STATISTICAL MODELLING OF RIVERINE NUTRIENT SOURCES AND RETENTION IN THE LAKE PEIPSI DRAINAGE BASIN

A. Vassiljev* and P. Stålnacke**

*Tallinn Technical University, Ehitajate tee 5, Tallinn, 19086, Estonia (E-mail: avasil@cc.ttu.ee)
**Jordforsk – Norwegian Centre for Soil and Environmental Research, N-1432 Ås, Norway
(E-mail: per.stalnacke@jordforsk.no)

ABSTRACT
Implementation of the Water Framework Directive calls for methodologies and tools to quantify nutrient losses from diffuse sources at a river basin district scale. In this paper we elucidate the possibility to use a statistical model for source apportionment and retention of nutrients in a large transboundary drainage basin (44,000 km^2). The model-approach uses non-linear regression for simultaneous estimation of e.g. source strength, i.e. export coefficients to surface waters, for the different specified land-use or soil categories and retention coefficients for pollutants in a drainage basin. The model was tested on data from 26 water quality stations with corresponding subbasin data, i.e., land cover, point sources and atmospheric deposition, from the Estonian part of the Lake Peipsi drainage basin. Results from the model showed that it was statistically possible to derive reliable export coefficients (i.e. unit-area loads) for nitrogen on agricultural land and forests. Moreover, it were with simple empirical functions shown that lake retention was approximately 30-35% for both nitrogen and phosphorus and that the riverine retention was low for both nitrogen and phosphorus (approx. 10%). Results show that the MESAW model is a simple and powerful tool for simultaneous estimation of sources and retention of nutrient loads in a river basin.

KEYWORDS: Export coefficients, Lake Peipsi, land use, nutrients, retention, source apportionment

INTRODUCTION

Prerequisites for successful environmental management of river basins include the collection of basic environmental statistics and quantitative assessments of the riverine loads including estimation of the pollution sources and retention in the drainage basin. The nutrient level and fluxes at a specific location in a river network depends on the pollution sources in the upstream area, and the transfer, retention, and loss of nutrients in the soil, groundwater, and surface water network. This is a complex function of biological, physical, and chemical processes. Therefore, models are needed to analyse how these processes influence nutrient fluxes from pollution sources to river outlets over large spatial and temporal scales. Several models for so-called source apportionment and retention have already been developed worldwide. However, basins in Eastern European are often regarded as 'data-poor' or 'information-poor' and characterised by highly varying quantity and quality of input data. This is particularly problematic in international or transboundary waters, where the amount of data may differ between countries both in terms of quantity and quality. As most of the existing models require very detailed and spatially consistent input data, their applicability may be limited. Thus, simple models or tools are needed to address the limitations in these basins. The development of such models is still in its infancy. In this paper we elucidate the possibility to use a statistical model (i.e., MESAW) for source apportionment and retention of nutrients in a drainage basin characterised as both data-rich (on the Estonian side) and data-poor (on the Russian side), and the magnitude of the nutrient loads and sources has been uncertain for a long time.

DATA BASE AND METHODS

The MESAW-model is a statistical model for source apportionment of the riverine transport of pollutants (Grimvall & Stålnacke, 1996). This model-approach uses non-linear regression for simultaneous estimation of source strength (i.e. export coefficients to surface waters) for the different land use or soil categories and retention coefficients for pollutants in a river basin or lakes. The basic principles and major steps in the procedure are as follows: (i) estimation of riverine loads at each water quality monitoring site (ii) subdivision of the entire drainage basin into subbasins, defined by the monitoring sites for water quality and their upstream-downstream relationships (describing the river system) (iii) derivation of statistics on e.g. land use, soil type, lake area, point source emissions and other relevant data for each subbasin (iv) using a general non-linear regression expression with loads at each subbasin as the dependent/response variable and subbasin characteristics as covariates/explanatory variables. More precisely, load at the outlet of an arbitrary sub-basin can with MESAW be estimated from the following general expression (Lidèn *et al.*, 1999).

$$L_i = \sum_{j=1}^{n} (1 - R_{j,i}) L_j + (1 - R) S_i + (1 - R) P_i + (1 - R) D_i + e_i \tag{1}$$

where
L_i = load at outlet of sub-basin i;
L_j = load at outlet of nearest upstream sub-basin j;
$R_{j,i}$ = retention on the way from outlet of sub-basin j till outlet of subbasin i;
n = number of subbasins located nearest upstream;
S_i = total losses from soil to water in subbasin i;
P_i = point source discharges to waters in subbasin i;
D_i = atmospheric deposition on surface waters in subbasin i;

R = retention in subbasin i.
e_i = statistical error term.

The load at each subbasin can be decomposed into contributions from sources located in subbasins further upstream (the first term in formula above) and contributions from sources located within the subbasin under consideration (the S_i, P_i and D_i terms). It should be particulalry noted that the parameterisation of the model is flexible and can be study-area specific. The model is fitted by minimising the sum of squares for the difference in observed and estimated load. In this study, P_i and D_i was assumed to be known and S_i was assumed to be a simple function of land use according to $S_i = (b_1 a_{1i} + b_2 a_{2i} + b_3 a_{3i})$, where a_{1i}, a_{2i} and a_{3i} respectively denote the area of agricultural land (arable land and pastures, forests and other land (mainly bogs, and urban areas) in the subbasins i, and b_1, b_2 and b_3 are unknown export coefficients (i.e. emission coefficients, unit-area loads) for the three land use categories.

Nutrients are normally retained temporally or permanently in watercourses. Retention, in the model expressed as a summary expression for all hydrological and biogeochemical processes that may decrease or the transport or losses of nutrients, can be parameterised by any empirical function. In this study, we divided the retention into retention in lakes and river retention (i.e. instream retention). Both types of retention can be expressed according to the following general formula:

$$R = 1 - \frac{1}{1 + par * fact} \tag{2}$$

par = unknown parameter estimated by the model,
fact= empirical function based on the input data available (e.g. lake area).

Retention was parameterised by the simplest possible function (i.e. *fact*). More precisely, we assumed that retention in lakes was a direct function of the lake area, and riverine retention a function of the drainage area.
Retention from an arbitrary subbasin *m* to the river mouth (R_{mouth}) can be derived from:

$$R_{m,mouth} = 1 - \prod_{j=1}^{k} (1 - R_j) \tag{3}$$

where
$R_{m,\ mouth}$ = retention from the outlet of the subwatershed m on the way to the mouth of the whole river;
k = number of subbasins downstream subbasin m;
R_j = the values of retention within the different subbasin downstream subbasin m.

The estimated source strength (e.g. export coefficients for agricultural land) and retention parameters were finally used to calculate the contribution from each source and subbasin to the riverine load at the mouth(s). Further details regarding the general matrix expression for source strength and retention functions can be found in Grimvall&Stålnacke (1996).
Lake Peipsi is the fourth largest lake in Europe (Fig. 1). The drainage basin is approximately 44,000 km^2; 36% in Estonia, 57% in Russia and 7% in Latvia. Agricultural land and forests cover 42% and 40% of the total drainage basin, respectively. In this paper, we restrict the analyses to data from the Estonian part of the basin (15,700 km^2; Table 1).

Time series of total-N and total-P concentrations and data on runoff were obtained from the Estonian Environmental Information Centre. The study period was fixed to 1993-2000 and data from a total of 22 gauging stations, 26 water quality monitoring sites and precipitation sites, were collected. The same data holder also provided us with subwatershed delineation maps, a database on the point source emissions, and data on atmospheric deposition. The latter data was set to 500 kg km^{-2} for N and 5 kg km^{-2} for P. The digital CORINE land cover map was used to derive land use statistics for each of the 26 subbasins (Table 1), defined by the sites for water quality monitoring.

Water discharge at the ten water quality sites that lacked measurements were extrapolated from the most adjacent upstream sites with flow measurements. Since water discharge may vary with land use and precipitation, we have to take that into account in the extrapolation. More precisely, we firstly analysed the relationship between water discharge and land use by using the MESAW model, i.e., by replacement of load by water discharge in formula 1. The following formula was used to obtain corrected water flow.

$$Flow_{cor} = Flow * \frac{\sum_{i=1}^{3} flow_i * Area\%_{i,1}}{\sum_{i=1}^{3} flow_i * Area\%_{i,2}} \tag{4}$$

where:
Flow – water flow received using ratio between area of subwatersheds without measurements of water flow and representative subwatershed,
$flow_i$ – water flow from i-th land cover area (mm), obtained by MESAW,
$Area\%_{i,1}$ – area occupied by i-th land cover type in the subwatershed without measurements,

$Area\%_{i,2}$ – area occupied by by i-th land cover type in the subwatershed with measurements.

Annual load of tot-N and tot-P at each water quality site were finally calculated by multiplying daily water discharge with the observed or linearly interpolated concentrations. The time-averaged loads at each site are shown in Table 1.

Figure 1. *Lake Peipsi and its drainage basin with examples of water quality sites and major cities. Source: Royal Institute of Technology, Sweden (http://www.mantraeast.org/gis/).*

RESULTS AND DISCUSSION

Results showed that water flow from forests at all sites and years were lower than water flow from agricultural land (ratio on average 0.67), due to higher evapotranspiration in forests due to higher leaf-area (mainly coniferous forest). However, the ratio between flow from agricultural land and flow from forest was found not to be constant over the annual cycle. More precisely, a more thorough analysis showed that this ratio was higher during the agricultural growing season (May – October; Fig. 2) and the cold season (November-April) was characterised by almost the same water flow from agricultural land and forests.

The results from the model runs, which were conducted for each year separately, showed that the unit-area losses of nitrogen and phosphorus from agricultural land varied between 6.9-15 kg N ha^{-1} and 0.17-0.63 kg P ha^{-1}. Interestingly, our estimates corroborate well with the results of monitored losses from a small agricultural catchment (i.e., Oostriku, 29.7 km^2). In addition, our results showed that the nitrogen losses from agricultural land were almost four times higher than the corresponding losses from forested land (Table 2). The annual estimated emission coefficients (i.e. unit-area loads) were then used to establish relationships between the emission coefficients and water discharge individual years for each land use category (Table 2).

As already mentioned in the previous section, retention in lakes and retention in the river systems were parameterised separately. The area of lake was used as covariate for retention in lake and area of subwatershed for retention in the river system. For nitrogen, the parameter for the lake retention were found to be highly statistically significant all years ($p<0.01$) while the corresponding analyses of retention in river systems were much more uncertain ($p>0.05$). For phosphorus, the estimated parameters were much more uncertain ($p<0.05$). This is perhaps not surprising since it is well-known that nutrient retention capacity a river basin is dependent from other factors: trophic status, depth of the waterbody, water residence time, nitrogen loading, loading of organic matter, denitrification activity (primarily dependent on sufficient amounts of reducable organic substrates, low oxygen concentrations and high temperature). Regardless of the simple parameterisation of retention and the uncertain parameter estimates (especially for phosphorus) our results clearly indicate that retention in lakes is substantial for both nitrogen and phosphorus (Table 3). Nitrogen retention in lakes in the North Atlantic Ocean region has in literature been reported to range from 20 to 80% (Howarth et al., 1996). In the Nordic/Baltic region, lake retention is generally regarded as high. Jansson and co-workers (1994), proposed that productive lakes might remove up to 50% of total N-input. Studies carried out in Sweden show 50 % retention of total nitrogen in two eutrophic lakes with water residence time 2,5 years (Ahlgren et al., 1994). Svendsen and Kronvang (1993) estimated that the sedimentation and denitrification rate in Danish lakes vary between 33-48%. Information in the Baltic States is much scarcer. It is no doubt that lakes generally act as nutrient sinks, especially lakes in a steady state (equilibrium). However, at certain circumstances, lakes may act as a nutrient source rather than a sink. For example, Svendsen et al. (1995) showed negative P-retention for a shallow lowland lake in Denmark.

Table 1. Drainage area, time-averaged annual Tot-N and Tot-P loads, land cover distribution and urban point sources in subbasins of the Estonian part of the Lake Peipsi drainage basin.

Subbasin	Area (km^2)	N-load (tonnes yr^{-1})	P-load (tonnes yr^{-1})	Agriculture land (%)	Forest (%)	Other land (%)	Lakes (%)	Point sources (tonnes N yr^{-1})	Point sources (tonnes P yr^{-1})
Võhandu – Vagula	495	138	5.4	16.3	75.3	6.5	2.0	0.74	0.18
Võhandu – Himmiste	848	292	22.1	24.0	71.2	4.2	0.7	2.58	0.63
Võhandu Räpina	1144	375	24.6	19.2	65.8	13.9	0.4	37.93	6.55
V-Emajõgi – Tõlliste	1054	535	26.6	28.2	65.3	5.4	1.0	49.07	8.43
V-Emajõgi – Pikasilla sild	1270	601	26.2	15.1	77.5	7.4	0.0	49.63	8.53
Õhne Tõrva	266	120	4.3	17.7	64.3	15.9	1.8	2.61	0.44
Õhne Suislepa	577	318	10.0	31.3	62.0	6.4	0.2	6.10	1.22
Tänassilma Oiu	454	281	12.8	26.9	62.3	10.5	0.2	17.21	2.87
Emajõgi Rannu-Jõesuu	3374	827	35.6	24.5	46.1	5.2	23.8	83.60	14.38
Emajõgi Tartu	7828	3294	96.6	29.5	57.3	12.1	1.0	160.56	31.92
Emajõgi Kavastu	8539	3828	147.2	40.8	47.1	11.1	0.3	322.18	56.93
Pedja Jõgeva	665	437	6.4	27.8	64.2	7.9	0.0	1.90	0.82
Pedja Tõrve	776	537	10.7	35.7	49.7	13.8	0.8	24.47	4.47
Põltsamaa Rutikvere	861	694	7.8	33.4	54.4	11.7	0.5	4.91	1.85
Porijõgi Reola	241	84	3.0	17.7	79.8	2.1	0.4	1.09	0.19
Ahja Kiidjärve	336	122	5.2	23.5	73.0	3.4	0.2	2.26	0.57
Ahja Lääniste	930	345	21.5	25.5	66.1	8.2	0.1	16.54	5.76
Kääpa väljavool Kose paisjärvest	282	102	2.8	16.8	72.3	9.9	1.0	0.15	0.04
Avijõgi Mulgi	366	311	3.7	18.8	73.3	7.8	0.0	0.53	0.09
Ranna-pungerja Roostoja	313	219	3.3	9.6	65.3	25.1	0.0	42.61	0.46
Tagajõgi Tudulinna	252	162	3.1	2.3	77.9	19.7	0.1	0.00	0.00
Alajõgi Alajõe	140	109	1.7	7.3	74.5	18.0	0.2	0.00	0.00

Regarding riverine (or instream) retention, the most important factors are assimilation by algae and aquatic macrophytes and especially gaseous losses via denitrification. P is removed by adsorption onto streambed sediments, sedimentation, and through uptake by algae and aquatic macrophytes. In literature, special attention has been given to erosion processes (soil and bank erosion) and P adsorption and desorption processes in streams. The adsorption onto bed sediments is regarded to be the major mechanism for P retention. In the Ruhr River, Imhoff (1996) estimated an instream P-retention of 50%. However, observations that main river channel can act as sources rather than sinks have also been observed. Svendsen et al. (1995) showed in a Danish lowland stream a negative P-retention on an annual basis due to resuspension of retained material during high flows and stream bank erosion. On the other hand it was in the same study shown that the retention of DRP (dissolved reactive P) constituted up to 60% of the DRP input to the stream channel, most likely due to P uptake in benthic and pelagic algae rather than P uptake by aquatic macrophytes. The impact of instream physical and biological processes on the regulation of P fluxes through river systems is still poorly understood. Therefore, it is not surprising that we in this study with simple empirical functions were less successful.

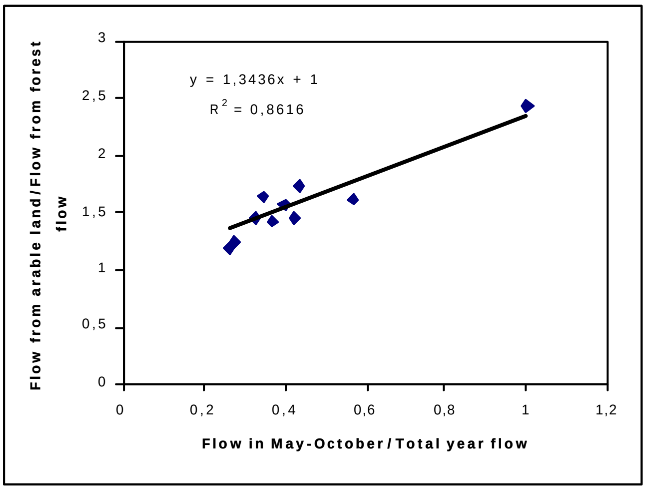

Figure 2. *Relationship between the ratio of water discharge from agricultural land and forests vs. the relative water discharge during May-October.*

Table 2. Relationships between emission coefficients (kg ha^{-1}) of nitrogen and phosphorus and water discharge (Q in mm yr^{-1}) from different land-use categories. Range of minimum and maximum flow was 119 and 300 mm, respectively.

	NITROGEN	PHOSPHORUS
Agricultural land	$\beta_1 = 0.0368Q + 3.76$ (R^2=0.74)	$E = 0.0019Q - 0.06$ (R^2=0.96)
Forest	$\beta_2 = 0.0095Q + 0.56$ (R^2=0.57)	$E = 0.00012Q + 0.022$ (R^2=0.49)

For nitrogen, riverine retention estimates are much more scattered. A recent literature review by Haag & Kaapenjohann (2001) stated that the nitrate-N retention in rivers most likely are in the range of 1-5%, although values of 20% and 30% also has been reported by Hill (1997) and Billen et al. (1991), respectively. In the Nordic/Baltic region, instream retention is regard as low. Arheimer (1998) indicated in calculations for entire southern Sweden an instream retention of 2% or maximum 7% (B. Arheimer, pers.comm.). In the Kasari river in Estonia, in-stream retention was found to account for less than 10% (Lidén et al., 1998). Thus, the relatively low derived retention in rivers found in this study (Table 3) may be regarded as rather non-controversial. Instream retention can also occur in open channels in the agricultural landscape. In fact, it has in catchments in the Baltic States been shown that up to 60% of the load at tile drain outlets are lost on its way to the mouth of the open channel (first-order stream). In this study, we have most likely not been able to capture this possible phenomena but will be further examined later.

Table 3. Estimated lake and river system retention of nitrogen and phosphorus

Type of retention	Nitrogen retention, (%)	Phosphorus retention, (%)
Lakes	33	35
River system	11	14

Based on the model results, we for 1993 found that more than 60% of the total nitrogen originated from agricultural land (Table 4). Forest, despite the large areal coverage (Table 1), contributed with less than 30% of the nitrogen load. The corresponding results for phosphorus showed that approximately 40% of the load originated from agricultural land while point sources accounted for approximately 42%.

The main uncertainty in model runs was found to be for phosphorus. This was rather expected since phosphorus loads also depends on other factors than land use (e.g., soil type). Despite these discerned uncertainties, it seems that the MESAW model is a reliable tool for simultaneous estimation of sources and retention in a river basin. It was also evident that MESAW can be used to investigate of the water flow from different type of land cover. In addition, MESAW can be used to identify measurements that are outside the general patterns and relationships (i.e., outliers). The main advantages with the model are: (i) the simple structure of the model (ii) the simple input data (iii) all unknown parameters are derived from empirical data, and (iv) that information from all water quality monitoring sites are used in an optimal way. The main advantage of the MESAW is that it gives results on the base of all available measured data which is better than to apply emission coefficients received from literature; normally even extrapolated from other regions or up-scaled from small watersheds. To its character, MESAW, have many common features with the more famous SPARROW model developed in U.SA. (Smith et al., 1997; Alexander et al., 2000). It should be pointed out that the examples given in this article is performed deliberately with very simple input data and parameterisation. This implies that we have a good possibility to apply the model for the entire Lake Peipsi region, i.e. also on the Russian side where data is far more limited.

Table 4. Sources of the total nitrogen and phosphorus load in 1993 from Estonian rivers to the Lake Peipsi.

Source	Agricultural land	Forest	Other	Point sources	Deposition on the lakes
Nitrogen (%)	61	27	2	8	2
Phosphorus (%)	40	16	1	42	1

CONCLUSIONS

Results show that less than 10% of the nitrogen load from Estonian rivers to Lake Peipsi originates from wastewater (point pollution sources); approximately 60% of the load come from agriculture and 30% originates from forests and other diffuse sources. Of the phosphorus load, over 40% come from point pollution sources and almost 40% from agriculture via the rivers in the catchment area. Estimation of retention was found to be large in lakes: 30-35% in lake Võrtsjärv for nitrogen and phosphorus. Riverine retention of phosphorus was found to difficult to assess but estimated to less than 15% while riverine retention of nitrogen was found to somewhat smaller: i.e., approximately 10%.

Results show that the MESAW model is a simple tool for simultaneous estimation of sources and retention in a river basin due to (i) the simple structure of the model (ii) the simple input data (iii) that all unknown parameters are derived from empirical data, and (iv) that information from all water quality monitoring sites are used in an optimal way. Results also showed that it could be powerful to analyse the relationships between water discharge and land use. For example, in this study we showed that water discharge from agricultural land generally were higher than from forests, especially during the growing season.

ACKNOWLEDGEMENTS

The authors are grateful to the Estonian Environmental Information Centre for providing us with the major part of the data set. The European Union (contract No. EVK1-CT-2000-00076) are acknowledged for their financial contribution.

REFERENCES

Ahlgren, I., Sörensen, F., Waara, T. and Vrede, K. (1994). Nitrogen budgets in relation to microbial transformations in lakes. *Ambio* **23**: 363-366.

Alexander, R.B., Smith, R.A. and Schwarz, G.E. (2000). Effect of stream channel size on the delivery of nitrogen to the gulf of Mexico. *Nature* **403**: 758-761.

Arheimer, B. (1998). *Riverine nitrogen: analysis and modelling under Nordic conditions*. PhD thesis. Linkoping Studies in Arts and Science **185**.

Billen, G., Lancelot, C. and Maybeck, M. (1991). N, P, and Si retention along the aquatic continuum from land to ocean. In*: Ocean Marigin Processes in Global Change*. Mantoura, R.F.C., et al. (Eds.). Wiley & Sons, Chichester, pp: 19-44.

Grimvall, A. and Stålnacke, P. (1996). Statistical methods for source apportionment of riverine loads of pollutants. *Environmetrics* **7**: 201-213.

Haag, D. and Kaupenjohann, M. (2001). Landscape fate of nitrate fluxes and emissions in Central Europe – a critical review of concepts, data, and models for transport and retention. *Agriculture, Ecosystems and Environment* **86**: 1-21.

Howarth, R.W., Billen, G., Swaney, D., Townsend, A., Jaworski, L., Lajtha, K., Downing, J.A., Elmgren, R., Caraco, N., Jordan, T., Berendse, F., Kudeyarov, V., Murdoch, P. and Zhao-Liang, Z. (1996). Regional nitrogen budgets and riverine N and P fluxes for the drainages to the North Atlantic Ocean: natural and human influences. *Biogeochemistry* **35**: 75-139.

Imhoff, K.R. (1989). Balance and elimination strategies for heavy metals, phorpshorus, ammonia and AOX in the Ruhr catchment. *In: River Basin Management – V. (Laikari, H. Ed.). Proceedings of an IAWPRC Conference,* Rovaniemi, Finland, July 31-August 4, 1989, pp. 285-293.

Hill, A.R. (1997). The potential role of in-stream and hyperheic environments as buffer zones. *In: Haycock, N.E., Burt, T.P., Goulding, K.W. and Pinay, G. (Eds.), Buffer zones; their processes and potential in water protection*. Quest Environment, Hertfordshire. pp.115-127.

Jansson, M.,Andersson, R.,Berggren, H. & Leonardson, L. (1994). Wetlands and lakes as nitrogen traps. *Ambio* **23**: 320-325.

Lidèn, R., Vasilyev, A,. Loigu, E., Stålnacke, P. and Wittgren, H.B. (1998). Nitrogen source apportionment using a physically and statistically based model. *Ecological Modelling* **114**: 235-250.)

Smith, R.A., Schwarz, G.E. and Alexander, R.B. (1997). Regional interpretation of water-quality monitoring data. *Water Resources Research* **33**: 2781-2798.

Svendsen, L-M. and Kronvang, B. (1993). Retention of nitrogen and phosphorus in a Danish lowland river system: implication for the export from the watershed. *Hydrobiology* **251**: 123-135.

Svendsen, L-M., Kronvang, B., Kristensen, P. and Græsbøl, P. (1995). Dynamics of phorpshorus compounds in a lowland river system: importance of retention and non-point sources. *Hydrological Processes* **9**: 119-142.