Landscape metrics as indicators of river water quality at catchment scale

Evelyn Uuemaa, Jüri Roosaare and Ülo Mander*

Institute of Geography, University of Tartu, Vanemuise 46, 51014 Tartu, Estonia.

*Corresponding author: E-mail: *mander@ut.ee*

Received 31 December 2005; accepted in revised form 7 December 2006

Nordic Hydrology **Vol 38 No 2 pp 125–138** © IWA Publishing 2007

Abstract We investigated the relationship between land use parameters and FRAGSTATS-based landscape metrics (Edge Density, Patch Density, Mean Shape Index, Mean Euclidean Nearest Neighbour Index, Contagion, Patch Richness Density and Shannon's Diversity Index) and nutrient/organic-matter-based water quality indicators (BOD_7 and COD_{KMnO4} values, total-N and total-P concentrations in water) in 24 catchments with various land use patterns in Estonia, using the CORINE Land Cover Map (1:100 000). Multiple regression analysis showed that, for BOD_7, total-N and total-P, the most important predictor was the proportion of urban areas, but landscape metrics also had a significant relationship with water quality. Mean Shape Index and Contagion were the most important predictors for COD_{KMnO4}. The knowledge that land use and landscape configuration impact water quality can be used in establishing and implementing water management plans in Europe.

Keywords Catchments; land use; landscape metrics; water quality

Introduction

The influence of land use on water quality in streams is scale-dependent and varies over time and space (Behrendt *et al.* 2002; Buck *et al.* 2004). Numerous studies have found landscape structure to be the main factor influencing nutrient and organic matter runoff from watersheds. This has been shown at the global scale (Turner *et al.* 2003) as well as at the regional and local scales for catchments of predominantly agricultural use (Young *et al.* 1987; McDowell *et al.* 2001; Cao *et al.* 2003), for urban (Wickham *et al.* 2002) and forested areas (Wickham *et al.* 2003), and for heterogeneous multifunctional landscapes (Stålnacke *et al.* 1999; Arheimer and Brandt 2000; Mander *et al.* 2000; Chen *et al.* 2002; Steinhardt and Volk 2003). In relation to material export, different landscape metrics have been performed for the description of landscape structure in catchments: areas of landscape elements and the distances of landscape elements to water bodies (Thierfelder 1998); topography elements (Jones *et al.* 2001), the presence of riparian zones (Kuusemets and Mander 2001), wetlands (Trepel and Palmeri 2002) and various diversity indices (Jones *et al.* 2001; Chen *et al.* 2002; Gergel *et al.* 2002). Several studies have shown stream water quality to be effectively detectable using remote sensing data (Griffith *et al.* 2002; Davenport *et al.* 2003). Johnson *et al.* (2001) have found that using only landscape measurements obtained solely from remotely sensed data can explain about 75% of the water quality variability in catchments.

Hundreds of landscape metrics indices have been proposed by various researchers to analyse landscape structure. Most of them are covered by the FRAGSTATS computer program (McGarigal *et al.* 2002). Since the emergence of FRAGSTATS (the first version appeared in 1993, but widescale use began with public domain v.2.0), the measures and

doi: 10.2166/nh.2007.002

methods incorporated into this software have been very widely used. In landscape ecology, FRAGSTATS is the *de facto* standard for calculating landscape metrics. Therefore we limited ourselves to the possibilities offered by FRAGSTATS. Surprisingly, we could not find any papers analysing the relationship between material transport from catchments and FRAGSTATS-based landscape indices.

The aim of our study was to ascertain the relationship between the quantitative and qualitative parameters of the spatial structure of landscape and nutrient/organic-matter-based water quality indicators in catchments with various land use patterns.

Material and methods

Water quality data

We used the water quality data from the database of the Estonian Environmental Monitoring Programme database. Fifty-seven catchments are included in the Environmental Monitoring Programme. Of these 57 catchments, we could only use 24, because many catchments extended almost all the way to Russia or Latvia. We decided to use multiple regression analysis to detect relationships between water quality data and landscape metrics because then results are most easily interpretable. Therefore we could not use more than one subcatchment of the catchment, or else the data points would not have been independent. Also the computational limits of FRAGSTATS restricted the choice of catchments. Of the monitoring programme data, we used BOD_7 and COD_{KMnO4} values, and total-N and total-P concentrations in water samples from the closing weirs of the studied rivers ($mg\,l^{-1}$). Johnson *et al.* (1997) suggest that concentration data should be used when examining landscape influence on water quality. For annual concentrations, we calculated arithmetical means for the years 1996–1998 (the land cover maps were made in this period). However, we presume that the selected catchments present a representative sample of all Estonian catchments.

The disadvantage of these data was their dependence on point pollution sources (towns, factories). However, the relation between the Biological Oxygen Demand (BOD_7) and the Chemical Oxygen Demand (determined based on potassium permanganate; COD_{KMnO4}) helps us to distinguish between anthropogenic (mostly point pollution) sources and natural/semi-natural sources of pollution. In particular, high BOD_7 values indicate the presence of point-pollution sources (urban and industrial areas, settlements), whereas the COD_{KMnO4} value is high in natural areas with a high percentage of swamps, fens and bogs (Behrendt *et al.* 2002). Therefore the COD_{KMnO4} is closely comparable with the widely used Dissolved Organic Carbon (DOC) value (Qualls and Richardson 2003).

Land cover data

To determine relationships between landscape metrics and water quality, we used the land cover maps of 24 Estonian catchments (Figure 1) that were included in the Estonian Environmental Monitoring Programme.

The data used for the calculation of landscape metrics was derived in raster form from the CORINE Land Cover Map of Estonia (1:100 000). The CORINE database allows the use of this method in other parts of Europe. The spatial resolution was 30 m, and landscape metrics were calculated using FRAGSTATS on the original CORINE Land Cover types that are listed in the first column of Table 1.

In order to ascertain the relationships between land use and water quality, we reclassified CORINE land cover types into four general groups: (1) the proportion of natural areas (NA) (forests, grasslands); (2) the proportion of agricultural land use (ALU); (3) the proportion of fens, bogs and mires (FBM); and (4) the proportion of urban land use (ULU) (Table 1). These

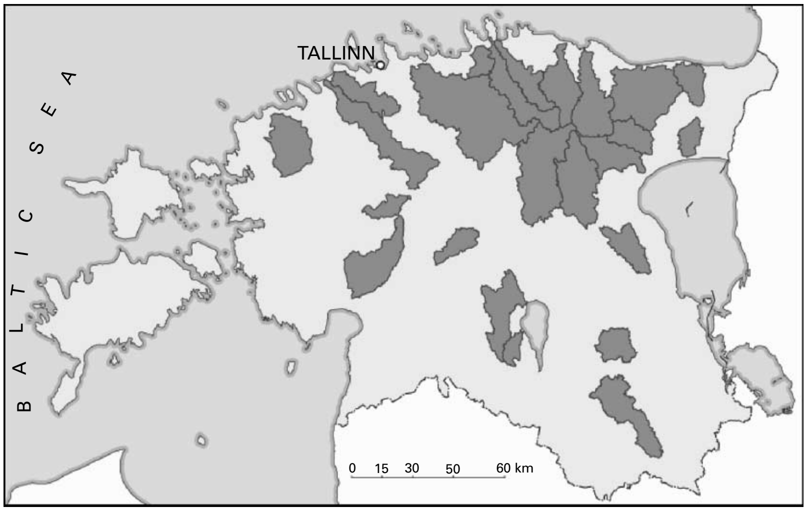

Figure 1 Location of study catchments

land use proportions were used in the regression analysis. Mining lands, dump sites and peat bogs were classified as Other and were not used in the analysis.

Landscape metrics

The landscape metrics were analysed using the computer programme FRAGSTATS (McGarigal *et al.* 2002). Many of the landscape metrics are correlated with each other. Therefore we performed a correlation analysis and picked those landscape metrics that did not correlate significantly with the others (Uuemaa *et al.* 2005). There was only one exception – patch density, which correlated with edge density, but is very often used. We used the following landscape metrics: Edge Density (ED), Patch Density (PD), Mean Shape

Table 1 Land use and land cover types in study catchments

CORINE Land Cover Map 1:100 000	New classification
Lakes, ponds	Natural areas
Water courses	Natural areas
Non-irrigated arable land	Agricultural land use
Urban lands	Urban land use
Mining lands	Other
Dump site	Other
Inland marshes	Fens, bogs and mires
Bogs	Fens, bogs and mires
Peat bogs	Other
Salt marshes	Fens, bogs and mires
Green urban areas	Urban land use
Sport and leisure facilities	Urban land use
Fruit trees and berry plantations	Agricultural land use
Pastures	Agricultural land use
Coniferous forest	Natural areas
Mixed forest	Natural areas
Natural grassland	Natural areas
Moors and heath land	Natural areas
Sparsely vegetated areas	Natural areas

Index (SHAPE_MN), Mean Euclidean Nearest Neighbour Index (ENN_MN), Contagion (CONTAG), Patch Richness Density (PRD) and Shannon's Diversity Index (SHDI; Table 2).

Statistical analysis

According to the Kolmogorov–Smirnov test for normality, all of the variables under consideration were normally distributed, except for one variable – total-P. Stepwise multiple regression analysis was used to analyse the dependent variables of water quality, and landscape metrics and land use proportions as independent variables. Regressions were tested using an ANOVA test and the normality of residuals. A casewise plot of residuals was used to calculate the number of possible outliers. The regression models were validated through an examination of their predictability using an independent data set. In order to achieve this, we randomly selected four catchments out of 24 and performed stepwise multiple regression analysis on 20 catchments. A statistically significant regression model was found for every water quality parameter, and these models were validated on the four catchments that were initially left out of the analysis. The procedure was performed six times, and in each analysis we randomly selected four new catchments for the validation of the models. Repetitions were avoided, and therefore six different combinations were possible. In Table 3, 'No. of random selection' indicates the number of selections when a catchment was used for validation. For all predicted values, we calculated 95% prediction intervals using STATISTICA 6.0.

The probability to enter variables into the stepwise regression model was set at $p < 0.01$ and the probability to remove was set at $p < 0.05$. For the statistical analysis of data, the computer program STATISTICA 6.0 was used. The level of significance of $\alpha = 0.05$ was accepted in all cases.

Results and discussion

BOD_7

Regressions of landscape metrics and land use proportions explained up to 82% of the observed variation in BOD_7 (Table 4), and the percentage of outliers was relatively low. ULU was the most important predictor in all regression equations, except in the *VI* model. Catchments with high ULU had higher values of BOD_7. PD was also present in four equations out of six and had high β values. As the PD showed a negative correlation with BOD_7, lower amounts of organic matter are washed out of the catchments with high PD values. In the *VI* regression equation, the most important predictors are ALU, NA and FBM. According to the regression model, the catchments with high ALU, NA and FBM should have low BOD_7 values. This is probably because catchments with high ALU, NA and FBM have higher ULU. The importance of ULU in regression equations shows the role of point-pollution sources in BOD_7 values. It seems, however, that landscape complexity also significantly influences BOD_7 values.

Predicted and measured values of BOD_7 for control catchments are presented in Table 5. In model validation, the difference between predicted and measured values of BOD_7 were small (Table 5). Only in the case of the Purtse and Pudisoo catchments did the observed value not lie within the 95% prediction limits. The model underestimates the value of BOD_7 for the Purtse catchment, because there are lots of factories. For Pudisoo, the model calculated a higher value than that measured (Table 5), because there is high FBM (Table 3), and it has relatively high β values (Table 4).

COD_{KMnO4}

For COD_{KMnO4}, different random selections gave relatively different equations (Table 4), which explained up to 94% of the variation in COD_{KMnO4} values. The percent of outliers was

Table 2 List of landscape metrics used (McGarigal *et al.* 2002)

Landscape metrics	Description
Edge Density (ED)	The total length of all edge segments per ha for the landscape under consideration (unit: m/ha). ED = 0 when there is no edge in the landscape, and ED increases when the landscape is more complex (i.e. contains more edge).
Patch Density (PD)	The number of patches per unit of area. PD = 0 when there are no patches in the landscape, and PD increases when the landscape is more complex (i.e. contains more patches).
Mean Shape Index (SHAPE_MN)	A patch-level shape index averaged over all patches in the landscape: $$SHAPE\_MN = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(p_{ij}/2\sqrt{\pi \times a_{ij}}\right)/N$$ where P_{ij} is the perimeter and a_{ij} is the area of patch ij, and N is the total number of patches in the landscape (unitless). SHAPE_MN = 1 when the patches are all squares, and increases without limit as the shape of the patches becomes more irregular.
Mean Euclidean Nearest Neighbour Index (ENN_MN)	A patch-level distance (m) to the nearest neighbouring patch of the same type, based on the shortest edge-to-edge distance is averaged over all patches in the landscape: $$ENN\_MN = \sum_{i=1}^{m}\sum_{j=1}^{n} h_{ij}/N$$ where h_{ij} is distance from patch ij to nearest neighbouring patch of the same type (class), based on edge-to-edge distance between patches, computed from cell centre to cell centre, and N is the total number of patches in the landscape (unit: m). ENN_MN approaches 0 as the distances to the nearest neighbours decreases.
Contagion (CONTAG)	Indicates the aggregation of patches: $$CONTAG = \left[1 + \sum_{i=1}^{m}\sum_{k=1}^{m}\left[(P_i)\left(g_{ik}\Big/\sum_{k=i}^{m} g_{ik}\right)\right] \times \left[\ln(P_i)\left(g_{ik}\Big/\sum_{k=1}^{m} g_{ik}\right)\right]\Big/2\ln(m)\right](100\%)$$ where P_i is the proportion of the landscape occupied by patch type i; g_{ik} is the number of adjacencies between pixels of patch types (classes) i and k based on the *double-count* method; and m is the number of patch types (classes) in the landscape (including landscape border if present) (unit: %). CONTAG approaches 0 when the patch types are maximally disaggregated, and equals100 when all patch types are maximally aggregated; i.e. when the landscape consists of a single patch.
Patch Richness Density (PRD)	The number of patch types per unit area (unit: patches per 100 ha). PRD increases as the number of different patch types (i.e. patch richness) increases.
Shannon's Diversity Index (SHDI)	Based on information theory, indicates the patch diversity in the landscape: $$SHDI = -\sum_{i=1}^{m}(P_i \cdot \ln P_i)$$ where P_i is the proportion of the landscape occupied by patch type i (unitless). SHDI = 0 when the landscape contains only 1 patch (i.e. no diversity). SHDI increases as the number of different patch types (i.e. patch richness) increases and/or the proportional distribution of area among patch types becomes more equitable.

Table 3 Landscape metrics and land use proportions calculated for all catchments used in the analysis

No. of random selection	Catchment	ED	PD	SHAPE_MN	ENN_MN	CONTAG	PRD	SHDI	NA	ALU	FBM	ULU
IV	Alajõgi	38.61	0.73	2.10	628.87	60.67	0.07	1.60	73.82	15.09	10.43	0.21
II	Avijõgi	36.09	0.76	2.00	802.19	59.28	0.03	1.68	72.10	26.66	0.54	0.64
III	Jägala	36.43	0.80	2.01	861.75	60.99	0.01	1.96	59.09	30.79	8.32	1.15
V	Kääpa	43.22	1.00	2.04	509.37	61.00	0.04	1.67	68.41	28.67	1.47	0.07
IV	Keila	39.10	0.96	1.92	792.41	60.81	0.02	1.86	41.72	46.83	8.64	1.92
I	Kunda	35.79	0.83	1.98	946.02	61.97	0.03	1.82	58.07	36.57	3.76	0.69
II	Loobu	41.18	1.07	1.90	995.71	58.33	0.04	1.88	50.49	42.86	5.32	0.92
III	Pedja	32.53	0.68	1.96	849.43	61.70	0.02	1.81	61.58	33.84	2.97	0.66
II	Põltsamaa	36.91	0.85	1.94	679.98	59.96	0.02	1.87	50.10	41.71	5.85	1.40
V	Porijõgi	40.03	0.87	1.99	602.07	65.74	0.04	1.35	44.74	54.17	0.37	0.30
VI	Pudisoo	36.28	0.88	1.96	590.79	62.00	0.08	1.61	71.77	17.10	9.28	1.59
I	Pühajõgi	35.39	0.83	1.92	756.59	60.52	0.06	1.86	59.35	31.60	0.38	7.52
V	Purtse	33.57	0.69	2.01	774.00	59.60	0.02	2.06	63.07	23.59	8.14	3.09
II	Saarjõgi	39.66	0.82	2.08	575.97	57.21	0.05	1.68	77.14	20.82	1.87	0.00
III	Sauga	36.61	0.80	2.04	676.73	55.10	0.02	2.01	50.39	30.61	12.48	0.90
VI	Seljajõgi	34.28	0.85	1.89	756.44	64.68	0.03	1.64	29.62	66.08	0.27	3.71
VI	Tagajõgi	37.69	0.82	2.05	650.72	57.76	0.04	1.74	82.98	7.26	9.31	0.33
I	Tänassilma	46.51	1.11	2.09	508.58	60.35	0.03	1.74	49.98	43.34	4.64	0.38
VI	Tarvastu	38.44	0.87	1.88	834.20	62.57	0.08	1.37	40.52	57.09	0.00	2.39
IV	Vääna	36.92	0.89	1.90	1013.85	60.52	0.05	1.94	43.41	44.99	3.83	6.46
V	Valgejõgi	40.36	1.10	1.90	891.29	56.03	0.03	2.06	57.33	28.47	10.74	2.69
III	Velise	41.54	0.99	1.96	518.16	59.29	0.07	1.65	71.46	21.32	6.84	0.02
IV	Vihterpalu	37.64	0.80	2.09	632.24	59.07	0.03	1.92	62.75	16.56	18.71	0.56
I	Võhandu	40.24	0.89	2.01	586.56	63.72	0.03	1.61	50.76	45.40	0.83	0.96

Table 4 Regression equations and main statistics for water quality characteristics

No. of random selection	Regression equation	R^2	Adjusted R^2	F	P	See	Outliers %	β
I	$BOD_7 = -3.48PD + 0.25\,ULU + 4.7$	0.50	0.44	8.33	0.003	0.58	10%	PD $\beta = -0.52$; ULU $\beta = 0.52$
II	$BOD_7 = 0.21ED - 7.68PD + 1.3SHDI + 0.35ULU - 1.86$	0.76	0.69	11.72	0.00016	0.49	5%	ED $\beta = 0.78$; PD $\beta = -1.0$; SHDI $\beta = 0.30$; ULU $\beta = 0.83$
III	$BOD_7 = 0.18ED\text{-}7.87PD + 1.65SHDI + 0.32\,ULU - 1.15$	0.82	0.77	16.61	0.000022	0.43	0%	ED $\beta = 0.61$; PD $\beta = -1.0$; SHDI $\beta = 0.35$; ULU $\beta = 0.74$
IV	$BOD_7 = 0.23ED\text{-}8.28PD + 1.46SHDI + 0.37ULU - 2.57$	0.74	0.68	10.94	0.000235	0.48	0%	ED $\beta = 0.94$; PD $\beta = -1.2$; SHDI $\beta = 0.34$; ULU $\beta = 0.77$
V	$BOD_7 = 4.78SHAPE\_MN + 0.35ULU - 7.98$	0.74	0.71	12.71	0.000011	0.37	0%	SHAPE_MN $\beta = 0.51$; ULU $\beta = 1.06$
VI	$BOD_7 = 0.12CONTAG - 0.32NA - 0.37ALU - 0.36FBM + 27.55$	0.77	0.71	12.76	0.000101	0.48	5%	CONTAG $\beta = 0.34$; NA $\beta = -3.8$; ALU $\beta = -4.6$; FBM $\beta = -2.0$
I	$COD_{KMnO4} = 64.52SHAPE\_MN - 0.87\,CONTAG - 58.93$	0.86	0.84	51.79	0.000000	2.43	0%	SHAPE_MN $\beta = 0.75$; CONTAG $\beta = -0.37$
II	$COD_{KMnO4} = -0.80ED + 54.25SHAPE\_MN - 1.83CONTAG - 9.74SHDI + 66.35$	0.87	0.83	24.14	0.000002	2.47	0%	ED $\beta = -0.45$; SHAPE_MN $\beta = 0.64$; CONTAG $\beta = -0.79$; SHDI $\beta = -0.33$
III	$COD_{KMnO4} = 56.71SHAPE\_MN + 2.84NA + 2.7\,ALU + 3.1FBM + 3.97ULU - 377.7$	0.86	0.81	17.14	0.000016	2.57	0%	SHAPE_MN $\beta = 0.75$; NA $\beta = 6.81$; ALU $\beta = 7.20$; FBM $\beta = 2.60$; ULU $\beta = 1.40$
IV	$COD_{KMnO4} = -0.7ED + 63.93SHAPE\_MN - 1.26\,CONTAG + PRD87.55 - 12.49$	0.89	0.86	29.69	0.000001	2.12	0%	ED $\beta = -0.43$; SHAPE_MN $\beta = 0.72$; CONTAG $\beta = -0.61$; PRD $\beta = 0.32$
V	$COD_{KMnO4} = -0.64ED + 47.52SHAPE\_MN - 1.75CONTAG - 13.37SHDI + 0.146FBM + 0.9ULU + 70.69$	0.94	0.86	31.24	0.000000	1.78	5%	ED $\beta = -0.32$; SHAPE_MN $\beta = 0.57$; CONTAG $\beta = -0.64$; SHDI $\beta = -0.35$; FBM $\beta = 0.36$; ULU $\beta = 0.30$
VI	$COD_{KMnO4} = 1.49ED - 48.86PD - 1.02CONTAG - 0.31NA - 0.41ALU + 94.38$	0.87	0.83	19.21	0.000008	2.31	0%	ED $\beta = 0.89$; PD $\beta = -1.1$; CONTAG $\beta = -0.44$; NA $\beta = -0.59$; ALU $\beta = -0.81$

Table 4 – *continued*

No. of random selection	Regression equation	R^2	Adjusted R^2	F	P	See	Outliers %	β
I	N-Total $= 0.46$ULU $+ 1.83$	0.49	0.46	17.05	0.00063	0.78	0%	ULU $\beta = 0.7$
II	N-Total $= -0.2$ED $+ 11.11$SHAPE_MN $- 0.34$CONTAG $+ 0.67$NA $+ 0.74$ALU $+ 0.63$FBM $+ 0.98$ULU $- 60.06$	0.81	0.70	7.24	0.001572	0.63	0%	ED $\beta = -0.59$; SHAPE_MN $\beta = 0.69$; CONTAG $\beta = -0.76$; NA $\beta = 7.74$; ALU $\beta = 9.97$; FBM $\beta = 2.77$; ULU $\beta = 1.79$
III	N-Total $= 0.37$ULU $+ 1.92$	0.46	0.42	15.03	0.001106	0.86	0%	ULU $\beta = 0.68$
IV	N-Total $= -0.17$ED $+ 0.02$ALU $+ 8.0$	0.48	0.42	7.89	0.003764	0.79	0%	ED $\beta = -0.58$; ALU $\beta = 0.41$
V	N-Total $= 0.4$ULU $+ 1.97$	0.50	0.48	18.23	0.000461	0.82	0%	ULU $\beta = 0.709$
VI	N-Total $= 0.34$ULU $+ 1.89$	0.49	0.46	17.38	0.000577	0.72	0%	ULU $\beta = 0.7$
I	P-Total $= -0.05$ED $+ 1.3$PD $+ 1.71$SHAPE_MN $+ 1.86$PRD $+ 0.006$ALU $- 2.62$	0.80	0.73	11.28	0.00016	0.04	0%	ED $\beta = -2.1$; PD $\beta = 2.13$; SHAPE_MN $\beta = 1.72$; PRD $\beta = 55$; ALU $\beta = 1.29$
II	P-Total $= -0.00035$ENN_MN $+ 0.06$ULU $+ 0.24$	0.66	0.62	16.52	0.000103	0.08	5%	ENN_MN $\beta = -0.40$; ULU $\beta = 0.96$
III	P-Total $= -0.0003$ENN_MN $+ 0.058$ULU $+ 0.21$	0.64	0.60	15.19	0.000165	0.08	5%	ENN_MN $\beta = -0.35$; ULU $\beta = 0.91$
IV	P-Total $= 0.065$ULU $- 0.0068$	0.74	0.73	52.59	0.000001	0.07	15%	ULU $\beta = 0.86$
V	P-Total $= 0.051$ULU $+ 0.01$	0.63	0.61	30.66	0.000029	0.08	5%	ULU $\beta = 0.79$
VI	P-Total $= 0.043$ULU $+ 0.0086$	0.55	0.53	22.4	0.00017	0.08	10%	ULU $\beta = 0.75$

Table 5 Results of model estimation. Results where the predicted value did not lie within the 95% prediction intervals are in bold

No of random selection	Catchment name	BOD_7 Predicted	BOD_7 Measured	BOD_7 Residual	COD_{KMnO_4} Predicted	COD_{KMnO_4} Measured	COD_{KMnO_4} Residual	Total-N Predicted	Total-N Measured	Total-N Residual	Total-P Predicted	Total-P Measured	Total-P Residual
I	Kunda	1.97	2.13	0.16	14.45	11.66	− 2.79	2.15	3.07	0.92	**0.14**	**0.035**	**− 0.105**
I	Pühajõgi	3.67	4.08	0.41	11.86	14.06	2.2	5.29	4.04	− 1.25	**0.099**	**0.57**	**0.471**
I	Tänassilma	0.93	1.53	0.6	**23**	**12.93**	**− 10.07**	2	1.87	− 0.13	0.14	0.085	− 0.055
I	Võhandu	1.85	1.66	0.19	15.13	8.09	− 7.04	2.27	0.86	− 1.41	0.072	0.037	− 0.035
II	Põltsamaa	2.09	1.51	− 0.58	14.38	9.57	− 4.81	3.09	3.1	0.01	0.085	0.035	− 0.05
II	Avijõgi	2.1	2.23	0.13	21.37	16.16	− 5.21	3.6	3.33	− 0.27	− 0.0036	0.026	0.0296
II	Saarjõgi	2.17	1.94	− 0.23	26.34	23.25	− 3.09	3.64	1.48	− 2.16	0.037	0.03	− 0.007
II	Loobu	1.09	1.29	0.2	11.49	11.28	− 0.21	2.6	2.18	− 0.42	**− 0.055**	**0.053**	**0.108**
III	Jägala	2.57	1.77	− 0.8	17.43	14.99	− 2.44	2.34	2.4	0.06	0.01	0.051	0.041
III	Pedja	2.42	1.68	− 0.74	11.54	13.29	1.75	2.16	2.83	0.67	− 0.01	0.036	0.046
III	Sauga	2.66	2.29	− 0.37	**6.05**	**26.77**	**20.72**	2.25	1.63	− 0.62	0.056	0.076	0.02
III	Velise	1.1	1.71	0.61	15.42	15.05	− 0.37	1.92	1.32	− 0.6	0.05	0.021	− 0.029
IV	Alajõgi	2.64	2.43	− 0.21	23.98	24.42	0.44	1.72	2.39	0.67	0.067	0.037	− 0.03
IV	Keila	1.89	1.9	0.01	7.74	12.65	4.91	2.57	3.38	0.81	0.12	0.09	− 0.03
IV	Vääna	3.74	3.44	− 0.3	11.47	14.83	3.36	**2.89**	**4.77**	**1.88**	**0.41**	**0.1**	**− 0.31**
IV	Vihterpalu	2.47	1.79	− 0.68	22.85	25.03	2.18	1.93	1.87	− 0.06	0.03	0.036	0.006
V	Kääpa	1.78	1.9	0.12	**11.42**	**17.851**	**6.431**	2	1.82	− 0.18	0.014	0.04	0.026
V	Porijõgi	1.64	1.43	− 0.21	6.95	8.64	1.69	2.09	1.58	− 0.51	0.025	0.056	0.031
V	Purtse	**2.73**	**4.51**	**1.78**	19.52	19	− 0.52	3.19	2.52	− 0.67	0.17	0.029	− 0.141
V	Valgejõgi	2.02	1.32	− 0.7	**16.59**	**11.1**	**− 5.49**	3.03	1.65	− 1.38	0.15	0.042	− 0.108
VI	Pudisoo	**2.78**	**1.39**	**− 1.39**	12.73	14.16	1.43	2.43	1.77	− 0.66	0.077	0.08	0.003
VI	Seljajõgi	1.66	2.12	0.45	**1.07**	**8.63**	**7.56**	**3.14**	**5.01**	**1.87**	**0.17**	**0.35**	**0.18**
VI	Tagajõgi	2.3	2.66	0.36	22.49	26.3	3.81	2.0	2.5	0.5	0.023	0.047	0.024
VI	Tarvastu	1.35	1.88	0.53	9.16	8.94	− 0.22	2.69	3.01	0.32	0.11	0.069	0.041

very low (Table 4). For COD_{KMnO4}, the most important predictors were SHAPE_MN and CONTAG, which had the highest β values. Lower amounts of humic and fluvic acids are washed out of the catchments with low SHAPE_MN values and high CONTAG values (fragmented landscape structure). Land use proportions were also important predictors in regression equations for COD_{KMnO4}. ALU, NA and FMB had high (β values, which show that organic matter losses are higher from natural areas, swamps, fens and bogs (Behrendt *et al.* 2002). Therefore the COD_{KMnO4} values are higher when fens and natural areas account for a high proportion of the catchment's land use, and landscape is fragmented.

Validation results were not as good in the case of COD_{KMnO4}, although the R^2 was always greater than 0.8. Regression model *I* strongly overestimated COD_{KMnO4} values for the Tänassilma and Võhandu catchments (Table 5). Both catchments have a very fragmented landscape (high SHAPE_MN values), but not as high a proportion of fens, bogs, mires and natural areas (Table 3), which are mostly the source of humic and fulvic acids. The SHAPE_MN and CONTAG values most likely reflect the catchments' land use as natural areas, and fens, bogs and mires tend to have a more complex and less compact shape (Uuemaa *et al.* 2005). Regression model *III* very heavily underestimates the COD_{KMnO4} value in the Sauga catchment (Table 5). The Sauga catchment has a very high FBM (Table 3), from which most of the humic and fulvic acids come. However, ALU and NA have the highest β values apart from FBM. The *V* regression model gives a higher COD_{KMnO4} value for Valgejõgi, and a lower COD_{KMnO4} value than measured for the Kääpa catchment. There is strong groundwater input in the Valgejõgi catchment, which probably dilutes the concentration of COD_{KMnO4}. The Kääpa catchment has a very low ULU, which is not an important source for humic and fulvic acids, but is included in the model as a predictor for COD_{KMnO4}. Model *VI* underestimates the COD_{KMnO4} value for the Seljajõgi catchment (Table 5), because there is high ALU (Table 3) that has high β values, although NA is a more important source for humic and fulvic acids.

Total-N

Regression equations explained up to 81% of the observed variation of total-N (Table 4). In equations *I, III, V* and *VI*, the only predictor for total-N was ULU, which explained up to 50% of the observed variation of total-N. Although the R^2 is not very high, the regressions have significant ANOVA tests, and the percent of outliers was zero in all cases (Table 4). Figure 2 shows that, with increasing ULU values, the total-N value also increases. The good correlation between total-N and ULU indicates the insufficient wastewater treatment in the catchments (Ahearn *et al.* 2005). This is a critical problem, especially in catchments that belong to the Gulf of Finland basin. In equation *IV* there is, in addition to ALU, also an ED predictor for total-N – with increasing ED values, the total-N concentration decreases (Figure 3). ALU and NA had very high β values in equation *II*, which refers to the important role of land use in total-N runoff (Arheimer and Liden 2000; Iital *et al.* 2003; Uuemaa *et al.* 2005).

The results of model estimation were good (Table 5). Only two total-N values did not lie between the 95% prediction intervals. Model *IV* calculates a lower total-N value for the Vääna catchment, and model *VI* underestimates nitrogen runoff from Seljajõgi catchment. In both catchments, the high loads of nitrogen come from point source pollution. Regression equation *IV* does not take into account ULU, but there are several towns in the Vääna catchment. Therefore the model underestimates the value of total-N. *VI* gives a lower total-N value than the measured model, which does not take into account ALU, which is very high in the Seljajõgi catchment (Table 3).

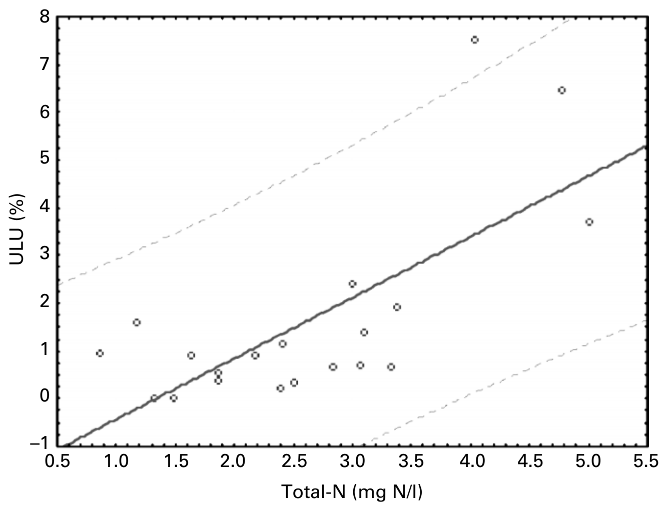

Figure 2 Relationship between ULU and total-N (V equation). The solid line is the linear regression equation. The dashed lines indicate the 95% prediction intervals

Total-P

For total-P, ULU had the highest β values. Regression equations where the only predictor was ULU explained up to 74% of the measured variation, and ANOVA tests were significant (Table 4). The percentage of outliers was relatively high, probably because values of total-P were not normally distributed. The close correlation between total-P and ULU points to the problem with phosphorus removal from wastewater originating from industries and towns (Sliva and Williams 2001; Yin *et al.* 2005). Figure 4 shows that total-P is positively correlated with ULU. The relationship between ULU and total-P is very close, because the Seljajõgi and Pühajõgi catchments have high values of ULU and total-P. These two have a

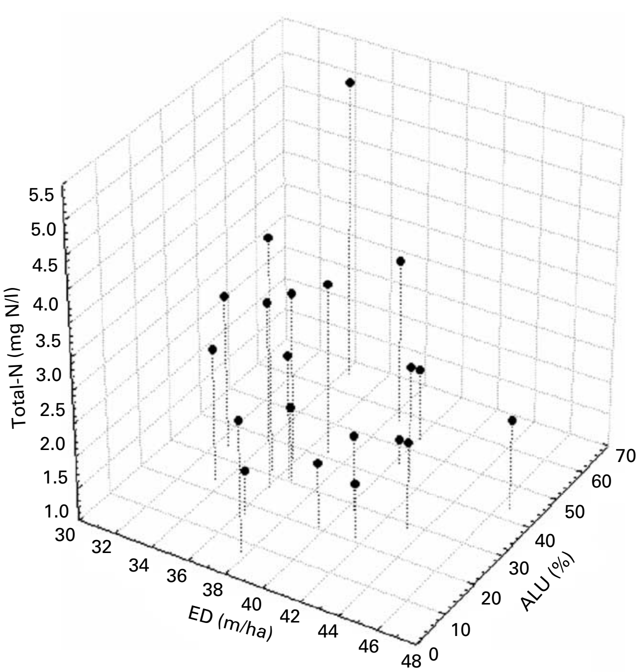

Figure 3 Relationship between ED, ALU and total-N

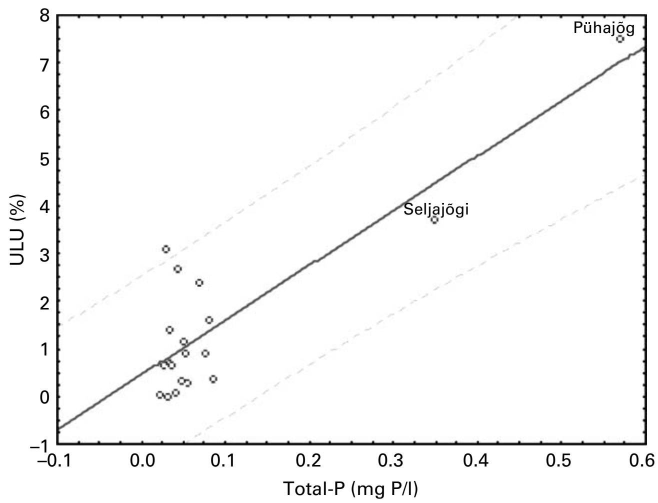

Figure 4 Relationship between ULU and total-P (*IV* equation). The solid line is the linear regression equation. The dashed lines indicate the 95% prediction intervals

lot of industries and high population densities, which are the main source of phosphorus. The removal of phosphorus in wastewater treatment plants is evidently insufficient. If these two catchments are left out of the analysis, there is no relationship between ULU and total-P. As these analyses were performed with water quality data from the years 1996–1998, for the year 2003 the value of total-P in the Pühajõgi catchment has decreased almost four times. However, the relationships between total-P and ULU may lead to the wrong conclusion, because urban areas are impervious, and drainage is frequently routed to wastewater treatment plants (which may or may not be in the same basin), then discharged to local rivers as point sources (Ahearn *et al.* 2005). Furthermore, Ahearn *et al.* (2005) found that population density describes water quality better than the percentage of urban areas.

The results of the model estimation were not very good (Table 5). The difference between measured and predicted total-P values were quite great considering the actual total-P values. Regression model *I* strongly underestimated the total-P value for the Pühajõgi catchment. In that regression equation, ULU is not included as a predictor for total-P, but there is high ULU (7.5%) in the Pühajõgi catchment. In the case of the Kunda catchment, the ED is relatively low, and in the regression equation the β value for ED is very high. In regression model *II*, the Loobu catchment has a high ENN_MN value that causes a very low predicted value of Total-P. Regression model *IV* overestimates the value of total-P for the Vääna catchment. The ULU is high in the Vääna catchment, but most of the wastewater coming from Tallinn (part of Tallinn belongs to the Vääna catchment) is routed to the wastewater management plant and then discharged to the sea. Model *VI* gives a lower value for total-P than measured, because there are problems with wastewater management.

Conclusions

Land use proved to be most important predictor for water quality, but landscape structure also had a significant role in predicting the values of water quality in catchments. For BOD_7, total-P and total-N ULU was evidently the most significant predictor, because in catchments that belong to the Gulf of Finland basin, organic matter, nitrogen and phosphorus runoff is strongly influenced by point-pollution sources. There are also problems with waste water treatment in many Estonian catchments. In addition to ULU, ALU and ED seemed to play an important role in predicting values of total-N. For BOD_7, PD was also an important predictor. Catchments with complex landscape configuration have lower organic matter runoff. Lower amounts of humic and fulvic acids are washed out of catchments with complex landscape structures (low CONTAG and high SHAPE_MN). We can conclude that, even with the same land use, landscape configuration has an important role in organic matter and nutrient runoff from catchments.

Although the regression models used in this study should not be used in other catchments, the methods can be applied anywhere in Europe. The CORINE Land Cover Map may be used for landscape metrics calculation, and water quality data is also available.

Knowledge of the impact of land use and landscape configuration on water quality should be taken into account in land-use planning in watersheds. According to the EU Water Framework Directive, river basin management plans should be established, and the land–water relationship can be easily used by planners there.

Acknowledgements

This study was supported by the EU 5 FP RTD project PRIMROSE (EVK1-2000-00728) "Process Based Integrated Management of Constructed and Riverine Wetlands for Optimal Control of Wastewater at Catchment Scale", Estonian Science Foundation grant no. 6083, and Target Funding Projects no. 0182534s03 of the Ministry of Education and Science, Estonia.

References

Ahearn, D.S., Sheibley, R.W., Dahlgren, R.A., Anderson, M., Johnson, J. and Tate, K.W. (2005). Land use and land cover influence on water quality in the last free-flowing river draining the western Sierra Nevada, California. *J. Hydrol.*, **313**(3–4), 1–14.

Arheimer, B. and Brandt, M. (2000). Watershed modelling of non-point nitrogen from arable land to the Swedish coast in 1985 and 1994. *Ecol. Eng.*, **14**(4), 389–404.

Arheimer, B. and Liden, R. (2000). Nitrogen and phosphorus concentrations for agricultural catchments; influence of spatial and temporal variables. *J. Hydrol.*, **227**(1–4), 140–159.

Behrendt, H., Huber, P., Kornmilch, M., Opitz, D., Schmoll, O., Scholz, G. and Uebe, R. (2002). *Nutrient Emissions into River Basins in Germany*, Federal Environmental Agency, BerlinUBA-Texte, 23/00: 1-266.

Buck, O., Niyogi, D.K. and Townsend, C.R. (2004). Scale-dependence of land use effects on water quality of streams in agricultural catchments. *Environ. Pollut.*, **130**(2), 287–299.

Chen, L.D., Fu, B.J., Zhang, S.R., Qiu, J., Guo, X.D. and Yang, F.L. (2002). A comparative study on nitrogen-concentration dynamics in surface water in a heterogeneous landscape. *Environ. Geol.*, **42**(4), 424–432.

Cao, W., Hong, H., Yue, S., Ding, Y. and Zhang, Y. (2003). Nutrient loss from and agricultural catchment and landscape modeling in Southeast China. *B. Environ. Contam. Tox.*, **71**(4), 761–767.

Davenport, I.J., Silgram, M., Robinson, J.S., Lamb, A., Settle, J.J. and Willig, A. (2003). The use of earth observation techniques to improve catchment-scale pollution prediction. *Phys. Chem. Earth*, **28**(33–36), 1365–1376.

Gergel, S.E., Turner, M.G., Miller, J.R., Melack, J.M. and Stanley, E.H. (2002). Landscape indicators of human impacts to riverine systems. *Aquat. Sci.*, **64**(2), 118–128.

Griffith, J.A., Martinko, E.A., Whistler, J.L. and Price, K.P. (2002). Preliminary comparison of landscape pattern-normalized difference vegetation index (NDVI) relationships to central plains stream conditions. *J. Environ. Qual.*, **31**(3), 846–859.

Iital, A., Stålnacke, P., Deelstra, J., Loigu, E. and Pihlak, M. (2003). Effects of large-scale changes in emissions on nutrient concentrations in Estonian rivers in the Lake Peipsi drainage basin. *J. Hydrol.*, **304**(1–4), 261–273.

Johnson, G.D., Myers, W.L. and Patil, G.P. (2001). Predictability of surface water pollution loading in Pennsylvania using watershed-based landscape measurements. *J. Am. Wat. Res. Assoc.*, **37**(4), 821–835.

Johnson, L.B., Richards, C., Host, G.E. and Arthur, J.W. (1997). Landscape influences on water chemistry in midwestern stream ecosystems. *Freshwater Biol.*, **37**(1), 193–208.

Jones, K.B., Neale, A.C., Nash, M.S., Van Remortel, R.D., Wickham, J.D., Riitters, K.H. and O'Neill, R.V. (2001). Predicting nutrient and sediment loadings to streams from landscape metrics, A multiple watershed study from the United states Mid-Atlantic Region. *Landscape Ecol.*, **16**(4), 301–312.

Kuusemets, V. and Mander, Ü. (2001). Ecotechnological measures to control nutrient losses from catchments. *Wat. Sci. Technol.*, **40**(10), 195–202.

Mander, Ü., Kull, A. and Kuusemets, V. (2000). Nutrient flows and land use change in a rural catchment: A modelling approach. *Landscape Ecol.*, **15**(3), 187–199.

McDowell, R., Sharpley, A. and Folmar, G. (2001). Phosphorus export from an agricultural watershed, Linking source and transport mechanisms. *J. Environ. Qual.*, **30**(5), 1587–1595.

McGarigal, K., Cushman, S.A., Neel, M.C. and Ene, E. (2002). *FRAGSTATS, Spatial Pattern Analysis Program for Categorical Maps*. Computer software program produced by the authors at the University of Massachusetts, Amherst. Available at: http://www.umass.edu/landeco/research/fragstats/fragstats.html.

Qualls, R. and Richardson, C.J. (2003). Factors controlling concentration, export, and decomposition of dissolved organic nutrients in the Everglades of Florida. *Biogeochemistry*, **62**(2), 197–229.

Sliva, L. and Williams, D.D. (2001). Buffer zone versus whole catchment approaches to studying land use impact on river water quality. *Water Res.*, **35**(14), 3462–3472.

Stålnacke, P., Vagstad, N., Tamminen, T., Wassmann, P., Jansons, V. and Loigu, E. (1999). Nutrient runoff and transfer from land Gulf of Riga. *Hydrobiologia*, **410**, 103–110.

Steinhardt, U. and Volk, M. (2003). Meso-scale landscape analysis based on landscape balance investigations, problems and hierarchical approaches for their solution. *Ecol. Model.*, **168**(3), 251–265.

Thierfelder, T. (1998). The morphology of landscape elements as predictors of water quality in glacial/boreal lakes. *J. Hydrol.*, **207**(3–4), 189–203.

Trepel, M. and Palmeri, L. (2002). Quantifying nitrogen retention in surface flow wetlands for environmental planning at the landscape-scale. *Ecol. Eng.*, **19**(2), 127–140.

Turner, R.E., Rabalais, N.N., Justic, D. and Dortch, Q. (2003). Global patterns of dissolved N, P and Si in large rivers. *Biogeochemistry*, **64**(3), 297–317.

Uuemaa, E., Roosaare, J. and Mander, Ü. (2005). Scale dependence of landscape metrics and their indicatory value for nutrient and organic matter losses from catchments. *Ecol. Indic.*, **5**(4), 350–369.

Wickham, J.D., O'Neill, R.V., Riitters, K.H., Smith, E.R., Wade, T.G. and Jones, K.B. (2002). Geographic targeting of increases in nutrient export due to future urbanization. *Ecol. Appl.*, **12**(1), 93–106.

Wickham, J.D., Wade, T.G., Riitters, K.H., O'Neill, R.V., Smith, J.H., Smith, E.R., Jones, K.B. and Neale, A.C. (2003). Upstream-to-downstream changes in nutrient export risk. *Landscape Ecol.*, **18**(2), 195–208.

Yin, Z-Y., Walcott, S., Kaplan, B., Cao, J., Lin, W., Chen, M., Liu, D. and Ning, Y. (2005). An analysis of the relationship between spatial patterns of water quality and urban development in Shanghai, China. *Comput. Environ. Urban*, **29**(2), 197–221.

Young, R.A. Onstad, C.A., Bosch, D.D. and Anderson, W.P. (1987). *AGNPS Agricultural Non-Point Source Pollution Model-A Watershed Analysis Tool*. USDA, Conservation Research Report 35.