Sanna Varanka

MULTISCALE INFLUENCE OF ENVIRONMENTAL FACTORS ON WATER QUALITY IN BOREAL RIVERS

APPLICATION OF SPATIAL-BASED STATISTICAL MODELLING
SANNA VARANKA

MULTISCALE INFLUENCE OF ENVIRONMENTAL FACTORS ON WATER QUALITY IN BOREAL RIVERS
Application of spatial-based statistical modelling

Academic dissertation to be presented with the assent of the Doctoral Training Committee of Technology and Natural Sciences of the University of Oulu for public defence in the OP auditorium (L10), Linnanmaa, on 15 January 2016, at 12 noon

UNIVERSITY OF OULU, OULU 2016
Rivers create unique habitat for aquatic life and provide ecosystem services for humans. Thus, degradation of river water quality is a serious, global problem. Water quality is the outcome of anthropogenic and natural landscape factors and the interaction of these two. To improve water quality, robust and quick methods are needed to study the complex, spatio-temporally dependent relation between water quality and environment conditions across extensive areas.

This thesis aimed to study the relationship between water quality (total phosphorus and nitrogen, pH, water colour and dissolved oxygen) and environmental factors in boreal rivers combining grid-based data and statistical methods. The study comprised of 34 Finnish rivers with their catchments. First, the effect of natural and human-induced environmental factors on water quality was studied. Then, (a) the ability of the characteristics of different spatial scales around the river channel and under different discharge conditions to predict water quality was explored and (b) the suitability of the applied statistical methods (generalized linear and additive models, partitioning methods, non-metric multidimensional scaling) in water quality studies was evaluated.

As expected, the results highlighted the impact of agricultural activities on water quality as nutrients and pH increased, together with the cover of agricultural activities. However, when studied as a group, natural factors explained water quality better than land use/cover. Lakes were strongly related to decreased nutrients and water colour. The effect of fine-grained soils on nutrients and pH was positive. In the scale studies, nutrients and water colour were best explained by the characteristics of the entire catchment but pH was mostly predicted by the characteristics of the 50 m riparian zone. The connection between water quality and environment was strongest during high-flow discharge periods.

The results encourage the use of the applied methods, showing that the combination of grid-based data and advanced statistical methods provide an efficient first-filter estimate of water quality-environment relations. Spatial-based statistical modelling provides a crucial framework for river, water resources and land use management. The applied methods can also be seen as essential tools when predicting the impacts of global change on water quality.

**Keywords:** boreal region, catchment, catchment physiography, climate, GIS, hydrogeography, land use/cover, river, spatial scale, statistical modelling, temporal scale, water quality
Varanka, Sanna, Ympäristötekijöiden vaikutus boreaalisten jokien vedenlaatuun eri mittakaavoissa. Paikkatietoon perustuvan tilastollisen mallinnuksen soveltaminen
Oulun yliopiston tutkijakoulu; Oulun yliopisto, Luonnontieteellinen tiedekunta
Oulun yliopisto, PL 8000, 90014 Oulun yliopisto

Tiivistelmä

Tässä väitöskirjassa oli tavoitteena tutkia vedenlaadun (kokonaisfosfori ja -typpi, pH, värikuvoja ja liukoverkkoja) ja ympäristön yhteyttä boreaalissa vyöhykkeellä käyttäen paikkatietoaineistoa ja tilastollisia menetelmiä. Tutkimusalueena oli 34 suomalaista jokatuotea-alue. Ensinnäkin tutkittiin luonnollisten ympäristötekijöiden ja ihmistoiminnan vaikutusta vedenlaatuun. Tavoitteena oli myös selvittää, miten jokien ympärillä olevien eriakoisien vyöhykkeiden ominaisuuksien ja vaihteleva virtaamaluku selittävät vedenlaatuun. Lopuksi arvioitiin käytettyjen tilastollisten menetelmien (yleistetyt lineaariset ja additiiviset mallit, hajonnan osittaminen, ordinaatioanalyysi) soveltuvuutta vedenlaatututkimuksissa.


Asiasanat: boreaalinen alue, GIS, hydrogeografia, ilmasto, joki, maankäyttöÖ-peite, spatioaalinen mittakaava, temporaalinen mittakaava, tilastollinen mallinnus, valuma-alue, valuma-alueen yhteys, vedenlaatu
To my family.
Acknowledgements

This has been a rather long journey, one which would not have been possible without the help and support of many people. First of all, I extend my gratitude to my supervisors and co-authors Jan Hjort and Miska Luoto. Your ideas, help, support and comments have been invaluable and have guided me throughout the different steps of my thesis. I would also like to give my warmest thanks to Jarno Rusanen for his support during these years. My thesis was financially supported by the Department of Geography, University of Oulu. In addition, the financial support provided by the University of Oulu Graduate School (UniOGS) in the spring of 2015 was particularly crucial because it enabled me to concentrate fully on the synopsis.

I am grateful to my pre-examiners, Dr. Petri Ekholm and Dr. James John Rothwell for their constructive and encouraging comments, which helped me to finalise the synopsis. I would also like to thank Paula Kuusisto-Hjort for her invaluable comments and Aaron Bergdahl for language editing.

It has been a privilege to work with so many friendly people during my postgraduate studies at the Department of Geography. I express my warmest feelings to all of you. I am thankful for the former and the present members of the Physical geography research group for your support and companionship. Janne Alahuhta, Terhi Ala-Hulkko, Marja Anttonen, Mathieu Marmion, Henna Sormunen and Helena Tukiainen are especially acknowledged. Finally, I wish to thank all the other people that I have had the privilege to meet and work with during these years.

I am thankful to my friends and relatives for their support. In addition, I owe a number of scratches and titbits to my dear cat, Hessu, who usually slept next to my computer while I was working with the synopsis. This made the work so much less stressful. Above all, I am grateful to my wonderful parents, Tuula and Esko and to my sister, Johanna. You have supported my life and studies since the beginning. You have always have belief in me, even when I have lost mine. I can never thank you enough.

Ylivieska, 8 November, 2015

Sanna Varanka
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>Akaike information criterion</td>
</tr>
<tr>
<td>BIC</td>
<td>Bayesian information criterion</td>
</tr>
<tr>
<td>DEM</td>
<td>Digital elevation model</td>
</tr>
<tr>
<td>SYKE</td>
<td>Finnish Environment Institute</td>
</tr>
<tr>
<td>FMI</td>
<td>Finnish Meteorological Institute</td>
</tr>
<tr>
<td>GAM</td>
<td>Generalized additive model</td>
</tr>
<tr>
<td>GIS</td>
<td>Geographic information system</td>
</tr>
<tr>
<td>GLM</td>
<td>Generalized linear model</td>
</tr>
<tr>
<td>HP</td>
<td>Hierarchical partitioning</td>
</tr>
<tr>
<td>NDVI</td>
<td>Normalized difference vegetation index</td>
</tr>
<tr>
<td>NMDS</td>
<td>Non-metric multidimensional scaling</td>
</tr>
<tr>
<td>SAC</td>
<td>Spatial autocorrelation</td>
</tr>
<tr>
<td>VIF</td>
<td>Variance inflation factor</td>
</tr>
<tr>
<td>WFD</td>
<td>Water Framework Directive</td>
</tr>
<tr>
<td>VP</td>
<td>Variation partitioning</td>
</tr>
</tbody>
</table>
List of original articles

This thesis is based on the following articles, which are referred throughout the text by their Roman numerals:


Author’s contributions

I Sanna Varanka (SV) and Miska Luoto (ML) were responsible for the study idea and study design. SV processed the data, performed the statistical analyses and interpreted the results. SV was responsible for preparing the manuscript, while ML commented and contributed.

II ML was responsible for the study idea and SV together with ML planned the study design. SV processed the data, conducted the statistical analyses and interpreted the results. SV was responsible for preparing the manuscript, while ML and Jan Hjort (JH) commented and contributed.

III SV came up with the study idea and SV, jointly with, JH planned the study. SV processed the data, executed the statistical analyses and interpreted the results. SV was responsible for preparing the manuscript, while JH commented and contributed.
# Table of contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>9</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>11</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>13</td>
</tr>
<tr>
<td>List of original articles</td>
<td>15</td>
</tr>
<tr>
<td>Table of contents</td>
<td>15</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>17</td>
</tr>
<tr>
<td>2 Aims of the study</td>
<td>19</td>
</tr>
<tr>
<td>3 Study background</td>
<td>21</td>
</tr>
<tr>
<td>4 Study area</td>
<td>33</td>
</tr>
<tr>
<td>5 Material and methods</td>
<td>37</td>
</tr>
<tr>
<td>6 Results and Discussion</td>
<td>43</td>
</tr>
<tr>
<td>3.1 River and catchment area</td>
<td>21</td>
</tr>
<tr>
<td>3.2 Physico-chemical indicators of water quality</td>
<td>22</td>
</tr>
<tr>
<td>3.2.1 Phosphorus and nitrogen</td>
<td>23</td>
</tr>
<tr>
<td>3.2.2 pH</td>
<td>24</td>
</tr>
<tr>
<td>3.2.3 Water colour</td>
<td>25</td>
</tr>
<tr>
<td>3.2.4 Oxygen</td>
<td>26</td>
</tr>
<tr>
<td>3.3 Factors affecting water quality</td>
<td>26</td>
</tr>
<tr>
<td>3.3.1 Natural and anthropogenic factors</td>
<td>26</td>
</tr>
<tr>
<td>3.3.2 Spatial and temporal variability</td>
<td>29</td>
</tr>
<tr>
<td>3.4 GIS and statistical modelling</td>
<td>30</td>
</tr>
<tr>
<td>5.1 Water quality variables</td>
<td>37</td>
</tr>
<tr>
<td>5.2 Environmental variables</td>
<td>38</td>
</tr>
<tr>
<td>5.3 Spatiality and temporality</td>
<td>39</td>
</tr>
<tr>
<td>5.4 Statistical methods</td>
<td>39</td>
</tr>
<tr>
<td>6.1 Environmental variables in explaining variation in water quality</td>
<td>43</td>
</tr>
<tr>
<td>6.1.1 Specific environmental variables</td>
<td>43</td>
</tr>
<tr>
<td>6.1.2 Environmental variable groups</td>
<td>44</td>
</tr>
<tr>
<td>6.1.3 Geomorphological variables</td>
<td>47</td>
</tr>
<tr>
<td>6.2 Spatial scales and temporal periods in studying environment- water quality relationship</td>
<td>49</td>
</tr>
<tr>
<td>6.2.1 Spatial scales</td>
<td>49</td>
</tr>
<tr>
<td>6.2.2 Temporal periods</td>
<td>51</td>
</tr>
<tr>
<td>6.3 Spatial-based statistical modelling in water quality studies</td>
<td>52</td>
</tr>
</tbody>
</table>
1 Introduction

Fluvial ecosystems are significant with rich, varied biota as running water generates unique environment and habitat for aquatic biota (Allan & Castillo 2007). Rivers and streams also provide important utilities and ecosystem services for humans (de Groot et al. 2010). As a consequence, rivers and streams are possibly one of the most exploited ecosystems by human action (Malmqvist & Rundle 2002). Eutrophication, pollution and global change impact freshwaters. The effects of ongoing climate change is expected to impact rivers through several processes such as changing land use and land cover (hereafter land use) as well as the hydrological regime (Meybeck 2003; Palmer et al. 2008; Whitehead et al. 2009a). As a result, water quality is expected to degrade (e.g. Whitehead et al. 2009a; Henriques et al. 2015) and freshwater biodiversity is expected to decrease (Heino et al. 2009). These are threats not only to the diversity of nature but to humans as well. For example, river water is used for drinking worldwide and thus degrading water quality is a serious problem.

Concern about and studies of freshwater ecosystems, their condition and water quality have increased during the last decades (e.g. Meybeck & Helmer 1989; Richards et al. 1996; Sponseller et al. 2001; Chang 2008; Lindell et al. 2010; Withers et al. 2014; Bowes et al. 2015). The Federal Water Pollution Control Act (commonly referred as Clean Water Act) in the United States (Public Law 95–217 1977) and the Water Framework Directive (WFD) in Europe (European Commission 2000) are examples of legislative instruments aiming to restore and rehabilitate water resources. The WFD was established in 2000 aiming to achieve good surface water status within 15 years in the European Union (EU) member states (European Commission 2000). However, it is likely that this will not be achieved within the original timetable (e.g. Allan 2012; Richter et al. 2013; Smith et al. 2014). Therefore, the member states will face a considerable task in the future that includes protecting, restoring and enhancing of surface waters (European Commission 2000). These actions, as well as the expected causes of climate change, will require increasing levels of knowledge about the factors and their interactions influencing freshwater ecosystems together with cost-efficient modelling techniques. These would help to identify, for instance, the factors that actually affect water quality naturally and anthropogenically, which is essential in protecting and restoring water resources and ecosystems.

The relationship between environmental conditions and water quality has been strongly related to the characteristics of the entire catchment (e.g. Nielsen et al.
and a riparian zone (e.g. Chang 2008). In addition, the water quality-
vironment connection is considered to be pronounced during high-flow periods
(e.g. Zhang et al. 2014). Land use (e.g. Carroll et al. 2013), and especially
agricultural activities (e.g. Evans et al. 2014), are important factors affecting water
quality. In addition, natural environment factors such as soil (e.g. Ågren & Löfgren
2012) and bedrock (e.g. Brown et al. 2011) properties are also essential drivers of
water quality. However, studies that simultaneously consider water quality
determinants from different environmental variable groups, like land use, geology
and climate, are scarce (Jarvie et al. 2002). Similarly, there are few studies based
on water bodies and their catchments across extensive environmental gradients
(Stendera & Johnson 2006; Nielsen et al. 2012). Instead, studies are usually local
or regionally based on a few sub-catchments (e.g. Meyendonckx et al. 2006;
Gonzales-Inca et al. 2015) or main catchments (e.g. Young et al. 2005; Lindell et
al. 2010). In addition, multicollinearity, correlation between environmental
variables (Dormann et al. 2013), is a common methodological problem
encountered in environmental studies, and it can lead to spurious results in water
quality studies. Therefore, there is a need for robust approaches for the analyses
and predictions of the water quality-environment relationship across extensive
areas.

Hydrological cycles in boreal landscapes are strongly controlled by cold
climatic conditions and seasonal variations in climate. Snowfall, frozen soils and
spring floods are essential factors affecting hydrological regime (Valtanen et al.
2014), and associated water quality in boreal rivers. The spatial scale in which
water quality determinants are studied is also important as environmental
characteristics and therefore processes controlling water quality vary with scale.
For instance, boreal rivers are commonly surrounded by agriculture, whereas
forests are located further from the river channel. Overall, the water quality-
environment relationship in boreal catchments is complex and requires reliable and
cost-efficient research methods. This is increasingly important as the effects of
climate change in these high-latitude areas are expected to be substantial (IPCC
2015) through changes in the hydrological regime causing, together with land use,
water quality degradation (IPCC 2014). However, good availability of
environmental data, geographic information system (GIS) and complementary
statistical methods are expected to provide new perspectives and valuable insights
into water quality studies, complementing local-scale process-based explorations.
2 Aims of the study

The main study aim of this thesis was to investigate the relationships between water quality and environmental factors in boreal rivers combining GIS data and analyses with statistical methods, which are novel in water quality studies. First, the goal was to determine the most important environmental variables explaining variation in water quality in boreal rivers. Especially, the role of agricultural activities, lake percentage and soil properties in a catchment area affecting water quality was explored. Second, the relationship between water quality and environmental variables from several groups simultaneously (land use, physiography and climate) was modelled in order to find out the most important variable group determining variation in water quality. Third, the possibility of explaining water quality using solely geomorphological variables was studied. Fourth, there was interest in the importance of different spatial scales, especially the entire catchment, explaining water quality. Fifth, the possible differences between water quality-environment relationships during high-flow and low-flow discharge periods were explored. Finally, the goal was to evaluate the suitability of advanced statistical methods when modelling the relationship between water quality and environmental factors across extensive areas.

More precisely, the following hypotheses were explored:

H1: Agriculture, lake percentage and clayish soil are the most important environmental variables explaining water quality in boreal rivers.

H2: Land use is the most influential environmental variable group in determining water quality in boreal rivers.

H3: Geomorphological variables can be used to predict water quality in boreal rivers at a catchment scale.

H4: Environmental characteristics at the entire catchment scale predict water quality best in boreal rivers.

H5: The water quality-environment relationship is strongest during high-flow periods.

H6: Statistical modelling methods combined with GIS data provide new perspectives in the catchment-scale water quality studies across extensive areas.
3 Study background

3.1 River and catchment area

According to the Finnish Water Act (587/2011), a river is defined as flowing water, having a catchment area of at least a 100 km². According to Wetzel (2001) a catchment is an area of land that is drained by tributary streams merging into a main channel, river. Catchment area can also be defined as a topographically determined region where precipitation falls before draining into a river. Tributary streams and catchments are nested in hierarchically. Therefore, low-order streams have their own catchment within a larger one. (Allan & Castillo 2007). A common technique in stream ordering was proposed by Horton (1945) and modified by Strahler (1957). Although there are overlapping terms in use, drainage basin and watershed area are both used to refer to a catchment area (Wetzel 2001).

Rivers and their catchments are inextricably and reciprocally connected. River biota (e.g. Richards et al. 1996; Sponseller et al. 2001) and water quality (e.g. Stendera & Johnson 2006) are strongly influenced by the surrounding landscape and catchment characteristics. Dissolved substances and suspended material that enter rivers are derived mainly from their catchment area (Wetzel 2001) and therefore this area has a major impact on an entire stream ecosystem on multiple scales (Allan 2004). On the other hand, rivers affect their environment shaping it through erosion, transportation and accumulation.

The linkage between rivers and catchments has long been recognised (e.g. Hynes 1975; Vannote et al. 1980). The river continuum concept (RCC) by Vannote et al. (1980) acknowledges the role of allochthonous organic matter contributions to instream ecosystems. It has been an important conceptual framework for river ecologists, but it is considered as a rather a simplistic view of the patterns of connectedness and variations in water and material flows in rivers (Fausch et al. 2002; Wiens 2002). Nowadays, rivers are commonly seen as internally heterogeneous landscapes, riverscapes, which are strongly connected to the surrounding landscape (Wiens 2002; Allan 2004). More specifically, rivers are seen as ecological representations and combinations of broad-scale variations in energy, matter, and habitat structure in addition to local discontinuous zones and patches (Carbonneau et al. 2012). The concept of a riverscape was introduced by Leopold & Marchand (1968) and its use has been focused in the 2000s (e.g. Fausch et al. 2002; Wiens 2002; Allan 2004; Carbonneau et al. 2012; Roux et al. 2015). The
need to explore riverscapes and their processes has produced the rapid development of techniques and methods like the fluvial information system (FIS) (Carbonneau et al. 2012) and GIS applications (Roux et al. 2015).

3.2 Physico-chemical indicators of water quality

The spatial and temporal composition of water is the outcome of numerous physical, chemical and biological components and water quality varies together with these. Water quality can be studied through countless indicators depending on the purpose, each of which reveal something about the state of a freshwater ecosystem. For example, the amount of phosphorus (e.g. Withers & Jarvie 2008), nitrogen (e.g. Vitousek et al. 1997) and oxygen (Abdul-Aziz & Ishtiaq 2014) in water, the pH value (Rothwell et al. 2010a; 2010b) and alkalinity (e.g. Stets et al. 2014) are key indicators when exploring the state of an ecosystem. Water colour (Vinogradoff & Oliver 2015) and turbidity (Galbraith & Burns 2007) are visual indicators of water quality and are related to material drifting or floating in water. Investigating the composition and concentration of dissolved major ions in water is also important as those have been related both to natural and anthropogenic catchment factors such as land use (Jarvie et al. 2002; Lindell et al. 2010). In addition, electrical conductance of water (Lecomte et al. 2009) is an approximate measure of total dissolved ions (Allan & Castillo 2007). As metals in water can be toxic to the aquatic biota (Malmqvist & Rundle 2002), it is important to explore their concentrations in water as well. If water is used for drinking (Capral 2010) or recreational purposes such as swimming (Wade et al. 2008), it should be analysed especially in the case of faecal indicator organisms.

Good surface water status in the WFD is specified in terms of good ecological and chemical status. Therefore, the use of biological elements, such as the composition and abundance of aquatic flora, benthic invertebrates and fish, are essential in defining the state of freshwater ecosystems (European Commission 2000). However, as suitable physico-chemical water quality supports biological elements, these indicators give an important general view of the state of freshwater ecosystems. Therefore, nutrients, pH, water colour and dissolved oxygen, were selected to define water quality and these indicators will be considered below.
3.2.1 Phosphorus and nitrogen

Phosphorus (P) and nitrogen (N) are essential nutrients for sustaining life defining mostly the biological productivity in ecosystems (e.g. Elser et al. 2007). Therefore, nutrients are important indicators of productivity in, and the ecological status of (Aroviita et al. 2012), river ecosystems. However, over-enrichment of nutrients often results eutrophication in surface waters (e.g. Carpenter et al. 1998; Dodds 2006; Withers et al. 2014). The amount of nutrients entering rivers varies both spatially and temporally due to the combination of a catchment’s characteristics like bedrock, hydrological regime and anthropogenic impact (Withers & Jarvie 2008; Bechmann 2014) (Fig. 1). All possible forms of nitrogen and phosphorus are analysed to get the amount of total nitrogen and total phosphorus in a sample. In Finland, total phosphorus and nitrogen concentrations are usually highest in rivers flowing in the southern and western coasts (Räike et al. 2003; Niemi & Raateland 2007).

Phosphorus in running water occurs in either inorganic or organic form and the most significant form of inorganic phosphorus is orthophosphate (PO$_{4}^{3-}$). In addition, phosphorus is either dissolved in water or attached to particulate material like detritus. (Wetzel 2001; Allan & Castillo 2007). In many rivers, phosphorus is often the nutrient in shortest supply (Bowes et al. 2003) and an increase in phosphorus is often followed by eutrophication (Gelbrecht et al. 2005; Jordan et al. 2005). Natural sources of phosphorus include catchment geology and the geochemistry of river sediments, whereas agriculture is a major anthropogenic source (Withers & Jarvie 2008) (Fig. 1). In the cycling of phosphorus, it is retained by organic material and then subsequently remobilised by decomposition (Allan & Castillo 2007). In addition, phosphorus usually sinks to the bottom of freshwater and attaches to the sediments (Hejzlar et al. 2009). Phosphorus is released from sediments owing to bioturbation and anaerobic conditions, which can cause internal loading especially in lakes and ponds (Wetzel 2001; Allan & Castillo 2007).

Nitrogen gas (N$_{2}$) is abundant in the atmosphere but it is largely inaccessible to most organisms in this form (Vitousek et al. 1997). In freshwaters nitrogen occurs, for example, as dissolved inorganic nitrogen like nitrate (NO$_{3}^{-}$), nitrite (NO$_{2}^{-}$) and ammonium (NH$_{4}^{+}$) and as dissolved or particulate organic nitrogen (Allan & Castillo 2007). Nitrogen in its many forms enters surface waters from atmospheric deposition, nitrogen fixation and terrestrial inputs (Vitousek et al. 1997) (Fig. 1). Denitrification, the microbial production of N$_{2}$, is a key process by which nitrogen is removed from surface waters (Seitzinger et al. 2006; Hejzlar et
Further losses of nitrogen occur through sedimentation and biological uptake (Hejzlar et al. 2009). As organic materials decompose, nitrogen is remobilised and therefore available for further uptake and assimilation (Allan & Castillo 2007). A significant anthropogenic source of nitrogen is agriculture (e.g. Lepistö et al. 2006; Nielsen et al. 2012; Bechmann 2014).

Fig. 1. Simplified theoretical framework about the factors affecting water quality. Arrows describe connections between environmental factors or processes and their relations with the water quality variables. Arrows with a broken line illustrate the output processes of nutrients. Arrows, which point out directly to a specific water quality variable describe relationship between this water quality variable and the connected factor. Arrows, which point out to the box including all water quality variables illustrate that all of them are affected by this factor or process. The lines between the water quality variables describe the close connection between them. \(^1\) the referred variable, factor or perspective was included in the study.

### 3.2.2 pH

pH is a measure of acidity. It is commonly defined as the negative logarithm of the hydrogen ion (H\(^+\)) activity in a given solution (Wetzel 2001). In acid waters, carbon dioxide (CO\(_2\)) and carbonic acid (H\(_2\)CO\(_3\)) dominate and at the higher pH values, bicarbonate (HCO\(_3^-\)) and carbonate (CO\(_3^{2-}\)) are present (Allan & Castillo 2007). Therefore, pH is closely connected to alkalinity, a measure of the acid buffering
components in a solution (Neal et al. 2002). Acidity is often followed by releasing of metals from the sediments, which has deleterious effects on aquatic biota causing, for example, large fish kills (Toivonen & Österholm 2011). Diurnal changes in pH are regulated by the CO\textsubscript{2} level in water, which is consumed by photosynthesis during the day and increased by respiration at night (Fig. 1) (Neal et al. 2002). Therefore, the pH value is usually increases during the day and in summer. On the other hand, microbial activity is greatest during summer causing decomposition of organic matter within the bed sediments, which leads to the elevated production of CO\textsubscript{2} (Neal et al. 2002).

Waters can be naturally acidic but human impact has exacerbated acidification, for example, through atmospheric deposition caused by sulphur dioxide (SO\textsubscript{2}) and nitrogen (NO\textsubscript{x}) emissions (Meybeck 2003; Stets et al. 2014). In Finland, natural acidity can often be explained by acid sulphate soils, which are common in western coastal areas (Toivonen & Österholm 2011). The lowest pH values in Finland are measured in rivers flowing through these soils together with the areas covered by peatlands (Niemi and Raateland 2007). In addition, acidity has been connected to increased discharge (Saarinen et al. 2010) and runoff (Toivonen et al. 2013). Connections between bedrock composition and the pH value of surface waters have also been discovered (e.g. Young et al 2005; Brown et al. 2011) (Fig. 1). Like nutrients, pH is important when evaluating the ecological status of river ecosystems (Aroviita et al. 2012).

### 3.2.3 Water colour

The colour of water is the result of light being scattered from the water. Scattering of short wavelengths, such as blue, is greater compared to longer wavelengths, such as red, which are mostly absorbed, and therefore clear water often looks bluish. (Wetzel 2001). Observing the colours can be subjective and therefore, it has been standardised by comparing a water sample with a known concentration of coloured solution (Niemi & Raateland 2007). The platinum-cobalt method is commonly used and clear water is equal to 0 Pt (Wetzel 2001). The colour of water is affected by dissolved and suspended material (Niemi & Raateland 2007). In addition to humus, the end product of decomposed organic matter, the water colour index (CO) has been connected to wastewaters (Niemi & Raateland 2007), carbon-rich soils, peatlands and wetlands (Vinogradoff & Oliver 2015) (Fig. 1). As the proportion of peatlands and the production of humus in Finland is considerable, water colour is an important indicator of water quality. According to Niemi and Raateland (2007),
water colour value is typically lowest in rivers in northern Finland and highest in western coastal areas and in the south.

3.2.4 Oxygen

Oxygen (O₂) dissolved in water indicates the general health of aquatic ecosystems (e.g. Abdul-Aziz & Ishtiaq 2014) and aquatic organism are dependent on oxygen (Harvey et al. 2011). Moreover, dissolved oxygen is a key factor determining water quality (Abdul-Aziz & Ishtiaq 2014). The primary oxygen inputs from atmosphere and photosynthesis and the outputs through respiration and decomposition of organic material defines the concentration of oxygen in water (Best et al. 2007; Harvey et al. 2011) (Fig. 1). In addition, dissolution of oxygen decreases together with increasing water temperature and salinity (Best et al. 2007).

The cycle of dissolved oxygen varies both spatially and temporally because it is subject to variations in hydro-climatic and biochemical factors such as discharge and organic waste (Abdul-Aziz & Ishtiaq 2014). Essential causes for diurnal variations in oxygen concentrations are photosynthesis during the day and respiration during the night (e.g. Harvey et al. 2011) (Fig. 1). In addition, oxygen is an important factor modulating the nutrient cycle in water (e.g. Seitzinger et al. 2006; Harris et al. 2015). For instance, eutrophication often causes anaerobic conditions. This is followed by the release of phosphorus from the bottom sediments, subsequently leading to internal loading especially in lakes (e.g. Spears et al. 2007). In the study by Niemi & Raateland (2007), oxygen concentration was relatively high in all studied rivers. This refers, that problems with low oxygen concentrations in Finland are only local. However, Niemi & Raateland (2007) also concluded that there is a tendency towards higher concentrations in northern rivers, which also was in accordance with generally better water quality in these rivers.

3.3 Factors affecting water quality

3.3.1 Natural and anthropogenic factors

Water quality varies naturally because of natural catchment characteristics. For example, the type of bedrock (e.g. Holloway et al. 1998; Brown et al. 2011), soil (e.g. Sliva & Williams 2001; Ågren & Löfgren 2012) and vegetation (Malmqvist & Rundle 2002) influence the composition of water (Fig. 1). This is seen through
several water quality indicators including nutrients and the pH value. Lake area in catchments has been related to reduced nutrient loads entering rivers because of retention mechanisms (e.g. Malve et al. 2012; Stålnacke et al. 2015) such as sorption to sediments or to organic matter (Hejzlar et al. 2009). The relationship between different water quality indicators and the relief (i.e. slope) of catchments has been also indicated (e.g. Chang 2008). This relation can be explained, for example, by water residence time (McGuire et al. 2005) and flow regime (Lecomte et al. 2009) as greater flow rates due to steep slopes can increase erosion (Chiverton et al. 2015) which in turn increase the rates of particulate matter entering rivers (Sliva & Williams 2001). Biochemical processes such as denitrification in aquatic and terrestrial ecosystems influence water quality through the amount of nitrogen in water (Seitzinger et al. 2006). In addition, the interaction between surface waters and groundwater is important to consider as groundwater can help to maintain good surface water quality (Lerner & Harris 2009) but it can also cause contamination of surface waters (Rozemeijer & Broers 2007). Overall, water quality is naturally controlled by many biological, physical, and chemical processes.

Although rivers are affected by natural environment conditions, anthropogenic pressure on rivers is significant (Fig. 1). Rivers attract settlement as they offer utilities for humans. As a consequence, abstraction of water, pollution, energy production, river channelisation and damming affect the entire river ecosystems by altering water quality and the hydrology of rivers and changing river habitat (Fig. 2a) (Malmqvist & Rundle 2002). Atmospheric deposition of nitrogen (Vitousek et al. 1997) and metals like mercury (Hg) and lead (Pb) (Pacyna et al. 2009), as an example, are emitted both from natural and anthropogenic sources. Industrial emissions are a major anthropogenic input of atmospheric pollution and deposition, which has led to acid rain and acidification of surface waters especially in areas with low buffering capacity (Meybeck 2003). In addition, ongoing climate change is expected to generally deteriorate water quality as it causes changes in land use (Henriques et al. 2015) and hydrological regime (Arnell et al. 2015). Changes in discharge can alter the mobility and dilution of contaminants. For example, according to Whitehead et al. (2009b) lower flow rates can result in less dilution and therefore in higher concentrations of contaminants downstream from point source pollutants. Climate change is also expected to increase water temperature. This would affect chemical reaction kinetics as most chemical and microbiological reactions run faster at higher temperatures, which affects water quality (Whitehead et al. 2009a).
Fig. 2. Examples of typical landscapes that rivers flow through in Finland illustrating environmental factors affecting water quality. a) Urban landscape in Oulu around the Oulujoki river, which has been strongly modified and dammed for energy production (Wikimedia Commons, Estormiz). Rural landscape is common for rivers on the western coast of Finland. b) Fields around the Siikajoki river. c) Agriculture close to the Kalajoki river. d) Vaaditkoski rapid in the Merikarvianjoki river is surrounded by fields and forests (Wikimedia Commons, pjg).

Land use is a major human-induced and modified landscape feature affecting water quality (e.g. Herlihy et al. 1998; Zampella et al. 2007; Carroll et al. 2013). Agricultural activities (e.g. Evans et al. 2014), urban land use (e.g. Pratt & Chang 2012), forests (e.g. Singh & Mishra 2014) and peat bogs (e.g. Vassiljev & Blinova 2012) affect water quality (Fig. 1 & 2). Land use modifications like drainage ditching (Ecke 2009) and clear-cutting in forests can have significant negative impacts on water quality (Palviainen et al. 2015). Non-point sources of anthropogenic inputs are commonly linked with agriculture, which has especially been connected to increased nutrient concentrations (e.g. Lepistö et al. 2006; Evans...
et al. 2014), which leads to eutrophication (e.g. Withers et al. 2014). In addition, agricultural activities, like tillage and grazing (Knox 2001), and the increased cover of impervious surfaces due to urbanisation reduce infiltration and increase surface runoff (Paul & Meyer 2001). Therefore, anthropogenic pressure on water quality is both direct, such as nutrient enrichment, and indirect, such as changed natural processes and hydrological regimes (Fig. 1).

3.3.2 Spatial and temporal variability

Surface water quality varies spatially and temporally (Miller et al. 2014) together with the processes affecting water quality (Fig. 1). Environmental conditions and pollution sources vary through a river. For example, industrial wastewaters are typical point sources of anthropogenic inputs (Withers & Jarvie 2008) degrading water quality. However, agricultural activities are typically located along the river channel (Fig. 2b–d) causing significant non-point pollution to rivers. Landscape factors such as land use patterns have the tendency to be dependent (Overmars et al. 2003) due to geographic proximity (Legendre & Legendre 1998). The distribution of different land use patterns is crucial to consider as different factors can affect water quality at different scales. As processes behind water quality vary, conclusions about the most important spatial scale in relation to water quality have varied as well. The variation in different water quality indicators has been explained the most by the characteristics of the entire catchment (e.g. Sliva & Williams 2001; Nielsen et al. 2012). However, Chang (2008) and Roberts & Prince (2010) concluded that the areas closest to the river channel largely explain water quality. On the other hand, in the study of Meyendonckx et al. (2006) environmental conditions near the river were not critical factors in water quality modelling.

Riparian areas are important spatial feature in relation to water quality. They are considered as a transition zone between terrestrial and aquatic ecosystems (Luke et al. 2007; Soininen et al. 2015) and they have a significant role in nutrient and energy flux between these two systems (Naiman & Décamps 1997; McClain et al. 2003). The typically recommended riparian zone width varies between six metres based on stream shading (DeWalle 2010) and 100 m based on processes that need protection (Allan et al. 1997). Processes in the riparian zone are essential in relation to water quality (e.g. Dosskey et al. 2010) as water to rivers flows through this area just before entering the river. For example, riparian vegetation reduces the amount of nutrients entering rivers (e.g. Vought et al. 1995; Sahu & Gu 2009; Mayer et al. 2010) by direct chemical uptake in root zones (Dosskey et al. 2010).
Riparian vegetation also increases infiltration and stabilises streambanks and soil, which reduces soil erosion and the loading pressure on rivers (Dosskey et al. 2010). Therefore, riparian zone management has become an important part in watershed (Allan et al. 1997) and river management (Gumiero et al. 2013). However, depending on the conditions in the riparian area, it can also influence water quality negatively. Therefore, it is important to know if areas close to the river channels determine water quality.

Seasonal variability in water quality is the result of interactions between many processes caused by seasonal variations in climate (Fig. 1). Surface runoff has been connected to variations in water quality, for example, through nutrients (Withers & Jarvie 2008; Malve et al. 2012) and acidity (Toivonen et al. 2013). In addition, high-flow and low-flow discharge periods are reflected in water quality (Woli et al. 2008; Carrol et al. 2013; Zhang et al. 2014). Surface runoff, or overland flow, is generated when infiltration is limited by low soil permeability or its saturation causing water to flow over the landscape surfaces increasing discharge in the receiving rivers (Winter 2001; Dosskey et al. 2010). As surface runoff can cause soil erosion and the delivery of eroded sediments and contaminants into rivers, it can also lead to water pollution. The rates of runoff and discharge are dependent on the rate of precipitation and evapotranspiration (Winter 2001), which change throughout the year. In addition, soil hydraulic and hydrological processes are not constant throughout the year (Bormann & Klaassen 2008) and the infiltration capacity of soil varies according to the season and soil characteristics (Horton 1945). The ability of vegetation to uptake nutrients and retain water, and therefore to reduce surface runoff, changes throughout the year as well (Bonan & Shugart 1989). Changes in these processes are highlighted in cold climate areas. Winter time in these areas, with freezing soil and surface waters as well as precipitation in the form of snow, alters hydrological processes significantly (Korhonen 2007; Valtanen et al. 2014).

3.4 GIS and statistical modelling

GIS provides a powerful modelling environment for analysing a wide range of spatial phenomena and data. It is also a very useful and common tool in water quality studies (e.g. Osborne & Wiley 1988; Herlihy et al. 1998; Smart et al. 2001; Jarvie et al. 2002; Meyndonckx et al. 2006; Rothwell et al. 2010b; Tu 2013; Gonzales-Inca et al. 2015). By using digital maps and GIS data it is possible to determine, for example, the composition of land use in a specific spatial unit,
commonly in a catchment area. In addition, it enables the calculation of different derivative variables such as topographical indices (e.g. slope angle) from digital elevation models (DEMs) (cf. Alahuhta et al. 2011a; Pratt & Chang 2012) and remote sensing-based vegetation indices such as the normalized difference vegetation index (NDVI) (Soininen & Luoto 2012). Consequently, GIS data offers an efficient framework for analysing water quality-environment relations at different scales, especially when combined with statistical modelling.

Statistical modelling has many benefits when analysing environmental phenomena. First of all, it increases the objectivity of the studies as hypotheses can be tested quantitatively. In addition, statistical modelling enables simplification of complex systems (Hjort & Luoto 2013) and therefore provides understanding about the environmental systems and processes. This is a great advantage in water quality studies as water quality is the outcome of complex environmental processes and factors. One benefit in statistical modelling is that it can be also used for predicting how climate change will affect environmental phenomena (cf. Fronzek et al. 2010; Alahuhta et al. 2011b). Moreover, catchments can be extensive and therefore not completely achievable for field studies. However, the combination of multivariate statistical techniques and efficient data procurement and management techniques enables the investigation of extensive and remote areas (e.g. Guzzetti et al. 1999) and water resources. Using statistical modelling it is also possible to study phenomena across scales from local to global (Hjort & Luoto 2013), which is a great advantage as the determinants of water quality can vary according to scales.

Statistical modelling has been used in exploring factors affecting water quality. For instance, simple linear regression (Woli et al. 2008; Evans et al. 2014) and multiple regression analysis (Osborne & Wiley 1988; Herlihy et al. 1998; Uuemaa et al. 2007; Rothwell et al. 2010a; Pratt & Chang 2012) has been common. In addition, the use of ordination methods such as redundancy analysis (Johnson et al. 1997; Sliva & Williams 2001) and principal component analysis (Galbraith & Burns 2007; Andersson & Nyberg 2009) has been popular in water quality studies. However, asymmetric and complex phenomena and responses in water quality-environment relations are expected. Traditional least square regression analysis has strict statistical assumptions to fulfil (McCullagh & Nelder 1989). One solution to this problem is to use developed statistical regression methods such as generalized linear models (GLMs) or generalized additive models (GAMs). These methods are more flexible than traditional least square regression methods as these can, for instance, handle non-linear responses (Wood & Augustin 2002; Venables & Ripley 2002). In addition, non-metric dimensional scaling (NMDS) has been
considered to have great advantages over other ordination methods (McCune & Grace 2002). Variation partitioning (VP) (Borcard et al. 1992) and hierarchical partitioning (HP) (Chevan & Sutherland 1991) are statistical methods, which have been considered suitable for tackling collinearity in data (e.g. Heikkinen et al. 2004; Luoto 2007), which is a usually problem in environmental data. Overall, there is a comprehensive variety in statistical methods to choose from to explore environmental phenomena. This is also a great opportunity for water quality studies as spatial-based statistical modelling can provide reasonably accurate and cost-efficient methods, for example, for the WFD-related water quality and quantity studies.
4 Study area

The rivers and their catchments chosen in the study are located in Finland, northern Europe, between 60° and 68° N latitudes (Fig. 3). The study area is comprised of 34 rivers with their catchments covering the majority of the entire land area of Finland. Therefore, it is a comprehensive assemblage of the major Finnish rivers. The study area in each article is comprised of 32 rivers with minor variation in the rivers between articles (Appendix 1). All rivers are flowing into the Baltic Sea, more precisely into the Gulf of Bothnia and the Gulf of Finland.

In Finland, there are over 640 rivers with a catchment area over 100 km² and with at least a 10 km-long continuous channel (Niemi et al. 2004). The majority of present Finnish watercourses have arisen after the latest glacial period, within the last 10,000 years (Hyvärinen & Kajander 2005). Land uplift has caused some significant changes in flow directions and it still affects Finnish water systems (Kuusisto 2004). The nature of rivers in Finland varies. In the north, they are large compared to the small coastal rivers with few lakes in their catchment. Rivers in central Finland are mostly discontinuous as they connect numerous lakes in the area, called the Finnish lake plateau (Hyvärinen & Kajander 2005).

In Köppen–Geiger climate classification, Finland’s climate is classified as cold with no dry seasons (Df) (Kottek et al. 2006). The climate is fully humid (Kottek et al. 2006) with mostly cold summers (Dfc) (Peel et al. 2007). Key factors affecting the climate of Finland are the latitudinal gradient, maritime climate from the Atlantic Ocean and the continental climate from Eurasia (Atlas of Finland 1987; Tikkanen 2005). This boreal climate is characterised by decreasing mean annual air temperature (from circa 5 °C to −2 °C between 1981–2010) and precipitation (from circa 750 mm to 450 mm) from the south-western coast to northern Finland (Pirinen et al. 2012). Climatic conditions play a significant role in influencing the hydrological functioning of rivers in Finland (Korhonen & Kuusisto 2010). For example, floods and the highest discharge of the year, as a consequence of snow melting, characterise all Finnish rivers but particularly rivers located in coastal areas.

The bedrock in the study area was formed during the Precambrian orogenies (>1.9 billion years) (Atlas of Finland 1990) and is mostly acid and comprised of plutonic rocks and metamorphic schists and gneisses. Surficial ground material (soil) was formed as a consequence of the latest glacial period (Atlas of Finland 1990). The study area is mostly covered by glacigenic deposits (till), peat and fine-
grained soils, such as clay and silt deposits. The relief varies between the especially flat western coast and glacially eroded mountain relief in the north.

Biogeographically, the study area is located in the boreal vegetation zone, extending from the hemiboreal zone to the northern boreal zone. The majority of the study area is covered by forests, which are typically coniferous (Fig. 2b & d). Forests are specifically concentrated on hilly, till-covered areas in the north and central parts of the study area. Mires are concentric bogs in the south and aapa mires in the middle and in the north (Atlas of Finland 1988). Agriculture is a significant form of land use in the study area and it is usually located in the clayish, flat areas near the coast and rivers (Fig. 2b–d). The amount of settlements and infrastructural areas in the catchments varies but the most densely populated areas are located in southern Finland. Precise information about the studied rivers and their catchment can be found in Appendix 1.
Fig. 3. Study catchments and the sites for water quality sampling and discharge observation in Finland between 60° and 68° N latitudes. In the legend, roman numerals after the study catchments refer to study articles. In the map, the number in each catchment refers to the Finnish catchment system.
5 Material and methods

5.1 Water quality variables

Water quality was studied through basic indicators of physico-chemical water quality, namely total phosphorus (P), total nitrogen (N), pH, water colour (CO) (I, II, III) and dissolved oxygen (O) (I) (Fig. 4). Water quality data covered the years 1995–2005 (I, II) and 2000–2012 (III) and it was collected from the Hertta-database, which is maintained by the Finnish Environment Institute (SYKE). Water quality data comprised from nearly 7 000 (I, II) to more than 8 000 (III) water samples.

Fig. 4. A schematic summary of the study steps including the used data, scales and the main methods. Roman numerals refer to study articles. ¹ the referred variable was included in the physiographical variable group, ² the referred variable was included in the geomorphological variables.

Sampling and chemical analyses have been carried out by the laboratories of the Finnish Environmental Authorities or other laboratories using accredited methods based on European norms and standards set by the International Organization for Standardization. Water samples have been collected year around approximately 10 to 25 times in a year and most frequently during high-flow periods. The sampling
sites are located, with few exceptions, close to the river outlet to the sea (Fig. 3). Median values from the water quality variables were used to diminish the effect of great fluctuations in water quality, typical in Finnish rivers (Niemi 2010). Precise information about the water quality data in 1995–2005 can be found in the Appendix of the article I.

5.2 Environmental variables

Water quality was explained by environmental factors, which are rather easy to define from a local to global scale. More precisely, the environmental factors consisted of data from bedrock and soil (I, II), mean basin slope (I, II, III), climate (I, III), NDVI (III), lake percentage (I, III) and land use (I, III) in a catchment (Fig. 4 & Appendix 1). These environmental variables were computed using tools for spatial analyst in ArcGIS versions for Desktop (Esri Corp., Redlands, CA, USA). In addition, discharge median (I) and mean (III) at each observation site was calculated from the data acquired from SYKE’s Hertta-database. The five soil (bedrock outcrop, till, sand-fine sand, clay-silt and peat) and three bedrock types (acid, intermediate and mafic rocks) variables, as percentage covers for each study catchment, were derived from digital maps (1: 1 000 000) of Quaternary deposited and pre-Quaternary rocks (Atlas of Finland 1990). The mean basin slope as degrees was derived from a DEM at 25 m resolution. The climate data at 1 km (III) and 10 km (I) resolution were derived from the Finnish Meteorological Institute (FMI). In article III, climate data was downscaled from the original 1 km grid to a 0.25 km grid by using kriging interpolation to achieve spatially more accurate climatic data. Physiographical variable group (I) and geomorphological factors (II) both included variables from soil deposits, bedrock and mean basin slope. In addition, the physiographical variable group included lake percentage and discharge. NDVI is the most commonly used parameter for quantifying productivity and aboveground biomass (Lillesand et al. 2004). It has been considered to predict planktonic and taxa richness in boreal water systems (Soininen & Luoto 2012) and it is connected to nutrients (Elser et al. 2007). Therefore, NDVI was used as an estimate of catchment productivity. NDVI is based on satellite images collected in 1999–2002 during the growing season (Soininen & Luoto 2012), provided by SYKE and orthorectified by the Swedish National Land Survey (METRIA). Lake area and land use variables, as percentage cover for each catchment, were derived from the Finnish Corine Land Cover (CLC) 2006 data. Spatial resolution for the land use variables, lake area and NDVI was 25 m. The land use data was
reclassified into five categories; agriculture, pasture, forest, peatbogs and wetlands and urban areas.

5.3 Spatiality and temporality

The entire catchment was one of the spatial scales in article III. In addition, buffer zones of 50 m, 100 m, 200 m, 500 m and 1000 m around the river channel were computed. In the temporal study, the entire year was divided into four periods, which were defined according to observed variability in discharge of the studied rivers. The entire year was the fifth temporal period. The spring-winter minimum (hereafter winter minimum) was observed in January, February or March and the spring maximum in April or May. The summer-early autumn minimum (summer minimum) was observed in June, July, August or September and the autumn-winter maximum (autumn maximum) in October, November or December. This division follows annual discharge variability in Finnish rivers as two high-flow periods are seen; one in spring as snow melts and one in autumn after the growing season. Environmental data in articles I and II covered the entire catchment and year.

5.4 Statistical methods

The statistical analyses were started by testing the possible multicollinearity in the data because it can lead to false results as variables that are more causal in reality are excluded from the regression model and inaccurate predictor variables are identified (Dormann et al. 2013). This was done by using the Spearman’s rank correlation test (I, II, III) as the discovered relations between the environmental variables were mostly non-linear, and Spearman’s rank correlation ($r_s$) does not expect linearity between variables (Bonett & Wright 2000). If the $r_s$ between environmental variables were $> 0.9$ (I, II) and $> 0.85$ (III), the variable with lower coefficient of determination ($R^2$) (I, II) and lower $r_s$ (III) with water quality variables was excluded from the further statistical analyses. In article I, the $r_s$ values within the original variable groups (physiography, climate and land use) were calculated. Overall, the following variables were not included in further analyses: the size of the catchment (I, III), acid rocks (I), mafic rocks (II), peat (I), urban areas (III), forests (III), and mean (annual) temperature (III). Furthermore, preconditions in the modelling were improved by log-transforming total phosphorus and nitrogen (I, II, III) as well as water colour (I, III).
In article III, collinearity was also considered by using variance inflation factors (VIF). These indicate how much of a regressor’s variability is explained by the rest of the regressors included in the model because of correlation among these regressors (Craney & Surles 2002). The environmental variable was removed from the final model, if its VIF was > 3 (e.g. Beelen et al. 2013). In addition, spatial autocorrelation (SAC) in the residuals of GAMs (III) was studied by calculating the Moran’s I values (e.g. Dormann et al. 2007; Chang 2008; Hjort et al. 2012) for each model with the program ROOKCASE (Sawada 1999). Phenomena and variables are spatially autocorrelated when measures are spatially dependent (Dormann et al. 2007). If objectives nearby are more alike than subjects far away, SAC is positive (Overmars et al. 2003). This can be problematic and lead to misleading results as, for instance, regression techniques assume that the modelled events are independent (e.g. McCullagh & Nelder 1989).

Statistical modelling was performed using the R statistical environment (R Development Core Team). The relations between water quality and the environmental variables were explored using different modelling techniques (Fig. 4). GLMs are mathematical extensions of linear models able to handle non-linear relationships and different types of statistical error distributions (McCullagh & Nelder 1989; Venables & Ripley 2002). Technically, GLMs are relatively close to linear regressions but more flexible and better suited for analysing environmental relationships (Sokal & Rohlf 1995). GLMs were calibrated using backward elimination procedures (McCullagh & Nelder 1989) and the variable selection criterion ($p < 0.05$) was based on the F-ratio test (Borcard et al. 1992; Venables & Ripley 2002) (I). GAMs are semi-parametric extensions of GLMs and the only assumption made is that the functions are additive and that the components are smooth (Hastie & Tibsirani 1990; Wood 2006). As GAMs also permit complex additive response shapes or a combination of the two within the same model, they are not limited to linear response shapes (Wood & Augustin 2002). In article III, variables to the models were selected using forward selection procedures according to statistical significance ($p < 0.01$). As GAMs are considered to be more flexible statistical models than GLMs (Hjort & Luoto 2013), a more conservative $p$-value for GAMs was utilized.

NMDS (II) is a data reduction technique (Kantvilas & Minchin 1989; McCune & Grace 2002) and it is also considered as a robust ordination method in multivariate analysis (Minchin 1987). NMDS is also well-suited to non-normal (Dodson et al. 2005) and non-linear data, which is a major advantage when compared to other ordination methods (McCune & Grace 2002) such as principal
component analysis. NMDS can also be used to handle correlation in a variable set (Dormann et al. 2013). Therefore, NMDS was applied to describe the general relationships between the geomorphological and water quality variables by using MetaMDS procedure with the Gower dissimilarity measure (Gower 1971). First, the geomorphological vectors were fitted to ordination and the significance of each vector ($p < 0.05$) was assessed with 1000 permutation tests. Then, smooth surfaces of water quality variables were fitted to ordination using GAMs with thin plane splines (Wood 2000). NMDS analyses were performed using the Vegan package (Oksanen 2015) of R (version 2.14.2) statistical environment (R Development Core Team 2012).

VP (I) and HP (I, II) were employed to study the importance of environmental variables as groups (VP) and independently (HP) in relation to the water quality variables. In addition, these statistical methods have been considered as a possible solution for collinearity problems (e.g. Hatt et al. 2004; Heikkinen et al. 2004). VP (Borcard et al. 1992; Anderson & Gribble 1998) divided the variation of water quality variables among three environmental variable groups (physiography, climate and land use) using GLMs. HP (Chevan & Sutherland 1991; Luoto 2007) determined the independent contribution of environmental variables at the 95% confidence limit (Mac Nally 2000). Although VP and HP have been considered suitable methods, for example in aquatic ecology (Alahuhta et al. 2011a) and biodiversity research (Heikkinen et al. 2005), they are rare (but see Stendera & Johnson 2006; Morrice et al. 2008) in studies modelling the relationship between water quality and environmental conditions.
6 Results and Discussion

6.1 Environmental variables in explaining variation in water quality

6.1.1 Specific environmental variables

Agriculture was highlighted as a major factor influencing water quality in boreal rivers. Especially, total phosphorus and nitrogen were related positively to agriculture in GLMs (I) and GAMs (III). In addition, pH value was observed to increase together with the cover of pastures in the catchments (I, III). These support the hypothesis about the importance of agricultural activities on water quality (H1). These findings are also consistent with earlier studies, which have concluded that agriculture is an important source of nutrients entering surface waters (e.g. Granlund et al. 2005; Nielsen et al. 2012; Evans et al. 2014). Increases in the amount of nitrogen in water have been related to specialised agriculture and animal husbandry (Ekholm et al. 2007), and phosphorus to runoff from pastures (Withers & Jarvie 2008). In addition, the use of manure as a fertilizer causes a major nutrient leaching risk from soils despite that its use as fertilizer is controlled by the Nitrate Directive (Government Decree on the Restriction of Discharge of Nitrates from Agriculture into Waters 931/2000, 5 §). Despite the minority (0.3%) of pastures in the study area, its relationship with the pH value can indicate, for example, the effect of animal waste on water quality. Animal waste can increase biological production and, as a consequence, increase the pH value of surface waters. The cover of pastures was also positively connected to phosphorus (article III: Table 2). Agricultural activities like tillage can cause mineralisation as microbial activity increases (Evans et al. 2014) and reduce the capacity of soils to infiltrate (Knox 2001), which increases nutrient leaching from soils. Water protection policy by the EU has decreased agricultural nutrient loading to surface waters during the latest decades (Aakkula & Leppänen 2014; Velthof et al. 2014). For example, the Finnish Agri-environmental support has been important in diminishing the impact of agricultural activities on surface water quality (Aakkula & Leppänen 2014).

The first hypothesis (H1) was also related to the significance of lake percentage and clayish soils in the catchments in explaining water quality in boreal rivers. The findings of this thesis support these assumptions. As lake percentage in the catchments increased, nutrient concentrations and water colour decreased (I, III). Nevertheless, river catchments with higher nutrient concentrations were
characterised by fine-grained soils namely clay-silt soils (I, II), whereas catchments with less nutrients were characterised by coarse-grained soils (II). The results indicate that lakes are major retention basins complementing the conclusions of earlier studies (e.g. Arheimer & Lidén 2000; Lepistö et al. 2006; Stålnacke et al. 2015). A higher lake percentage means longer water residence times in a catchment (Hejzlar et al. 2009) and therefore more time will be available for sedimentation and biological uptake (e.g. Lepistö et al. 2006). In addition, denitrification in lakes is an important removal process of nitrogen from surface waters (e.g. Hejzlar et al. 2009). The connection between clayish soils and water quality is related to the infiltration capacity of soils. Fine-grained soils can retain water more than coarse-grained soils, which decreases their capacity to infiltrate water (Horton 1945). This increases surface runoff, erosion rates and therefore the flowing of nutrients into rivers. This conclusion is supported by Sliva & Williams (2001) as they explained the positive relationship between ammonium ($\text{NH}_4^+$) and clay-silt soil with surface runoff and soil type. Clayish soils in the study area are naturally eutrophic marine and lacustrine deposits (Åström & Rönnback 2005), which is one likely reason for the observed clayish soil-water quality relationship. Clay minerals and organic matter have a good potential to adsorb nutrients since they generally have a rather high, though varying, cation exchange capacity (Sharma et al. 2015).

### 6.1.2 Environmental variable groups

The second hypothesis ($H_2$) assumed that land use would be the most influential environmental variable group in determining water quality in boreal rivers. Generally, it was shown that land use variables, especially agricultural activities, are important factors affecting water quality. However, when the variation in water quality variables was divided among three environmental variable groups, namely catchment physiography, land use and climate, using VP, it was revealed that the largest pure effect was that of physiography (I: Fig. 2) not that of land use, which do not support the hypothesis ($H_2$). However, the second most significant pure effect was that of land use whereas the role of climatic seemed rather weak. The shared effects of physiography and land use were also notable.

The impact of the physiographical variable group on water quality was seen through several variables, such as lake percentage, mean basin slope, and soil deposits (I, II, III) whereas the influence of land use was mainly seen through agricultural activities (I, III). In previous studies, urban coverage has been considered important affecting water quality (e.g. Sliva & Williams 2001;
Meyendonckx et al. 2006), for example, through point sources such as septic tanks (Miller et al. 2011) and industrial discharge (Tu 2013). The observed minor role of urban areas (I) is likely due to the fact that settlements and infrastructural areas are mostly scattered in Finland and single point source pollutants were not revealed at the scale of the study. In addition, wastewater purification in Finland has reduced the amount of pollution entering surface waters (Räike et al. 2003).

Studies on the impact of forests on water quality have produced different results. Lepistö et al. (2006) found that forests are a major source of nitrogen. In addition, negative correlations between forest cover and water quality have been observed (e.g. Allan et al. 1997; Ye et al. 2009; Miller et al. 2011; Tu 2013), which also supports the findings of this study (I). Trees and other vegetation can decrease surface runoff and increase infiltration and water retention capacity as well as prevent erosion by stabilising soil. These factors can decrease the rates of particulate matter with the adsorbed nutrients draining into surface waters, which impacts positively water quality. The relationship between forests and water quality is also affected by the disturbance and the age of the forest as Singh & Mishra (2014) connected undisturbed and old forests to improved water quality through water quality indicators such as pH and turbidity. On the other hand, an increase in the cover of forest in a catchment likely follows a reduction in the proportion of agricultural areas, which would result in a spurious negative relation between forests and water quality. Forests in Finland are rather heavily managed and therefore affected by humans. All forest management practices and especially clear-cutting can impact water quality, for example, by increasing nutrient leaching into surface waters (Löfgren et al. 2014). Overall, it is obvious that forests have an impact on water quality but their response is also affected by other factors such as study area, forest type and age together with management practises.

Water quality was least explained by the group of climatic variables, mean annual temperature and precipitation (I: Fig. 2). Studies focusing on the possible consequences of a changing climate on water quality are rather common (e.g. Whitehead et al. 2009b; Arnell et al. 2015) whereas studies on the impact of current air temperature on water quality are rare (but see Larned et al. 2004). Current precipitation has been used more often as a water quality determinant than air temperature (e.g. Jarvie et al. 2002; Larned et al. 2004; Rothwell et al. 2010a; Zhang et al. 2014). Climate variability caused by ENSO (El Niño Southern Oscillation) has also been associated with variation in water quality (Scarsbrook et al. 2003). Varying impact of precipitation on all studied water quality variables except water colour was discovered (I, III). For example, although the variation in
oxygen dissolved in water were poorly related to the studied environmental variables, it was most strongly explained by precipitation (I). Although the $R^2$ in GLMs was low, 29.4%, (I: Table III) the modelled connection could reflect hydro-climatical variations, such as river flow, which can affect to the concentration of dissolved oxygen (Abdul-Aziz & Ishtiaq 2014). Multicollinearity restricted the studies on the role of temperature on water quality but some positive impacts on nutrients were discovered (I: Table III). Although, the impact of the current climate was not highlighted, it cannot be set aside in water quality studies. This is supported by the fact that climate change is expected to increase both precipitation and temperature in boreal areas (IPCC 2015), enhance hydrological regime and cause changes in land use.

Although the discovered importance of land use according to VP contradicts, to some degree, with earlier studies (e.g. Sliva & Williams 2001) highlighting the physiographical variable group, it is obvious that land use is also an essential factor determining water quality. The strong shared effects of physiography and land use explaining water quality (I: Fig. 2) support the fact that physiographical catchment factors, particularly soil and topography, and different land use forms are connected. For instance, the topsoil of agricultural land is usually clayish (Ekholm et al. 2000), as naturally eutrophic clayish soils are exploited in agriculture. In addition, acid sulphate soils in the western coast of Finland cause acidity in surface waters. As these acid soils are drained, often for agricultural purposes, metal sulphides in the soils are oxidised and sulphate is released causing a decrease in the pH value and the forming of acid sulphate soils (Åström & Rönnback 2005; Toivonen & Österholm 2011). In addition, the impact of climate occurs indirectly through other environmental factors. For instance, conditions for agriculture improve together with air temperature. Agricultural activities and urban areas change hydrological regime, which affect surface runoff and therefore nutrients and pollution entering rivers. Decomposition of organic matter in soils is dependent upon soil temperature and moisture (Davidson & Janssens 2006). Therefore, climate can impact water quality through decomposition processes. Altogether, the impact of different environmental determinants, such as land use and soil deposits, on water quality is difficult to separate since usually these affect water quality together. In addition, spatial scale can affect the dominating factor. However, the results of this thesis show that not only land use and agriculture, but also physiographical factors such as soil properties should be considered when predicting water quality in boreal rivers.
6.1.3 Geomorphological variables

In order to study more specifically the influence of natural catchment factors on water quality, it was hypothesised (H3) that geomorphological variables can be used to predict water quality in boreal rivers at a catchment scale. Geomorphological variables, soil deposits, bedrock and mean basin slope, were significant factors affecting water quality in boreal rivers (II), which is strongly in accordance with the third hypothesis. Clear trends between water quality and geomorphological variables were exposed using NMDS as predominant soil properties in the study area, till, peat and clay-silt soil, correlated strongly ($R^2 \geq 0.616$) (II: Table II) with water quality. In addition, the highest relative independent contributions in HP were over 40% (II: Fig. 4). Especially the cover of clay-silt soil in the catchments appeared to be an important factor predicting water quality as higher nutrient concentrations in water was related to the higher proportion of clay-silt soil in the catchments (I, II). The impact of clay-silt soils on water quality has been discussed in more detail in chapter 6.1.1.

In addition to clay-silt soil, till, peat and sand was observed to influence water quality (I, II). Till was related to decreased nutrient concentrations and pH value (II). Peat was mostly associated with decreased nutrient concentrations in water (I, II) together with decreased pH (I, II) and increased water colour (II). On the contrary, sand was related to the increased pH value and decreased water colour (I, II). Sand and peat affected negatively dissolved oxygen concentration but the coefficient of determination ($R^2$) was low (I, Table III). Compared to fine-grained soils, till and sand are naturally rather nutrient-poor soil types, and their capacity to infiltrate is better, which decreases surface runoff to rivers and thus the amount of nutrients and other material entering rivers. Therefore, water is clearer and contains less nutrients compared to areas dominated by fine-grained soil. As peatlands are an important source of organic acids, they are often associated with decreased pH values in river waters (e.g. Mattson et al. 2007), which is consistent with the results presented here. In addition, decomposition of organic matter consumes oxygen, which can affect oxygen concentration in surface waters. According to Lepistö et al. (2006) peatlands within the river basin can cause both nitrogen leaching and retention. Drained peatlands cause significant nutrient loading (Vassiljev & Blinova 2012), whereas nutrient loading from natural peatlands is controlled by microbial activity and plant uptake (Marttila 2010). Peatlands in Finland are mostly ditched, thereby offering a transport pathway for dissolved and suspended material to enter rivers. However, both peat deposits (II) and peatbogs (III) were mostly related to
decreased nutrient concentrations, especially nitrogen, in river water. This can be explained by plant uptake and processes such as denitrification (Seitzinger et al. 2006) and ammonia volatilisation (Hantush et al. 2013), which represent a loss pathway for nitrogen.

A clear connection between catchment geology and the pH value of river water was observed as mafic rocks were positively related (I) and acid rocks negatively related (II) to pH. This observed relationship between rock types and pH, produced with GLMs (I), NMDS (II) and HP (I, II), is consistent with earlier studies (e.g. Young et al. 2005; Brown et al. 2011). Surface waters in Finland are naturally acid because of acid soil deposits and the dominance of acid rocks. However, the results demonstrate that catchments with mafic rocks have more alkaline waters compared to catchments with no mafic rocks. The connection between bedrock and water quality generally could be stronger without the soil deposit over the bedrock, as soil coverage reduces the direct contact between surface water and bedrock. Nevertheless, till deposit, which is the most common soil in Finland, has a rather local origin. In other words, the mineralogy and chemical constituent of glacial till deposits reflect local bedrock (Lahermo et al. 1996). Thus, part of the effect of bedrock on water quality is seen through soil deposits.

Mean basin slope has often been related to variation observed in different water quality indicators (e.g. Lecomte et al. 2009; Young et al. 2005; Ye et al. 2009; Rothwell et al. 2010b; Pratt & Chang 2012). In this study, river catchments with steeper and varying slope were mostly associated with the higher pH value and decreased water colour (e.g. II: Fig. 3). The discovered relationship between pH and slope contradicts, for example, the outcome of Rothwell et al. (2010b). However, Ye et al. (2009) found a positive correlation between alkalinity and catchment slope. High alkalinity also refers to the higher pH value because of a better buffering capacity against acidity. In addition, Galbraith & Burns (2007) discovered a negative connection between catchment slope and water colour. The effect of mean basin slope on nutrient concentrations were rather weak and unclear (I, II), whereas other studies (Galbraith & Burns 2007; Chang 2008; Rothwell et al. 2010b) have associated decreasing amount of nutrients in water with increased slope. However, the effect of topography is complex and occurs together with and through other factors. For instance, surface runoff on steeper slopes is faster and infiltration is lower compared to less-steep slopes (Chiverton et al. 2015), which increases erosion rates and therefore, for example, nutrients entering surface waters. On the other hand, undeveloped, steep but forested hillsides may adsorb nutrients more than gentle ones (Chang 2008). Topography often also constrains the
locations of different land use forms (Ye et al. 2009) such as agriculture. In addition, the proportion of peat areas often increases together with flat topography (II: Fig. 2), which can be reflected in the discovered relationship between pH and mean basin slope. As Chang (2008) concluded, catchment slope is a secondary factor affecting water quality. Site-specific characteristics are primary factors, and therefore, the connections between catchment slope and water quality vary in different studies.

6.2 Spatial scales and temporal periods in studying environment-water quality relationship

6.2.1 Spatial scales

When studying the importance of different spatial scales in relation to water quality, it was hypothesised (H4) that the environmental characteristics at the entire catchment scale are the most significant predictors of water quality in boreal rivers. All water quality variables except pH were best explained in GAMs when environmental data from the entire catchment was considered (III: Table 2), which endorse the stated hypothesis. After the entire catchment, water quality was mostly affected by the environmental characteristics at the finest, 50 m, buffer scale. pH was mostly explained by the catchment characteristics at this finest buffer scale but least explained by the conditions of the broadest scale, entire catchment. As the width of riparian buffer can be as large as 100 m (Allan et al. 1997), the 50 m buffer zone around the river channel in this study can be considered as the riparian zone. Overall, the findings show that both of these scales, the entire catchment and the riparian zone, should be considered in water quality modelling. This is in line with the recommendation of Amiri & Nakane (2008) about the integration of land use from the entire catchment and riparian zone to build robust water quality models. On the other hand, as Hunsaker & Levine (1995) and Nielsen et al. (2012) concluded that the most important spatial scale in relation to nutrient concentrations is the entire catchment, Chang (2008) and Roberts & Prince (2010) concluded that particularly the characteristics of the riparian zone influence nutrient loadings entering the streams.

The correspondence between land use in close proximity to a stream and in the entire catchment influence the spatial scale at which an effect is detected (Allan 2004). Therefore, if land use close to a river corresponds to land use at the entire
catchment, important relations between water quality and spatial scales can remain unrevealed. It is likely that the distribution of different environmental factors around the catchment affect water quality. For instance, it is crucial whether agricultural activities are concentrated close to rivers or far from rivers. Longer distances from the loading point to a river means more time for infiltration and retention and therefore, an area close to a river would determine water quality. As studies have produced different results about the most important spatial scale in relation to water quality, it seems that the site-specific characteristics (Nielsen et al. 2012) have an essential role in determining the importance of the riparian zones and other areas close to rivers in relation to water quality. Chang (2008) highlighted especially the role of topography and soil deposits.

Riparian zone management has become an important application in watershed and river management (Allan et al. 1997, Gumiero et al. 2013) as processes operating at different kinds of riparian zones such as wetlands (Kreiling et al. 2013) and vegetation areas (Dosskey et al. 2010; Gonzales-Inca et al. 2015) have an essential role in protecting rivers and streams from nutrients and pollutants. The efficiency to retain pollutants and nutrients depends, for example, on the width, slope and soil types of the riparian buffer zone. For instance, the riparian buffer zone should be wider for areas with steeper slopes and fine-grained soils compared to flat areas with coarse-grained soils (Shan et al. 2014). Conditions and processes operating at the riparian zone can also cause impairment of water quality. For example, the proximity of agricultural activities and rivers in the study area can cause nutrients and manure wastes transported to rivers. In addition, the geochemical characteristics of riparian zone are significant in predicting stream water sensitive to acidity (Smart et al. 2001). In this study, especially pH was associated with the riparian zone (III: Table 2) but the variables in the GAMs were mainly related to increases in the pH value. Many rivers particularly in Finland’s western coast are strongly affected by acid soils causing acidification. Demonstrating this would require knowledge about the acid sulphate soils in the catchments, which was not included in this study. However, the results indicate that the variation in pH is controlled by factors and processes operating at the local scale. Although connections between the riparian buffer zones and water quality were discovered, more research is needed in the future.
6.2.2 Temporal periods

According to the fifth hypothesis (H₅), the relationship between water quality and environment is strongest during high-flow periods. The findings obtained by GAMs are consistent with this assumption as the seasonal rhythm of discharge divided the year into high-flow and low-flow periods, from which the high-flow periods during spring or autumn were strongest related to the variation in water quality variables (III, Table 3). Despite this, the discharge variable itself was not clearly observed as a statistically significant variable affecting water quality likely because of a few rivers with exceptionally high discharge. Wet seasons (Carroll et al. 2013), high-flow periods (Gonzales-Inca et al. 2015), spring floods (Buck et al. 2004) and floods caused by snow melting (Woli et al. 2008) have been associated with increased nutrient inputs to surface waters. On the contrary, Zhang et al. (2014) associated greater precipitation and discharge, with lower nutrient concentrations and pollution explaining it by the dilution effect, as greater discharge has a greater capacity to dilute pollution from point sources. Altogether, it seems obvious that seasonal discharge variability causes changes also in water quality.

Cold climate is an essential reason behind the discovered discharge variability and its relation to water quality. In winter, precipitation falls as snow, soil and surface waters are frozen. In summer, precipitation is at its highest in Finland, but the majority evaporates (Korhonen & Kuusisto 2010). In addition, during the summer months vegetative water retention occurs and soil is permeable. Therefore, during these two low-flow periods surface runoff to rivers is smaller and the connection between river and its catchment weaker compared to high-flow periods. During high-flow periods, however, increased surface runoff and discharge result from snow melting in spring and from decreased evapotranspiration in autumn. The capacity of soil to infiltrate is also limited due to partly frozen and saturated soil (Korhonen 2007). In addition, the ability of vegetation to retain water is limited at the beginning and at the end of the growing season. The connection between the surrounding catchment and river water quality increases during these high-flow periods as increased surface runoff erodes landscape surface and carries eroded material into rivers. For example, nutrient loading from agriculture to surface waters occurs mostly outside the growing season (Granlund et al. 2005; Puustinen et al. 2007) when fields lack vegetation, which would prevent erosion and retain nutrients (Ekholm et al. 2007). High-flow periods also increase the acidity caused by acid sulphate soils (Saarinen et al. 2010; Toivonen et al. 2013). The results of this study indicate a connection between pH and flow regime (III, Table 3), but
conclusions about the acidity of rivers during high-flow discharge periods cannot
be made.

Although the effect of the discharge variable on water quality was unclear, the
influence of precipitation was more pronounced as it was included as a statistically
significant variable in some GLM and GAM models (I, III) and in HP (I). However,
its impact was not straightforward as the direction of the effect varied. For example,
$pH$ first increased together with precipitation until the $pH$ value started to decrease
as opposite to precipitation, which can refer to connection between increased flow
conditions and the sensitivity of rivers to acidify. The impact of precipitation is,
however, dependent on many other catchment processes and factors, such as land
use (Jarvie et al. 2002) and soil moisture (Bowes et al. 2015), which can explain
the various modelled relations between precipitation and water quality variables.
In addition, the intensity and quantity of rain are important (Zhang et al. 2014), as
even a short, hard rain can increase both surface runoff and erosion for short periods,
which in turn affects water quality. Environmental impacts on water quality can
also occur with delay long after the disturbance (Allan 2004). For example,
Rankinen et al. (2007) concluded that nitrogen accumulated in soil during dry
periods was washed away during rainy seasons. However, Meyendonckx et al.
(2006) explained the observed low connection between precipitation and nutrient
concentrations by spatially limited precipitation data.

All other water quality variables except nitrogen were related to NDVI,
especially during high-flow periods (III). NDVI indicates primary productivity
(Cramer et al. 1999), but it seems that NDVI has not been previously used as a
direct water quality determinant. However, is has been connected to nutrients
entering aquatic ecosystems (Elser et al. 2007; Soininen & Luoto 2012). NDVI has
been also considered to predict the richness and composition of aquatic
communities (Soininen et al. 2015). Therefore, the discovered direct relationship
between water quality and NDVI was not unexpected. It was also shown that the
effect of NDVI depends on the seasonal variations in river flow conditions as the
connection between water quality variables and NDVI during the high-flow period
in autumn was highlighted (III: Table 3). Moreover, this refers to a delay between
catchment productivity in the growing season and its effect upon water quality.

6.3 Spatial-based statistical modelling in water quality studies

To explore the suitability of the methods used in the study, it was hypothesised ($H_0$)
that statistical modelling methods combined with GIS data provide new
perspectives in the catchment-scale water quality studies across extensive areas. Overall, the findings strongly support the stated hypothesis encouraging the use of the applied methods in studies of water quality determinants. For example, substituting traditional regression analysis by GLMs (I) and GAMs (III) enables studying of non-linear and intricate relationships between water quality and environmental factors in a more realistic manner. Due to the complexity of the relationships and processes impacting water quality, it does not seem as suitable as possible to force the data into scales such as linearity. In addition, multivariate regression methods take into consideration the influences of several explanatory variables at the same time. This helps to understand the joint effects of environmental factors affecting water quality, which is an important advantage over, for example, simple regression. On the other hand, applied approaches also simplify the connections helping to understand complex phenomena. According to Guisan et al. (2002), the use of flexible GLMs and GAMs can aid the development of more representative models and increase understanding of ecological systems. These are also commonly used methods, for example, in geomorphology (e.g. Hjort & Luoto 2013) and aquatic ecosystems (e.g. Alahuhta et al. 2011a; 2011b). The findings support the use of GLMs and GAMs also in studies of water quality determinants, which has been rather rare (but see Morrice et al. 2008; Gonzales-Inca et al. 2015).

Multicollinearity between environmental variables is a common phenomenon, which was also seen in this study (e.g. II, Table III). Traditional solutions to overcome problems caused by multicollinearity have been to summarise variation in the explanatory variables into composite variables by ordination methods, such as principal component analysis, or to exclude some of the most intercorrelated variables (Quinn & Keough 2002). Excluded variables have been selected according to VIF (Zampella et al. 2007; Rothwell et al. 2010a; Gonzales-Inca et al. 2015) or according to correlation coefficients (Young et al. 2005; Uuemaa et al. 2008). In this thesis, both of these methods were utilised. However, the commonly used threshold value, $r_s = 0.7$ (e.g. Dormann et al. 2013), for a critical collinearity was not suitable in this case as causally important explanatory variables would have been excluded from the analyses. However, the use of partitioning methods was discovered to be useful for tackling multicollinearity between environmental variables (I, II). For example, VP revealed that physiographical variables as a group explained studied water quality variables better compared to land use variables. Partitioning methods are especially suitable when used in a complementary manner together with regression methods (Heikkinen et al. 2004) such as GLMs and GAMs.
Although Dormann et al. (2013) preferred GLMs and threshold-based pre-selection of study variables over methods designed for collinearity, according to this thesis partitioning methods provided considerable value into the studies of water quality determinants.

NMDS (II) was used to obtain a general overview of the relationship between the geomorphological and water quality variables. NMDS is a commonly used ordination technology in ecology (e.g. Tolotti et al. 2012; Robinson et al. 2014). However, it seems that it has not been previously used in studying the relationship between water quality and environmental variables. Unlike in a majority of the ecological studies (e.g. Campbell & McIntosh 2013), the aim was not to cluster the rivers according to their location or water quality. Instead, the geomorphological and water quality variables were fitted to ordination in order to visualise their relations. The results of NMDS were mostly in accordance with the results from other methods used in this thesis. However, NMDS revealed a more general view from the explored relations by taking into account all water quality variables. Overall, NMDS is a convenient and quick method for visualisation and presenting graphically the general relations between water quality and environmental factors. The results also support application of NMDS more widely outside the field of ecological studies.

The undetermined variation (I: Fig. 2) of dissolved oxygen and the ability of the environmental variables to explain the variation in oxygen concentration (I: Table III) differed obviously from that of other water quality variables. As oxygen is strongly influenced by biochemical processes, which were not included in this thesis, these findings are unsurprising. However, this difference is a strong suggestion that the applied statistical approaches are highly suitable when studying the relationship between water quality and environmental factors across extensive areas. This is also supported by the fact that the results mostly support the findings of other studies on the water quality-environment relationship despite the used study or modelling method. Therefore, the results encourage the use of GLMs, GAMs and partitioning methods more widely in water quality studies. In addition, the use of NMDS would provide useful perspectives and could be applied in visualising water quality-environment relationships. Combining statistical methods with increasingly available data sources together with GIS data and techniques, we are able to build rapidly inexpensive and reliable models. These models provide a valuable first-filter estimate about the complex water quality-environment relations across extensive areas before more time-consuming and data-hungry dynamic models or extensive fieldwork. The applied approach also helps specify crucial
issues for further study. These are important advantages, for example, in the WFD-related water quality studies. In addition, as land use often causes water degradation, the need to protect and restore the quality of water resources impacts the potential land use and land management practices (Weatherhead & Howden 2009). Therefore, the wide range of land use management planning could utilise spatial-based statistical modelling of water quality. In addition, these applied statistical methods can be seen as essential tools in estimating the water quality in different catchments in response to global change and in identifying areas that are sensitive to this ongoing process.

### 6.4 Sources of uncertainties

#### 6.4.1 Data-related uncertainties

The processing of datasets, such as collecting and calculation, has been subject to uncertainties. Moreover, water quality was studied through a rather modest number of water quality indicators. The selection of water quality indicators together with the studied rivers was restricted and guided by the possibility to get comprehensive data. Especially, the lack of water quality measurements taken during low-flow periods (III) excluded many water quality indicators and rivers from the study. The selected water quality variables are, however, important indicators of environmental phenomena that are closely connected to water quality in Finland, such as eutrophication and acidification. Median values from the water quality variables were used instead of, for example, arithmetic mean. This was decided since mean is more sensitive to extreme values (von Hippel 2005) and northern rivers are characterised by extreme conditions and great transient fluctuations (Niemi 2010).

Environmental factors were selected according to the expected causality in relation to water quality variables. The selected ensemble could be considered extensive. However, topographic wetness index (TWI) (see Andersson & Nyberg 2009; Gonzales-Inca et al. 2015) could have been a valuable variable as soil moisture in the catchments could have been related to soil permeability and infiltration. Extreme climatic conditions such as maximum or minimum temperatures or heavy, short rains might have also been related to water quality. However, especially if based on a single observation, generalisation of the results would be questionable. Mean monthly values correlated greatly with monthly
maximum values (III), which indicates that the effect of using extreme values would have had an unsubstantial effect on the results. Therefore, mean annual (I, III) and mean monthly (III) temperature and precipitation were applied.

A majority of the water quality sampling sites were located in the outlet of the river. Sampling sites along the river channel could have given a specific picture of the spatiality of water quality. Specifically, it could have revealed single point load sources. Nevertheless, as the study covered a total of 34 rivers with their catchments across Finland, it was more appropriate to utilise water quality information as close to the outlet of river as possible. Therefore, environmental factors throughout the entire catchment (I, II, III) or riparian buffer zones (III) were utilised. For each river, discharge data from an observation site that mirrored the water quality sampling site as closely as possible was used instead of, for example, discharge data along the river channel. However, the results demonstrate that this approach is suitable when the relationship between discharge and water quality indicators are studied using a rather coarse scale. Another possible uncertainty source was resolution of the data, which can affect the inferences of the analysis (Luoto & Hjort 2006). Climate variables were downscaled from a coarser resolution (III), which might have underestimated the role of climate. However, as the study was implemented on a rather course scale, covering catchments across the country, it could be assumed that resolution had a rather minor influence on the discovered water quality-environment relations.

6.4.2 Methodological uncertainties

Spatial-based statistical modelling is an effective way to explore environmental relations but it is subjected to many uncertainties starting from the choice of the most applicable modelling technique. In addition, although GLMs are flexible methods, they might not be flexible enough, as the processes and factors impacting water quality form a complex system. Likewise, GLMs are not always able to capture the shape of the relations between responses and environmental variables (Guisan & Zimmermann 2000). Non- and semi-parametric methods, such as GAMs, are more flexible and suitable methods for exploring complex processes (Hjort & Luoto 2013). On the other hand, GAMs can be very complex and therefore difficult to interpret (e.g. Venables & Ripley 2002). GAM is also a data-driven technique, which may produce overestimated predictions (e.g. Hjort & Luoto 2013). In this study, GAMs were not difficult to interpret and the predictions did not different substantially from GLMs. In addition, GLMs and GAMs have been considered to
produce accurate results (e.g. Elith et al. 2006; Marmion et al. 2008), which also reinforce the suitability of these methods in water quality studies.

Neither of the applied partitioning methods produce a model, which can be considered as a weakness. HP also assumes that explanatory and response variables are linearly related (Heikkinen et al. 2005), which might have affected the identification of causal environmental variables determining water quality variables. Since the results of HP were consistent with the outcomes of the other methods, this did not prove to be a problem. VP can also include variables with polynomial relationships (Heikkinen et al. 2004), which is an advantage compared to HP. When applied like in this study, NMDS reveals the relationships between responses and environmental processes on a coarse scale. However, this is a great benefit when the goal is to get a general and rapid view of the studied relations. Like HP and VP, NMDS is a particularly suitable method when used together with other methods. At best, results support each other, which helps interpretation and strengthens the conclusions to be made.

Uncertainties can also result, for example, from model and variable selection criteria, problems with fitting, over-dispersion, multicollinearity and SAC (Heikkinen et al. 2006). Variable selection was done manually, which was time-consuming, but produced more transparent models compared to automated model selection methods. Selecting variables manually also enabled controlling the causes of multicollinearity in the data. Final models were produced according to a $p$-value based on the F-test. Other model selection criteria were also tested such as the Akaike information criterion (AIC) (Akaike, 1974) and the Bayesian information criterion (BIC) (Schwarz, 1978), which can consider model fit and complexity simultaneously (Heikkinen et al. 2006). AIC and BIC were not used because they appeared to be too liberal criteria in the used data. A different kind of model selection criterion could have produced partly different results. However, it is likely that regardless of the used model selection criteria, the included important environmental variables in the models, such as agriculture, would have been the same. Over-fitted models are extremely complex including too many explanatory variables and they are sensitive to fit random noise in the data. This decreases a model’s ability to predict studied phenomena. (Hjort & Luoto 2013). Although GAMs are sensitive to over-fitting, GAMs were not complex, nor did they include many explanatory variables (III). In addition, over-dispersion in the data (P, N, CO) was corrected using log-transformation (I, II, III).

Multicollinearity in the data was taken into account during the entire process, and especially when selecting the modelling techniques and interpreting the results.
Despite this, it is possible that collinearity between environmental variables affected the results in the way that it was not acknowledged. However, it was considered important not to further reduce the amount of environmental variables. It would have distorted and restricted the testing of the study hypotheses. In addition, it is not possible to solve multicollinearity completely (Dormann et al. 2013). SAC is also a common statistical property in geographical phenomena. It can prevent the identification of plausible relationships between studied phenomena and environmental correlates (e.g. Bini et al. 2009). The presence of SAC in model residuals may also increase the rates of type I error (e.g. Hjort et al. 2012). SAC in the residuals of GAMs (III) were rather low as the Moran’s I values in the models, with a lag distance of 50 km, varied between −0.29 and 0.36 (e.g. Dormann et al. 2007) with one exception, 0.51. Moreover, nearly all of the Moran’s I values were statistically non-significant ($p > 0.01$) (III: Table 2 & 3). Therefore, the potential effect of residual SAC on the results can be considered rather low.

Many choices, which are connected to statistical modelling, are subjective. There is no all-inclusive guidebook for decisions concerning, for example, data, modelling technique, variable selection or a final, accepted model. Despite this and other sources of uncertainties, it is concluded that the findings of this thesis show that the applied statistical methods are highly potential methods and provide new insights into water quality studies.
7 Concluding remarks

This thesis aimed to study the relations between water quality and environmental factors in boreal rivers at multiple spatial scales and temporal periods using GIS data and novel statistical methods in water quality studies. Based on the results, six main conclusions are drawn according to the study hypotheses. Conclusions about the modelled water quality-environment relationships are shown in Table 1. Moreover, few general perspectives and future research needs are highlighted.

1. Water quality is the outcome of numerous factors and processes acting in catchments. However, water quality was most strongly affected by agricultural activities (I, III), lake percentage (I, III) and the cover of clay-silt soils (I, II) in the river catchments, which support the first hypothesis (H1). These three factors were highlighted regardless of the modelling technique used or study scale. For example, nutrient concentrations increased as the cover of agricultural areas and clay-silt soil in the catchments increased. In addition, the pH value was positively related to the proportion of pastures in the catchments. Instead, nutrient concentrations and water colour decreased together with increased lake percentage. Nevertheless, water quality variables were also related to other environmental factors such as precipitation and rock types.

2. The group of physiographical variables explained independently more of the variation in water quality than the land use group. Consequently, the findings contradict with the second hypothesis (H2). This highlights the importance of including natural catchment factors in water quality studies in addition to land use properties.

3. Geomorphological catchment factors can be used to predict water quality at a catchment scale, supporting the third hypothesis (H3). For example, nutrients were strongly affected by clay-silt soil and topography was related to pH and water colour.

4. Water quality was best explained when the environmental characteristics at the entire catchment scale were considered, which is in accordance with the fourth hypothesis (H4). However, also the finest scale, 50 m around the river channel, was highlighted. For instance, the characteristics of the 50 m riparian zone around the river channel played the greatest role on pH. This indicates that both spatial scales should be considered in water quality modelling.

5. Seasonal variations in discharge affect water quality. The connection between environment and water quality was stronger during high-flow (spring and
autumn) periods when compared to low-flow periods (summer and winter). This is consistent with the fifth hypothesis (H5).

6. The combination of GIS-based data and applied statistical methods introduce new perspectives in the catchment-scale water quality studies. For example, HP and VP approaches enabled the exploration of the complex water quality-environment relations in intercorrelated environmental data. In addition, NMDS improved visual examination and presentation of these relations. Generally the results were not dependent on the method used. Thus, the findings support the sixth hypothesis (H6) and encourage the use of these methods, showing that spatial-based statistical modelling provides a significant first-filter estimate of the water quality-environment relations at a catchments-scale across extensive areas.

Table 2. Conclusion about the highlighted environmental variables, spatial scales and temporal discharge periods in relations to water quality variables in boreal rivers (+ positive effect, – negative effect, +/-, +/- non-linear effect). Lake refers to lake percentage, slope refers to mean basin slope, precipit. refers to precipitation and peat refers to class peatbogs and wetlands. In addition, entire refers to entire catchment and high-flow means high-flow discharge period. Not means that this was not studied.

<table>
<thead>
<tr>
<th>Environmental factors</th>
<th>Total phosphorus</th>
<th>Total nitrogen</th>
<th>pH</th>
<th>Water colour</th>
<th>Dissolved oxygen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physiography/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geomorphology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lake</td>
<td>–Lake</td>
<td>–Lake</td>
<td>+/-Slope</td>
<td>–Lake</td>
<td>–Sand-fine</td>
</tr>
<tr>
<td>Clay-silt soil</td>
<td>+</td>
<td>+</td>
<td>Sand</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Peat soil</td>
<td>+/-</td>
<td>–</td>
<td>Sand-fine sand</td>
<td>–Slope</td>
<td>+/-Till</td>
</tr>
<tr>
<td>Slope</td>
<td></td>
<td></td>
<td>Bedrock outcrop</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Acid rock</td>
<td></td>
<td></td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mafic rock</td>
<td></td>
<td></td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Land use/cover</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agriculture</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pastures</td>
<td>+</td>
<td>+/-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peat</td>
<td></td>
<td>–/–</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Climate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Precipit.</td>
<td>+/-</td>
<td>+/-</td>
<td></td>
<td></td>
<td>+/-Precipit.</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+/-NDVI</td>
</tr>
<tr>
<td>NDVI</td>
<td>+</td>
<td>+/-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spatial scale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Catchment</td>
<td>High-flow</td>
<td>High-flow</td>
<td>50 m</td>
<td>Catchment</td>
<td>Not</td>
</tr>
<tr>
<td>Temporal period</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The results of this thesis support the fact that river water quality is governed by interlinked processes and factors. The modelled relationships between water quality indicators and environmental catchment characteristics are also largely in accordance with the results from other studies. These serve as strong evidences that
the applied approach, which combined GIS data and developed statistical methods, is very suitable to modelling water quality determinants. For example, the results indicate that management practices should be targeted temporally to the key periods and spatially to the key to areas, which is necessary from both an economical and practical point of view. The significance of agricultural activities in relation to water quality also suggests that despite the improvements in agricultural technology and practice, it is necessary to develop new, more efficient technology in order to reduce leaching of nutrients into surface waters.

More catchment-scale water quality investigations across extensive areas are needed as a majority of the studies has focused on a few catchments or sub-catchments (e.g. Ekholm et al. 2000; Rankinen et al. 2013; Gonzales-Inca et al. 2015). These local-scale studies provide valuable knowledge about the water quality-environment relationships in specific catchments and environmental settings. However, the influence of environment on water quality varies between catchments due to the multivariate nature of the phenomenon. Studies concerning river catchments across environmental gradients reveal the relationship between water quality and environmental conditions in extensive areas and describe general trends. This is highly important when predicting the impacts of global change on water quality. Therefore, both local-scale studies and studies across extensive areas are needed, as they complement each other and provide broader perspectives about the water quality-environment relations in a changing world.

Dynamic nutrient models such as SWAT (e.g. Arnold & Fohrer 2005; Sahu & Gu 2009), ICECREAM (e.g. Granlund et al. 2007) and INCA (e.g. Whitehead et al. 1998; Rankinen et al. 2013) have been successfully used in simulating nutrient leaching from different sources. These studies have produced significant knowledge about the water quality-environment relations and processes behind them. However, these approaches are often site-specific and require rather detailed data or extensive calibration, or both, which takes some time and is a burden to often limited budgets (Chang et al. 2015). This thesis showed that spatial-based statistical analysis, and especially advanced statistical methods, can be applied in water quality research to quickly achieve reliable models at a coarse spatial scale. These methods provide an initial estimate of the environmental factors driving different water quality indicators before future studies and actions, such as dynamic models and fieldwork, are needed. This promotes, for example, water resource management and land use management practices. The applied approach is also advantageous when predicting the impacts of climate change on water quality and identifying the areas and catchment types that are sensitive to expected changes. In
addition, machine learning techniques, such as artificial neural networks, have a good potential for use in water quality research (Chang et al. 2015) and should be explored more for use in predicting water quality. Finally, water quality studies cannot be solely based on fine-scale process models or coarse-scale statistical approaches. Instead, it would be valuable to combine both of these scales and methods to a hybrid approach when studying the water quality-environment relationship across spatial scales and temporal periods. The hybrid approach could produce reliable results about this relationship at different scales together with the mechanisms for the underlying processes. This approach would provide more accurate and rapid knowledge about complicated environmental phenomena without the need for heavy datasets. Therefore, this perspective could be a significant turning point in catchment science.
References


Errata

Article I:

Pages 1038–1039, between the pages is written “With regard to…joint effects of all predictors (fraction g; 46.7%, 62.7%, 29.7%).”

It should be changed to “With regard to…joint effects of all predictors (fraction g; 46.7%, 67.2%, 29.7%).”

Page 1040, in the results of variation partitioning in Figure 2a the joint effect of climate and land use (fraction f) is 1.0%. It should be changed to 3.1%.

Page 1040, in the results of variation partitioning in Figure 2b the joint effects of all predictors (fraction g) is 62.7%. It should be changed to 67.2%

Page 1040, in the results of variation partitioning in Figure 2d the joint effect of all predictors (fraction g) is 29.7%. It should be changed to 0.0%.

Pages 1041–1042, in the middle of the paragraph is written “Our results also show that the relationship between…lake percentage and nitrogen is strong. As the lake percentage in the catchment increases, the water colour and nitrogen decreases”

It should be changed to “Our results also show that the relationship between…lake percentage and phosphorus is strong. As the lake percentage in the catchment increases, the water colour and phosphorus decreases.”
## Appendix

### Appendix 1

Information about the studied rivers and their catchments through studied environmental factors used in the thesis.  

1. river was included in articles I, II, 2 river was included in article III, 3 data covers the years 2000–2012, 4 data is derived from the Finnish Corine Land Cover (CLC) 2006, 5 data covers the years 1995–2005. Abbreviations: No = the official catchment number in the Finnish catchment system, Not = this was not studied, Q = discharge, Slope = mean basin slope, AC = acid, IM = intermediate and MA = mafic rocks, BR = bedrock outcrops, CS = clay-silt soil, PE = peat deposit, SS = sand-fine sand, TI = till, AG = agriculture, PA = pasture, FO = forests, PW = peat bogs and wetlands, UR = urban areas, PRE = precipitation, TEMP = temperature.

<table>
<thead>
<tr>
<th>Study river, No</th>
<th>Size (km²)</th>
<th>Q³ (m³/s)</th>
<th>Slope (°)</th>
<th>Bedrock (%)</th>
<th>Soil deposits (%)</th>
<th>Land use/cover (%)</th>
<th>Lake (%)</th>
<th>Climate³</th>
<th>NDVI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kemijoki, 65</td>
<td>49470</td>
<td>560</td>
<td>1.89</td>
<td>75-1-20</td>
<td>6-0-25-11-58</td>
<td>0.6-0-2-74-9-19-5-0.9</td>
<td>3.8</td>
<td>0.5</td>
<td>0.36</td>
</tr>
<tr>
<td>Simojoki, 64</td>
<td>3160</td>
<td>46</td>
<td>0.91</td>
<td>96-0-4</td>
<td>7-0-31-6-56</td>
<td>1.6-0-2-69-1-22-4-1.1</td>
<td>5.6</td>
<td>1.6</td>
<td>0.35</td>
</tr>
<tr>
<td>Kuivijoki, 63</td>
<td>1360</td>
<td>17</td>
<td>0.59</td>
<td>85-0-15</td>
<td>6-0-38-8-48</td>
<td>1.6-0-3-64-4-30-2-1.2</td>
<td>2.3</td>
<td>1.9</td>
<td>0.36</td>
</tr>
<tr>
<td>Iijoki, 61</td>
<td>14190</td>
<td>173</td>
<td>1.66</td>
<td>92-0-8</td>
<td>6-0-24-13-57</td>
<td>1.3-0-4-73-4-18-0-1.4</td>
<td>5.5</td>
<td>1.5</td>
<td>0.38</td>
</tr>
<tr>
<td>Kiiminkijoki, 60</td>
<td>3810</td>
<td>44</td>
<td>0.96</td>
<td>93-0-5</td>
<td>2-0-42-10-46</td>
<td>2.0-0-4-71-9-20-9-2.0</td>
<td>2.9</td>
<td>2.2</td>
<td>0.38</td>
</tr>
<tr>
<td>Oulujoki, 59</td>
<td>22510</td>
<td>263</td>
<td>1.44</td>
<td>94-0-6</td>
<td>5-1-28-8-58</td>
<td>2.3-0-4-74-5-8-8-2.1</td>
<td>11.9</td>
<td>2.1</td>
<td>0.41</td>
</tr>
<tr>
<td>Temmesjoki, 58¹</td>
<td>1180</td>
<td>9³</td>
<td>0.30</td>
<td>100-0-0</td>
<td>0-12-37-20-31</td>
<td>17.6-0.3-60-4-19-1-2.6</td>
<td>0.4</td>
<td>542³</td>
<td>Not</td>
</tr>
<tr>
<td>Siljakajoki, 57</td>
<td>4280</td>
<td>42</td>
<td>0.50</td>
<td>86-2-12</td>
<td>1-8-44-10-37</td>
<td>9.7-0-3-75-5-10-3-2.2</td>
<td>2.0</td>
<td>593</td>
<td>0.44</td>
</tr>
<tr>
<td>Pyhäjoki, 54</td>
<td>3710</td>
<td>30</td>
<td>0.62</td>
<td>83-3-14</td>
<td>7-7-28-3-55</td>
<td>11.4-0.4-75-5-4-3-3.4</td>
<td>4.9</td>
<td>594</td>
<td>0.45</td>
</tr>
<tr>
<td>Kalajoki, 53</td>
<td>4220</td>
<td>34</td>
<td>0.60</td>
<td>85-2-13</td>
<td>3-11-23-5-58</td>
<td>17.0-0.2-73-0-4-4-3.7</td>
<td>1.7</td>
<td>603</td>
<td>0.45</td>
</tr>
<tr>
<td>Lestijoki, 51</td>
<td>1370</td>
<td>12</td>
<td>0.66</td>
<td>89-2-9</td>
<td>1-9-30-6-54</td>
<td>11.2-0.1-69-1-10-9-2.7</td>
<td>5.9</td>
<td>633</td>
<td>0.32</td>
</tr>
<tr>
<td>Perhonjoki, 49</td>
<td>2520</td>
<td>21</td>
<td>0.69</td>
<td>85-1-14</td>
<td>1-6-34-7-52</td>
<td>11.9-0.1-70-0-12-3-2.9</td>
<td>2.8</td>
<td>620</td>
<td>0.37</td>
</tr>
<tr>
<td>Ähtävänjoki, 47</td>
<td>2050</td>
<td>16</td>
<td>0.96</td>
<td>89-2-5</td>
<td>3-10-25-4-5-8</td>
<td>15.1-0.1-66-0-4-5-4.7</td>
<td>9.7</td>
<td>605</td>
<td>0.29</td>
</tr>
<tr>
<td>Lapuanjoki, 44</td>
<td>4120</td>
<td>34</td>
<td>0.90</td>
<td>97-0-3</td>
<td>9-18-24-4-45</td>
<td>22.7-0.1-64-4-5-1-5.2</td>
<td>2.6</td>
<td>568</td>
<td>0.24</td>
</tr>
<tr>
<td>Kyrönjoki, 42</td>
<td>4830</td>
<td>44</td>
<td>0.87</td>
<td>91-0-9</td>
<td>6-22-25-4-43</td>
<td>25-6-0-1-61-5-6-1-5.6</td>
<td>1.1</td>
<td>591</td>
<td>0.13</td>
</tr>
<tr>
<td>Study river, No</td>
<td>Size (km²)</td>
<td>Q³ (m³/s)</td>
<td>Slope (°)</td>
<td>Bedrock (%)</td>
<td>Soil deposits (%)</td>
<td>Land use/cover (%)</td>
<td>Lake (%)</td>
<td>Climate³ PRE TEMP (mm) (ºC)</td>
<td>NDVI</td>
</tr>
<tr>
<td>----------------</td>
<td>------------</td>
<td>-----------</td>
<td>-----------</td>
<td>-------------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>---------</td>
<td>-----------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Laihianjoki, 41</td>
<td>510</td>
<td>3</td>
<td>0.79</td>
<td>Not</td>
<td>Not</td>
<td>28.0-0.1-61.5-3.3-6.7</td>
<td>0.4</td>
<td>583 4.5 0.09</td>
<td></td>
</tr>
<tr>
<td>Närpiönjoki, 39</td>
<td>990</td>
<td>9</td>
<td>0.88</td>
<td>92-8-0</td>
<td>1-22-21-3-53</td>
<td>21.8-0.1-66.5-6.0-5.0</td>
<td>0.6</td>
<td>614 4.5 0.18</td>
<td></td>
</tr>
<tr>
<td>Lapväärtanjoki, 37</td>
<td>1100</td>
<td>12</td>
<td>0.90</td>
<td>89-6-5</td>
<td>3-12-26-11-48</td>
<td>13.8-0.1-75.7-7.0-3.2</td>
<td>0.2</td>
<td>662 4.6 0.25</td>
<td></td>
</tr>
<tr>
<td>Karvianjoki, 36</td>
<td>3440</td>
<td>15</td>
<td>1.11</td>
<td>96-3-1</td>
<td>10-12-26-7-45</td>
<td>12.9-0.2-71.6-7.3-4.2</td>
<td>3.8</td>
<td>652 4.7 0.25</td>
<td></td>
</tr>
<tr>
<td>Kokemänenjoki, 35</td>
<td>26990</td>
<td>222</td>
<td>2.23</td>
<td>87-2-11</td>
<td>13-18-14-9-46</td>
<td>15.9-0.2-64.7-2-0.6-6</td>
<td>10.7</td>
<td>631 4.5 0.36</td>
<td></td>
</tr>
<tr>
<td>Eurajoki, 34</td>
<td>1340</td>
<td>8</td>
<td>1.11</td>
<td>81-0-19</td>
<td>17-25-13-13-32</td>
<td>23-4-0.2-54.4-1.9-7.2</td>
<td>12.8</td>
<td>615 5.4 0.13</td>
<td></td>
</tr>
<tr>
<td>Aurajoki, 28</td>
<td>870</td>
<td>7</td>
<td>1.73</td>
<td>78-1-21</td>
<td>35-42-10-5-8</td>
<td>37.2-0.4-47.7-2-3-12</td>
<td>0.2</td>
<td>656 5.6 0.35</td>
<td></td>
</tr>
<tr>
<td>Paimionjoki, 27</td>
<td>1090</td>
<td>7</td>
<td>1.71</td>
<td>64-1-35</td>
<td>22-46-5-7-19</td>
<td>43-0-0.4-45.7-0.9-8.4</td>
<td>1.6</td>
<td>652 5.4 0.38</td>
<td></td>
</tr>
<tr>
<td>Ukkelanjoki, 25</td>
<td>570</td>
<td>5</td>
<td>2.39</td>
<td>86-6-8</td>
<td>21-51-4-8-16</td>
<td>43-4-1-64-4-0.6-9.3</td>
<td>0.6</td>
<td>671 5.4 0.40</td>
<td></td>
</tr>
<tr>
<td>Kiskońjoki, 24</td>
<td>1050</td>
<td>6</td>
<td>3.85</td>
<td>76-10-14</td>
<td>35-22-10-16-17</td>
<td>23-5-0.4-61.8-1-2-7-4</td>
<td>5.7</td>
<td>703 5.6 0.37</td>
<td></td>
</tr>
<tr>
<td>Karjaanjoki, 23</td>
<td>2050</td>
<td>17</td>
<td>3.87</td>
<td>49-30-21</td>
<td>22-21-7-16-34</td>
<td>18-5-5-59.2-0.8-9.2</td>
<td>1.4</td>
<td>681 5.4 0.40</td>
<td></td>
</tr>
<tr>
<td>Siuntionjoki, 22</td>
<td>480</td>
<td>1³</td>
<td>4.34</td>
<td>80-11-9</td>
<td>34-30-7-14-15</td>
<td>25-8-0.4-58.9-0.7-14</td>
<td>4.5</td>
<td>675 5.3 0.3³ Not</td>
<td></td>
</tr>
<tr>
<td>Vantaanjoki, 21</td>
<td>1690</td>
<td>16</td>
<td>2.77</td>
<td>57-26-17</td>
<td>18-35-10-14-23</td>
<td>24-1-0.5-52.0-0.6-20.7</td>
<td>2.1</td>
<td>643 5.3 0.25</td>
<td></td>
</tr>
<tr>
<td>Mustijoki, 19</td>
<td>780</td>
<td>6</td>
<td>2.70</td>
<td>71-2-27</td>
<td>14-30-12-8-36</td>
<td>31-1-0.3-57.3-0.8-9.2</td>
<td>1.4</td>
<td>626 5.2 0.39</td>
<td></td>
</tr>
<tr>
<td>Poroönjoki, 18</td>
<td>1240</td>
<td>11</td>
<td>2.88</td>
<td>90-1-9</td>
<td>19-36-5-13-27</td>
<td>31-9-0.3-55.9-0.3-10.3</td>
<td>1.3</td>
<td>633 5.1 0.43</td>
<td></td>
</tr>
<tr>
<td>Ilosanjoki, 17²</td>
<td>310</td>
<td>2</td>
<td>3.54</td>
<td>Not</td>
<td>Not</td>
<td>26-2-0.4-62.5-0.4-7.2</td>
<td>3.3</td>
<td>665 5.3 0.48</td>
<td></td>
</tr>
<tr>
<td>Koskenkylänjoki, 16</td>
<td>900</td>
<td>8</td>
<td>2.77</td>
<td>97-0-3</td>
<td>21-36-3-13-27</td>
<td>31-0.0-0.3-57.7-0.7-6.3</td>
<td>4.1</td>
<td>651 5.1 0.39</td>
<td></td>
</tr>
<tr>
<td>Kymi, 14</td>
<td>37160</td>
<td>67</td>
<td>2.29</td>
<td>93-2-5</td>
<td>10-5-9-7-69</td>
<td>7-3-0.2-68.2-1-7-4.5</td>
<td>18.1</td>
<td>639 4.0 0.41</td>
<td></td>
</tr>
<tr>
<td>Virojoki, 11</td>
<td>380</td>
<td>4</td>
<td>2.47</td>
<td>100-0-0</td>
<td>9-4-18-13-56</td>
<td>14-9-0.1-76.1-1.5-4.3</td>
<td>3.2</td>
<td>698 4.9 0.44</td>
<td></td>
</tr>
</tbody>
</table>
Original articles


Reprinted with permission from John Wiley & Sons, Ltd (I, II).

Original publications are not included in the electronic version of the dissertation.
650. Shao, Xiuyan (2015) Understanding information systems (IS) security investments in organizations
652. Tolkkinen, Mikko (2015) Biodiversity and ecosystem functioning in boreal streams: the effects of anthropogenic disturbances and naturally stressful environments
655. Li, Ying (2015) Users’ information systems (IS) security behavior in different contexts
657. Lappalainen, Katja (2015) Modification of native and waste starch by depolymerization and cationization: utilization of modified starch in binding of heavy metal ions from an aqueous solution
663. Pakanen, Minna (2015) Visual design examples in the evaluation of anticipated user experience at the early phases of research and development
664. Hyry, Jaakko (2015) Designing projected user interfaces as assistive technology for the elderly

Book orders:
Granum: Virtual book store
http://granum.uta.fi/granum/
Sanna Varanka

MULTISCALE INFLUENCE OF ENVIRONMENTAL FACTORS ON WATER QUALITY IN BOREAL RIVERS

APPLICATION OF SPATIAL-BASED STATISTICAL MODELLING