

Spatio-temporal aspects of the environmental factors affecting water quality in boreal rivers

Sanna Varanka^{1*} and Jan Hjort¹

¹Geography Research Unit, University of Oulu, PO Box 3000, FI-90014 Oulu, Finland
sanna.varanka@gmail.com
+358 44 956 6420

Abstract

River water quality is the outcome of multiple processes and factors, which vary spatially and temporally. In this study, the key spatial and temporal scales and the most important environmental factors explaining river water quality at these scales were analysed by generalized additive models (GAMs). Water quality was studied through total phosphorus (median = 61.9 $\mu\text{g l}^{-1}$) and nitrogen (1388.1 $\mu\text{g l}^{-1}$), pH (6.7) and water colour (143.3 mg Pt l^{-1}). Environmental factors covered variables from land use/cover, climate and other landscape characteristics. The spatial scales used were the closest 50 m, 100 m, 200 m, 500 m and 1000 m buffer zones around the river channel, in addition to the entire catchment area. Temporality was studied through the entire year and four periods, which were determined by the natural variation in discharge.

In the comparison of spatial scales of environmental factors, the variation in phosphorus, nitrogen and water colour was best explained using environmental data from the broadest scale, the entire catchment. In contrast, the variation in pH was best explained using data from the closest (50 m) buffer zone, the riparian area. In modelling temporal scales, the variation in water quality variables was best explained during discharge maximum periods and when the environmental data covering the entire year was considered. Nutrients were related specifically to agriculture, water colour to lake percentage, and pH to pastures. The results showed the suitability of GAMs in water quality studies.

Keywords: generalized additive models, riparian area/zone, river, spatial scale, temporal scale, water quality

1 Introduction

Water resources and surrounding landscapes are inseparably connected and river water quality reflects the landscape through which river flows (Allan 2004). One major landscape characteristics in a catchment area is land use/cover (hereafter land use). This strongly human-induced landscape feature has a significant role in determining water quality (e.g. Galbraith and Burns 2007; Tu 2013). However, natural landscape factors, such as soil properties (Varanka et al. 2015) and lake percentage (Varanka and Luoto 2012), in a catchment area has been seen to affect water quality. Overall, river water quality is the outcome of complex interactions of physical, biological and hydrological processes in a catchment area, which are not constant in time and space.

The relationship between stream conditions and landscape characteristics is considered scale-dependent (Allan 2004). Furthermore, surface water quality is subject to substantial spatial and temporal variability (Miller et al. 2014) together with processes affecting water quality. For example, seasonal variability in soil hydraulic and hydrological properties has been observed (Bormann and Klaassen 2008). The ability of vegetation to retain water and uptake nutrients varies in different seasons (Bonan and Shugart 1989) as well as the capacity of soil to infiltrate water (Horton 1945). Moreover, precipitation and evapotranspiration change through the year. These influence surface runoff, which is an essential factor affecting discharge (Winter 2001) and water quality (e.g. Withers and Jarvie 2008; Toivonen et al. 2013). In addition, in cold climate areas precipitation falls as snow, soil and surface waters freeze during winter time, which influences hydrological processes (e.g. Korhonen 2007; Valtanen et al. 2014).

Water quality is not constant through the river (e.g. Chang 2008; Evans et al. 2014) as landscape characteristics and pollution sources vary spatially. Land use, especially agriculture (e.g. Evans et al. 2014) is a major cause of negative changes in water quality. Commonly, agriculture is located along the river channel. Therefore, effective water protection and management requires knowledge about the factors that are affecting water quality at different spatial scales. Conclusions about the most important spatial scales in relation to water quality vary. Some have concluded that the entire catchment determinates water quality the most (e.g. Sliva and Williams 2001) while others have found out that the conditions of the area close to river channel largely explain water quality (Chang 2008; Roberts and Prince 2010). In addition, riparian zone management has become a common application in watershed management (Allan et al. 1997) as riparian area represents a transition zone between terrestrial and aquatic ecosystems (Luke et al. 2007). The riparian area has an essential role in nutrient and energy flux between these two ecosystems (Naiman and Décamps 1997; McClain et al. 2003) and therefore its relationship with surface water is significant (Naiman and Décamps 1997). Processes operating in the riparian zone have significant impact on water quality in receiving water systems (e.g. Smart et al. 2001; Dosskey et al. 2010). Therefore, if the riparian zone is determining river water quality, knowledge about this is important.

By combining spatial perspective with temporality it is possible to achieve specific information about the relationship between water quality and landscape characteristics, which is essential for protecting and managing surface water resources. We applied generalized linear models (GAMs) to explore these complex relationships. GAM is an advanced statistical regression method, a semi-parametric extension of generalized linear model (GLM) (Wood 2006). Although GAM is considered more flexible than traditional least square regression methods (Wood and Augusting 2002; Hjort and Luoto 2013), it is rarely used in water quality studies. More precisely, the goal of this study was to explore at which spatial and temporal scales variation in water quality is best explained using catchment's environmental characteristics. In addition, we studied

which environmental variables are the best determinants of water quality in boreal rivers at different spatial and temporal scales. This was done by developing GAMs for 32 rivers in Finland, Northern Europe, with comprehensive water quality and environmental data.

2 Study area

The study area is located in Finland, northern Europe between 60° and 68° N latitudes. The area covers over half of Finland's land area and it is comprised of 32 rivers and their catchments (Fig. 1). The studied rivers flow into the Gulf of Bothnia and the Gulf of Finland in the Baltic Sea. The relief in the study area is mostly rather flat and mean elevation is about 100 m above sea level. The bedrock is a part of the Fennoscandian shield. It is mostly acid and comprised of plutonic rocks and metamorphic schists and gneisses. The most common surficial ground material (soil) in the study area is till, followed by peat and clayish deposits.

According to Köppen–Geiger climate classification Finland has snow climate, which is fully humid (=Df). There is no dry season and temperature minimum is -3 °C or lower (Kottek et al. 2006). Decreasing mean annual air temperature (from ca. 5 °C to -2 °C between 1981–2010) and precipitation (from ca. 750 mm to 450 mm) from the south-western coast to northern Finland characterise this boreal climate (Pirinen et al. 2012). Precipitation is mainly stored in snow cover in winter (Korhonen and Kuusisto 2010) and surface waters are ice-covered approximately from November to April (Korhonen and Haavanlammi 2012). Snow melting and the breaking up of ice in April or in May can cause significant floods. At the same time, discharge is usually at its highest due snow melting and frozen soil. Lowest discharge values are usually measured in summer and late winter.

Biogeographically, the study area is located in the boreal vegetation zone. Forests cover over half of the study area and are mainly coniferous. Mires are concentric bogs in the south and aapa mires in the middle and northern parts of the study area (Atlas of Finland 1988). Agriculture covers approximately one fifth of the area studied. Urban areas are concentrated in the largest cities in the southern part of Finland. Maps with examples of catchments characteristics are provided in Appendix 1.

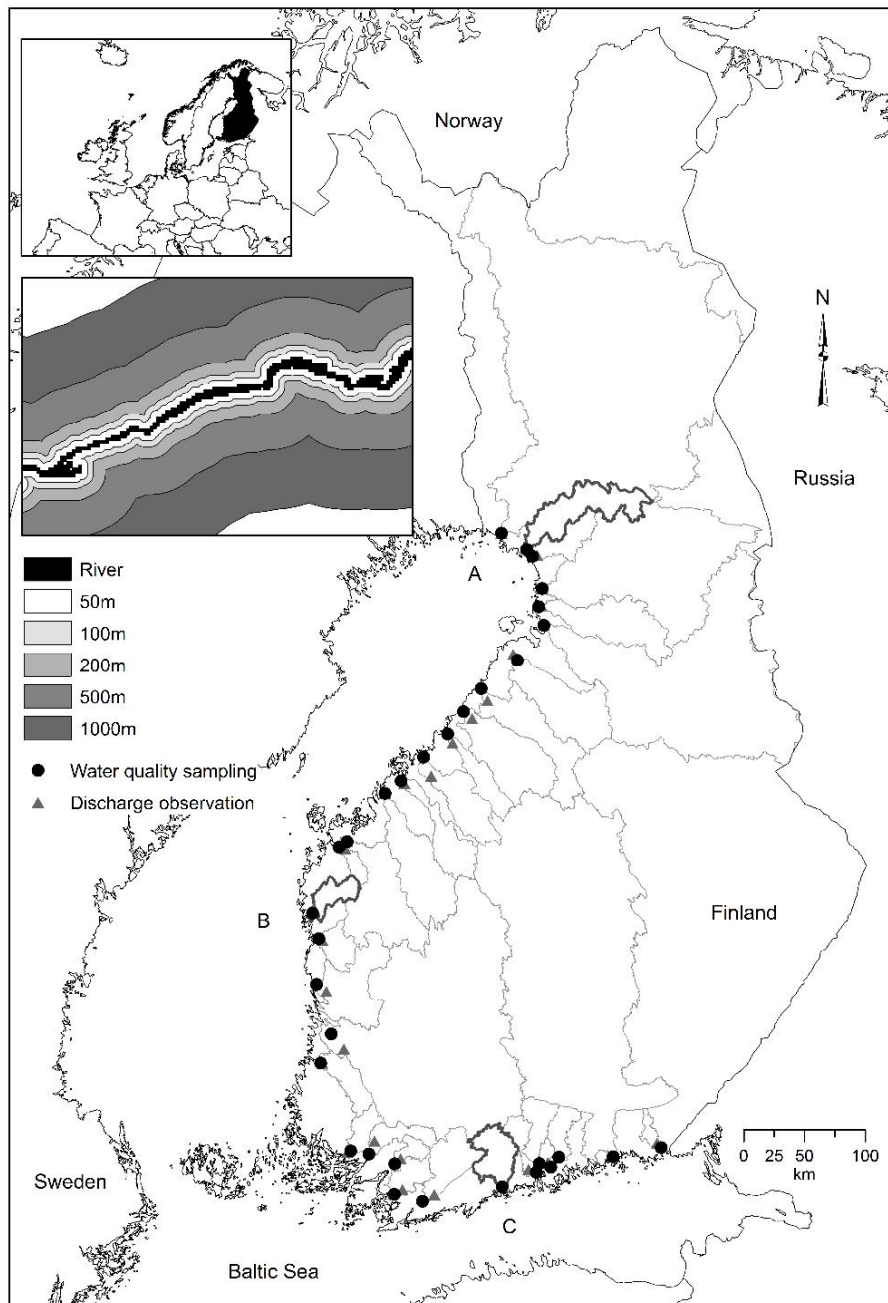


Fig 1 Location of the study catchments and the sites for water quality sampling and discharge observation in Finland between 60° and 68° latitudes. Picture in the left illustrates the buffer zones used in the study. Rivers A (Simojoki), C (Närpiönjoki) and D (Vantaanjoki) have been used as examples of discharge in Finnish rivers (Fig. 2).

3 Material and methods

3.1 Study variables

Water quality was studied through total phosphorus (P), total nitrogen (N), pH and water colour (CO) (Table 1). These physico–chemical variables were selected to the study because they are commonly used in water quality research (e.g. Galbraith and Burns 2007) and central water quality parameters in the study area. The water quality data was gathered from the Finnish Environment Institute's (SYKE) database called Hertta (Finnish Environment Institute, 2016). Water samples were taken and analysed by the experts from the laboratories of the Finnish Environmental Authorities or from other accredited laboratories. Certification of sampling personnel and accredited analytical methods produce reliable and comparable data. Accreditation and certification are based on international standards (Finnish Accreditation Service 2016). Water samples were taken over 8100 times in all seasons during the years 2000–2012 from 32 rivers. Most of the sampling sites are located close to the river outlet to the sea (Fig. 1). Taking into account the geographical and temporal coverage of the data and the fact that developing analytics and updating method standards are ongoing processes, it is not possible to give detailed description about the analytical methods used in the laboratories. However, the water quality data was skewed and the amount of samplings in different season varied. Therefore, median values from the water quality variables were chosen instead of mean values. Median values for each river, which were used in the analysis, were calculated directly for the study periods.

The environmental variables used in this study consist of land use, climate and other landscape characteristics (Table 1). All environmental variables, except discharge, were prepared using tools for spatial analyst in ArcGIS 10.2.1 for Desktop (Esri Corp., Redlands, CA, USA). The discharge data was derived from the Hertta-database. As mean is officially used measure of discharge in Finland (Korhonen and Haavanlammi 2012), annual and monthly means of discharge at each observation site were used in the analyses (Fig. 1). Land use variables and lake area in the river catchments as percentage were derived from Finnish Corine Land Cover (CLC) 2006 using the *Zonal area* function. Mean basin slope as degrees was derived from a digital elevation model (DEM) from the National Land Survey of Finland and it was defined using *Slope* and *Zonal Statistics as table* -functions. Climate variables, mean (annual) precipitation and temperature, and mean NDVI (normalized difference vegetation index) were calculated by the *Zonal statistics as Table* function. The climate data was downscaled from the original 1km grid database, from the Finnish Meteorological Institute, to 250 m grid by using kriging interpolation (cf. Alahuhta et al. 2011). NDVI indicates catchment productivity as it is the most used parameter for quantifying the productivity and aboveground biomass of ecosystems (Lillesand et al. 2004). NDVI is based on satellite imaged collected in 1999–2002 during the growing season (Soininen and Luoto 2012), provided by the Finnish Environment Institute and orthorectified by Swedish National Land Survey (METRIA). Spatial resolution for the land use variables, lake percentage, mean basin slope and NDVI was 25 m.

Table 1 List of water quality and environmental variables, their abbreviations, units and statistical descriptions. Statistical descriptions have been defined from the data used in the analysis. Marked (^E) variables were excluded from generalized additive modelling due to high bivariate correlation with other environmental variables (see the chapter Statistical analysis for further details).

Water quality	Abbreviations	Unit	Minimum	Mean-Median	Maximum
Total phosphorus	P	$\mu\text{g l}^{-1}$	16.0	53.0–61.9	190.0
Total nitrogen	N	$\mu\text{g l}^{-1}$	350.0	1100.0–1388.1	3200.0
pH	pH	mg Pt l^{-1}	5.1	6.7–6.7	7.5
Water colour	CO		30.0	160.0–143.3	220.0
Environmental variables					
Land use					
Urban ^E	UR	%	0.9	5.7–4.8	20.7
Agriculture	AG	%	0.6	18.2–16.4	43.3
Pasture	PA	%	0.1	0.3–0.3	1.6
Forest ^E	FO	%	44.4	64.6–65.3	76.1
Peat bogs and wetlands	PE	%	0.3	6.8–4.3	30.2
Climate					
Mean (annual) temperature ^E	TEMP	$^{\circ}\text{C}$	0.5	4–4.5	5.6
Mean (annual) precipitation	PRE	mm	558	633–635	703
Other					
Catchment size ^E	SIZE	km^2	310	6370–1860	49400
Lake percentage	LA	%	0.2	4.5–3.1	18.1
Discharge	Q	$\text{m}^3 \text{s}^{-1}$	2.1	55.0–15.7	560.1
Mean basin slope	SLO	$^{\circ}$	0.5	1.7–1.3	3.9
Normalized difference vegetation index	NVDI		0.094	0.341–0.369	0.484

3.2 Spatial and temporal scales

To compile environmental data at different spatial scales, we build buffer zones of different size around the river channel: 50 m, 100 m, 200 m, 500 m and 1000 m. In addition, the entire catchment area was one of the spatial scales (Fig. 1). For temporal study, we divided the year into four periods according to observed variability in discharge of the studied rivers. In addition, the entire year was the fifth period. The spring-winter minimum (hereafter winter minimum) covered January, February and March. The spring maximum covered April and May. The summer-early autumn minimum (summer minimum) included June, July, August and September. The autumn-winter maximum (autumn maximum) covered October, November and December. This division follows the annual discharge variability in Finnish rivers as there is seen two high-flow periods (Fig. 2). The highest discharge is found in spring, usually in April or May. The second but lower

peak in discharge is found in late autumn after the growing season. Discharge is usually lowest in February–March just before snow melting. The second minimum is found in summer, when evapotranspiration is strong (Korhonen 2007). In the temporal study, we used monthly means of discharge. For example, the value for spring maximum discharge was monthly mean from April or May depending when discharge was highest. Water quality and climate values for each river were calculated from the months corresponding the observed minimum and maximum mean discharge.

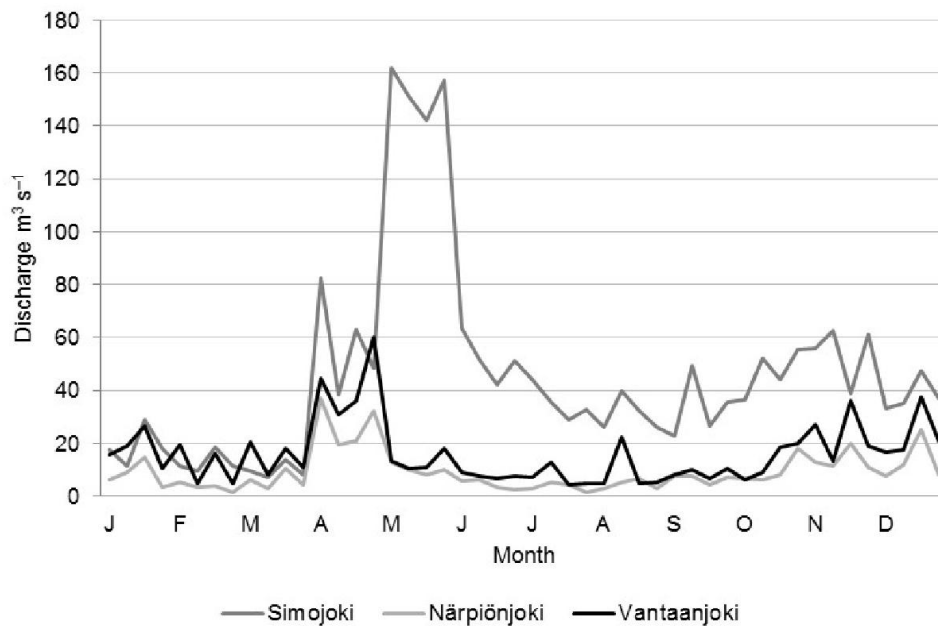


Fig 2 Hydrographs for monthly mean discharge in three different rivers: Simojoki in the north, Närpiönjoki in the west and Vantaanjoki in the south. Monthly mean discharge has been calculated in three year periods in 2000–2012. Locations of the river catchments are indicated in Fig. 1.

3.3 Statistical analysis

Collinearity in a statistical model refers to linearly related, non-independent predictor variables (Dormann et al. 2013). Statistical analyses were started by the Spearman's rank correlation tests to see if there was multicollinearity in the data. If the Spearman's rank correlation coefficients (r_s) between the environmental variables were > 0.85 in one or more spatial scales or temporal discharge periods, variables with lower correlation coefficients between water quality variables were excluded from the statistical analyses. The excluded environmental variables were the size of the catchment, mean (annual) temperature, urban areas and forests in a catchment area (Table 1). The exclusion covered all spatial scales and temporal discharge periods.

The relationships between the water quality and environmental variables were analysed using GAMs. GAMs are semi-parametric extensions of generalized linear models (GLMs) and the only assumption made is that the functions are additive and that the components are smooth (Hastie and Tibshirani 1990; Wood 2006). Unlike traditional least square regression analysis, GAMs are not limited to linear response shapes as they permit as well complex additive response shapes or a combination of the two within the same model (Wood and Augustin 2002). When compared to GLMs, GAMs are more data driven (Hastie and Tibshirani 1990) and able to reveal more complicated relations (Hjort and Luoto 2013). Therefore, GAMs provide significant possibility to explore the complex relationships between water quality and environmental variables. GAMs relate the expected value (μ) to the explanatory variables (x_j) by

$$g(\mu) = \alpha + \sum_{j=1}^p f_j(x_j)$$

where g is a link function, α is the constant and f_j are unspecified smooth functions. In practice, the f_j are estimated from the data by using techniques developed for smoothing scatterplots (Hastie and Tibshirani 1990).

In GAMs, the environmental variables were fitted to water quality using smoothing spline with degrees of freedom (d.f.), which was allowed to vary between one and three. If d.f. is one, it refers to linear relationship and d.f. three refers to non-linear relationship. GAMs were built using forward procedures. Explanatory variables were added to the models according to their Spearman's correlation coefficient with the water quality variables. The variable with the highest correlation was added first and the variable with the lowest correlation last. The variable was included to the final model if it was statistically significant ($p < 0.01$), otherwise it was excluded from the model. All statistical analyses were performed using the R (version 2.14.2) statistical environment (R Development Core Team 2012). Due to over-dispersion in the data, P, N and CO were log-transformed before modelling.

During the modelling, we examined the presence of possible outliers through residuals and influence/hat matrix of the model (Wood 2006). If some of the residuals were considered to be outliers, the model was recalibrated without these observations. Based on the results of the calibration (i.e. residual properties, D^2 and sign of the regression coefficient) highlighted observations were included or excluded from the final step of model calibration. In addition, multicollinearity of calibrated GAMs was explored using variance inflation factors (VIF). If the VIF of an environmental variable was above the strict limit of three, the variable was removed from the final model (e.g. Beelen et al. 2013). In addition, we studied spatial autocorrelation (SAC) in the residuals of GAMs. SAC was estimated by calculating the Moran's I values (e.g. Dormann et al. 2007; Chang 2008; Hjort et al. 2012) in the residuals of the final models with the Microsoft Excel add-in ROOKCASE (Sawada 1999). A lag distance of 50 km between the nearest sites of water quality sampling was used. Spatial dependency and therefore SAC (Dormann et al. 2007) is a common statistical property in geographical phenomena. SAC is positive when subjects close to each other are more similar than subjects far away (Overmars et al 2003). The presence of SAC in model residuals may increase the rates of type I error (Dormann et al. 2007), which can lead to flawed results.

4 Results

The following results of GAMs for different spatial scales are summarised in Table 2. Total phosphorus, total nitrogen and water colour data was best explained utilising environmental variables at the broadest modelling scale, the entire catchment. Instead, pH was best explained at the finest, 50 m, scale. After the entire catchment, the finer spatial scales had slightly more influence on the amount of total phosphorus in the river water compared to the broader scales. The ability to explain the pH value decreased as the size of the spatial zone increased. In the case of water colour, the situation was reverse and the explanation ability was smallest at the finest scales. However, the results of total nitrogen were complex after the optimum, the entire catchment scale. Each water quality variable was mostly explained by the same environmental variables at different spatial scales. Total phosphorus and total nitrogen were affected by agricultural activities (the direction of effect was positive, +) and lake percentage (–) in the catchments. In addition, total phosphorus was explained by the cover of pastures (+). The cover of pastures (+), mean basin slope (non-linear effect, +/-) and precipitation (+/-) were the three most important variables explaining the pH value of water. The variation in water colour was explained mostly by lake percentage, which affected negatively.

Table 2 The results of generalized additive modelling including the environmental variables selected to the final calibrated models (+ positive effect, – negative effect, +/- non-linear effect). Spatial scales are ranked according to the explained deviance (D^2) (%) from highest to lowest percent. Table also shows the Moran's I value for each model with a lag distance of 50 km. Statistical significance codes: *** $p < 0.001$, ** $p < 0.01$. For abbreviations, see Table 1.

	Spatial scale	(%)	Environmental variables			Moran's I value
P	Entire catchment	88.4	+AG***	–LA***		0.24
	50 m	85.8	+AG***	+PA***		0.00
	200 m	85.4	+AG***	+PA***	–LA**	0.13
	100 m	85.0	+AG***	+PA***		0.01
	500 m	82.4	+AG***	+PA***	–LA**	0.12
	1000 m	81.3	+AG***	–LA***	+PRE***	0.09
N	Entire catchment	95.1	+/-AG***	–LA***	+/-PRE***	–0.16
	50 m	82.0	+AG**	–/+PE**		0.28
	500 m	80.9	–PE***	–LA***		0.10
	1000 m	80.7	+AG***			–0.09
	200 m	80.6	+AG**	–/+PE**		0.12
	100 m	74.5	+AG***			0.16
pH	50 m	74.9	+SLO***	+NDVI***		–0.04
	100 m	68.4	+/-PRE***	+/-PA**		0.13
	200 m	68.3	+/-PRE***	+/-PA**		0.11
	500 m	67.4	+PA***	+/-PRE***		0.11
	1000 m	67.1	+/-SLO**	+PA**		0.05
	Entire catchment	63.6	+PA**	+/-SLO**		0.07
CO	Entire catchment	91.7	–LA***	–SLO***		0.36**
	500 m	77.2	–LA***			0.03
	1000 m	76.2	–LA***			0.07
	100 m	71.8	–LA***			0.21
	200 m	70.2	–LA***			0.01
	50 m	51.1	–LA***			0.05

The following results of GAMs for different temporal scales are summarised in Table 3, if not mentioned otherwise. The variation in total phosphorus, total nitrogen and pH data was best explained during maximum flow conditions in spring or in autumn. In addition, the variation in total nitrogen was well explained when the environmental data from the entire year was considered. The variation in water colour data was slightly better explained when the entire year was considered instead of the discharge maximum period in autumn. The models for minimum flow conditions in winter and in summer gained the lowest coefficients of determination for all water quality variables except for water colour. In general, the models for temporal discharge periods highlighted the same environmental variables as the comparison of spatial scales (Table 2 and 3). Agricultural activities (+, +/-) and lake percentage (-) explained best the variation in total phosphorus and total nitrogen. The pH value of water was mostly affected by the cover of pastures (+) and mean basin slope (+/-). In addition to lake percentage (+), water colour was mostly explained by mean basin slope (-). Normalized difference vegetation index and precipitation were also statistical significant environmental variables in few models but the direction of the effect changed according to the water quality variable.

Table 3 The results of generalized additive modelling including the environmental variables selected to the calibrated models (+ positive effect, - negative effect, +/- non-linear effect). Temporal scales are ranked according to the explained deviance (D^2) (%) from highest to lowest percent. Table also shows the Moran's I value for each model with a lag distance of 50 km. Statistical significance codes: *** $p < 0.001$, ** $p < 0.01$. For abbreviations, see Table 1.

	Temporal scale	(%)	Environmental variables			Moran's I value
P	Autumn max	90.9	+AG***	-LA***	+NDVI**	0.08
	Spring max	89.0	+AG***	+NDVI**	-PRE**	0.07
	Entire year	88.4	+AG***	-LA***		0.24
	Summer min	88.0	+AG***	-LA***		-0.08
	Winter min	84.6	+AG***	-LA**		-0.15
N	Spring max	95.5	+/-AG***			0.03
	Entire year	95.1	+/-AG***	-LA***	+/-PRE***	-0.16
	Autumn max	93.0	+AG***	-LA***	+/-PRE**	0.01
	Summer min	86.0	+AG***	-/+LA**		-0.29
	Winter min	84.3	+AG***			-0.16
pH	Spring max	72.6	+/-SLO***	+PA**		0.16
	Autumn max	67.2	+PA***	+/-NDVI**		0.20
	Entire year	63.6	+PA**	+/-SLO**		0.07
	Winter min	51.4	+/-PRE***			0.21
	Summer min	43.5	+PA***			0.23
CO	Entire year	91.7	-LA***	-SLO***		0.36**
	Autumn max	91.0	-LA***	+/-NDVI***	+AG**	0.06
	Summer min	89.4	-LA***	-SLO***		0.10
	Winter min	66.9	-LA***	-SLO***		-0.13
	Spring max	52.0	-LA***			0.51**

SAC in the residuals were rather low as the Moran's I values varied mostly between -0.29 and 0.36, with one exception 0.51 (cf. Dormann et al. 2007; Hjort et al. 2012) (Table 2 and 3). In addition, only three of the residual SAC values were statistically significant ($p < 0.01$) (Table 2 and 3).

5 Discussion

5.1 Spatial scale

Water quality was best explained when the environmental data from the entire catchment or from the closest area, 50 m, around the river channel was considered. This is consistent with Amiri and Nakane (2008) who recommended the integration of land use from the entire catchment and riparian buffer zone to build robust water quality models. Instead, Sliva and Williams (2001) and Nielsen et al. (2012) highlighted the characteristics of the entire catchment in relation to water quality. According to Allan (2004), the spatial scale at which an effect is noticed is affected by how closely land use in the immediacy of river represents land use in the entire catchment. In the study area, agriculture is mainly concentrated around the river channels and the cover of forests increase further from rivers. The distribution of different landscape characteristics in a catchment is likely to affect water quality. For example, the longer the distance from the loading point to the water resource, the more infiltration and retention will occur and the characteristics of the area close to the water resource would be the most important affecting water quality. The conclusions of Roberts and Prince (2010) and Chang (2008) support this as the amount of nutrients was connected to the area close to river channel instead of the entire catchment.

The important relationship between riparian zones and water resources is recognised (e.g. Naiman and Décamps 1997; Luke et al. 2007). Common recommendations on the riparian buffer width vary between ten meters to 100 m based on processes that need protection (Allan et al. 1997). Thus, at least the two narrowest spatial scales in this study could be considered as riparian zones. The importance of the finest scales in relation to water quality was not clearly pronounced. Nevertheless, nutrients and particularly pH were related to the finest, 50 m, scale. Agricultural activities, especially in the western and south-western coast of Finland, are usually located close to rivers (Appendix 1). The proximity of agriculture and rivers in the study area can increase transportation of nutrients and manure wastes to rivers because of, for example, shorter time for infiltration and retention compared to the remoteness of agricultural activities and rivers. In addition, Smart et al. (2001) connected riparian areas to acidity as they concluded that riparian areas are important in predicting stream waters sensitive to acidity. The forming of acid sulphate soils, typical especially in Finland's western coast (Toivonen et al. 2013), is causing the low pH values in rivers flowing through these areas (Niemi and Raateland 2007). However, revealing the connection between acid waters and acid soils would require knowledge about the acid sulphate soils in the catchments. Likewise, bedrock geology in the catchments was excluded from the study but it can have significant role on water quality (Brown et al. 2011). For example, carbonates have been connected to increased pH value along rivers (e.g. Barth et al. 2003). In addition, the importance of near-river areas can depend on site-specific characteristics not included in studies with coarse spatial scale (Nielsen et al. 2012).

5.2 Temporal scale

Although the discharge variable itself was not a significant variable in this study, its seasonal rhythm divided the year into high-flow and low-flow periods, which explained variation in water quality differently. According to Shrestha and Kazama (2007), water quality is impacted by seasonal variability in discharge. In

addition, Vuorenmaa et al. (2002) concluded that usually the main reason for nutrient losses in Finland is fluctuation in discharge. During discharge maximum periods in Finland, infiltration is limited due to partly frozen or saturated soil (Korhonen 2007). Snow melting, particularly in spring, can cause significant floods. Moreover, the ability of vegetation to retain water is limited in the beginning and end of the growing season. Increased surface runoff erodes landscape surface and carries eroded material to surface waters, which impair water quality. Therefore, variations in water quality are better explained during high-flow than low-flow periods.

The results of the temporal exploration are supported by other studies. Buck et al. (2004) connected spring floods and Woli et al. (2008) floods caused by snow melting with increased nutrient runoff to rivers. In addition, wet seasons (Carroll et al. 2013) and high-flow periods (Gonzales-Inca et al. 2015) have been associated with increased nutrient inputs to surface waters. However, Zhang et al. (2014) concluded that during high-flow periods, water quality is better compared to low-flow periods. In this study, it was more difficult to explain the water quality during low-flow periods compared to high-flow periods. In winter, soil and surface waters are frozen and precipitation falls as snow. Whereas in summer, precipitation increase but majority evaporates (Korhonen and Kuusisto 2010), vegetation retains and soil infiltrates water. Therefore, during these low-flow periods surface runoff to rivers and the connection between a river and the surrounding catchment is smaller compared to high-flow periods.

5.3 Environmental variables explaining water quality

Agriculture is a central source of nutrients entering surface waters (e.g. Granlund et al. 2005; Evans et al. 2014). The results of this study are consistent with these studies. Increased nitrogen leaching from agricultural areas has been associated with specialised agriculture and regional intensification of animal husbandry (Ekholm et al. 2007). However, the spreading of manure to the fields is forbidden from late autumn until April due to implementation of Nitrates Directives (Government Decree on the Restriction of Discharge of Nitrates from Agriculture into Waters 931/2000, 5 §) but it is again allowed during spring floods. Manure as a fertilizer causes higher nitrogen balance, which increases leaching risk (Bechmann 2014). Cultivated land and pastures are also important sources of phosphorus entering rivers (Withers and Jarvie 2008). Nutrient losses from agricultural activities are the result of complex interactions between soil processes, vegetation, climate and management practises (Granlund et al. 2007). For example, nitrogen losses vary due to seasonal fluctuations in mineralisation (Bechmann 2014) and the ability of stream ecosystems to retain nutrients varies spatio-temporally (Fisher et al. 1998). Moreover, nutrient loading from agricultural areas to surface waters occurs mostly outside the growing season (Granlund et al. 2005) when surface runoff and discharge is higher compared to summer. In addition, outside the growing season, fields lack vegetation, which would prevent erosion and uptake nutrients (Ekholm et al. 2007). These conclusions are consistent with the discovered connections between water quality and the temporal discharge periods in this study.

Precipitation had some effect on the water quality variables. However, as precipitation is closely connected to other spatio-temporally varying factors and processes in a catchment, such as infiltration, the influence of precipitation on water quality is not straightforward. In addition, the intensity and quantity of rain are important affecting water quality in receiving surface waters (Zhang et al. 2014) as they affect erosion, surface runoff and discharge. Rainy seasons can also wash away nitrogen accumulated in soil during dry periods (Rankinen et al. 2007) as environmental impacts on water quality can occur with a delay long after

the appearance of the disturbance (Allan 2004). However, Zhang et al. (2014) explained the negative correlation between precipitation and pollution index by the dilution effect of increased flow conditions. In this study, precipitation was most related to pH and total nitrogen with a non-linear relationship. The acidity of Finnish rivers located in western coastal areas has been connected to seasons with increased discharge (Saarinen et al. 2010) and runoff (Toivonen et al. 2013). Although the variation in pH were strongest related to the high-flow periods, statistically significant relationship between pH and discharge variable was not observed in this study. However, pH first increased together with precipitation until the pH values started to decrease contrary to precipitation and it is possible that this can indicate connection between increased flow conditions and sensitivity to acidify.

Rivers and lakes in a catchment area are closely connected. Rivers deliver, for instance, nutrients from landscapes to lakes, which makes rivers an important factor between landscape activity and lake water quality (Nielsen et al. 2012). In addition, lakes are important retention basins, for example, because of sedimentation (Arheimer and Lidèn 2000), biological uptake (Lepistö et al. 2006) and denitrification (Hejzlar et al. 2009). The positive impact of lakes in a catchment on water quality was also seen in this study. Moreover, mean basin slope appeared to be an important feature influencing water quality. However, slope is considered as a secondary predictor affecting water quality through other environmental characteristics (e.g. Chang 2008; Varanka et al. 2015). NDVI has not been previously used directly to explain water quality in river systems. In this study, all other water quality variables except nitrogen were related positively to NDVI and it was connected to the maximum period of discharge especially after the growing season. This indicates a clear delay between the catchment productivity during the growing season and its effect upon water quality.

5.4 GAMs in water quality studies

Statistical modelling is rather common in studying the water quality-environment relationship. For example, simple linear regression analysis (e.g. Woli et al. 2008; Evans et al. 2014) has been common. However, asymmetric and complex responses in water quality-environment relations are expected. Therefore, advanced statistical regression methods such as GAMs are more flexible than traditional least square regression methods as these can, for instance, handle non-linear responses (Wood and Augusting 2002). The results encourage the use of GAMs in water quality studies, which has been rare. GAMs are rather easy and quick to implement but they simplify the connections between water quality helping to understand this complex phenomenon. On the other hand, GAM is a data-driven technique, which can produce overestimated predictions (e.g. Hjort & Luoto 2013). GAMs can also be complex and difficult to interpret (e.g. Venables and Ripley 2002). In this study, GAMs were not difficult to interpret. GAMs have also been considered to produce accurate results (e.g. Elith et al. 2006), which reinforce the suitability of these methods in water quality studies. In addition, SAC in the residuals of GAMs were rather low. Therefore, the potential influence of the residual SAC for the interpretation of the results can be considered rather low.

6 Conclusions

In the exploration of the river water quality-environment relationship, it is important to consider its spatio-temporal aspects as water quality and processes vary in time and space. We studied at which spatial and temporal scale the variation in water quality in Finnish boreal rivers is best explained using catchment's environmental characteristics and which of them are the best determinants of water quality. These were studied using GAMs, which are rather rarely used in water quality studies. However, it was shown that GAMs provide robust insights when exploring the complex relationships between water quality and environmental characteristics across scales. For the spatial scales, the variation in water quality was best explained using the characteristics from the entire catchment or the finest, 50 m, scale. In the comparison of temporal scales, variation in water quality was best explained during discharge maximum periods. Temporal discharge periods had slightly greater influence on variation in water quality than spatial scales. The water quality variables were mostly explained by the same environmental determinants regardless of the used scales. Natural landscape factors, such as lake percentage, but especially agricultural activities in a catchment provide a good indicator of water quality. The results support land use and water resources management. Water quality-environment relationship is complex and requires cost-efficient and reliable research methods such as GAMs. Expected increase in precipitation and floods in addition to changes in land use as a consequence of ongoing global change highlights this importance. In addition to the use of advanced statistical methods, it is encouraged to expand the study scale to other rivers worldwide and the amount of water quality parameters as well to study the effects of, for example, background geology when studying water quality-environment relationships.

Acknowledgements

We would like to thank Janne Alahuhta for his ideas when planning the study. Jan Hjort acknowledges the Academy of Finland (grant numbers 267995 and 285040). Comments by an anonymous reviewer substantially improved this manuscript.

References

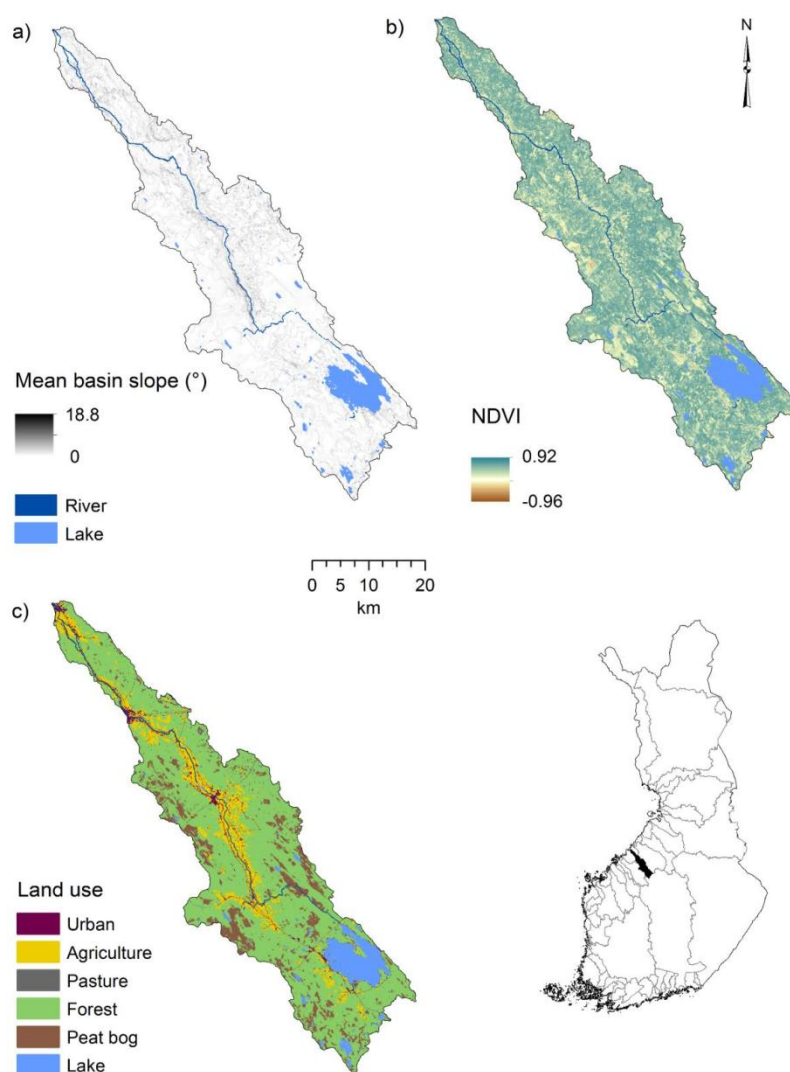
- Alahuhta J, Heino J, Luoto M (2011) Climate change and the future distributions of aquatic macrophytes across boreal catchments. *Journal of Biogeography* 38: 383–393
- Allan JD, Erickson DL, Fay J (1997) The influence of catchment land use on stream integrity across multiple spatial scales. *Freshwater Biology* 37:149–161
- Allan JD (2004) Landscapes and riverscapes: the influence of land use on stream ecosystems. *Annual Reviews of Ecology Evolution and Systematics* 35:257–284
- Amiri BJ, Nakane K (2008) Entire catchment and buffer zone approaches to modelling linkage between river water quality and land cover—a case study of Yamaquchi Prefecture, Japan. *Chinese Geographical Science* 18:85–92
- Arheimer B & Lidén R (2000) Nitrogen and phosphorus concentrations from agricultural catchments— influence of spatial and temporal variables. *Journal of Hydrology* 227:140–159
- Atlas of Finland (1988) Biogeography, Nature Conservation (in Finnish). Folio 141–143. Helsinki, National Board of Survey & Geographical Society of Finland
- Barth JAC, Cronin AA, Dunlop J & Kalin RM (2003) Influence of carbonates on the riverine carbon cycle in an anthropogenically dominated catchment basin: evidence from major elements and stable carbon isotopes in the Lagan River (N. Ireland). *Chemical Geology* 200: 203–216
- Bechmann M (2014) Long-term monitoring of nitrogen in surface and subsurface runoff from small agricultural dominated catchments in Norway. *Agriculture, Ecosystems and Environment* 198:13–24
- Beelen RM, Hoek G, Vienneau D et al (2013) Development of NO₂ and NO_x land use regression models for estimating air pollution exposure in 36 study areas in Europe – the ESCAPE project. *Atmospheric Environment* 72:10–23
- Bonan GB, Shugart HH (1989) Environmental factors and ecological processes in boreal forests. *Annual Review of Ecology and Systematics* 20:1–28
- Bormann H, Klaassen K (2008) Seasonal and land use dependent variability of soil hydraulic and soil properties of two Northern German soils. *Geoderma* 145:295–302
- Brown S, Lavkulich LM, Schreier H (2011) Developing Indicators for Regional Water Quality Assessment: An Example from British Columbia Community Watersheds. *Canadian Water Resources Journal* 36: 271–284
- Buck O, Niyogi DK, Townsend CR (2004) Scale-dependence of land use effects on water quality of streams in agricultural catchments. *Environmental Pollution* 130:287–299
- Carroll S, Liu A, Dawes L, Hargreaves M, Goonetilleke A (2013) Role of land use and seasonal factors in water quality degradations. *Water Resources Management* 27:3433–3440
- Chang H (2008) Spatial analysis of water quality trends in the Han River basin, South Korea. *Water Research* 42:3285–3304
- Dormann CF, McPherson JM, Araújo MB et al (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30:609–628
- Dormann CF, Elith J, Bacher S et al (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36:27–46

- Dosskey MG, Vidon P, Gurwick NP, Allan CJ, Duval TP, Lowrance R (2010) The role of riparian vegetation protecting and improving chemical water quality in streams. *Journal of the American Water Resources Association* 46:261–277
- Ekhholm P, Granlund K, Kauppila P, Mitikka S, Niemi J, Rankinen K, Räike A, Räsänen J (2007) Influence of EU policy on agricultural nutrient losses and the state of receiving surface waters in Finland. *Agricultural and Food Science* 16:282–300
- Elith J, Graham CH, Anderson RP et al. (2006) Novel methods improve predictions of species' distributions from occurrence data. *Ecography* 29:129–151
- Evans DM, Schoenholtz SH, Wigington PJ Jr, Griffith SM, Floyd WC (2014) Spatial and temporal patterns of dissolved nitrogen and phosphorus in surface waters of a multi-land use basin. *Environmental Monitoring and Assessment* 186:873–887
- Finnish Accreditation Service (2016) Accreditation and certification; objectives and the main differences. <https://www.finas.fi/sites/en/topical/articles/Pages/Accreditation-and-certification;-objectives-and-the-main-differences.aspx>. Accessed 7 December 2016
- Finnish Environment Institute (2016) Open data (in Finnish). http://www.syke.fi/fi-FI/Avoin_tieto/Ymparistotietojarjestelmat. Accessed 7 December 2016
- Fisher SG, Grimm NB, Martí E, Holmes RM, Jones JB (1998) Material spiralling in stream corridors: a telescoping ecosystem model. *Ecosystems* 1:19–34
- Galbraith LM, Burns CW (2007) Linking land-use, water body type and water quality in southern New Zealand. *Landscape Ecology* 22:231–241
- Gonzales-Inca CA, Kalliola R, Kirkkala T, Lepistö A (2015) Multiscale landscape pattern affecting stream water quality in agricultural watershed, SW Finland. *Water Resources Management* 29:1669–1682
- Government Decree on the Restriction of Discharge of Nitrates from Agriculture into Waters (931/2000, 5§). Environmental Protection Act (86/2000). Ministry of the Environment. <http://www.finlex.fi/en/laki/kaannokset/2000/en20000931.pdf>. Accessed 7 December 2016
- Granlund K, Räike A, Ekhholm P, Rankinen K, Rekolainen S (2005) Assessment of water protection targets for agricultural nutrient loading in Finland. *Journal of Hydrology* 304:251–260
- Granlund K, Bärlund I, Salo T, Esala M, Posch M (2007) The effect of decreasing fertilization on agricultural nitrogen leaching: a model study. *Agricultural and Food Science* 16:376–386
- Hastie TJ, Tibshirani RJ (1990) *Generalized additive models*. London, Chapman and Hall
- Hejzlar J, Anthony S, Arheimer B et al (2009) Nitrogen and phosphorus retention in surface waters: an inter-comparison of predictions by catchment models of different complexity. *Journal of Environmental Monitoring* 11:584–593
- Hjort J, Heikkinen RK, Luoto M (2012) Inclusion of explicit measures of geodiversity improve biodiversity models in a boreal landscape. *Biodiversity and Conservation* 21:3487–3506
- Hjort J, Luoto M (2013) Statistical methods for geomorphic distribution modeling. In: Shroder J Jr & Baas ACW (eds) *Treatise on Geomorphology*. Academic Press, San Diego, pp 59–73
- Horton RE (1945) Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology. *Bulletin of the Geological Society of America* 56:275–370
- Korhonen J (2007) Discharge and water level variations in lakes and rivers in Finland (in Finnish). *The Finnish Environment* 45/2007. Finnish Environment Institute, Edita Publishing, Helsinki

- Korhonen J, Kuusisto E (2010) Long-term changes in the discharge regime in Finland. *Hydrology Research* 41:253–268
- Korhonen J, Haavanlammi E (2012) Hydrological Yearbook 2006–2010. *The Finnish Environment* 8/2012. <http://hdl.handle.net/10138/38812>. Accessed 7 December 2016
- Kottek M, Grieser J, Beck C, Rudolf B, Rubel F (2006) World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift* 15:259–263
- Lepistö A, Granlund K, Kortelainen P, Räike A (2006) Nitrogen in river basins: Sources, retention in the surface waters and peatlands, and fluxes to estuaries in Finland. *Science of the Total Environment* 365:238–259
- Lillesand TM, Kiefer RW, Chipman JW (2004) Remote sensing and image interpretation. Wiley, New York
- Luke ST, Luckai NJ, Burke JM, Prepas EE (2007) Riparian areas in the Canadian boreal forest and linkage with water quality in streams. *Environmental Reviews* 15:79–97
- McClain ME, Boyer EW, Dent CL et al (2003) Biogeochemical hot spots and hot moments at the interface of terrestrial and aquatic ecosystems. *Ecosystems* 6:301–312
- Miller C, Magdalina A, Willows RI, Bowman AW, Scott EM, Burgess C, Pope L, Pannullo F, Haggarty R (2014) Spatiotemporal statistical modeling of long-term change in river nutrient concentration in England & Wales. *Science of the Total Environment* 466–467:914–923
- Naiman RJ, Décamps H (1997) The ecology of interfaces: riparian zones. *Annual Review of Ecology and Systematics* 28:621–658
- Nielsen A, Trolle D, Søndergaard M, Lauridsen TL, Bjerring R, Olesen JE, Jeppesen E (2012) Watershed land use effects on lake water quality in Denmark. *Ecological Applications* 22:1187–1200
- Niemi J, Raateland A (2007) River water quality in the Finnish Eurowaternet. *Boreal Environment Research* 12:571–584
- Overmars KP, de Koning GHJ, Veldkamp A (2003) Spatial autocorrelation in multi-scale land use models. *Ecological Modelling* 164:257–270
- Peel MC, Finlayson BL, McMahon TA (2007) Updated world map of the Köppen–Geiger climate classification. *Hydrology and Earth System Sciences* 11:1633–1644
- Pirinen P, Simola H, Aalto J, Kaukoranta JP, Karlsson P, Ruuhela R (2012) Climatological statistics of Finland 1981–2010. Finnish Meteorological Institute Reports 1
- Rankinen K, Salo T, Granlund K, Rita H (2007) Simulated nitrogen leaching, nitrogen mass field balances and their correlation on four farms in south-western Finland during the period 2000–2005. *Agricultural and Food Science* 16:387–406
- R Development Core Team (2012) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
- Roberts AD, Prince SD (2010). Effects of urban and non-urban land cover on nitrogen and phosphorus runoff to Chesapeake Bay. *Ecological Indicators* 10:459–474
- Saarinen T, Vuori K-M, Alasaarela E, Kløve B (2010) Long-term trends and variation of acidity, COD_{Mn} and colour in coastal rivers of Western Finland in relation to climate and hydrology. *Science of the Total Environment* 408:5019–5027

- Sawada M (1999) ROOKCASE: An Excel 97/2000 Visual Basic (VB) add-in for exploring global and local spatial autocorrelation. *Bulletin of the Ecological Society of America* 80:231–234
- Shrestha S, Kazama F (2007) Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. *Environmental Modelling & Software* 22:464–475
- Sliva L, Williams DD (2001) Buffer zone versus whole catchment approaches to studying land use impact on river water quality. *Water Research* 35:3462–3472
- Smart RP, Soulsby C, Cresser MS, Wade AJ, Townend J, Billett MF, Langan S (2001) Riparian zone influence on stream water chemistry at different scales: a GIS-based modelling approach, an example for the Dee, Scotland. *The Science of the Total Environment* 280:173–193
- Soininen J, Luoto M (2012) Is catchment productivity a useful predictor of taxa richness in lake plankton communities? *Ecological Applications* 22:624–633
- Toivonen J, Österholm P, Fröjdö S (2013) Hydrological processes behind annual and decadal-scale variations in the water quality of runoff in Finnish catchments with acid sulfate soils. *Journal of Hydrology* 487:60–69
- Tu J (2013) Spatial variations in the relationships between land use and water quality across an urbanization gradient in the watersheds of Northern Georgia, USA. *Environmental Management* 51:1–17
- Valtanen M, Sillanpää N, Setälä H (2014) Effects of land use intensity on stormwater runoff and its temporal occurrence in cold climates. *Hydrological Processes* 28:2639–2650
- Varanka S, Luoto M (2012) Environmental determinants of water quality in boreal rivers based on partitioning methods. *River Research and Applications* 28:1034–1046
- Varanka S, Hjort J, Luoto M (2015) Geomorphological factors predict water quality in boreal rivers. *Earth Surface Processes and Landforms* 40:1989–1999
- Venables WN & Ripley BD (2002) *Modern Applied Statistics with S. Statistics and Computing*. Springer, New York
- Vuorenmaa J, Rekolainen S, Lepistö A, Kenttämies K. & Kauppila P (2002) Losses of nitrogen and phosphorus from agricultural and forest areas in Finland during the 1980s and 1990s. *Environmental Monitoring and Assessment* 76:213–248
- Winter TC (2001) The concept of hydrological landscapes. *Journal of the American Water Resources Association* 37:335–349
- Withers PJA, Jarvie HP (2008) Delivery and cycling of phosphorus in rivers: a review. *Science of the Total Environment* 400:379–395
- Woli KP, Hayakawa A, Kuramochi K, Hatano R (2008) Assessment of river water quality during snowmelt and base flow periods in two catchment areas with different land use. *Environmental Monitoring and Assessment* 137:251–260
- Wood SN, Augustin NH (2002) GAMs with integrated model selection using penalized regression splines and applications to environmental modelling. *Ecological Modelling* 157:157–177
- Wood S (2006) *Generalized additive models: an introduction with R*. Chapman & Hall/CRC, Boca Raton
- Zhang Z, Chen Y, Wang P, Shuai J, Tao F, Shi P (2014) River discharge, land use change, and surface water quality in the Xiangjiang River, China. *Hydrological Processes* 28:4130–4140

Appendix



Appendix 1. Maps of a) mean basin slope, b) normalized difference vegetation index (NDVI) and c) land use for the Lestijoki river catchment area.