LIFE + Environment Policy and Governance
Project name: Participatory monitoring, forecasting, control and socio-economic impacts of eutrophication and algal blooms in River Basin Districts (GISBLOOM)
Project reference: LIFE09 ENV/FI/000569
Duration: October 1st, 2010 – September 30th, 2013

EU LIFE+
GisBloom

Deliverable

Sub-action 2.3

Report on nation-wide diffuse load equations for phosphorus and nitrogen, and comparison of the two modelling approaches for selected catchment

Beneficiary: Finnish Environment Institute (SYKE)
Contributors: Jaakkola, E., Ekholm, P., Tattari, S. (Finnish Environment Institute)

Due date of deliverable: 1.9.2012
Actual submission date: 31.8.2012

With the contribution of the LIFE financial instrument of the European Community

Content

Introduction

Today, the bulk of nutrient loading to surface waters originates from diffuse sources in Finland. The amount of nutrients transported from a catchment into a water body is a complex function of inherent catchment characteristics and various human activities, such as agriculture, urban runoff and forestry. During the past decades, there have been many nation wide studies to approximate the losses of total nitrogen (TN) and total phosphorus (TP) from differing areas (Vuorenmaa et al. 2002, Mattson et al. 2003, Rankinen et al. 2010). Since knowing the amount of nutrient losses and their origin is fundamental in implementing cost-efficient counter-measures, different models, such as VEMALA and VEPS, have been developed for Finnish conditions for quantifying nutrient loading and for performing source apportionment (Huttunen et al. 2008, Tattari & Linjama 2004). The above-quoted studies did not utilize all available water quality data in the estimating nutrient losses. Moreover, GIS data are improving as are methodology in studying different characteristics affecting the water quality (such as land use, soil, slope etc.). Thus, a new estimation of the losses of the main nutrients (TN, TP) from the Finnish catchments was made here based on a combination of best available GIS and statistical data with the observed nutrient losses measured at national and regional water quality monitoring sites. Simple empirical relations were formulated from these data, which makes it possible to predict the nutrient losses at any site in Finland, excluding the northernmost parts at this stage of the project. The study period 2000-2011 was chosen to represent the average conditions prevailing after Finland joined the European Union and the farmers started to follow the agri-environmental measures. The study proceeded in phases (Fig. 1): First, the nutrient fluxes were calculated for each measuring site where there were daily discharge data and enough observations on the water quality available. Then, the catchments were delineated and a set of catchment characteristics was calculated by using GIS methods. Finally, the empirical relationships between the nutrient load (dependent variable) and the specific catchment characteristics (independent variables) were formulated from the data, and the results were compared with the VEMALA model in the Aurajoki catchment.

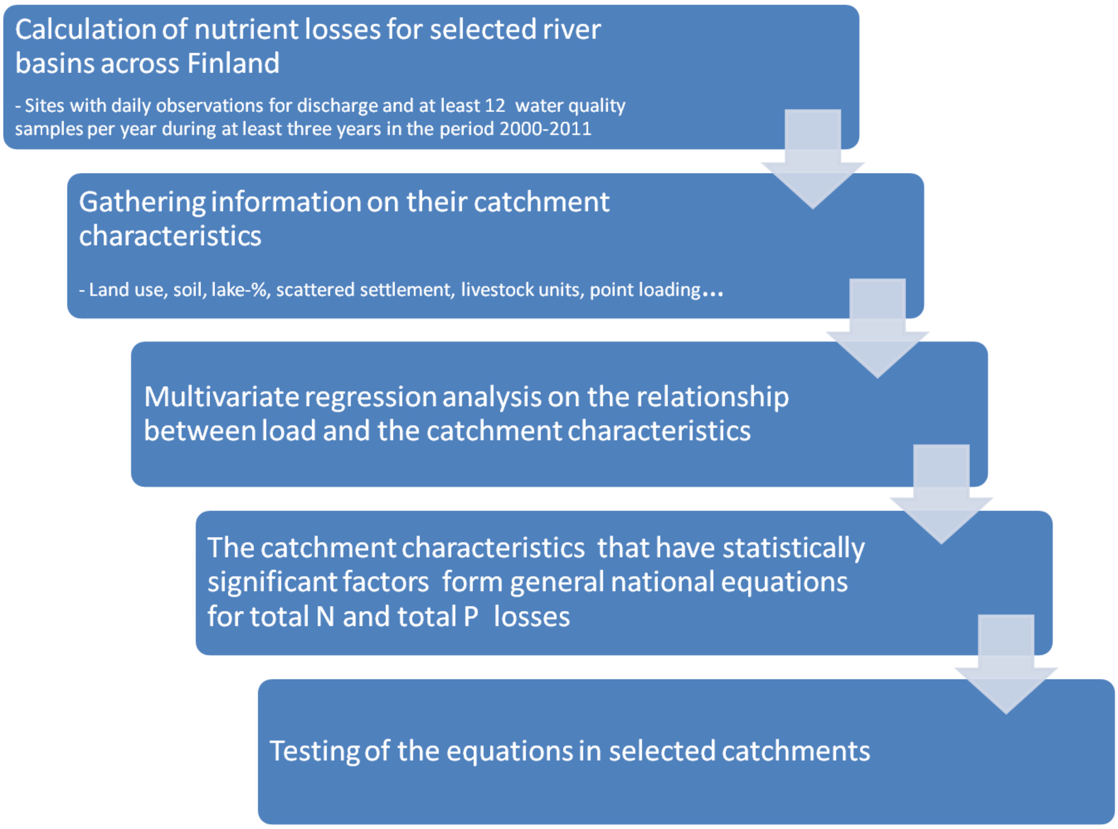

Figure 1. Flow chart of the different phases in the study.

Material and methods

Selecting suitable monitoring stations

The sub-action was initiated by examining national data bases and selecting the water quality sites that were sampled frequently enough for TN and TP, and where discharge was also measured. The data were taken from the Hertta system, a database maintained by the Finnish Environment Institute (SYKE). Information content of the system increases all the time as the implementation is being progressed and new data is added. The water quality data is in the PIVET database and discharge data in the HYDRO database. From the PIVET database, all the sampling sites that had at least 12 samples taken and analyzed for TP yearly, during at least 3 years in the period 2000–2011 were picked (n = 257). It was assumed, that if TP has been analyzed, there is also a result for TN. Only 123 of these sites were marked as monitoring sites of the River Basin Districts (RBD), which the work was started with at this stage of the project. Later on, all the suitable stations will be included. The list of RBD sites was shortened to represent only those sites that had a reliable hydrological station nearby, of which catchment located entirely inside Finland, and where the catchment delineation was possible to perform with reasonable accuracy. This led to a list of 68 applicable sites. Checking of hydrological stations was primarily done by a map survey, as there were no means to perform it by linking information from PIVET to HYDRO in this respect. Before calculating the nutrient losses, a quality checking was done for the PIVET data by drawing scatter plots for the substances that usually correlate with each other (such as TP with TDP or SS, $NO_{23}N$ with TN) for each site. The seemingly outlying observations were sent for checking to the corresponding Centre for Economic Development, Transport and the Environment, and the corrected or deleted samples are shown in Appendix 1. Not all the questionable values could be checked, as most of the laboratories preserve the original measurement results only for five years.

The catchments were delineated using a test version of the flow direction grid prepared by a project VALUE (http://www.fgi.fi/tutkimus/tiedot_aiheesta.php?projekti=108). It is based on DEM (25 x 25) m which is owned by the National Land Survey of Finland (MML). The catchment delineation was done by ArcMap's Hydrology tools, the snapping distance being 125 m at most (i.e., the distance that program uses in finding the grid with a highest flow accumulation, which then forms the pour point of the catchment). The borders of small catchments were gathered elsewhere, as the DEM 25 x 25 m is too inaccurate for them. Some of the selected catchments were sub basins of bigger catchments. Therefore, the catchments were divided into two sets. The first set was used to establish the equations between nutrient losses and catchment characteristics and included "independent" catchments, i.e. areas that were not nested. (n = 36, Figure 2, Table 1). The second set included the lower reach sub-basins of some of the head-water catchments in the first set as well as some independent catchments (n = 32). This set was used for verification of the equations. At this stage, the model represents the area below the Arctic Circle. At the next stage of the project, also the Northern catchments will be added to the model.

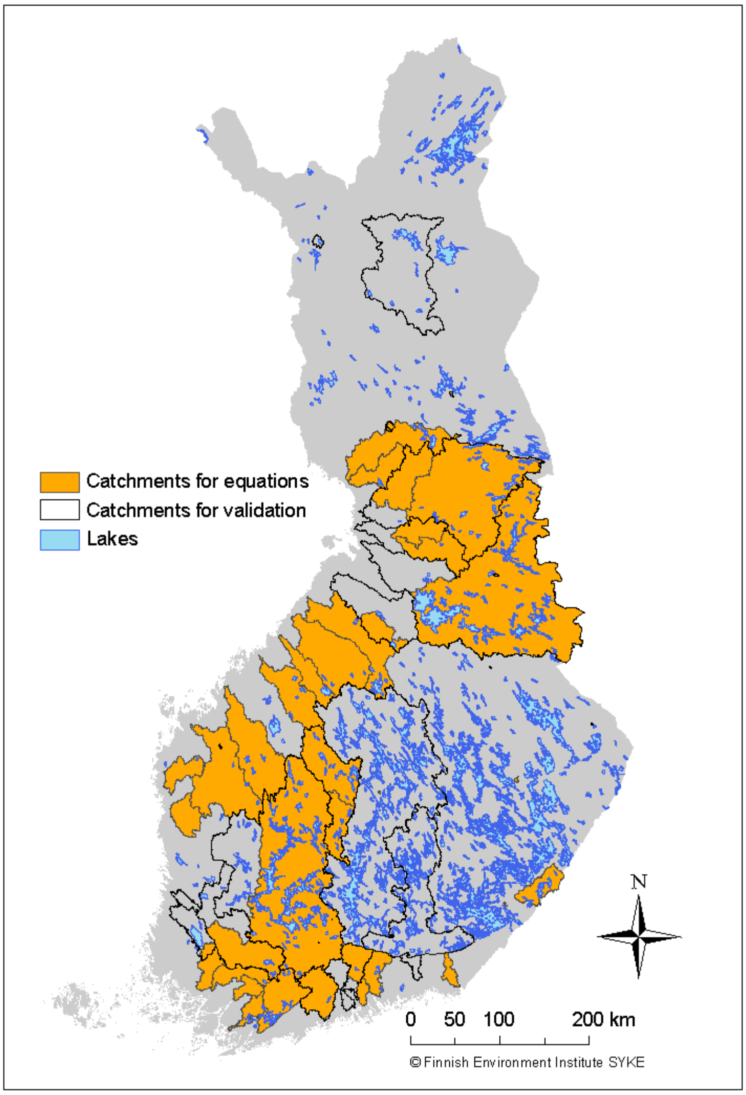

Figure 2. Catchments where the data for the equations, and their validation was derived from.

Table xx. PIVET stations that were used in deriving the equations, and their corresponding hydrological stations.

Pivet id	Pivet station name	Hydro. st. nro	Hydrological station name
42344	Yläneenjoki P2 Vanhakart	3400130	Yläneenjoki
332	Ilolanjoki 1,3	1700500	Ilolanjoki, Ilolan silta
11310	Virojoki 006 3020	1100500	Salmen silta
6169	Uske 16 Salon yp va6101	2500400	Kaukolankoski
6047	Kisko 14 Vanhak va6111	2400400	Koski
6219	Aura 54 ohikulku va6401	2800700	Halinen
227	Koskenkylänjoki 6030	1601100	Koskenkylänjoki, Niinikoski
417	Porvoonjoki 35,5	1800500	Vakkola
4247	Närpiö rts. vp 9200	3900800	Allmänningsforsen
28066	Lamujoki Jylhänranta	5700130	Lamujoki
6193	Pajo 44 Isosilta va6301	2700250	Juvankoski
4081	Myllykanava vp 9100	3700300	Perus
10060	Hiitolanjoki rajav 002	300450	Kangaskoski
28564	Nuorittajoki suu	6000200	Nuorittajoki
29462	Kuivajoki rautatiesilta	6300210	Ravaska
26935	Lestijoki 10800 8-tien s	5100500	Saarenpää
25709	Jämsänjoki 4500	1405225	Patalankoski
605	Vantaa 4,2 6040	2101700	Oulunkylä
28526	Kiiminkijoki Joki-Kokko	6000100	Porkkalan silta
1069	Mustionjoki 4,9 15500	2300935	Peltokoski
29092	Siuruanjoki alap silta	6101600	Siuruanjoki, Leuvankoski
26740	Perhonjoki 10600	4900350	Kaitfors
6534	Lojo 64 Pori-Hki	3509410	Loimijoki, Maurialankoski
25872	Aittokoski 3300	1402150	Kiimasjärvi, Hietamankoski
36177	SIMOJOKI AS. 13500	6400410	Simo
27697	Pyhäjoki Hourunk 11400	5400410	Tolpankoski
26534	Lapuanjoki 9900	4400850	Uusikaarlepyy
27095	Kalajoki 11000	5300740	Niskakoski
4381	Skatila vp 9600	4201000	Skatila
28686	Iijoki Kipinä 13200	6101451	Haapakoski
7650	Nokiankoski 8200 alavirt	3507450	Nokia
31083	Jylhämä 12800	5903450	Oulujärvi, Jylhämä
22590	Kesselinpuro 35	51	Kesselinpuro
10566	Huhtisuonoja 382 44	44	Huhtisuonoja
10564	Latosuonoja 402 43	43	Latosuonoja
25269	Ruunapuro 71	71	Ruunapuro

Estimating TN and TP fluxes for each selected catchment

To be able to approximate the average yearly TP and TN losses (kg km^{-2} a^{-1}) for each river basin during a period 2000-2011, a so called '*concentration-discharge-model*' was used. The model finds a best possible daily combination for parameters a, b, and c, based on the daily discharge data and the observed concentrations in the river (Wartiovaara 1975, Rankinen et al. 2010):

$$C = a + b/Q + c \cdot Q \qquad (1)$$

where
C = modeled concentration of TP or TN in the river
a = parameter for "base flow" concentration
b = parameter for concentration that is not related to discharge, such as the sewage water
c = parameter for concentration that increases when discharge increases and vice versa, such as the diffuse loading
Q = observed discharge

In reality, there is also some stochastic variation in the observation sites. Thus, Wartiovaara (1975) emphasizes that the factors in the equation are bound for each observation site and - time, and to determine trustable factors, enough observations are needed. In this study, the criteria were at least 12 samples per year during at least three years, and daily discharge data for the entire study period. The data was fitted to the equation by SAS® 9.3 software.

The average yearly load was calculated based on the estimated daily concentrations and measured discharges.

Gathering information on the catchment characteristics

Next, the availability of GIS data was checked. The slope calculations were also based on the DEM (25 × 25) m which was previously used in delineating the catchment borders, as more precise DEM that would have covered the entire Finland was not available at the time.

Animal data was obtained from Evira (the Finnish Food Safety Authority) and Mavi (Agency for Rural Affairs). The data included coordinates of each farm and number of different animals, excluding the majority of horses though. The animal numbers were converted to livestock units and further to nutrients by the help of tables from Evira and MTT (Jouni Nousiainen). The table concerning the nutrient content of the manure did not include all the classes of animal data, and therefore some of the classes were given "no data" value (see Appendix 2). The animal data was utilized as point-features in a GIS program ArcMap (ESRI).

The point source loading information of TN and TP for the period 2000–2010 was collected from the Compliance Monitoring Data System (VAHTI) with the coordinate information and divided into four sectors: fish farming, industry, municipal waste waters and peat-industry. These data were also utilized as point-features in ArcMap.

The national, (25 x 25) m grid data of Corine Land Cover 2006 (CLC2006) was used to obtain the basic land use data (e.g. forests, wetlands and build areas) for the catchments. For the agricultural area, a field plot register, owned by Mavi, was used. The register, i.e. "identification system of the fields", is a nationwide register where all the field plots that have got area-based subsidies are digitized. The data for the year 2010 was used here. The dataset covers the information on the dominant plant and its share of the field plot, among other things.

The YKR database (Finnish Monitoring System of Urban Form and Spatial Structure), also located under the above mentioned Hertta system, is a tool for e.g. following long-term structural changes and studying the spatial and social structure. YKR is based on 250 x 250 m square information, i.e. one value is a sum of the same variable under 6.25 ha area. YKR is based on various sources of information (e.g. the Statistics Finland, the Population Register Centre and SYKE). Therefore, the numbers are not statistics but background data for monitoring and planning. SYKE has produced different spatial districts (such as urban areas and scattered settlements) that can be found from the YKR system. Here, the interest was in the population not connected to sewage plants. As there are no such data available directly, the best estimate for the period 2000-2011 was done by picking the YKR data of all the population living in scattered settlements in the year 2005. However, the resulting data might include population belonging to water cooperatives or similar.

Lake percentage and river length was calculated from data that is based on the shoreline data (*Rantaviiva-aineisto*) of MML. The shoreline data is part of the Terrain database (*Maastotietokanta*) of MML, and the position sharpness is 1:5 000-1:10 000. SYKE has re-classified and surveyed the data. The river network covers all the water ways with more than 10 km^2 catchment area. The data covers all of Finland excluding the northernmost parts.

Soil data was from the Geological Survey of Finland (GTK). We decided to use a rather simplified soil data (1:1 000 000) to enable a fluent data processing. The data holds all together 15 classes.

To gather the above mentioned data for the selected catchments (Figure 2), a macro was built with ArcMap Model Builder (Figure 3). The model calculates all the information for one catchment at a time.

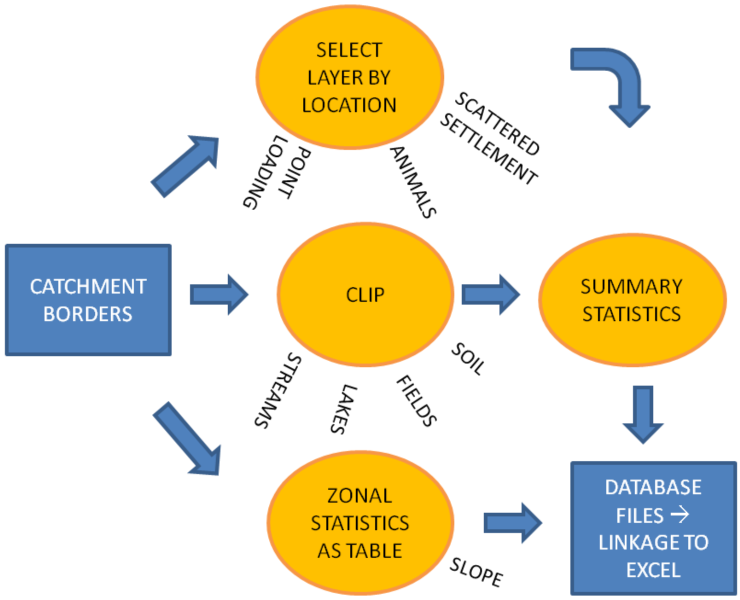

Figure 3. Imitation of the ArcMap model used to cut all the necessary information for the study catchments. Orange circles are ArcMap tools.

Studying the relationship between load and catchment characteristics

The effect of different catchment characteristics on diffuse TP and TN losses was first studied by a stepwise regression analysis to find out the most influencing catchment character. The results from the stepwise regression analysis were utilized in the next step, where another multivariate regression analysis was performed. There, the change in the dependent variable (TP or TN) is explained by one or more independent variables (different catchment characteristics C_1, C_2, …, C_n), the first variable being the one that is found the most explanatory one in the stepwise regression analysis:

$$\text{TP or TN} = p_1 \cdot C_1 + p_2 \cdot C_2 + \ldots + p_n \cdot C_n + \text{constant} \qquad (2)$$

where $p_1 \ldots p_n$ is a parameter estimate, which represent the effect of a specific catchment character on the nutrient load. If, for example, C_1 is the field percentage and it grows by one unit, the nutrient load will increase according to the parameter estimate p_1.

Results

The catchment characteristics

The catchment characteristics that were gathered at this stage of the project were further calculated to suitable units. In the end, following classes were utilized in the study: area of the entire basin (km^2) and area of the fields (%), slope of the entire area (%) and slope of the fields (%), calculated P and N in manure ($kg\ km^{-2}\ a^{-1}$), fields' plant types (%), scattered settlement (inhabitants km^{-2}), streams ($km\ km^{-2}$), lakes (%), point loading of P and N divided into their sources ($kg\ km^{-2}\ a^{-1}$), CLC land use classes (%) and soil classes (%). Figure 4 shows the histograms of the most important classes.

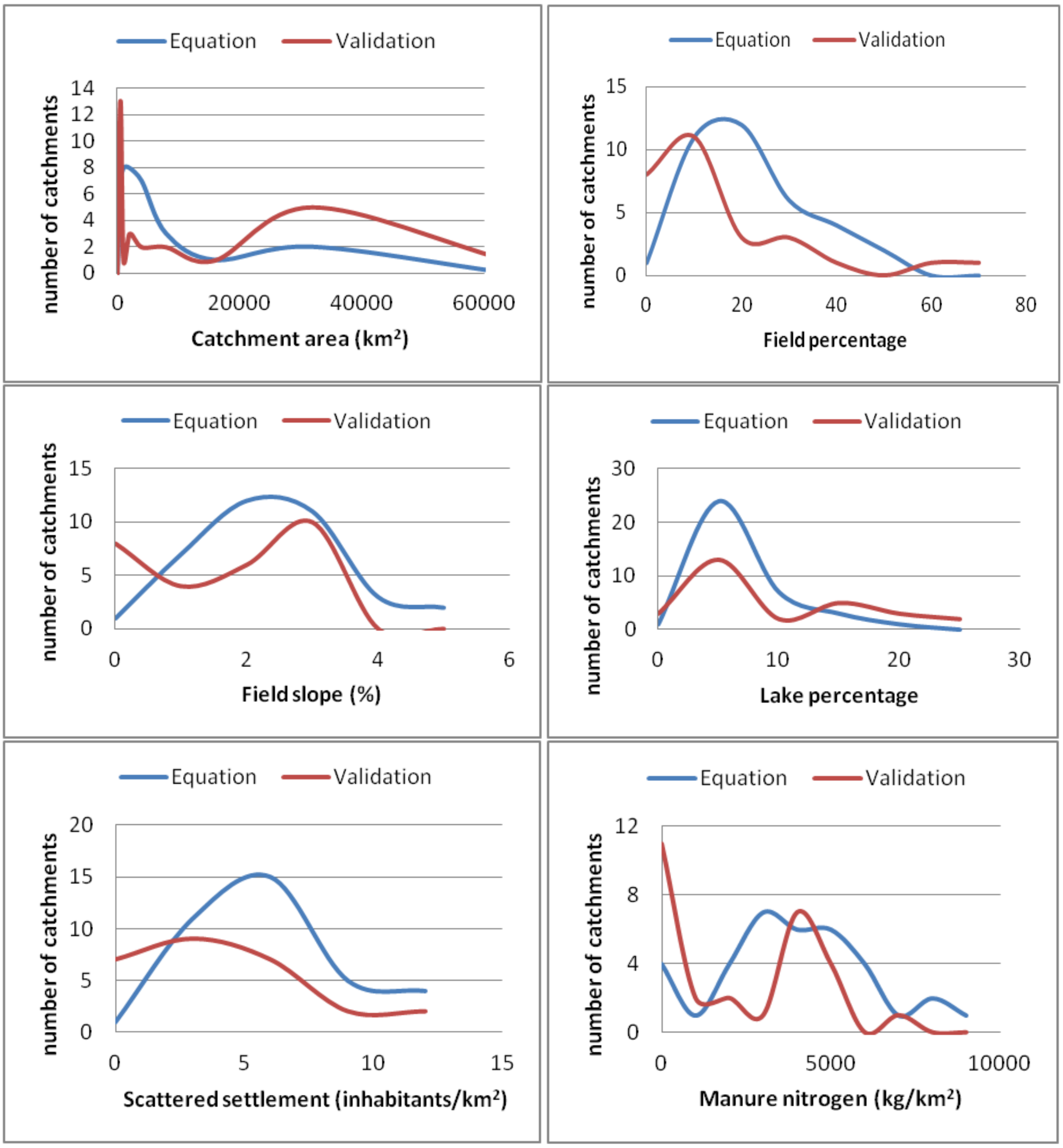

Figure 4. The distribution of some of the explaining variables in the study basins.

Load calculations

The average yearly load was calculated for each study basin based on the estimated daily concentrations from the equation (1) and measured discharges. The load (kg a^{-1}) was then divided by the land area of the basin enabling comparison amongst basins and excluding the effect of lakes, which is known to decrease nutrient losses in rivers by sedimentation processes rather than increasing them. At the end, the point load was subtracted from the total load. The distribution of the estimated diffuse loads is shown in figure 5.

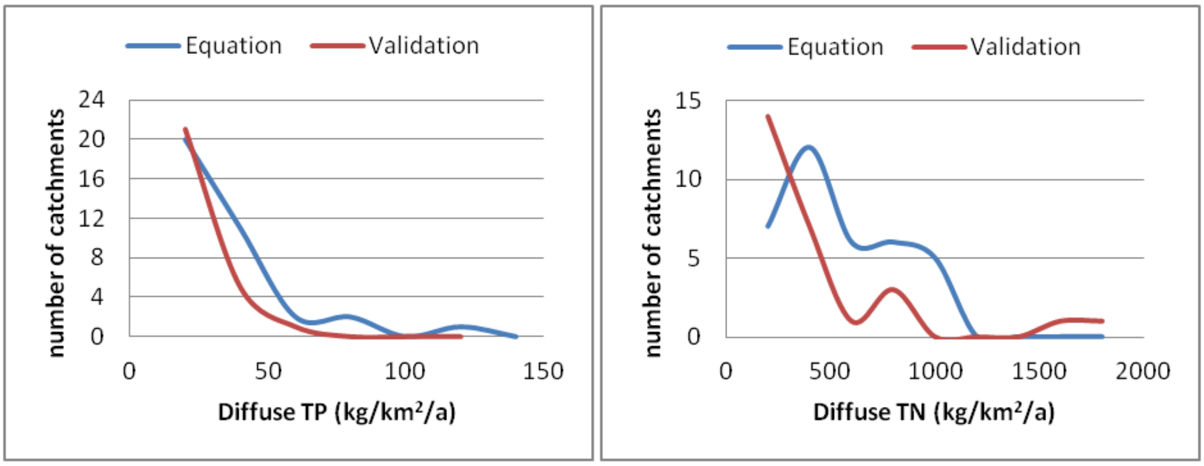

Figure 5. The distribution of the dependent variables (total phosphorus PTOT and total nitrogen NTOT) in the study basins. PTOT and NTOT loading is considered to originate from the land area only.

Next, the point load was subtracted from the total load so that the statistical analyses were done for the diffuse losses only.

Diffuse TP losses

When estimating the variables that explain the nutrient losses in rivers, a stepwise regression analysis was chosen as a starting point. For phosphorus, the following variables were included in the analysis: Catchment area (km^2), field slope (% rise), people living in scattered settlements (inhabitants km^{-2}), fields (%), P in manure (kg km^{-2}), root crops (%), spring cereals (%), grass (%), autumn cereals (%), garden plants (%), lakes (%), streams (km km^{-2}) built area (%), forest (%), peat (%) and clay (%). It turned out, that the share of clay soils explained the phosphorus loss from the land area (TP) with a highest coefficient of determination (0.71):

TP = 8.2 + 1.02 · Clay-% (n = 36)   (3)

The field percentage, correlating strongly with clay percentage (clay-% = 1.33 x Field-% - 4.3, r^2 = 0.88) explained 67% of the loading:

TP = 3.0 + 1.41 · Field-% (n = 36)   (4)

Thus, the equation (3) was chosen as a starting point for further regression analyses. The progress is shown step by step in Table 2.

Table 2. Parameter estimates and their significance for the tested variables explaining diffuse phosphorus loss from land, and the coefficient of determination of each Step.

	R^2		Step1 p	Step2 p	Step3 p	Step4 p	Step5a p	Step5b p	Step5c p	Step5d p	Step5e p	Step5f p
		Intercept	8.2 **	15.1 ***	54.0 **	63.9 **	66.6 **	65.1 **	64.0 **	65.0 **	63.8 **	63.7 **
Step1	0.71	Clay-%	1.0 ***	0.9 ***	0.7 ***	0.9 ***	0.9 ***	0.9 ***	0.9 ***	0.9 ***	0.9 ***	0.9 ***
Step2	0.77	Lake-%		-1.3 **	-1.3 **	-1.0 *	-0.9 +	-1.0 +	-0.9 *	-0.9 +	-1.0 +	-1.0 *
Step3	0.80	Forest-%			-0.5 *	-0.6 *	-0.6 *	-0.6 *	-0.6 *	-0.6 *	-0.6 *	-0.6 *
Step4	0.83	Scattered per area				-2.2 *	-2.3 *	-2.4 +	-2.2 *	-2.2 *	-2.2 *	-2.3 ns
Step5a	0.83	Manure P per area					0.0 ns	-	-	-	-	-
Step5b	0.83	Field slope (%)						0.7 ns	-	-	-	-
Step5c	0.83	Root crop-%							0.2 ns	-	-	-
Step5d	0.83	Streams per area								-2.3 ns	-	-
Step5e	0.83	Catchment area									0.0 ns	-
Step5f	0.83	Built area-%										0.1 ns

p = parameter estimate
Significance of the coefficients (p): + <0.1, * p < 0.05, ** p <0.01, *** p < 0.001 and ns = Not significant.

For the diffuse TP loading, four variables were found statistically significant at $p < 0.05$: clay percentage, lake percentage, forest percentage and scattered settlements. Together they explained 83% of the TP loading:

$$\text{TP} = 63.9 + 0.9 \cdot \text{Clay-\%} - 1.0 \cdot \text{Lake-\%} - 0.6 \cdot \text{Forest-\%} - 2.2 \cdot \text{Scattered per area} \quad (5)$$

Diffuse TN losses

For nitrogen, the following variables were fitted to the model: Catchment area (km^2), field slope (% rise), people living in scattered settlements (inhabitants km^{-2}), fields (%), N in manure (kg km^{-2}), root crops (%), spring cereals (%), grass (%), autumn cereals (%), garden plants (%), lakes (%), streams (km km^{-2}) built area (%), forest (%), peat (%) and clay (%). Unlike to diffuse TP losses, the stepwise process found field percentage best explaining the diffuse TN loss.
The field percentage explained 71% of all the loading:

$$\text{TN} = 172.2 + 17.3 \cdot \text{Field-\%} \ (n = 36) \quad (6)$$

This was chosen as a starting point for further regression analyses, of which progress is shown step by step in Table 3.

Table 3. The parameter estimates and their significance for the tested variables explaining diffuse nitrogen loss from land, and the coefficient of determination for each Step.

	R^2		Step1 p	Step2 p	Step3 p	Step4 p	Step5 p	Step6 p	Step7 p	Step8a p	Step8b p	Step8c p	Step8d p
		Intercept	175.2 ***	307.7 ***	268.5 ***	268.0 ***	276.7 ***	237.3 ***	345.8 ***	357.0 ***	350.5 ***	361.8 ***	345.7 ***
Step1	0.71	Field-%	17.3 ***	15.1 ***	5.6 +	11.0 ***	14.6 ***	14.2 ***	12.3 ***	12.2 ***	12.1 ***	12.3 ***	12.3 ***
Step2	0.84	Lake-%		-23.2 ***	-29.6 ***	-25.3 ***	-24.7 ***	-26.5 ***	-30.5 ***	-30.2 ***	-31.2 ***	-30.1 ***	-30.6 ***
Step3	0.89	Spring cereal-%			5.2 ***	5.8 ***	4.7 ***	4.5 ***	4.7 ***	4.7 ***	4.9 ***	4.7 ***	4.7 ***
Step4	0.92	Scattered per area				-29.0 **	-57.7 ***	-56.8 ***	-63.2 ***	-62.5 ***	-63.3 ***	-63.0 ***	-63.2 ***
Step5	0.94	Built area-%					20.5 **	23.2 **	24.2 ***	24.1 ***	23.8 ***	25.5 ***	24.2 ***
Step6	0.94	Manure N per area						0.0 *	0.0 *	0.0 *	0.0 **	0.0 **	0.0 *
Step7	0.95	Peat-%							-3.3 +	-3.5 ns	-3.4 +	-3.3 +	-3.3 +
Step8a	0.95	Field slope (%)								-3.8 ns	-	-	-
Step8b	0.95	Root crop-%									-3.4 ns	-	-
Step8c	0.95	Streams per area										-103.2 ns	-
Step8d	0.95	Catchment area											0.0 ns

p = parameter estimate
Significance of the coefficients: + <0.1, * p < 0.05, ** p <0.01, *** p < 0.001 and ns = Not significant.

For the diffuse nitrogen, many variables were found to be statistically significant, and in the end, it was possible to explain 94% of the loading with a 0.05 probability limit.

Thus, it can be concluded that based on our 36 catchments, 94% of the diffuse TN loading can be explained by the equation:

TN = 237.3 + 14.2 · Field-% - 26.5 · Lake-% + 4.5 · Spring cereal-% – 56.8 · Scattered per area + 23.2 · Built area-% + 0.01 · Manure N (7)

Comparison to VEMALA, and other model verification

In the VEMALA model, loads have been calculated by sources: Peat industry, point loading and atmospheric fallout forms one class, the rest of the classes being fields, housing and other. This information is calculated for each sub basin. In addition, the export from each sub basin is calculated. Here, our equation, applied to the Aurajoki catchment, was compared to the figure that was calculated from VEMALA by subtracting peat industry, point loading and atmospheric deposition from the load in the entire Aurajoki basin. According to VEMALA, the diffuse losses from Aurajoki (average yearly, also based on a period 2000–2011) was for TP 55.2 kg $km^{-2}a^{-1}$ and for TN 774 kg $km^{-2}a^{-1}$. With our equation, the diffuse nutrient losses in Aurajoki catchment were for TP 62.1 kg $km^{-2}a^{-1}$ and for TN 802 kg $km^{-2}a^{-1}$.

Our equations were also validated in a set of 27 different t catchments, where the diffuse load was calculated by two methods: Based on (i) the measured concentrations and discharges, i.e. *the concentration-discharge-model* (see Eq. 1) and (ii) our new equations. To have comparable figures, point loading was subtracted from the first mentioned. The scatter plots of the equations are shown in Fig. 6. For TP, the coefficient of determination is 0.59, and for TN 0.68. In some catchments, the load becomes negative with our new equation.

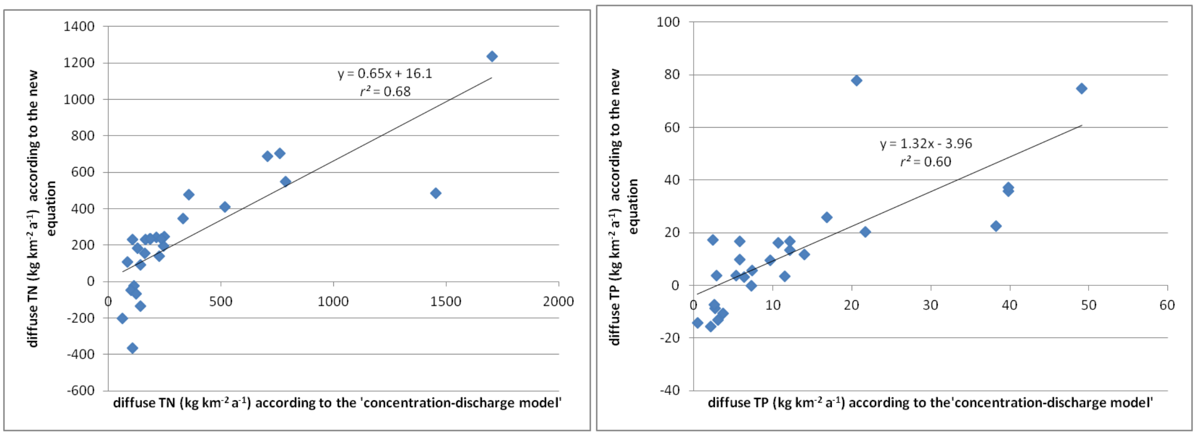

Figure 6. Our new equations plotted against the loads that have been calculated with the *concentration-discharge-model* (Eq. 1).

Discussion

Rankinen et al. (2010) found field and lake percentages being the most important variables when explaining the diffuse phosphorus and nitrogen losses. Due to the fact that field percentage strongly correlates with clay percentage, our preliminary results for both phosphorus and nitrogen can be considered to be in line with that of Rankinen et al. (2010).

According to our study, a catchment with 100% fields would have a background TP loss of 3.0 kg km^{-2} a^{-1}, while the loading from field area would be 141 kg km^{-2} a^{-1}. For TN, the corresponding values would be 172 kg km^{-2} a^{-1} and 1730 kg km^{-2} a^{-1}, respectively. The loading estimates for agricultural land area are in a good agreement with Vuorenmaa et al. (2002), who estimated that the spesific nutrient loss from agriculture in four river basins in southern Finland was on average 137 kg km^{-2} a^{-1} for TP and 1800 kg km^{-2} a^{-1} for TN during the period 1981-1997. In agricultural small catchments, the corresponding losses were estimated to be lower, 110 kg km^{-2} a^{-1} for TP and1500 kg km^{-2} a^{-1} for TN. In addition, above result is in line with Mattsson et al. (2003), who estimated the average annual TN losses from pristine forested areas to be 140 kg km^{-2}, and the average TP losses 5.4 kg km^{-2}.

In this kind of explanatory framework, it is important to recognize and understand the correlation between different variables. If, for example, increasing field percentage is already explaining the nutrient loss, it is difficult to recognize the effect of an increasing cultivation method or for example clay soil, which is a common soil type in the fields of southern Finland. In addition, in this kind of a regression analysis which is based on the sum of squares, the outliers may have such an effect on the result that can easily lead to incorrect interpretation of the results. For example, scattered settlements has a negative parameter estimate for both TP and TN, although based on a theory, it should increase the diffuse losses instead of decreasing them. Examining the data behind the figures more closely shows that the catchment Uskelanjoki formed an outlier. It has the biggest measured diffuse losses of both TP and TN, as well as the biggest clay and field percentage out of the data set. However, there is relatively few people living in scattered settlements, which skews the parameter estimate. If we take Uskelanjoki out from the analysis, scattered settlements are not found statistically significant.

Regardless of the above mentioned difficulties with the interpretation, our equations gave an estimate for the Aurajoki catchment that is close to the estimate given by the VEMALA model. However, when fitted to a bigger data set, the equation did not fully work: for example, the loads ended up being negative for the catchments that had large lake percentage. This is due to the fact that the validation set of catchments differ from the set of which the equations were derived from. The coefficient of determination was significantly better (0.73 for TP and 0.86 for TN), if all the catchments with more than 10% lakes and all the catchments with a catchment size smaller than 10 km^2 was erased from the dataset. Thus, the study should be continued by extending the data set of the suitable river monitoring sites and their catchments to the whole of Finland to have more diverse data set. Ideal would be to have big enough data set to be able to separate the catchments to different groups based on different characteristics, i.e., geographical location, soil type, lake percentage, catchment size or else like. Then, it would be possible to make more specified equations for differing areas, which could raise the reliability of the equations. In addition, new explanatory variables, such as the soil test P and more precise soil data could be added to the statistical analysis to obtain more information on the effect of agricultural areas and the texture of top soil on the nutrient losses.

Acknowledgements

The authors are grateful for numerous people from the Finnish Environment Institute who helped us to carry out this study; special thanks to Sari Mitikka, Arjen Raateland, Hannu Sirviö, Maria-Leena Hämäläinen and Antti Rehunen, who helped gathering the necessary data from different databases. The authors also wish to thank those people at the Centres for Economic Development, Transport and the Environment, who helped in checking the reliability of the water quality data.

References

Huttunen I., Huttunen M., Tattari S. & Vehviläinen B. 2008. Large scale phosphorus load modelling in Finland. XXV Nordic Hydrological Conference 2008. *NHP Report* No. 50, s. 548–556.

Mattsson, T., Finér, L., Kortelainen, P. & Sallantaus, T. 2003. Brook water quality and background leaching from unmanaged forested catchments in Finland. *Water, air, and soil pollution* 147: 275–297.

Rankinen, K., Ekholm, P., Sjöblom, H., Rita, H. & Vesikko, L. 2010. In Aakkula, J., Manninen, T. & Nurro, M. (ed.) 2010. Maatalouden ympäristötuen vaikuttavuuden seurantatutkimus (MYTVAS 3) – Väliraportti. Pp. 122–131. English summary: Follow-up study on the impacts of agri-environment measures (MYTVAS 3) – Mid-term report. Maa- ja metsätalousministeriön julkaisuja, Helsinki.

Tattari, S. & Linjama, J. Vesistöalueen kuormituksen arviointi. *Vesitalous* 3/2004: 26–30.

Wartiovaara, J. 1975. Jokien ainevirtaamista Suomen rannikolla. English summary: Amounts of substances discharged by rivers off the coast of Finland. *Vesientutkimuslaitoksen julkaisuja* 13.
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.

Appendix 1. Corrected water quality data.

Site	Date	Parameter	Corrected value
Yläneenjoki P2 Vanhakart	26.5.2011	FE	deleted
	2.10.2007	TDP	deleted
	28.4.2010	NTOT	deleted
Aura 54 ohikulku va6401	12.10.2011	NO23N	2500 µg l^{-1}
	14.4.2008	FE	deleted
Pajo 44 Isosilta va6301	31.3.2008	FE	deleted
	14.12.2004	all	deleted
	14.8.2006	SS	26 mg l^{-1}
	14.2.2000	NH4N	deleted
Kisko 14 Vanhak va6111	10.12.2003	NTOT	deleted
	13.11.2001	SS	deleted
	17.11.2008	SS	deleted
Savi 12 mittapato	17.4.2001	NTOT	deleted
Uske 16 Salon yp va6101	12.10.2011	NO23N	1600 µg l^{-1}
Löytäneenoja mittapato	7.4.2003	NO23N	deleted
	14.4.2003	NO23N	deleted
	4.11.2008	NTOT	5300 µg l^{-1}
Kojo 35 Pori-Tre	12.8.2002	PTOT	deleted
	6.7.2009	PTOT	12 µg l^{-1}
Huhtisuonoja 382 44	5.11.2008	NTOT	deleted
Latosuonoja 402 43	20.4.2011	NTOT	2700 µg l^{-1}
Virojoki 006 3020	6.6.2005	FE	deleted
	16.5.2001	FE	deleted
Iijoki Kipinä 13200	19.3.2002	PTOT	deleted
Kesselinpuro 35	24.4.2001	SS	6.3 mg l^{-1}

Appendix 2. Animal classes and their conversion to livestock units and manure nutrients.

Animal class nro	Animal class	Animal tot number in Finland	LSU factor	Ntot kg / animal (av. 2000-2009)	Ptot kg / animal (av. 2000-2009)
310	Lampaat	103970	0.15	9.7	1.5
320	Vuohet	5591	0.15	10.7	1.6
11100	Lypsylehmä 24kk-	286449	1	118.2	18.9
11200	Alle 24kk poikinut lypsylehmähieho	1093	0.6	49.9	6.1
12100	Emolehmä 24kk-	54725	1	67.6	12.9
12200	Alle 24kk poikinut emolehmähieho	608	0.6	49.9	6.1
13110	Sonni alle 6kk, lihantuotantoon	76278	0.375	41.0	5.7
13120	Sonni alle 6kk, jalostukseen	1175	0.375	41.0	5.7
13210	Sonni väh. 6kk, alle 8kk, lihantuotantoon	22078	0.375	41.0	5.7
13220	Sonni väh. 6kk, alle 8kk, jalostukseen	140	0.375	41.0	5.7
13310	Sonni väh. 8kk, alle 12kk, lihantuotantoon	49964	0.375	41.0	5.7
13320	Sonni väh. 8kk, alle 12kk, jalostukseen	383	0.375	41.0	5.7
13410	Sonni väh. 12kk, alle 16kk, lihantuotantoon	52873	0.6	62.3	8.0
13420	Sonni väh. 12kk, alle 16kk, jalostukseen	654	0.6	62.3	8.0
13510	Sonni väh. 16kk, alle 21kk, lihantuotantoon	39595	0.6	62.3	8.0
13520	Sonni väh. 16kk, alle 21kk, jalostukseen	282	0.6	62.3	8.0
13610	Sonni väh. 21kk, alle 24kk, lihantuotantoon	8767	0.6	62.3	8.0
13620	Sonni väh. 21kk, alle 24kk, jalostukseen	240	0.6	62.3	8.0
13710	Sonni väh. 24kk, lihantuotantoon	9231	1	62.3	8.0
13720	Sonni väh. 24kk, jalostukseen	2504	1	62.3	8.0
14110	Lehmävasikat ja hiehot alle 6kk, lihantuot./emol.	8370	0.375	30.9	4.2
14120	Lehmävasikat alle 6kk, jalostukseen	3	0.375	30.9	4.2
14130	Lehmävasikat alle 6kk, emolehmäksi	11440	0.375	30.9	4.2
14140	Lehmävasikat ja hiehot alle 6kk, lypsylehmiksi	58309	0.375	30.9	4.2
14210	Lehmävasikat ja hiehot 6kk-12kk, lihantuot./emol.	10112	0.375	30.9	4.2
14220	Lehmävasikat 6kk-12kk, jalostukseen	2	0.375	30.9	4.2
14230	Lehmävasikat 6kk-12kk, emolehmäksi	6377	0.375	30.9	4.2
14240	Lehmävasikat ja hiehot 6kk-12kk, lypsylehmiksi	55121	0.375	30.9	4.2
14310	Lehmävasikat ja hiehot 12kk-24kk, lihantuot./emol.	12439	0.6	49.9	6.1
14320	Hiehot 12kk-24kk (ei poikineet), jalostukseen	15	0.6	49.9	6.1
14330	Hiehot 12kk-24kk (ei poikineet), emolehmäksi	12853	0.6	49.9	6.1
14340	Lehmävasikat ja hiehot 12kk-24kk, lypsylehmiksi	107137	0.6	49.9	6.1
14410	Hiehot 24kk- (ei poikineet), lihantuotantoon	1679	1	49.9	6.1
14420	Hiehot 24kk- (ei poikineet), jalostukseen	3	1	49.9	6.1
14430	Hiehot 24kk- (ei poikineet), emolehmäksi	5212	1	49.9	6.1
14440	Hiehot 24kk- (ei poikineet), maidontuotantoon	24151	1	49.9	6.1
19100	Muu nauta yli 24 kk	1565	1	no data	no data
19200	Muu nauta 6-24 kk	151	0.6	no data	no data
19300	Muu vasikka alle 6 kk	37	0.375	35.9	4.9
20000	Hevoset	29803	1	58.7	9.7

41100	Lihasiat, elopaino 50-80 kg	290132	0.11	17.5	3.9
41200	Lihasiat, elopaino 80-110 kg	166411	0.11	17.5	3.9
42100	Tiineet, vähintään kerran porsineet emakot	100767	0.4	28.8	7.2
42400	Emakoiksi tarkoit. siat, alle 8 kk	34243	0.11	8.8	1.9
43000	Siitoskarjut, elopaino 50kg-	2560	0.11	19.7	5.1
44000	Siat, elopaino 20kg-50kg	282377	0.11	8.8	1.9
45000	Porsaat, elopaino alle 20kg	347816	0	no data	no data
51000	Munivat kanat väh. 20 vk (ei broileremot)	3404575	0.011	0.6	0.2
52000	Kukot	12884	0.005	0.9	0.2
53000	Broileremot väh. 18 viikkoa	419314	0.01	0.4	0.1
54000	Broilerit	4322014	0.002	0.4	0.1
55100	Kananpoikaset 16- alle 20 vk (pl. kukkop. ja br.)	830850	0.002	0.4	0.1
56100	Ankat	1011	0.005	0.0	0.0
56200	Kalkkunat	271946	0.005	1.5	0.4
56300	Hanhet	1124	0.005	no data	no data
56400	Fasaanit	5828	0.005	no data	no data
56500	Sorsat	4084	0.005	no data	no data
56700	Helmikanat	61	0.005	no data	no data
56800	Strutsit	124	0.005	no data	no data