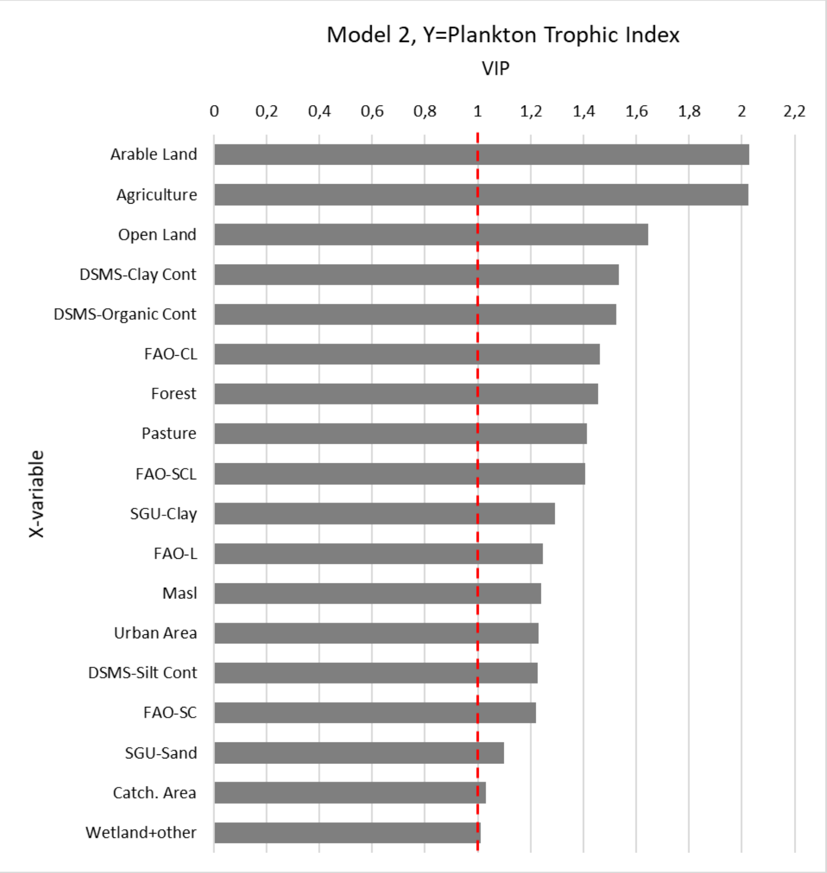

Figure 10. Model 2, catchment variables with VIP >1 from PLS for Plankton Trophic Index (PTI) with 523 catchments included and two factors explained. Red dashed line indicated VIP =1.

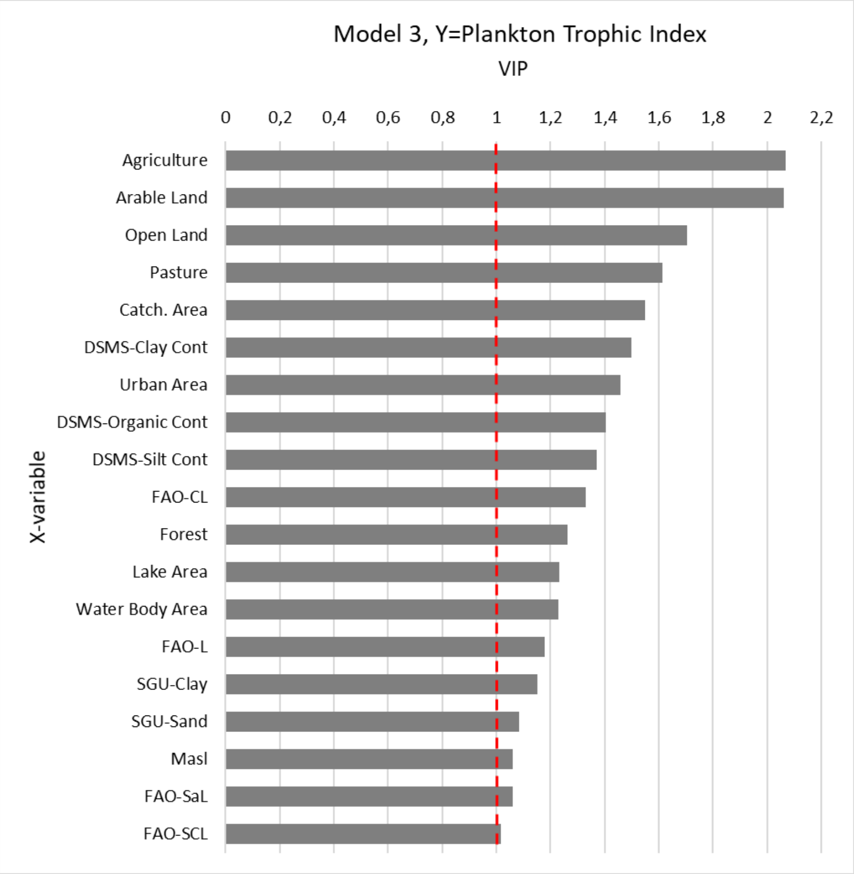

Figure 11. Model 3, catchment variables with VIP >1 from PLS for Plankton Trophic Index (PTI), 310 catchments included with less than 10 % agricultural land and one factor explained. Red dashed line indicated VIP =1.

Table 7 show results from the multiple regressions analysis for PTI with associating explanatory variables for each model. The results for the PTI without TP as variable in the multiple regressions, show that catchment variables together explained 44-48 % of the lake PTI, see *Table 7* for explanatory variables for each regression. The share of agricultural land was important in all these three regressions (excl. TP), in model 1 and 2 it is the most important variable for explaining PTI according to the effect test (*Table 12*, Appendix 2). Five catchment variables were important for the model 3 regression (excl. TP, *Table 7*) with the share of the clay soil on non-agricultural areas (SGU-Clay) as the most important variable (*Table 12*, Appendix 2). Note that the forested area and total catchment area are important for this regression (model 3) but not for the other two regressions without TP.

PTI can be explained with 65 % when lake TP and share of agricultural land are included in a multiple regression (*Table 7*). For lakes with less than 10 % agricultural land in the catchment, 66 % of PTI can be explained with share of agricultural land, TP, lake area and FAO-CL. In all three multiple regressions where the TP is included, it was the most important variable for explaining PTI (effect test in *Table 12*, Appendix 2). For all six regressions, with and without TP as a variable, the share of agricultural land is important for explaining variation of the lake PTI. All PTI multiple regression selected by JMP (minimum BIC) and by criteria can be found together with the associated test in *Table 12* Appendix 2, and for step history results see *Table 13* .

Table 7. *Multiple regression with R^2 and significance level for the Plankton Tropic Index (PTI). Input variables were derived from PLS with VIP >1 for each model, without and with total phosphorus (TP) as variable. Observations show the number of catchments used in multiple regressions and the numbers in parentheses show observations used in stepwise regressions. Multiple regressions selected by JMP (minimum BIC) are shown in bold text and by criteria as Italic text.*

Model	Excl/Incl TP	R^2	Observ.	Multiple Regression
M1	Excl. TP	0.44***	523 (250)	PTI=-0.319+0.425*log(DSMS-Organic Cont+1)+ 0.481*log(Agriculture+1)+0.328*log(FAO-SC+1)
	Incl. TP	0.65***	303 (161)	PTI=-0.898+0.326*log(Agriculture+1)+0.613*logTP
M2	Excl. TP	0.48***	523 (523)	PTI=-0.399+0.235*log(Urban Area+1)+ 0.526*log(Agriculture+1)+0.183*log(SGU-Clay+1)
	Incl. TP	0.65***	303 (303)	PTI=-0.898+0.326*log(Agriculture+1)+0.613*logTP
M3[1]	Excl. TP	**0.47*** **	**310 (310)**	**PTI=-0.0144-0.00632*Forest+ 0.0802*logCatch.Area+0.232*log(Urban Area+1)+ 0.344*log(Agriculture+1)+0.258*log(SGU-Clay+1)**
	Incl. TP	0.66***	199 (199)	PTI=-0.850+0.100*logLake Area+ 0.190*log(Agriculture+1)+0.561*log(FAO-CL+1)+ 0.594*logTP

[1] M3 represent catchment with <10 % agriculture. * ($p < 0.05$) ** ($p < 0.01$) *** ($p < 0.001$)

4.6 Total Phosphorus

PLS analysis for the TP as a Y-variable show that model 2 had the highest explanation degree for variation in lake TP, with 68.1 % (Y) and 67.5 % (X) for the catchment variables, whereas the other two models explained less (*Table 8*). The number of included catchments differ between the models as well the number of factors explaining the variation.

Table 8. *PLS analysis for total phosphorus (TP) as Y and catchment variables as X. Included total number of X variables, selected catchments and numbers of factors explaining the variation for X and Y in the model.*

Model	No. of X Variables Incl.	Catchments	No. of Factors	Variation Explained for Cumulative X (%)	Variation Explained for Cumulative Y (%)	No. of VIP > 1
M1	64	161	3	44.5	55.4	26
M2	47	303	10	67.5	68.1	22
M3*	47	199	3	34.5	66.5	22

* M3 represent catchment with <10 % agriculture.

The important map variables (VIP >1) for TP common in all three models relates to land use (water, agricultural and open land), soil properties on arable land (FAO-texture, organic- and clay content), non-agricultural soil distribution (SGU- soil texture of clay and sand), lake location (longitude and altitude (masl)) and also the size of both lake (water body area and lake area) and catchment area (*Figure 12*, *13* and *14*). In model 1, map variables had the highest VIP value and only four variables from the arable soil survey were important (*Figure 12*). None of the MI-forest chemistry variables had VIP over 1. The PLS output for models 2 and 3 illustrates that important variables are about the same but differ in order of appearance, see *Figure 13* and *Figure 14*.

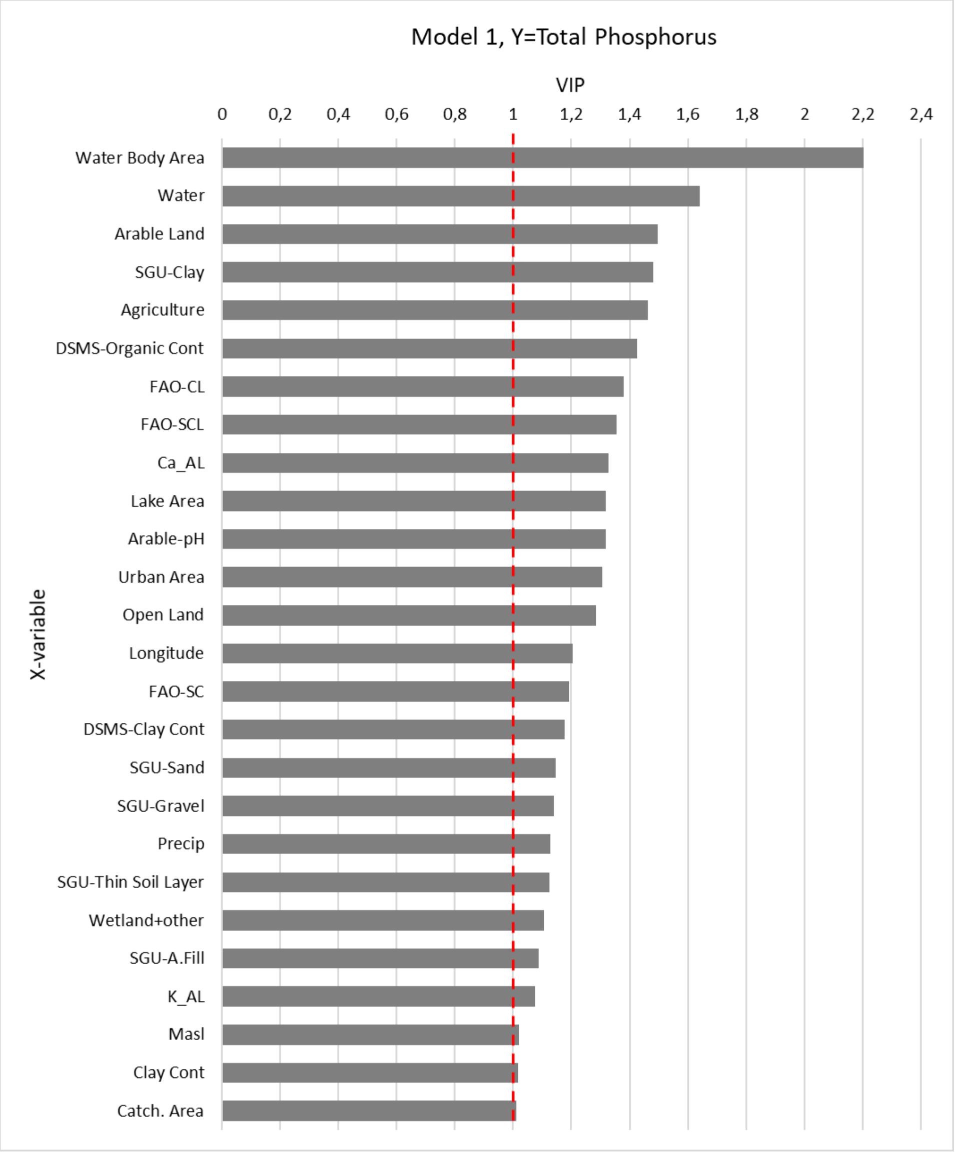

Figure 12. Model 1, catchment variables with VIP >1 from PLS for total phosphorus (TP) with 161 catchments included and three factors explained. Red dashed line indicated VIP =1.

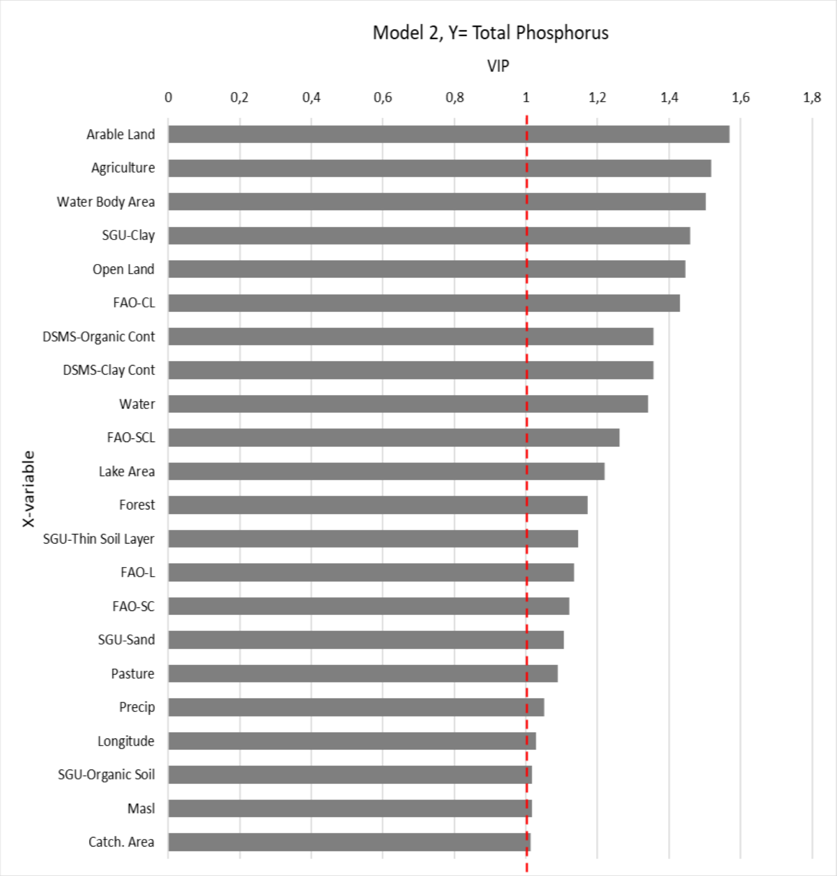

Figure 13. Model 2, catchment variables with VIP >1 from PLS for total phosphorus (TP) with 303 catchments included and 10 factors explained. Red dashed line indicated VIP =1.

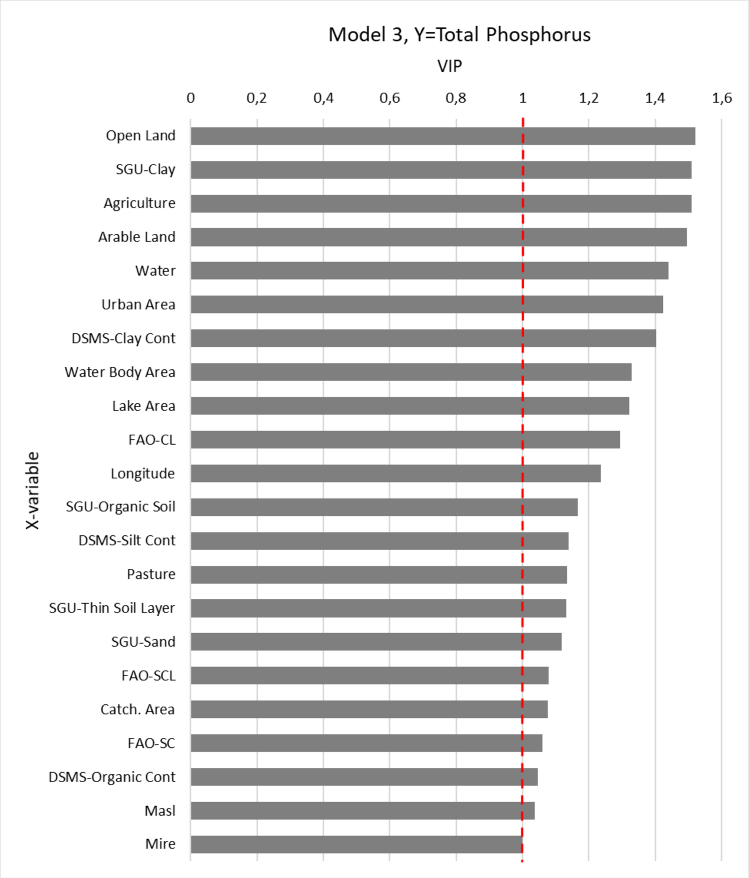

Figure 14. Model 3, catchment variables with VIP >1 from PLS for total phosphorus (TP), 199 catchments included with less than 10 % agricultural land and three factors explained. Red dashed line indicated VIP =1.

The multiple regression based on the catchment characteristics could explain 51-55 % of the variations in TP concentrations in studied lakes, depending on the model and associated variables with VIP over 1, see *Table 9*. Generally, for all three multiple regressions, different map data such as share of clay and organic soils on non-agricultural land, as well as the area of water body, and share of agricultural and urban areas (not model 2) were indicated as important variables and correlated with lake TP. However, the strongest explanatory variable for TP in model 1 was clay content on arable land (DSMS-Clay Cont) and for model 3 the share of non-agricultural clay (SGU-Clay), while in model 2 the share of agricultural land was most important, see effect test *Table 14* in Appendix 3. In cases where multiple regression was selected by criteria, it was found that the number of variables included could be reduced by marginal reduction of the explanation degree, see *Table 14* in Appendix 3 for all regression produced and *Table 15* for step history.

Table 9. *Multiple regression with R^2 and significance level for the total phosphorus (TP). Input variables were derived from PLS with VIP >1 for each model. Observations show the number of catchments used in multiple regressions and the numbers in parentheses show observations used in stepwise regressions. Multiple regressions selected by JMP (minimum BIC) are shown in bold text and by criteria as Italic text.*

Model	R^2	Observ.	Multiple Regression
M1	0.51***	303 (185)	logTP=0.800+0.0141*DSMS-Clay Cont-0.168*logWater Body Area+0.732*log(DSMS-Organic Cont+1)+ 0.251*log(Urban Area+1)+0.212*log(Agriculture+1)
M2	0.52***	303 (303)	logTP=0.535-0.137*logWater Body Area+ 0.472*log(Agriculture+1)+0.308*log(SGU-Organic Soil+1)+ 0.358*log(SGU-Clay+1)
M3[1]	0.55***	199 (199)	logTP=0.518-0.141*logWater Body Area+0.478*log(Urban Area+1)+ 0.380*log(Agriculture+1)+0.306*log(SGU-Organic Soil+1)+ 0.365*log(SGU-Clay+1)

[1] M3 represent catchment with <10 % agriculture. * ($p < 0.05$) ** ($p < 0.01$) *** ($p < 0.001$)

5 Discussion

This study has investigated the correlation between catchment properties with phytoplankton and TP in 487 lakes (corresponding to 523 waterbodies), in the southern part of Sweden (*Figure 3*). The catchments were characterised by in total 64 different variables covering land use, soil distribution and texture, soil chemistry, climate, lake location and size of lake and catchment. Since a large number of lakes and wide range of associating catchment properties were examined in this study, the result gives a broad indicator of interactions between properties in the catchment and their connection to lake phytoplankton and TP.

5.1 Lake Phytoplankton and Total Phosphorus Correlate with Similar Catchment Describing Variables

5.1.1 Important Catchment Variables from PLS Analysis

In the first step in PLS analysis, variables were selected and identified as important (VIP >1) for the three dependent variables. This study can from the PLS analysis generally conclude that proportion of agricultural land is an important catchment variable for explaining phytoplankton tot bio (*Figure 6*, *7* and *8*), PTI (*Figure 9*, *10* and *11*) and TP (*Figure 12*, *13* and *14*) regardless of which model was applied. Even when only lakes with less than 10 % agricultural land in their catchment were considered (model 3), the proportion of agricultural land was still shown as an important variable in all PLS analysis. Generally, the catchment variables that were indicated as important (VIP >1) are similar for all models and the three tested dependent variables (Y-variable) and differs mostly in order of appearance. The similarities between the models were seen even when the models differed to some extent regarding the number of included catchments. This confirms the robustness of selected catchment variables identified as important for the description of the phytoplankton and TP in this dataset.

Besides the share of agricultural land, other important explanatory variables were soil clay content, FAO-fine soil texture classes, lake location (longitude and altitude) and in several cases also urban area and climate (precipitation). It is not surprising that both agricultural land and urban areas are important variables in the PLS models, since both of these land use categories contribute to a large proportion of nutrient losses to surrounding water, not least phosphorus (Ejhed *et al.* 2016; Statistics Sweden 2018). For small agricultural catchment (2-35 km^2) studied by Kyllmar *et al.*, (2014), a strong positive correlation was found between the clay content of arable soils and the stream outlet TP concentration. This relationship could be a reason why the importance of fine soil texture was found to be important for the PLS models in the current study. The effect of the lake's position along the longitudinal axis could be explained by the difference in water-chemical properties in the west-east gradient, due to higher acidity deposition on the west coast (Fölster 2018) as well as a gradient in precipitation volumes. Importance of altitude (masl) can probably be related to water chemistry differences, where lakes at lower altitude have generally higher P concentration (Nõges 2009).

5.1.2 Evaluation of Catchment Variables used in Multiple Regressions

Selected important variables from the PLS with VIP >1 were used to build the multiple regressions and were shown to be statistically significant. There is generally a small difference in which variables were found as explanatory in the multiple regression for the three examined dependent variables (tot bio, PTI and TP). The differences depended partly on the identified variables of importance during the PLS analysis (VIP >1) but also on the differences in the number of catchments included in the PLS models. Without TP as an independent variable for tot bio (*Table 5*) and PTI (*Table 7*), a positive correlation was shown to proportion of agricultural land in the catchment for all three models with one exception for tot bio in model 3. The share of agricultural land is also the strongest explanatory variable for model 1 and 2 multiple regressions (tot bio: *Table 10*, Appendix 1 and PTI: *Table 12*, Appendix 2). However, importance of agricultural land does not apply to tot bio in model 3, representing catchments with less than 10 % agricultural land. In this case, the share of forest is the strongest explanatory variable for phytoplankton tot bio with negative correlation in the multiple regression (*Table 10*, Appendix 1), although even here the share of agricultural land was important for the PLS analysis (*Figure 8*). Lakes dominated by forest land use are then predicted to have lower volume of phytoplankton. Losses of nutrient to surrounding water are thus also considerably lower for forest than agricultural areas (Ejhed *et al.* 2016). Phytoplankton growth is thus also affected by other factors in the lake, for instance content of humic substances (Maileht *et al.* 2013).

For estimation of the TP concentration in the lake, the multiple regressions show that the share of agricultural land is also an important explanatory variable and correlates positively to TP (*Table 9*). For large Swedish catchment areas a strong positive correlation between the share of arable land and the TP concentration in the watercourse outlet has been found (Bol *et al.* 2018). The reason that phytoplankton (tot bio and PTI) correlate well with the share of agricultural land in the catchment is then probably an indirect effect of higher TP concentration in the lake as the share of agricultural land increases.

Correlation of catchment properties with total biovolume of phytoplankton

Multiple regression for lake tot bio showed that four catchment variables together significantly can predict 29-33 % of the variation in tot bio, depending on model, and with small difference in the explanatory variables (*Table 5*). Both models 1 and 2 showed a negative correlation between tot bio and water body area, and a positive correlation to diffuse- and point sources (agriculture and urban area). The reason for lower volume of phytoplankton in lakes with a larger water body area could be due to the properties in the lake affecting both the growth of phytoplankton and the concentration of P. For European lakes, it has been seen that lakes with longer residence time have lower concentration of nutrients (TP and nitrogen compounds) as well as lower chl-*a* concentrations (Nõges 2009). Although residence time is not investigated in this study, the larger lakes probably have longer residence time allowing sedimentation processes to take place in the lake. The fact that the volume of phytoplankton is lower in lakes with larger water body area might be consequently an effect of lower P concentrations. A negative correlation between the lake TP and water body area was also found in this study (*Table 9*). Another explanation for the negative correlation of water body area to both lake TP and tot bio could be a dilution effect. Considered that the lakes with a larger surface area have also larger lake volume, which was shown for European lakes (Nõges 2009).

Other explanatory variables for tot bio in the multiple regressions in models 2 and model 3, were the share of finer soils on non-agricultural land (SGU-Clay) and the clay content of arable (DSMS-Clay Cont), respectively (*Table 5*). Both these variables showed a positive correlation with the tot bio, i.e. higher tot bio is expected at higher clay contents in the catchment area. Longitude was also positive correlated to tot bio (model 1 and 3) in the multiple regression, probably due to the lake chemical gradient in west-east direction in Sweden. For instance, an acidity gradient where lakes in the eastern part of Sweden are less acidified (Fölster 2018) with higher pH values due to lower deposition of acidifying agents. Additionally, the phytoplankton composition is also controlled by other water chemistry parameters such as alkalinity and water colour (Maileht *et al.* 2013).

However, the multiple regression showed that catchment properties together with TP could not further explain the lake tot bio in this study. The strongest significantly correlation (model 2) found for tot bio was the correlation to lake TP where 66 % (N= 303) of the variation was explained (*Figure 5A*, *Table 5*). In the case with fewer catchments (N= 161, model 1) the correlation between tot bio and TP was weaker (R^2= 0.61, *Table 11* in Appendix 1). In this case the correlation could significantly be improved by inclusion of MI-pH (forest soil survey pH) to 65 % (N= 185, *Table 5*). The fact that lake pH is correlated with phytoplankton composition and biomass is well recognised (Brönmark & Hansson 2005). The positive correlation to MI-pH thus shows that phytoplankton biomass increases with higher pH values in surrounding forest soils.

For all examined phytoplankton indicators in this study, the strongest explanation was found for chl-*a* where 78 % (N= 303) of the variation was significantly explained by the lake TP (*Figure 5B*). That chl-*a* is showing a strong response to TP is in line with other studies (Phillips *et al.* 2008; Carvalho *et al.* 2012; Lyche-Solheim *et al.* 2013; Lindegarth *et al.* 2016). The strong correlations of chl-*a* with TP is the reason that no further investigation was made for chl-*a.* Also, the strong correlation, 75 % (N= 321, *Figure 5*D), between chl-*a* and tot bio means that the explanatory catchment variables found for the tot bio to a large extent also can explain chl-*a*.

Catchment variables improved prediction of plankton trophic index

According to the results of the multiple regressions, PTI was the only phytoplankton indicator where catchment properties (the share of agricultural land) significantly increase the explanation degree of the index variation together with TP to 65 % (N= 303, *Table 7*), where the concentrations of TP alone explain 56 % (N= 303, *Figure 5C*) of PTI. For lakes with less than 10 % agricultural land in the catchment, model 3, TP together with proportion of three catchment variables (FAO-CL, lake area and agriculture) could significantly explain the variation of PTI with 66 % (N= 199, *Table 7*). All these variables were positively correlated to PTI. However, TP is still the best predictor of PTI, and this strong positive relationship to TP was also found in other studies (Phillips *et al.* 2012; Lyche-Solheim *et al.* 2013; Lindegarth *et al.* 2016).

This study shows that three catchment variables (share of agricultural land, share of urban area and SGU-Clay) were able to significantly explain 48 % (N= 523, *Table 7*) of the variation in the lake PTI, having all positive correlation to PTI. Almost as strong correlations were also found for the other two models, model 1 (44 %, N= 523) and model 3 (47 %, N= 310), although with some differences in the explanatory catchment variables and the number of included catchments and lakes. Regardless of the model, with and without TP as an independent variable, the share of the agricultural land was indicated as a statistically important variable for PTI with positive correlation in the multiple regressions. This means that higher PTI values, i.e. species more tolerant to high nutrient levels, can be expected in lakes with a larger share of agricultural land in the catchment. Agriculture is also considered one of the major sources of nutrient losses to surrounding water (Ejhed *et al.* 2016).

Soils of importance for PTI were all related to fine texture soils, such as non-agriculture clay (SGU-Clay), as well as fine textured arable soils, silty clay (FAO-SC) and clay loam (FAO-CL) (*Table 7*). For the model 3 regression, lake surface area was one of the four variables (including TP) that significantly influenced PTI, with a positive correlation. The importance of lake surface area has also been found for PTI in Phillips *et al.,* (2012), where higher PTI were expected with increasing lake area. Without TP in model 3, the catchment size was found to be an important variable. In the whole dataset with 523 waterbodies, the PTI (-0.9 to 1.9) covered both the oligotroph and eutrophic scale. The fact that the PTI values ranged across the entire scale in this dataset means that the connections found in this study thus apply to lakes with both high and low PTI values.

How much can catchment variables together predict lake total phosphorus?

Certain catchment properties were significantly correlated with lake TP explaining 51-55 % of the variation, depending on the regression model used (*Table 9*). Explanatory variables differ slightly between the models and in total four to five variables were of importance for the regressions. It can also be concluded from the regressions that the proportion of the agricultural land and urban area (not for model 2) are positively correlated to the lake TP whereas TP was negatively correlated to the water body area. High level of P can then be expected in smaller lakes with a high proportion of agricultural and urban land use in the catchment. Both these land use categories are known as important sources of nutrient losses to surrounding water (Ejhed *et al.* 2016; Statistics Sweden 2018). Although lakes with less than 10 % agricultural land in associated catchment were studied separately, the share of agricultural land was still an important variable concerning the concentration of TP in the lake water. Negative correlation between water body area and TP has also been identified for tot bio and water body area in this dataset (mentioned above) and probably the same explanation could also be applied for TP. Namely, that TP is to some extent influenced by the water body area and in turn tot bio is correlated to TP levels. However, it is important to remember that phytoplankton growth is dependent on dissolved nutrients in the water column (Brönmark & Hansson 2005) whereas TP includes both dissolved P and P bound into organisms and particles.

Other important catchment variables for the description of lake TP identified by the regressions were related to soil properties in both agriculture and non-agricultural areas (*Table 9*). For instance, a positive correlation to TP was found to the share of organic and clay soil (non-agricultural areas) as well the soils' clay and organic content (arable areas) in the catchment. Since the soils with a higher clay content have a greater possibility to form macropores and through that contribute to P transport and leaching (Eriksson *et al.* 2014; Johnsson *et al.* 2016), this could then be one explanation why an increased soil clay content in the catchment contribute to increased concentration of TP in the lake. That clay content in the catchments arable soil is positively correlated to concentration of TP in the water is in line with Kyllmar *et al.*, (2014). Consequently, over 50 % of the variation in TP levels in the lake could significantly be explained with the help of properties in the catchment area (*Table 9*), confirming that the terrestrial properties in the surrounding area have a large impact on the lake's water quality and are important to consider for management to prevent nutrient losses and avoid eutrophication.

The regulation of P levels in lakes is complex and depends on both external process in the catchment, explored in this study, but also on internal processes within the lake such as retention of nutrients and internal loading of P from the sediment (Brönmark & Hansson 2005; Søndergaard *et al.* 2013). Catchment properties not investigated directly in this study which may affect transport and nutrient losses to surrounding water are for instance soil vulnerability for soil erosion (Djodjic & Markensten 2018), crop distribution and field slope (Johnsson *et al.* 2016). Additionally, the data regarding nutrient losses from individual sewers as well as wastewater treatment plants (Statistics Sweden 2018; Swedish EPA 2018) could be taken into consideration. An inclusion of these variables in regression models might lead to further increase in our understanding of the relationship between catchment characteristics and lake properties.

5.2 Evaluation of Input Data and Statistical Models

The importance of both soil chemistry data from soil surveys and available continuous map data were evaluated in this study, as shown in PLS model 1. Soil chemistry samples covering arable as well as forest soils (MI-data) were less important for the three studied Y-variables since some of these variables ended up in the lower half of important variables with VIP >1 (*Figure 6*, *9* and *12*). This is also confirmed in associated multiple regressions (*Table 7* and *Table 9*) where only in one case variables related to the soil surveys turns out to be statistically significant, i.e. MI-pH variable for the phytoplankton total volume (*Table 5*). The influence of climate (air temperature and precipitation) has not been detected as important for the multiple regressions although precipitation was important in most PLS analysis. Too small variation of used average values (period 1961-1990) within the climate variables in the dataset can be one reason. Optimally, the values for temperature and precipitation should cover the same period as the lake data (2000-2012) used. For instance, the growth of phytoplankton is influenced by the temperature during the growing season (Brönmark & Hansson 2005) and for Swedish lakes, the duration of the ice-free season also influence the phytoplankton biomass (Weyhenmeyer *et al.* 2013). In European lakes, climate variables were identified as important factors for PTI (Phillips *et al.* 2012).

The three produced and examined models were intended to cover the best available data. Model 1 includes all 64 catchment variables (both soil chemical data from soil surveys and map data) in the dataset but do not includes all lakes. The MI-forest soil survey data is distributed in a grid pattern over Sweden (Nilsson *et al.* 2015) and so even the arable soil survey samples but with increased samples density in more intensive agriculture regions (Paulsson *et al.* 2015). Consequently, smaller catchment areas are probably underrepresented in model 1 since there is lower probability that soil samples were collected compared to the larger catchment areas. On the other hand, model 2 includes all map data and the entire dataset of lakes. Importance of the agricultural land was examined in model 3, by selection criteria where only lakes with less than 10 % agricultural land in the catchment were included.

The PLS analysis show that model 2 explains the highest variation of the X-variables (catchment variables) and each Y-variable, for tot bio 43.7 % (N= 523, *Table 4*) of the variation were explained and 52.5 % for PTI (N= 523, *Table 6*) respective 68.1 % for TP (N= 303, *Table 8*). This indicated that a combination of these map variables explains the most variation and are especially important to consider in relation to phytoplankton and TP.

5.3 Limitations and Uncertainties

Several map layers have been used to describe the land use, soil texture and soil distribution as well as climate properties in the lake catchment area. The accuracy is thus controlled by the source's own spatial resolution where 50 x 50 m used for DSMS layers, 25 x 25 m SGU layer and climate 4000 x 4000 m. Reduced accuracy when converting land use (PLC 6) feature layer to 25 x 25 m raster, to be able to incorporate the GIS tool, can be considered negligible for the result since the other map sources used have the same or even lower resolution. Of course, some uncertainty can arise during the production of each individual layer. For instance Söderström & Piikki (2016) considered largest uncertainty in the organic DSMS layer.

Larger lakes with several waterbodies included in the dataset are represented by the lake's entire catchment area and not the sub-catchment. The specific data for the lake chemistry and biology are those represented for the water body area. For TP and the volume of phytoplankton is shown to correlate to the water body area and not to the entire lake surface area which could indicate that local properties within the waterbody influence the chemistry and biology. Additionally, catchment properties close to the waterbody might have higher influence on the waterbody compared to properties further away in the catchment and thus be more important for the management of large lakes. The specific catchment area for each waterbody within the large lakes can thus be of concern. This is especially important when assessing P losses on catchment scale, where only a small proportion of the catchment area contributes to the main losses, the 80:20 rule (Sharpley *et al.* 2009).

5.4 Future Implications

The main focus of this study was to evaluate the connections between properties in the catchment area and the lake phytoplankton and TP. To further enhance our understanding of the lake water quality, the correlation to other lake variables can be important to investigate further. The PCA analysis (*Figure 4*) show the correlation of all 93 variables in the dataset, in the lake as well on the land. It shows that phytoplankton indicators (tot bio, chl-*a* and PTI) are positively correlated to each other and TP. The PCA result can provide indications of water chemistry variables to investigate further but also how other catchment and lake variables correlate to each other. For phytoplankton, negative correlation is seen to lake depth (mean and max) and secchi depth while positive correlation to turbidity, alkalinity, conductivity, metals (mainly Cl, Na, K) and other nutrients compounds in the lake. Vulnerability and influence of alkalinity has also been concluded for PTI and chl-*a* (Phillips *et al.* 2008, 2012). Therefore, a possible continuation of this study could be to investigate the relationship between other water chemistry and catchment characteristics and if inclusion of such variables can increase the relationship found for the phytoplankton and TP in this study.

Since the effect of the agricultural land was shown to be important even when only lakes with less than 10 % of agricultural land were included, it could therefore be of interest to further develop and identify the threshold ("tipping point") where the importance of agricultural land no longer has a strong impact on phytoplankton and TP. It would also be interesting to investigate whether the connections found in this study also would applies to northern Sweden lakes which, unlike southern Sweden, are dominated by forest and are less impacted by agricultural and urban areas (Ejhed *et al.* 2016). In addition, the northern Sweden lake phytoplankton are to a greater extent N-limited compared to southern Sweden (Bergström *et al.* 2008).

In this large-scale study, no direct connection was found to the data from national soil chemistry surveys, although to the certain extent they were important in the PLS models. However, it is worth mentioning that the arable DSMS map layers are produced and to a high degree based in reference to soil samples from the field (Söderström & Piikki 2016). Further development could thus be to investigate if the arable soil chemistry properties are of greater importance for a subset of lakes within intensive agricultural areas and with a high proportion of agricultural land. It would be reasonable to assume that the properties of arable soils might have a more prominent role in catchments with high share of agricultural land.

However, findings in this study could also be used to develop the existing assessment methods for phytoplankton and P in lakes. For running water, assessment criteria regarding TP reference values is taken into account when there is more than 10 % of agricultural land in the catchment area (HVMFS 2013:19). Since the share of agricultural land also proves to be important for the studied lakes, consideration of the share of agricultural land might also be included in the assessment criteria for lake TP reference value. However, an investigation is needed to detect the agricultural limit for lakes since it is considered important for catchment with less than 10 % agricultural land, but in lakes also internal processes regulate P concentration which complicates the reference values of lakes which Fölster *et al.*, (2018) also has addressed.

This study's findings have hopefully also increased the understanding of the connection between lake biology and chemistry to the surrounding land and could be used for lake management and to prevent P losses. The performed multiple regressions could also be tested through dataset on lakes with measured values to validate predicted and measured values. Since the water sampling and analyses are expensive, the important map variables for phytoplankton and TP identified here could be used to fast screening lakes that may need measures to reach good ecological status and therefore need to be assessed according to the WFD. For the estimation of phytoplankton, the regression relationship found with catchment properties could be used as a broad indicator for lakes without measured TP, especially for PTI. This study still recommends, however, to measure TP in the lake for the best estimation and highest variation explained for tot bio. However, the measured TP should in combination with the share of agricultural land in the catchment be used for PTI estimation.

6 Conclusions

- The three investigated dependent lake variables (tot bio, PTI and TP) were all positive correlated to the proportion of agricultural land in the catchment area.
- Agricultural land was an important explanatory variable for PTI and TP, even when evaluating lakes with less than 10 % agricultural land.
- Lake water chemistry, TP, was shown to be the strongest explanatory variable for the lake tot bio and PTI as has been found before.
- PTI was the only phytoplankton indicator where catchment properties, share of agricultural land, together with TP support and increased the degree of estimation of PTI to 65 % from 56 % with TP alone. In lakes with less than 10 % agricultural land in the catchment area, 66 % variation of PTI were explained by positive correlation to TP and share of the three catchment variables: lake area, agriculture and FAO-CL.
- Without TP as an explanatory variable for PTI, catchment variables can together explain up to 48 % in the variation in the studied lakes, depending on model and included explanatory variables.
- Strongest correlation for tot bio was to lake TP alone with 66 % variation explained. Without TP, four catchment properties explain tot bio 29-33 %, depending on model and slightly differ in explanatory variables.
- Chl-*a* was the phytoplankton indicator that had strongest correlation with lake TP with 78 % variation explained.
- Catchment properties in the studied lakes could explain over 50 % of the lake TP variation. Explanatory variables depend on model, but overall the content and share of the clay- and organic matter in soil, share of agricultural and urban areas, as well as the area of the water body were the most important explanatory variables for TP.
- In this large-scale study, properties in the catchment area described by continuous map data were shown to be more important as explanatory variables compared to the data derived from national soil surveys.

References

Bergström, A.-K., Jonsson, A. & Jansson, M. (2008). Phytoplankton responses to nitrogen and phosphorus enrichment in unproductive Swedish lakes along a gradient of atmospheric nitrogen deposition. *Aquatic Biology*, vol. 4 (1), pp. 55–64. DOI: https://doi.org/10.3354/ab00099

Bol, R., Gruau, G., Mellander, P.-E., Dupas, R., Bechmann, M., Skarbøvik, E., Bieroza, M., Djodjic, F., Glendell, M., Jordan, P., Van der Grift, B., Rode, M., Smolders, E., Verbeeck, M., Gu, S., Klumpp, E., Pohle, I., Fresne, M. & Gascuel-Odoux, C. (2018). Challenges of Reducing Phosphorus Based Water Eutrophication in the Agricultural Landscapes of Northwest Europe. *Frontiers in Marine Science*, vol. 5. DOI: https://doi.org/10.3389/fmars.2018.00276

Brönmark, C. & Hansson, L.-A. (2005). *The Biology of Lakes and Ponds*. Second Edition. New York: Oxford University Press Inc.

Carvalho, L., Poikane, S., Solheim, A., Phillips, G., Borics, G., Catalan, J., De Hoyos, C., Drakare, S., Dudley, B., Järvinen, M., Laplace-Treyture, C., Maileht, K., McDonald, C., Mischke, U., Moe, J., Morabito, G., Nõges, P., Nõges, T., Ott, I. & Thackeray, S. (2012). Strength and uncertainty of phytoplankton metrics for assessing eutrophication impacts in lakes. *Hydrobiologia*, vol. 704, pp. 1–14. DOI: https://doi.org/10.1007/s10750-012-1344-1

Correll, D.L. (1998). The Role of Phosphorus in the Eutrophication of Receiving Waters: A Review. *Journal of Environmental Quality*, vol. 27 (2), pp. 261–266. DOI: https://doi.org/10.2134/jeq1998.00472425002700020004x

Cox, I. & Gaudard, M. (2013). Chapter 5- Predicting Biological Activity. *Discovering Partial Least Squares with JMP*. Cary, North Carolina, USA: SAS Institute Inc, pp. 75–104.

Directive 2000/60/EC *of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy*. *Official Journal of the European Communities*

Djodjic, F., Börling, K. & Bergström, L. (2004). Phosphorus Leaching in Relation to Soil Type and Soil Phosphorus Content. *Journal of Environmental Quality*, vol. 33, pp. 678–84. DOI: https://doi.org/10.2134/jeq2004.0678

Djodjic, F. & Markensten, H. (2018). From single fields to river basins: Identification of critical source areas for erosion and phosphorus losses at high resolution. *Ambio*, pp. 1–14. DOI: https://doi.org/10.1007/s13280-018-1134-8

Drakare, S. (2014). *Översyn av typologi för sjöar och vattendrag*. (Rapport 2014:2). Institutionen för Vatten och Miljö, Sveriges Lantbruksuniversitet.

Ejhed, H., Widén-Nilsson, E., Tengdelius Brunell, J. & Hytteborn, J. (2016). *Näringsbelastningen på Östersjön och Västerhavet 2014. Sveriges underlag till Helcoms sjätte Pollution Load Compilation*. (Havs- och vattenmyndighetens rapport 2016:12). Göteborg, Sverige.

Elser, J.J., Andersen, T., Baron, J.S., Bergström, A.-K., Jansson, M., Kyle, M., Nydick, K.R., Steger, L. & Hessen, D.O. (2009). Shifts in Lake N:P Stoichiometry and Nutrient Limitation Driven by Atmospheric Nitrogen Deposition. *Science*, vol. 326 (5954), pp. 835–837. DOI: https://doi.org/10.1126/science.1176199

Eriksson, A.K., Hesterberg, D., Klysubun, W. & Gustafsson, J.P. (2016). Phosphorus dynamics in Swedish agricultural soils as influenced by fertilization and mineralogical properties: Insights gained from batch experiments and XANES spectroscopy. *Science of The Total Environment*, vols 566–567, pp. 1410–1419. DOI: https://doi.org/10.1016/j.scitotenv.2016.05.225

Eriksson, J., Dahlin, S., Nilsson, I. & Simonsson, M. (2014). *Marklära*. Edition 1:3. Lund: Studentlitteratur AB.

ESRI (2017). *Spatial Analyst Supplemental Tools*. Environmental Systems Research Institute. Available at: https://www.arcgis.com/home/item.html?id=3528bd72847c439f88190a137a1d0e67 [2019-02-20]

ESRI Inc (2018). *ArcGIS desktop*. Version: 10.6.1. Redlands, California, USA: Environmental Systems Research Institute.

Fölster, J. (2018). Omdrevssjöarna i vattenförvaltningen. *Sötvatten 2017. Om miljötillståndet i Sveriges sjöar, vattendrag och grundvatten*. Göteborg: Havs- och vattenmyndigheten, pp. 20–23.

Fölster, J., Djodjic, F., Huser, B., Moldan, F. & Sonesten, L. (2018). *Bedömningsgrunder för fysikalisk-kemiska kvalitetsfaktorer i sjöar och vattendrag. Förslag till revidering av föreskrift HVMFS 2013:19*. (Rapport 2018:10). Uppsala: Institutionen för Vatten och Miljö, Sveriges Lantbruksuniversitet.

Holmgren, K. (2018). Djupkartor viktiga för miljöövervakning i sjöar. *Sötvatten 2017. Om miljötillståndet i Sveriges sjöar, vattendrag och grundvatten*. Göteborg: Havs- och vattenmyndigheten, pp. 24–26.

Huser, B.J. & Fölster, J. (2013). Prediction of Reference Phosphorus Concentrations in Swedish Lakes. *Environmental Science & Technology*, vol. 47 (4), pp. 1809–1815. DOI: https://doi.org/10.1021/es3040413

HVMFS 2013:19 *Havs- och vattenmyndighetens föreskrifter om klassificering och miljökvalitetsnormer avseende ytvatten*.

HVMFS 2018:17 *Havs- och vattenmyndighetens föreskrifter om ändring i Havs- och vattenmyndighetens föreskrifter (HVMFS 2013:19) om klassificering och miljökvalitetsnormer avseende ytvatten*.

Johansson, B. (2000). *Precipitation and Temperature in the HBV Model. A Comparison of Interpolation Methods.* (RH No.15). Norrköping: Swedish Meteorological and Hydrological Institute.

Johnsson, H., Mårtensson, K., Lindsjö, A., Persson, K., Andrist Rangel, Y. & Blombäck, K. (2016). *Läckage av näringsämnen från svensk åkermark-Beräkning av normalläckage av kväve och fosfor för 2013.* (SMED-Svenska MiljöEmissionsData Nr 189). Norrköping: Sveriges Meteorologiska och Hydrologiska Institut.

Kyllmar, K., Forsberg, L.S., Andersson, S. & Mårtensson, K. (2014). Small agricultural monitoring catchments in Sweden representing environmental impact. *Agriculture, Ecosystems & Environment*, vol. 198, pp. 25–35. DOI: https://doi.org/10.1016/j.agee.2014.05.016

Lindegarth, M., Carstensen, J., Drakare, S., Johnson, R.K., Nyström Sandman, A., Söderpalm, A. & Wikström, S.A. (Editors) (2016). *Ecological Assessment of Swedish Water Bodies; Development, harmonisation and integration of biological indicators.* (Final report of the research programme WATERS. Deliverable 1.1-4, WATERS report no 2016:10). Gothenburg: Swedish Institute for the Marine Environment.

Lyche-Solheim, A., Feld, C.K., Birk, S., Phillips, G., Carvalho, L., Morabito, G., Mischke, U., Willby, N., Søndergaard, M., Hellsten, S., Kolada, A., Mjelde, M., Böhmer, J., Miler, O., Pusch, M.T., Argillier, C., Jeppesen, E., Lauridsen, T.L. & Poikane, S. (2013). Ecological status assessment of European lakes: a comparison of metrics for phytoplankton, macrophytes, benthic invertebrates and fish. *Hydrobiologia*, vol. 704 (1), pp. 57–74. DOI: https://doi.org/10.1007/s10750-012-1436-y

Maileht, K., Nõges, T., Nõges, P., Ott, I., Mischke, U., Carvalho, L. & Dudley, B. (2013). Water colour, phosphorus and alkalinity are the major determinants of the dominant phytoplankton species in European lakes. *Hydrobiologia*, vol. 704 (1), pp. 115–126. DOI: https://doi.org/10.1007/s10750-012-1348-x

Microsoft Office (2016). *Microsoft Excel*. Microsoft Corporation.

Miljödata-MVM (2019). *National data host for lakes and watercourses, and national data host for agricultural land. Swedish University of Agricultural Sciences (SLU)*. Available at: http://miljodata.slu.se/mvm/ [2019-01-31]

Nilsson, T., Stendahl, J. & Löfgren, O. (2015). *Markförhållanden i svensk skogsmark – data från Markinventeringen 1993-2002*. (Rapport 19). Uppsala: Institutionen för Mark och Miljö, Sveriges Lantbruksuniversitet.

Nõges, T. (2009). Relationships between morphometry, geographic location and water quality parameters of European lakes. *Hydrobiologia*, vol. 633, pp. 33–43. DOI: https://doi.org/10.1007/s10750-009-9874-x

Novotny, V. & Olem, H. (1994). Chapter 12- Receiving Water Impacts. *Water Quality Prevention, Identification and Management of Diffuse Pollution*. New York: John Wiley & Sons, Inc., pp. 783–795.

Paulsson, R., Djodjic, F., Carlsson Ross, C. & Hjerpe, K. (2015). *Nationell jordartskartering. Matjordens egenskaper i åkermarken*. (Rapport 2015:19). Jönköping: Jordbruksverket.

Phillips, G., Lyche-Solheim, A., Skjelbred, B., Mischke, U., Drakare, S., Free, G., Järvinen, M., De Hoyos, C., Morabito, G., Poikane, S. & Carvalho, L. (2012). A phytoplankton trophic index to assess the status of lakes for the Water Framework Directive. *Hydrobiologia*, vol. 704, pp. 75–95. DOI: https://doi.org/10.1007/s10750-012-1390-8

Phillips, G., Pietiläinen, O.-P., Carvalho, L., Solimini, A., Lyche Solheim, A. & Cardoso, A.C. (2008). Chlorophyll–nutrient relationships of different lake types using a large European dataset. *Aquatic Ecology*, vol. 42 (2), pp. 213–226. DOI: https://doi.org/10.1007/s10452-008-9180-0

SAS Institute Inc (2018). *JMP Pro*. Version: 14.0.0. Cary, North Carolina, USA: Statistical Analysis Software (SAS).

Schindler, D.W. (1974). Eutrophication and Recovery in Experimental Lakes: Implications for Lake Management. *Science*, vol. 184 (4139), pp. 897–899. DOI: https://doi.org/10.1126/science.184.4139.897

Schoumans, O.F., Chardon, W.J., Bechmann, M.E., Gascuel-Odoux, C., Hofman, G., Kronvang, B., Rubæk, G.H., Ulén, B. & Dorioz, J.-M. (2014). Mitigation options to reduce phosphorus losses from the agricultural sector and improve surface water quality: A review. *Science of The Total Environment*, vols 468–469, pp. 1255–1266. DOI: https://doi.org/10.1016/j.scitotenv.2013.08.061

Sharpley, A.N., Kleinman, P.J.A., Jordan, P., Bergström, L. & Allen, A.L. (2009). Evaluating the success of phosphorus management from field to watershed. *Journal of Environmental Quality*, vol. 38 (5), pp. 1981–1988. DOI: https://doi.org/10.2134/jeq2008.0056

SMHI, (Swedish Meteorological and Hydrological Institute) (2019). *Ladda ner data från Svenskt Vattenarkiv | SMHI. Ladda ner data från Svenskt Vattenarkiv från SVARversion 2012_2*. Available at: https://www.smhi.se/klimatdata/hydrologi/sjoar-och-vattendrag/ladda-ner-data-fran-svenskt-vattenarkiv-1.20127 [2019-02-18]

Smith, V.H. (2003). Eutrophication of freshwater and coastal marine ecosystems a global problem. *Environmental Science and Pollution Research*, vol. 10 (2), pp. 126–139. DOI: https://doi.org/10.1065/espr2002.12.142

Smith, V.H. & Schindler, D.W. (2009). Eutrophication science: where do we go from here? *Trends in Ecology & Evolution*, vol. 24 (4), pp. 201–207. DOI: https://doi.org/10.1016/j.tree.2008.11.009

Söderström, M. & Piikki, K. (2016). *Digitala åkermarkskartan – detaljerad kartering av textur i åkermarkens matjord*. (Teknisk Rapport nr 37). Skara: Precisionsodling och Pedometri, Sveriges Lantbruksuniversitet.

Søndergaard, M., Bjerring, R. & Jeppesen, E. (2013). Persistent internal phosphorus loading during summer in shallow eutrophic lakes. *Hydrobiologia*, vol. 710 (1), pp. 95–107. DOI: https://doi.org/10.1007/s10750-012-1091-3

Statistics Sweden (2018). *Discharges to water and sewage sludge production in 2016. Municipal wastewater treatment plants, pulp and paper industry and some other industry*. (MI 22 SM 1801). Statistics Sweden.

Stendahl, J. (2019). *Markinventeringens databas*. Available at: https://www.slu.se/centrumbildningar-och-projekt/markinventeringen/dokumentarkiv/mibas/ [2019-03-08]

Swedish Agency for Marine and Water Management (2018). *Växtplankton i sjöar vägledning för statusklassificering*. (Havs- och vattenmyndighetens rapport 2018:39). Göteborg.

Swedish EPA, (Swedish Environmental Protection Agency) (2010). *Status, potential and quality requirements for lakes, watercourses, coastal and transitional waters. A handbook on how quality requirements in bodies of surface water can be determined and monitored*. (Handbook 2007:4, Edition 1). Stockholm.

Swedish EPA, (Swedish Environmental Protection Agency) (2018). *Rening av avloppsvatten i Sverige 2016*. (ISBN: 978-91-620-8808-8). Stockholm.

USDA, (United States Department of Agriculture) (2019). *Guide to Texture by Feel | NRCS Soils-Soil Texture Triangle*. Available at: https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/edu/?cid=nrcs142p2_054311 [2019-04-16]

Weyhenmeyer, G.A., Peter, H. & Willén, E. (2013). Shifts in phytoplankton species richness and biomass along a latitudinal gradient – consequences for relationships between biodiversity and ecosystem functioning. *Freshwater Biology*, vol. 58 (3), pp. 612–623. DOI: https://doi.org/10.1111/j.1365-2427.2012.02779.x

Widén-Nilsson, E., Djodjic, F., Englund, D., Liljeberg, M., Hellgren, S., Olshammar, M., Olsson, H., Orback, C. & Tengdelius Brunell, J. (2016). *Kartdata till PLC6. Underlagsrapport till Pollution Load Compilation 6 rörande markanvändning, vattenförekomstområden, regionsindelning, jordbruksmarkens jordart, lutning och fosforhalt samt medelvärdesberäkningar*. (SMED-Svenska MiljöEmissionsData, Rapport Nr 186 2016). Norrköping: Sveriges Meteorologiska och Hydrologiska Institut.

Willén, E. (2007). *Växtplankton i sjöar-Bedömningsgrunder*. (Rapport 2007:6). Uppsala: Institutionen för Miljöanalys, Sveriges Lantbruksuniversitet.

Zar, J.H. (1984). Chapter 14-Data transformations. In: Kurtz, B. (Editor) (ed.) *Biostatical Analysis*. New Jersey, USA: Prentice-Hall, Second Edition., pp. 236–243.

Özkan, K., Jeppesen, E., Søndergaard, M., Lauridsen, T.L., Liboriussen, L. & Svenning, J.-C. (2013). Contrasting roles of water chemistry, lake morphology, land-use, climate and spatial processes in driving phytoplankton richness in the Danish landscape. *Hydrobiologia*, vol. 710 (1), pp. 173–187. DOI: https://doi.org/10.1007/s10750-011-0996-6

Appendix 1

Multiple regression with associated tests and output can be seen in *Table 10* and step history report from stepwise regression analysis in *Table 11* for tot bio.

Table 10. *Multiple regression output for total biovolume phytoplankton with R^2. Parameter estimate and effect test with significance level for all variables. Input variables derived from PLS with VIP >1 for each model, without and with total phosphorus (TP) as variable. Number of observed catchments in multiple regression and observed in parentheses from stepwise regression. Multiple regression selected by JMP (minimum BIC) in bold text and by criteria in Italic text. Note that variables are transformed. For variable abbreviation see* Table 1 *and* Table 2.

					Parameters Estimates			Effect Test	
Model	Excl/Incl TP	R^2	Observ.	Variable	Estimate	t Ratio	Prob>|t|	F Ratio	Prob> F
M1	Excl. TP	**0.32*****	**523 (250)**	**Intercept**	**-1.142**	**-6.01**	**<.0001**		
				Longitude	**0.0548**	**4.37**	**<.0001**	**19.1**	**<.0001**
				Water Body Area	**-0.117**	**-4.29**	**<.0001**	**18.4**	**<.0001**
				Urban Area	**0.269**	**4.55**	**<.0001**	**20.7**	**<.0001**
				Agriculture	**0.566**	**12.34**	**<.0001**	**152.2**	**<.0001**
	Incl. TP	**0.65 *****	**185 (165)**	**Intercept**	**-2.133**	**-7.73**	**<.0001**		
				MI-pH	**0.215**	**3.12**	**0.0021**	**9.8**	**0.0021**
				TP	**1.136**	**17.37**	**<.0001**	**301.9**	**<.0001**
M2	Excl. TP	**0.34*****	**523 (523)**	**Intercept**	**-0.335**	**-8.07**	**<.0001**		
				Water Body Area	**-0.118**	**-4.43**	**<.0001**	**19.6**	**<.0001**
				DSMS-Organic Cont	**0.546**	**3.18**	**0.0016**	**10.1**	**0.0016**
				Urban Area	**0.259**	**4.45**	**<.0001**	**19.8**	**<.0001**
				Agriculture	**0.397**	**6.86**	**<.0001**	**47.1**	**<.0001**
				SGU-Clay	**0.253**	**5.36**	**<.0001**	**28.7**	**<.0001**
		*0.33****	*523 (523)*	*Intercept*	*-0.353*	*-8.54*	*<.0001*		
				Water Body Area	*-0.115*	*-4.29*	*<.0001*	*18.4*	*<.0001*
				Urban Area	*0.260*	*4.44*	*<.0001*	*19.7*	*<.0001*
				Agriculture	*0.500*	*10.33*	*<.0001*	*106.7*	*<.0001*
				SGU-Clay	*0.250*	*5.25*	*<.0001*	*27.6*	*<.0001*
	Incl. TP	**0.66*****	**303 (303)**	**Intercept**	**-1.277**	**-21.12**	**<.0001**		
				TP	**1.137**	**24.45**	**<.0001**	**597.7**	**<.0001**
M3	Excl. TP	**0.30*****	**310 (310)**	**Intercept**	**6.143**	**2.64**	**0.0087**		
				Longitude	**0.0247**	**0.9**	**0.3675**	**0.8**	**0.3675**
				DSMS-Clay Cont	**0.00915**	**3.72**	**0.0002**	**13.8**	**0.0002**
				Forest	**-0.0188**	**-7.83**	**<.0001**	**61.3**	**<.0001**
				Water	**-0.650**	**-5.73**	**<.0001**	**32.8**	**<.0001**

					Parameters Estimates			Effect Test	
Model	Excl/Incl TP	R^2	Observ.	Variable	Estimate	t Ratio	Prob>|t|	F Ratio	Prob> F
				Precip	**-1.631**	**-2.41**	**0.0167**	**5.8**	**0.0167**
		0.29***	*310 (310)*	*Intercept*	*0.605*	*1.76*	*0.0797*		
				Longitude	*0.0766*	*4.51*	*<.0001*	*20.3*	*<.0001*
				DSMS-Clay Cont	*0.00941*	*3.79*	*0.0002*	*14.4*	*0.0002*
				Forest	*-0.0179*	*-7.48*	*<.0001*	*56.0*	*<.0001*
				Water	*-0.631*	*-5.53*	*<.0001*	*30.6*	*<.0001*
	Incl. TP	0.65***	**199 (199)**	**Intercept**	**-1.602**	**-10.85**	**<.0001**		
				Water	**0.228**	**2.38**	**0.0182**	**5.7**	**0.0182**
				TP	**1.197**	**18.48**	**<.0001**	**341.4**	**<.0001**
		0.64***	*199 (199)*	*Intercept*	*-1.294*	*-18.05*	*<.0001*		
				TP	*1.142*	*18.65*	*<.0001*	*347.9*	*<.0001*

$* (p < 0.05)$ $** (p < 0.01)$ $*** (p < 0.001)$

Table 11. *Report from step history, for total biovolume phytoplankton with and without total phosphorus (TP). Mallow´s Cp criterion, p= number of parameters included intercept in selection, AICc=corrected Akaike´s Information Criterion, BIC=Bayesian Information Criterion. Entered action show variable includes in stepwise regression where "best" fit selected by JMP, minimum BIC, (Bold text) and by criteria (Italic text) for building multiple regression. Note that variables are transformed. For variable abbreviation see* Table 1 *and* Table 2.

	Step	Variable	Action	"Sig Prob"	Seq SS	R^2	Cp	p	AICc	BIC
M1 Excl. TP Observ. 250	***1***	***Agriculture***	***Entered***	***0.000***	***13.852***	***0.191***	***33.4***	***2***	***353.8***	***364.3***
	2	***Longitude***	***Entered***	***0.000***	***4.154***	***0.248***	***15.6***	***3***	***337.6***	***351.5***
	3	***Water Body Area***	***Entered***	***0.002***	***2.178***	***0.278***	***7.3***	***4***	***329.5***	***346.9***
	4	***Urban Area***	***Entered***	***0.016***	***1.226***	***0.295***	***3.5***	***5***	***325.7***	***346.5***
	5	K_AL	Entered	0.079	0.647	0.304	2.4	6	324.6	348.8
	6	DSMS-Clay Cont	Entered	0.070	0.679	0.313	1.2	7	323.4	351.0
	7	DSMS-Organic Cont	Entered	0.029	0.975	0.326	-1.5	8	320.6	351.6
	8	Mire	Entered	0.112	0.512	0.333	-1.9	9	320.2	354.5
	9	Mg_AL	Entered	0.310	0.208	0.336	-0.9	10	321.3	358.9
	10	FAO-CL	Entered	0.353	0.174	0.339	0.3	11	322.6	363.5
	11	MI-pH	Entered	0.430	0.126	0.340	1.7	12	324.1	368.4
	12	SGU-Thin Soil Layer	Entered	0.515	0.086	0.342	3.3	13	325.9	373.5
	13	SGU-Clay	Entered	0.420	0.132	0.343	4.7	14	327.5	378.3
	14	Clay Cont	Entered	0.616	0.051	0.344	6.4	15	329.5	383.5
	15	Best	Specific	.	.	0.295	3.5	5	325.7	346.5
M1 Incl. TP Observ. 161	***1***	***TP***	***Entered***	***0.000***	***29.339***	***0.605***	***15.3***	***2***	***120.3***	***129.3***
	2	***MI-pH***	***Entered***	***0.001***	***1.235***	***0.631***	***6.1***	***3***	***111.6***	***123.7***
	3	FAO-SCL	Entered	0.039	0.482	0.641	3.8	4	109.4	124.4
	4	Water Body Area	Entered	0.072	0.360	0.648	2.6	5	108.2	126.1

	Step	Variable	Action	"Sig Prob"	Seq SS	R^2	Cp	p	AICc	BIC
	5	DSMS-Organic Cont	Entered	0.197	0.183	0.652	2.9	6	108.6	129.4
	6	Open Land	Entered	0.286	0.125	0.654	3.8	7	109.6	133.3
	7	Mg_AL	Entered	0.356	0.094	0.656	5.0	8	111.0	137.5
	8	Ca_AL	Entered	0.179	0.198	0.660	5.2	9	111.3	140.7
	9	Clay Cont	Entered	0.242	0.149	0.663	5.8	10	112.2	144.3
	10	Precip	Entered	0.262	0.137	0.666	6.6	11	113.1	148.0
	11	Mire	Entered	0.238	0.151	0.669	7.2	12	114.0	151.6
	12	MI-C. Cont	Entered	0.100	0.291	0.675	6.6	13	113.5	153.7
	13	Best	Specific	.	.	0.631	6.1	3	111.6	123.7
M2 Excl. TP Observ. 523	*1*	***Agriculture***	***Entered***	***0.000***	***40.048***	***0.231***	***103.4***	***2***	***774.8***	***787.6***
	2	***SGU-Clay***	***Entered***	***0.000***	***10.181***	***0.290***	***57.8***	***3***	***735.3***	***752.2***
	3	***Urban Area***	***Entered***	***0.000***	***3.886***	***0.312***	***41.6***	***4***	***720.5***	***741.7***
	4	***Water Body Area***	***Entered***	***0.000***	***4.079***	***0.336***	***24.6***	***5***	***704.4***	***729.7***
	5	**DSMS-Organic Cont**	**Entered**	**0.002**	**2.205**	**0.349**	**16.2**	**6**	**696.3**	**725.9**
	6	FAO-SCL	Entered	0.019	1.207	0.356	12.6	7	692.7	726.5
	7	Precip	Entered	0.016	1.261	0.363	8.7	8	688.9	726.8
	8	Forest	Entered	0.025	1.070	0.369	5.7	9	685.8	728.0
	9	DSMS-Clay Cont	Entered	0.222	0.317	0.371	6.2	10	686.4	732.7
	10	FAO-SC	Entered	0.138	0.467	0.374	6.1	11	686.2	736.8
	11	FAO-L	Entered	0.458	0.117	0.374	7.5	12	687.8	742.4
	12	FAO-CL	Entered	0.575	0.067	0.375	9.2	13	689.6	748.4
	13	Open Land	Entered	0.730	0.025	0.375	11.1	14	691.6	754.5
	14	Longitude	Entered	0.801	0.014	0.375	13.0	15	693.6	760.7
	15	Masl	Entered	0.927	0.002	0.375	15.0	16	695.8	767.0
	16	Best	Specific	.	.	0.349	16.2	6	696.3	725.9
M2 Incl. TP Observ. 303	***1***	***TP***	***Entered***	***0.000***	***62.855***	***0.665***	***-4.8***	***2***	***181.5***	***192.6***
	2	FAO-L	Entered	0.079	0.324	0.669	-5.8	3	180.5	195.2
	3	FAO-SCL	Entered	0.134	0.235	0.671	-6.0	4	180.2	198.6
	4	Open Land	Entered	0.319	0.104	0.672	-5.0	5	181.3	203.3
	5	FAO-SC	Entered	0.321	0.103	0.673	-3.9	6	182.4	208.0
	6	Water Body Area	Entered	0.464	0.056	0.674	-2.5	7	184.0	213.2
	7	Urban Area	Entered	0.552	0.037	0.674	-0.8	8	185.7	218.5
	8	Longitude	Entered	0.556	0.036	0.675	0.9	9	187.5	223.9
	9	SGU-Clay	Entered	0.725	0.013	0.675	2.7	10	189.5	229.5
	10	Masl	Entered	0.599	0.029	0.675	4.5	11	191.4	234.9
	11	Forest	Entered	0.712	0.014	0.675	6.3	12	193.4	240.5
	12	Best	Specific	.	.	0.665	-4.8	2	181.5	192.6
M3 Excl. TP Observ. 310	***1***	***Longitude***	***Entered***	***0.000***	***9.539***	***0.112***	***103.5***	***2***	***448.6***	***459.7***
	2	***Forest***	***Entered***	***0.000***	***7.594***	***0.201***	***64.4***	***3***	***417.9***	***432.7***
	3	***Water***	***Entered***	***0.000***	***5.788***	***0.269***	***35.1***	***4***	***392.4***	***410.9***

	Step	Variable	Action	"Sig Prob"	Seq SS	R^2	Cp	p	AICc	BIC
	4	***DSMS-Clay Cont***	***Entered***	***0.000***	***2.807***	***0.302***	***21.9***	***5***	***380.2***	***402.3***
	5	**Precip**	**Entered**	**0.017**	**1.112**	**0.315**	**17.9**	**6**	**376.4**	**402.2**
	6	DSMS-Organic Cont	Entered	0.074	0.614	0.322	16.5	7	375.2	404.7
	7	Water Body Area	Entered	0.047	0.754	0.331	14.5	8	373.3	406.3
	8	SGU-Rock	Entered	0.032	0.866	0.341	11.8	9	370.7	407.3
	9	SGU-Clay	Entered	0.034	0.839	0.351	9.2	10	368.2	408.4
	10	Masl	Entered	0.114	0.461	0.357	8.7	11	367.7	411.5
	11	Agriculture	Entered	0.138	0.403	0.361	8.6	12	367.6	415.0
	12	SGU-Sand	Entered	0.287	0.207	0.364	9.4	13	368.6	419.5
	13	Lake Area	Entered	0.303	0.194	0.366	10.4	14	369.7	424.1
	14	FAO-SCL	Entered	0.332	0.172	0.368	11.5	15	371.0	428.9
	15	Urban Area	Entered	0.329	0.175	0.370	12.5	16	372.2	433.6
	16	Best	Specific	.	.	0.315	17.9	6	376.4	402.2
M3 Incl. TP Observ. 199	***1***	***TP***	***Entered***	***0.000***	***31.689***	***0.639***	***8.3***	***2***	***92.0***	***101.8***
	2	**Water**	**Entered**	**0.018**	**0.504**	**0.649**	**4.6**	**3**	**88.4**	**101.4**
	3	DSMS-Organic Cont	Entered	0.161	0.176	0.652	4.6	4	88.5	104.7
	4	FAO-CL	Entered	0.150	0.184	0.656	4.5	5	88.5	107.8
	5	SGU-Rock	Entered	0.035	0.389	0.664	2.1	6	86.1	108.6
	6	Forest	Entered	0.075	0.274	0.669	1.0	7	85.0	110.6
	7	SGU-Clay	Entered	0.185	0.151	0.672	1.3	8	85.3	114.0
	8	SGU-Sand	Entered	0.195	0.143	0.675	1.7	9	85.8	117.5
	9	Urban Area	Entered	0.256	0.110	0.677	2.4	10	86.7	121.5
	10	SGU-Thin Soil Layer	Entered	0.365	0.070	0.679	3.6	11	88.1	125.9
	11	Precip	Entered	0.447	0.049	0.680	5.1	12	89.7	130.6
	12	Masl	Entered	0.501	0.039	0.681	6.6	13	91.6	135.4
	13	Best	Specific	.	.	0.649	4.6	3	88.4	101.4

Appendix 2

Multiple regression with associated tests and output can be seen in *Table 12* and step history report from stepwise regression analysis in *Table 13* for PTI.

Table 12. *Multiple regression output for Plankton Trophic Index (PTI) with R^2. Parameter estimate and effect test with significance level for all variables. Input variables derived from PLS with VIP >1 for each model, without and with total phosphorus (TP) as variable. Number of observed catchments in multiple regression and observed in parentheses from stepwise regression. Multiple regression selected by JMP (minimum BIC) in bold text and by criteria in Italic text. Note that variables are transformed. For variable abbreviation see* Table 1 *and* Table 2.

					Parameters Estimates			Effect Test	
Model	Excl/Incl TP	R^2	Observ.	Variable	Estimate	t Ratio	Prob>|t|	F Ratio	Prob > F
M1	Excl. TP	0.30***	**314 (250)**	**Intercept**	**-0.306**	**-3.45**	**0.0006**		
				DPS	**0.168**	**2.27**	**0.0241**	**5.1**	**0.0241**
				DSMS-Organic Cont	**0.399**	**2.56**	**0.0111**	**6.5**	**0.0111**
				Agriculture	**0.349**	**4.55**	**<.0001**	**20.7**	**<.0001**
				FAO-SC	**0.336**	**4.56**	**<.0001**	**20.8**	**<.0001**
		0.44***	*523 (250)*	*Intercept*	*-0.319*	*-10.35*	*<.0001*		
				DSMS-Organic Cont	*0.425*	*3.28*	*0.0011*	*10.7*	*0.0011*
				Agriculture	*0.481*	*11.29*	*<.0001*	*127.6*	*<.0001*
				FAO-SC	*0.328*	*4.95*	*<.0001*	*24.5*	*<.0001*
	Incl. TP	0.66***	**303 (161)**	**Intercept**	**-0.708**	**-4.91**	**<.0001**		
				DSMS-Clay Cont	**0.00340**	**1.92**	**0.0555**	**3.7**	**0.0555**
				Forest	**-0.00172**	**-1.1**	**0.2742**	**1.2**	**0.2742**
				Mire	**-0.0580**	**-1.2**	**0.2321**	**1.4**	**0.2321**
				Agriculture	**0.253**	**5.08**	**<.0001**	**25.8**	**<.0001**
				SGU-Gravel	**0.183**	**1.53**	**0.1275**	**2.3**	**0.1275**
				TP	**0.579**	**12.16**	**<.0001**	**148.0**	**<.0001**
		0.65***	*303 (161)*	*Intercept*	*-0.898*	*-18.25*	*<.0001*		
				Agriculture	*0.326*	*8.84*	*<.0001*	*78.1*	*<.0001*
				TP	*0.613*	*13.51*	*<.0001*	*182.4*	*<.0001*
M2	Excl. TP	0.50***	**523 (523)**	**Intercept**	**-0.370**	**-12.33**	**<.0001**		
				DSMS-Organic Cont	**0.420**	**3.41**	**0.0007**	**11.6**	**0.0007**
				Urban Area	**0.258**	**6.08**	**<.0001**	**37.0**	**<.0001**
				Agriculture	**0.398**	**9.22**	**<.0001**	**84.9**	**<.0001**
				FAO-SCL	**0.204**	**3.2**	**0.0014**	**10.3**	**0.0014**
				SGU-Clay	**0.127**	**3.32**	**0.001**	**11.0**	**0.001**

					Parameters Estimates			Effect Test	
Model	Excl/Incl TP	R^2	Observ.	Variable	Estimate	t Ratio	Prob>|t|	F Ratio	Prob > F
		*0.48****	*523 (523)*	*Intercept*	*-0.398*	*-13.31*	*<.0001*		
				Urban Area	*0.235*	*5.53*	*<.0001*	*30.5*	*<.0001*
				Agriculture	*0.526*	*15.78*	*<.0001*	*249.0*	*<.0001*
				SGU-Clay	*0.183*	*5.38*	*<.0001*	*28.9*	*<.0001*
	Incl. TP	**0.68*****	**303 (303)**	**Intercept**	**-0.880**	**-16.8**	**<.0001**		
				Catch. Area	**0.0356**	**2.57**	**0.0106**	**6.6**	**0.0106**
				Urban Area	**0.150**	**2.51**	**0.0127**	**6.3**	**0.0127**
				Agriculture	**0.237**	**5.91**	**<.0001**	**34.9**	**<.0001**
				SGU-Clay	**0.146**	**3.63**	**0.0003**	**13.2**	**0.0003**
				TP	**0.538**	**11.11**	**<.0001**	**123.5**	**<.0001**
		*0.65****	*303 (303)*	*Intercept*	*-0.898*	*-18.25*	*<.0001*		
				Agriculture	*0.326*	*8.84*	*<.0001*	*78.1*	*<.0001*
				TP	*0.613*	*13.51*	*<.0001*	*182.4*	*<.0001*
M3	Excl. TP	**0.47*****	**310 (310)**	**Intercept**	**-0.0144**	**-0.11**	**0.913**		
				Forest	**-0.00632**	**-3.53**	**0.0005**	**12.4**	**0.0005**
				Catch. Area	**0.0802**	**4.11**	**<.0001**	**16.9**	**<.0001**
				Urban Area	**0.232**	**4.07**	**<.0001**	**16.5**	**<.0001**
				Agriculture	**0.344**	**5.76**	**<.0001**	**33.1**	**<.0001**
				SGU-Clay	**0.258**	**6.12**	**<.0001**	**37.4**	**<.0001**
	Incl. TP	**0.66*****	**199 (199)**	**Intercept**	**-0.814**	**-14.31**	**<.0001**		
				Lake Area	**0.0872**	**3.43**	**0.0007**	**11.8**	**0.0007**
				Urban Area	**0.187**	**2.44**	**0.0155**	**6.0**	**0.0155**
				Agriculture	**0.180**	**2.96**	**0.0035**	**8.7**	**0.0035**
				FAO-CL	**0.591**	**4.84**	**<.0001**	**23.4**	**<.0001**
				TP	**0.543**	**9.32**	**<.0001**	**86.8**	**<.0001**
		*0.66****	*199 (199)*	*Intercept*	*-0.850*	*-15.28*	*<.0001*		
				Lake Area	*0.100*	*4.00*	*<.0001*	*16.0*	*<.0001*
				Agriculture	*0.190*	*3.09*	*0.0023*	*9.5*	*0.0023*
				FAO-CL	*0.561*	*4.55*	*<.0001*	*20.7*	*<.0001*
				TP	*0.594*	*10.78*	*<.0001*	*116.1*	*<.0001*

* ($p < 0.05$) ** ($p < 0.01$) *** ($p < 0.001$)

Table 13. *Report from step history, for Plankton Trophic Index (PTI) with and without total phosphorus (TP). Mallow´s Cp criterion. p= number of parameters included intercept in selection, AICc=corrected Akaike´s Information Criterion, BIC=Bayesian Information Criterion. Entered action show variable includes in stepwise regression where "best" fit selected by JMP, minimum BIC, (Bold text) and by criteria (Italic text) for building multiple regression. Note that variables are transformed. For variable abbreviation see* Table 1 *and* Table 2.

	Step	Variable	Action	"Sig Prob"	Seq SS	R^2	Cp	p	AICc	BIC
M1 Excl. TP Observ. 250	***1***	***Agriculture***	***Entered***	***0.000***	***9.391***	***0.251***	***39.4***	***2***	***168.9***	***179.4***
	2	***FAO-SC***	***Entered***	***0.000***	***1.453***	***0.289***	***26.6***	***3***	***157.7***	***171.6***
	3	***DSMS-Organic Cont***	***Entered***	***0.001***	***1.258***	***0.323***	***15.8***	***4***	***147.7***	***165.1***
	4	**DPS**	**Entered**	**0.017**	**0.588**	**0.339**	**11.9**	**5**	**143.9**	**164.7**
	5	SGU-Gravel	Entered	0.080	0.309	0.347	10.7	6	142.9	167.1
	6	Mire	Entered	0.073	0.323	0.356	9.4	7	141.7	169.3
	7	K_AL	Entered	0.057	0.359	0.365	7.8	8	140.1	171.1
	8	FAO-SCL	Entered	0.038	0.421	0.376	5.5	9	137.9	172.1
	9	Open Land	Entered	0.107	0.253	0.383	4.9	10	137.3	175.0
	10	DSMS-Clay Cont	Entered	0.119	0.234	0.389	4.6	11	137.0	177.9
	11	Clay Cont	Entered	0.308	0.100	0.392	5.5	12	138.1	182.4
	12	Longitude	Entered	0.297	0.105	0.395	6.5	13	139.2	186.7
	13	Forest	Entered	0.204	0.155	0.399	6.9	14	139.8	190.5
	14	FAO-CL	Entered	0.224	0.141	0.403	7.5	15	140.5	194.5
	15	Best	Specific	.	.	0.339	11.9	5	143.9	164.7
M1 Incl. TP Observ. 161	1	***TP***	***Entered***	***0.000***	***11.195***	***0.490***	***47.1***	***2***	***40.2***	***49.3***
	2	***Agriculture***	***Entered***	***0.000***	***1.388***	***0.551***	***24.8***	***3***	***21.9***	***33.9***
	3	**SGU-Gravel**	**Entered**	**0.015**	**0.379**	**0.568**	**20.1**	**4**	**17.9**	**33.0**
	4	**DSMS-Clay Cont**	**Entered**	**0.014**	**0.375**	**0.584**	**15.6**	**5**	**13.9**	**31.8**
	5	**Mire**	**Entered**	**0.026**	**0.299**	**0.597**	**12.3**	**6**	**10.9**	**31.7**
	6	**Forest**	**Entered**	**0.007**	**0.421**	**0.615**	**7.0**	**7**	**5.6**	**29.3**
	7	Clay Cont	Entered	0.054	0.211	0.625	5.3	8	3.9	30.5
	8	Open Land	Entered	0.108	0.145	0.631	4.7	9	3.4	32.8
	9	MI-pH	Entered	0.128	0.129	0.637	4.5	10	3.3	35.4
	10	FAO-SCL	Entered	0.235	0.078	0.640	5.1	11	4.1	39.0
	11	FAO-SC	Entered	0.175	0.101	0.645	5.3	12	4.5	42.0
	12	SGU-Clay	Entered	0.260	0.070	0.648	6.1	13	5.5	45.7
	13	DPS	Entered	0.299	0.059	0.650	7.1	14	6.7	49.6
	14	K_AL	Entered	0.239	0.076	0.653	7.7	15	7.7	53.2
	15	FAO-CL	Entered	0.261	0.069	0.657	8.5	16	8.7	56.9
	16	Ca_AL	Entered	0.317	0.055	0.659	9.6	17	10.2	60.8
	17	Best	Specific	.	.	0.615	7.0	7	5.6	29.3
	1	***Agriculture***	***Entered***	***0.000***	***48.026***	***0.411***	***98.4***	***2***	***429.3***	***442.0***

	Step	Variable	Action	"Sig Prob"	Seq SS	R^2	Cp	p	AICc	BIC
M2 Excl. TP Observ. 523	2	***Urban Area***	***Entered***	***0.000***	***4.705***	***0.451***	***58.2***	***3***	***394.3***	***411.2***
	3	***SGU-Clay***	***Entered***	***0.000***	***3.384***	***0.480***	***29.8***	***4***	***367.9***	***389.1***
	4	**DSMS-Organic Cont**	**Entered**	**0.001**	**1.235**	**0.491**	**20.7**	**5**	**359.2**	**384.6**
	5	**FAO-SCL**	**Entered**	**0.001**	**1.157**	**0.501**	**12.3**	**6**	**351.0**	**380.6**
	6	Catch. Area	Entered	0.015	0.661	0.507	8.4	7	347.1	380.9
	7	Open Land	Entered	0.188	0.194	0.508	8.6	8	347.4	385.4
	8	FAO-L	Entered	0.193	0.189	0.510	8.9	9	347.8	389.9
	9	SGU-Sand	Entered	0.116	0.275	0.512	8.5	10	347.3	393.7
	10	FAO-CL	Entered	0.180	0.200	0.514	8.7	11	347.6	398.1
	11	Forest	Entered	0.387	0.083	0.515	9.9	12	348.9	403.6
	12	Masl	Entered	0.341	0.101	0.516	11.0	13	350.1	408.9
	13	DSMS-Silt Cont	Entered	0.389	0.083	0.516	12.3	14	351.5	414.4
	14	DSMS-Clay Cont	Entered	0.714	0.015	0.516	14.2	15	353.5	420.5
	15	FAO-SC	Entered	0.696	0.017	0.516	16.0	16	355.4	426.6
	16	Best	Specific	.	.	0.501	12.3	6	351.0	380.6
M2 Incl. TP Observ. 303	1	***TP***	***Entered***	***0.000***	***33.916***	***0.564***	***120.7***	***2***	***124.4***	***135.5***
	2	***Agriculture***	***Entered***	***0.000***	***5.418***	***0.654***	***36.0***	***3***	***56.3***	***71.1***
	3	**Catch. Area**	**Entered**	**0.000**	**0.866**	**0.669**	**24.1**	**4**	**45.5**	**63.9**
	4	**SGU-Clay**	**Entered**	**0.001**	**0.798**	**0.682**	**13.3**	**5**	**35.2**	**57.2**
	5	**Urban Area**	**Entered**	**0.013**	**0.396**	**0.688**	**9.0**	**6**	**31.0**	**56.6**
	6	FAO-SCL	Entered	0.088	0.183	0.691	8.1	7	30.1	59.3
	7	DSMS-Silt Cont	Entered	0.149	0.131	0.694	8.0	8	30.1	62.9
	8	FAO-L	Entered	0.155	0.127	0.696	7.9	9	30.1	66.5
	9	DSMS-Organic Cont	Entered	0.180	0.112	0.698	8.1	10	30.4	70.4
	10	FAO-SC	Entered	0.298	0.067	0.699	9.1	11	31.5	75.0
	11	FAO-CL	Entered	0.217	0.095	0.700	9.6	12	32.1	79.1
	12	Open Land	Entered	0.274	0.074	0.702	10.4	13	33.0	83.6
	13	Masl	Entered	0.447	0.036	0.702	11.8	14	34.6	88.7
	14	DSMS-Clay Cont	Entered	0.541	0.023	0.703	13.4	15	36.5	94.0
	15	Forest	Entered	0.522	0.026	0.703	15.0	16	38.3	99.3
	16	Best	Specific	.	.	0.688	9.0	6	31.0	56.6
M3 Excl. TP Observ. 310	***1***	***Agriculture***	***Entered***	***0.000***	***15.652***	***0.309***	***101.8***	***2***	***209.4***	***220.5***
	2	***Urban Area***	***Entered***	***0.000***	***4.299***	***0.394***	***53.7***	***3***	***170.8***	***185.6***
	3	***SGU-Clay***	***Entered***	***0.000***	***2.068***	***0.435***	***31.6***	***4***	***151.2***	***169.7***
	4	***Catch. Area***	***Entered***	***0.000***	***1.152***	***0.458***	***20.1***	***5***	***140.6***	***162.7***

	Step	Variable	Action	"Sig Prob"	Seq SS	R^2	Cp	p	AICc	BIC
	5	***Forest***	***Entered***	***0.001***	***1.080***	***0.479***	***9.5***	***6***	***130.2***	***156.0***
	6	FAO-CL	Entered	0.042	0.358	0.486	7.4	7	128.1	157.5
	7	DSMS-Organic Cont	Entered	0.122	0.205	0.490	7.0	8	127.8	160.8
	8	Open Land	Entered	0.108	0.221	0.495	6.4	9	127.2	163.9
	9	SGU-Sand	Entered	0.205	0.137	0.497	6.8	10	127.7	167.9
	10	FAO-SCL	Entered	0.370	0.068	0.499	8.0	11	129.0	172.8
	11	FAO-SaL	Entered	0.173	0.158	0.502	8.2	12	129.3	176.6
	12	DSMS-Clay Cont	Entered	0.390	0.063	0.503	9.4	13	130.7	181.6
	13	Lake Area	Entered	0.498	0.039	0.504	11.0	14	132.4	186.8
	14	Water Body Area	Entered	0.534	0.033	0.504	12.6	15	134.2	192.2
	15	DSMS-Silt Cont	Entered	0.553	0.030	0.505	14.2	16	136.1	197.5
	16	Best	Specific	.	.	0.479	9.5	6	130.2	156.0
M3 Incl. TP Observ. 199	***1***	***TP***	***Entered***	***0.000***	***14.558***	***0.502***	***100.4***	***2***	***49.0***	***58.7***
	2	***Agriculture***	***Entered***	***0.000***	***2.995***	***0.605***	***41.2***	***3***	***4.8***	***17.8***
	3	***FAO-CL***	***Entered***	***0.000***	***0.877***	***0.635***	***25.2***	***4***	***-8.9***	***7.2***
	4	***Lake Area***	***Entered***	***0.000***	***0.805***	***0.663***	***10.8***	***5***	***-22.5***	***-3.2***
	5	**Urban Area**	**Entered**	**0.016**	**0.293**	**0.673**	**6.8**	**6**	**-26.4**	**-4.0**
	6	SGU-Clay	Entered	0.091	0.141	0.678	5.9	7	-27.2	-1.7
	7	DSMS-Silt Cont	Entered	0.030	0.228	0.686	3.3	8	-30.0	-1.3
	8	FAO-SaL	Entered	0.299	0.052	0.688	4.2	9	-28.9	2.9
	9	DSMS-Clay Cont	Entered	0.216	0.073	0.690	4.7	10	-28.3	6.6
	10	FAO-L	Entered	0.439	0.029	0.691	6.1	11	-26.6	11.2
	11	DSMS-Organic Cont	Entered	0.400	0.034	0.692	7.4	12	-25.1	15.7
	12	SGU-Sand	Entered	0.462	0.026	0.693	8.9	13	-23.4	20.5
	13	Open Land	Entered	0.452	0.027	0.694	10.3	14	-21.6	25.1
	14	FAO-SCL	Entered	0.391	0.036	0.695	11.6	15	-20.1	29.6
	15	Catch. Area	Entered	0.626	0.012	0.696	13.4	16	-17.9	34.7
	16	Best	Specific	.	.	0.673	6.8	6	-26.4	-4.0

Appendix 3

Multiple regression with associated tests and output can be seen in *Table 14* and step history report from stepwise regression analysis in *Table 15* for TP.

Table 14. *Multiple regression output for total phosphorus (TP) with R^2. Parameter estimate and effect test with significance level for all variables. Input variables derived from PLS with VIP >1 for each model. Number of observed catchments in multiple regression and observed in parentheses from stepwise regression. Multiple regression selected by JMP (minimum BIC) in bold text and by criteria in Italic text. Note that variables are transformed. For variable abbreviation see* Table 1 *and* Table 2.

				Parameters Estimates			Effect Test	
Model	R^2	Observ.	Variable	Estimate	t Ratio	Prob>|t|	F Ratio	Prob > F
M1	**0.51*****	**303** **(185)**	**Intercept**	**0.800**	**25.01**	**<.0001**		
			DSMS-Clay Cont	**0.0141**	**8.06**	**<.0001**	**65.0**	**<.0001**
			Water Body Area	**-0.168**	**-7.69**	**<.0001**	**59.2**	**<.0001**
			DSMS-Organic Cont	**0.732**	**4.86**	**<.0001**	**23.6**	**<.0001**
			Urban Area	**0.251**	**4.00**	**<.0001**	**16.0**	**<.0001**
			Agriculture	**0.212**	**3.76**	**0.0002**	**14.1**	**0.0002**
M2	**0.58*****	**303** **(303)**	**Intercept**	**4.247**	**4.52**	**<.0001**		
			DSMS-Clay Cont	**0.00873**	**2.99**	**0.003**	**9.0**	**0.003**
			Forest	**-0.00892**	**-4.61**	**<.0001**	**21.3**	**<.0001**
			Water Body Area	**-0.170**	**-5.50**	**<.0001**	**30.3**	**<.0001**
			Catch. Area	**0.0512**	**2.43**	**0.0158**	**5.9**	**0.0158**
			DSMS-Organic Cont	**0.501**	**3.36**	**0.0009**	**11.3**	**0.0009**
			Water	**-0.328**	**-4.08**	**<.0001**	**16.6**	**<.0001**
			Agriculture	**0.0741**	**1.02**	**0.309**	**1.0**	**0.309**
			SGU-Organic Soil	**0.170**	**2.63**	**0.0089**	**6.9**	**0.0089**
			SGU-Clay	**0.143**	**2.09**	**0.0379**	**4.4**	**0.0379**
			Precip	**-0.900**	**-2.90**	**0.004**	**8.4**	**0.004**
	*0.52****	*303* *(303)*	*Intercept*	*0.535*	*8.22*	*<.0001*		
			Water Body Area	*-0.137*	*-6.41*	*<.0001*	*41.1*	*<.0001*
			Agriculture	*0.472*	*12.02*	*<.0001*	*144.4*	*<.0001*
			SGU-Organic Soil	*0.308*	*5.39*	*<.0001*	*29.0*	*<.0001*
			SGU-Clay	*0.358*	*8.69*	*<.0001*	*75.5*	*<.0001*
M3	**0.58*****	**199** **(199)**	**Intercept**	**0.724**	**6.12**	**<.0001**		
			Water Body Area	**-0.133**	**-5.22**	**<.0001**	**27.2**	**<.0001**

				Parameters Estimates			Effect Test	
Model	R^2	Observ.	Variable	Estimate	t Ratio	Prob>|t|	F Ratio	Prob > F
			Urban Area	**0.442**	**5.72**	**<.0001**	**32.7**	**<.0001**
			Open Land	**0.118**	**1.25**	**0.213**	**1.6**	**0.213**
			Water	**-0.180**	**-2.39**	**0.0181**	**5.7**	**0.0181**
			Agriculture	**0.263**	**3.43**	**0.0007**	**11.8**	**0.0007**
			FAO-SCL	**0.308**	**2.51**	**0.0131**	**6.3**	**0.0131**
			SGU-Organic Soil	**0.290**	**5.33**	**<.0001**	**28.5**	**<.0001**
			SGU-Clay	**0.287**	**5.4**	**<.0001**	**29.2**	**<.0001**
	*0.55****	*199 (199)*	*Intercept*	*0.518*	*8.75*	*<.0001*		
			Water Body Area	*-0.141*	*-5.66*	*<.0001*	*32.0*	*<.0001*
			Urban Area	*0.478*	*6.34*	*<.0001*	*40.2*	*<.0001*
			Agriculture	*0.380*	*6.68*	*<.0001*	*44.6*	*<.0001*
			SGU-Organic Soil	*0.306*	*5.62*	*<.0001*	*31.5*	*<.0001*
			SGU-Clay	*0.365*	*8.05*	*<.0001*	*64.8*	*<.0001*

* ($p < 0.05$) ** ($p < 0.01$) *** ($p < 0.001$)

Table 15. *Report from step history, for total phosphorus (TP) with Mallow´s Cp criterion, p= number of parameters included intercept in selection, AICc=corrected Akaike´s Information Criterion, BIC=Bayesian Information Criterion. Entered action show variable includes in stepwise regression where "best" fit selected by JMP, minimum BIC, (Bold text) and by criteria (Italic text) for building multiple regression. Note that variables are transformed. For variable abbreviation see* Table 1 *and* Table 2.

	Step	Variable	Action	"Sig Prob"	Seq SS	R^2	Cp	p	AICc	BIC
M1 Observ. 185	***1***	***Agriculture***	***Entered***	***0.000***	***6.165***	***0.243***	***112.2***	***2***	***112.1***	***121.68***
	2	***Water Body Area***	***Entered***	***0.000***	***2.732***	***0.351***	***72.5***	***3***	***85.9***	***98.52***
	3	***DSMS-Clay Cont***	***Entered***	***0.000***	***2.136***	***0.435***	***41.9***	***4***	***62.3***	***78.06***
	4	***Urban Area***	***Entered***	***0.001***	***0.938***	***0.472***	***29.6***	***5***	***51.9***	***70.77***
	5	***DSMS-Organic Cont***	***Entered***	***0.001***	***0.851***	***0.505***	***18.6***	***6***	***42.0***	***63.87***
	6	Ca_AL	Entered	0.026	0.347	0.519	15.3	7	39.0	63.90
	7	FAO-CL	Entered	0.035	0.303	0.531	12.7	8	36.5	64.47
	8	FAO-SCL	Entered	0.028	0.323	0.544	9.8	9	33.7	64.60
	9	Catch. Area	Entered	0.115	0.163	0.550	9.3	10	33.3	67.19
	10	SGU-A.Fill	Entered	0.128	0.152	0.556	8.9	11	33.1	69.94
	11	FAO-SC	Entered	0.071	0.211	0.564	7.7	12	31.9	71.65
	12	Open Land	Entered	0.054	0.237	0.574	6.1	13	30.3	72.87
	13	SGU-Gravel	Entered	0.272	0.076	0.577	6.9	14	31.3	76.78
	14	Water	Entered	0.296	0.069	0.579	7.9	15	32.5	80.80
	15	SGU-Thin Soil Layer	Entered	0.418	0.042	0.581	9.3	16	34.2	85.30
	16	Best	Specific	.	.	0.505	18.6	6	42.0	63.87

	Step	Variable	Action	"Sig Prob"	Seq SS	R^2	Cp	p	AICc	BIC
M2	***1***	***Agriculture***	***Entered***	***0.000***	***14.940***	***0.307***	***206.5***	***2***	***200.3***	***211.32***
Observ. 303	***2***	***SGU-Clay***	***Entered***	***0.000***	***5.444***	***0.419***	***126.7***	***3***	***148.9***	***163.60***
	3	***Water Body Area***	***Entered***	***0.000***	***2.821***	***0.477***	***86.4***	***4***	***119.1***	***137.42***
	4	***SGU-Organic Soil***	***Entered***	***0.000***	***2.253***	***0.524***	***54.6***	***5***	***93.0***	***114.99***
	5	**Water**	**Entered**	**0.003**	**0.674**	**0.538**	**46.5**	**6**	**86.1**	**111.75**
	6	**Forest**	**Entered**	**0.006**	**0.567**	**0.549**	**39.9**	**7**	**80.5**	**109.72**
	7	**Precip**	**Entered**	**0.002**	**0.698**	**0.564**	**31.5**	**8**	**72.8**	**105.63**
	8	**Catch. Area**	**Entered**	**0.007**	**0.518**	**0.574**	**25.7**	**9**	**67.5**	**103.85**
	9	**DSMS-Organic Cont**	**Entered**	**0.011**	**0.455**	**0.584**	**20.9**	**10**	**62.9**	**102.83**
	10	**DSMS-Clay Cont**	**Entered**	**0.003**	**0.603**	**0.596**	**13.8**	**11**	**55.9**	**99.39**
	11	Open Land	Entered	0.118	0.164	0.599	13.3	12	55.5	102.56
	12	FAO-SCL	Entered	0.101	0.180	0.603	12.6	13	54.9	105.46
	13	FAO-CL	Entered	0.121	0.160	0.606	12.2	14	54.6	108.64
	14	Lake Area	Entered	0.200	0.109	0.609	12.6	15	55.1	112.63
	15	Longitude	Entered	0.200	0.109	0.611	13.0	16	55.6	116.60
	16	Masl	Entered	0.335	0.062	0.612	14.0	17	56.9	121.33
	17	FAO-SC	Entered	0.492	0.031	0.613	15.6	18	58.7	126.54
	18	SGU-Thin Soil Layer	Entered	0.592	0.019	0.613	17.3	19	60.6	131.95
	19	SGU-Sand	Entered	0.614	0.017	0.614	19.0	20	62.7	137.39
	20	FAO-L	Entered	0.861	0.002	0.614	21.0	21	65.0	143.07
	21	Best	Specific	.	.	0.596	13.8	11	55.9	99.39
M3	**1**	**Open Land**	**Entered**	**0.000**	**5.511**	**0.227**	**174.6**	**2**	**101.3**	**111.07**
Observ. 199	***2***	***SGU-Clay***	***Entered***	***0.000***	***2.345***	***0.323***	***130.5***	***3***	***76.9***	***89.85***
	3	***SGU-Organic Soil***	***Entered***	***0.000***	***2.052***	***0.408***	***92.2***	***4***	***52.5***	***68.63***
	4	***Urban Area***	***Entered***	***0.000***	***1.751***	***0.480***	***59.8***	***5***	***28.8***	***48.13***
	5	***Water Body Area***	***Entered***	***0.000***	***1.161***	***0.527***	***38.9***	***6***	***11.8***	***34.27***
	6	***Agriculture***	***Entered***	***0.000***	***0.994***	***0.568***	***21.4***	***7***	***-4.0***	***21.55***
	7	**FAO-SCL**	**Entered**	**0.010**	**0.362**	**0.583**	**16.3**	**8**	**-8.8**	**19.86**
	8	**Water**	**Entered**	**0.018**	**0.295**	**0.595**	**12.5**	**9**	**-12.5**	**19.28**
	9	Mire	Entered	0.063	0.179	0.603	10.9	10	-13.9	20.92
	10	Longitude	Entered	0.043	0.210	0.611	8.8	11	-16.0	21.85
	11	SGU-Sand	Entered	0.081	0.153	0.618	7.8	12	-16.9	23.90
	12	FAO-SC	Entered	0.210	0.078	0.621	8.3	13	-16.3	27.51
	13	DSMS-Silt Cont	Entered	0.343	0.045	0.623	9.4	14	-14.9	31.83
	14	DSMS-Clay Cont	Entered	0.313	0.051	0.625	10.4	15	-13.7	36.02
	15	Masl	Entered	0.476	0.025	0.626	11.9	16	-11.8	40.76
	16	Lake Area	Entered	0.465	0.027	0.627	13.4	17	-10.0	45.47

Step	Variable	Action	"Sig Prob"	Seq SS	R^2	Cp	p	AICc	BIC
17	FAO-CL	Entered	0.647	0.011	0.627	15.2	18	-7.8	50.53
18	SGU- Thin Soil Layer	Entered	0.760	0.005	0.628	17.1	19	-5.4	55.72
19	Best	Specific	.	.	0.595	12.5	9	-12.5	19.28