Emergo
Studies in the Biogeophysical Environment

Statistical and neural network analysis of pesticide losses to surface water in small agricultural catchments in Sweden

Stina Adielsson

Picture from the Vemmenhög catchment (Photo: Jenny Kreuger)

Abstract

The aim of this thesis was to explain variations in pesticide leaching from the pesticide properties (DT_{50}, k_{oc}, log P_{ow}, S_w and combinations of these) using multiple linear regression and artificial neural networks. The data came mainly from Vemmenhög, a catchment nine square kilometres in size dominated by agriculture, located in the south of Sweden. The analysed period is May to November 1997-2003. The artificial neural network, a feed-forward back propagation network, did not work in this case. For the regression analysis, a stepwise selection was used. Analyses were performed both on data where all zero-losses were excluded and on data where substances used in low amounts were excluded. Excluding the pesticides that were applied in low amounts gave better results than excluding those with zero-loss. With loss rate as a response variable, it was possible to find significant functions explaining up to 99% of the variability for individual years. The combination of variables in the functions with the highest degree of explanation (r^2) differed for different years, but DT_{50}/k_{oc} was the most frequently occurring variable. Grouping the years, the best significant function for 1997-2003 (excluding pesticides used in low amounts) contained DT_{50}/k_{oc} and log P_{ow}, with an r^2 value of 70% (P<0.0001). It was generally not possible to use the formulas to predict pesticide loss for individual years, but it proved to be more reliable for the grouped years. The highest model efficiency found was 0.56. The result implies that a large part of the long-term leaching can be explained by pesticide properties.

Table of contents

1. Introduction

About 1700 tons of pesticides are sold each year in Sweden (www, Swedish environmental protection agency, 2004), which corresponds to some four million hectare doses applied to approximately half of the total arable acreage. A small fraction of these pesticides leach from the fields and end up in our streams and lakes. It is desirable to decrease this leaching. The Swedish Environmental Protection Agency has set up 15 environmental quality objectives of which three involve pesticide leaching. These goals are one expression of the general wish to decrease pesticide losses to the environment. Pesticides can have ecotoxicological effects on the aquatic environment and also impact human health through drinking water. To decrease losses it is important to improve understanding of when and why pesticides leach from Swedish agricultural soils. One way to achieve this is to investigate the leaching process itself from a mechanistic point of view. Another way is to try to explain variations in losses resulting from environmental factors and pesticide properties in a statistical sense. This thesis takes the second approach.

This project aims to rank the different factors, mainly pesticide properties, that contribute most to pesticide leaching, using multiple linear regression and neural networks. The analysis is performed mainly on data collected at Vemmenhög, a well-studied catchment in the south of Sweden.

Many efforts have been made to predict pesticide leaching from compound properties. An American study showed that k_{oc}, S_w and DT_{50} could not explain the occurrence of pesticides in groundwater. Instead pesticide properties such as molecular size, branching and functional group composition were chosen together with the catchment properties: organic carbon content, percentage sand and depth to the water table (Worrall and Kolpin, 2004). Sorensen et al. (2003) used partial order theory and found that the most important parameter describing pesticide loss was the per hectare dose, followed by sprayed area and adsorption to organic matter.

Pesticide loss to surface water was previously analysed in the Vemmenhög catchment. Kreuger and Törnqvist (1998) used multiple linear regression to look at the period 1990-1994. They found that the applied amount of pesticide was the most significant variable. It explained 50-85% of the variability for concentrations and transported amounts of pesticides in the Vemmenhög River. The concentrations and transported amounts linearly depend on the applied amounts of pesticide and thus this result is not surprising. The regression could be slightly improved by adding the intrinsic properties of the pesticides. With loss rate (percentage of dose) as dependent variable, log P_{ow} was found to be the most important pesticide property, explaining 26% of the variation (Kreuger & Törnqvist, 1998). Using the loss rate excludes the dependence on the applied amount and is therefore preferable.

Pesticides are transported to surface waters in two different ways. Either as point sources caused by spill or accidents, or as diffuse transport through the soil. It is possible that the lack of success in explaining pesticide losses from compound properties may be due to the strong influence of point sources. In Sweden, there have been several national programs to inform and educate farmers on how to avoid "spill" and to use plant protection products in the best possible way, also from the environmental point of view. In figure 1 the strong decrease in transported amounts

indicate that this work has given results in Vemmenhög. Figure 2 shows the applied amounts of pesticide for the same period. The amounts of pesticides that are transported to the stream have decreased considerably even though the applied amounts have not. We can therefore assume that point sources are now a minor component of losses in Vemmenhög. The diffuse leaching of pesticides seem to have dominated since about 1997. This prompted us to look at this area again, to try to find out what factors are controlling the losses of pesticides. The years that are of interest are from 1997 to 2003.
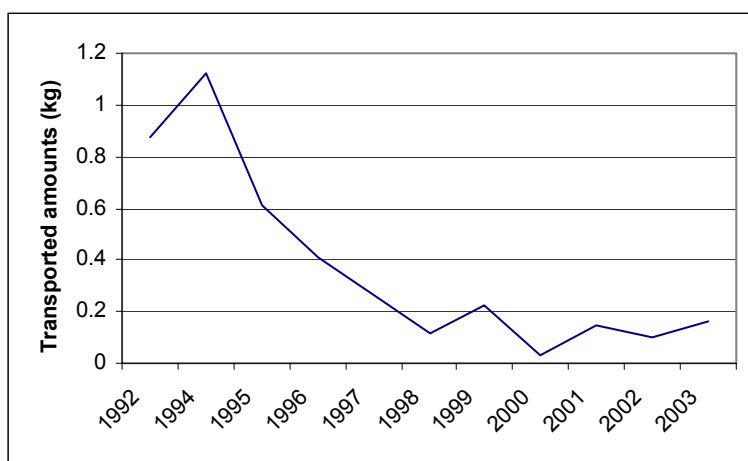
Figure 1. Amounts of spring applied pesticides transported during May-September 92, 94 and May-November for the remaining years.
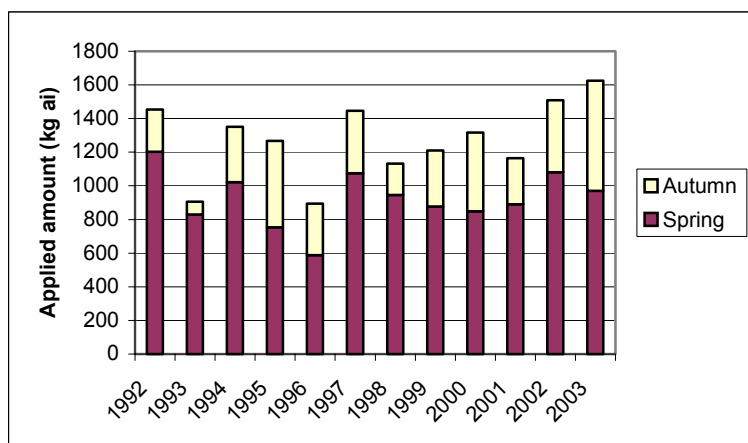
Figure 2. Total amounts of active ingredients of pesticides applied within the Vemmenhög catchment.

The loss rate, as percentage of the applied amount, was chosen as dependent variable. Predictor variables were the pesticide properties. In a first step, multiple linear regressions were performed for all the years separately. Secondly, the years were clustered into groups. Thereafter, the whole period was analysed with artificial neural networks (ANN), with a weather variable also included. A simple comparison was also made with lysimeter studies and three other study areas within the Swedish programme for environmental monitoring.

2. Material and methods

2.1. Vemmenhög –a short presentation

The Vemmenhög area in the southern most part of Sweden (55° 26' N; 13° 27' E) has been studied since the beginning of 1989, when water sampling started. The area is nine square kilometres in size of which 95% is arable land, mainly with sandy loam soils. The main production is for cereals but sugar beets and rape are also common. A more detailed description of the catchment is given in Kreuger (1998).

Most of the arable land (828 ha) is artificially drained and water samples were taken in the main culvert of the drainage system before it runs into the stream. A programmable automatic sampler was used to take water samples during May-November 1995-2003 and also during the winter 2001-2002. Water flow has been measured with an ISCO model 4250 since 1999 and calibration was done to previous instruments. Detailed information on sampling and analysis methods can be found in the yearly project reports (Kreuger, 1996, 1997; Kreuger & Hessel, 1998; Kreuger, 2000, 2002a, 2002b, 2003; Kreuger et al. 2003; Kreuger et al. 2004) and also in Kreuger (1998).

2.2. Vemmenhög –data selection

The pesticides included in this analysis are those applied during the spring and summer. Some pesticides are applied both during spring and autumn but those used mainly during the autumn have been excluded, for example cyfluthrin in 1997, glyphosate in all years, isoproturon in all years and metazachlor in 1999. Four substances (mecoprop-P 1998-2000 and 2001, esfenvalerate 1999, 2000, 2002 and 2003, cypermethrin 2001 and 2003 and finally deltamethrin 2002) were used mainly in spring, but occasionally also in autumn. For these substances, the time of application was compared to the results of the water samples and when they indicated an increase in concentration after autumn application, the substance was excluded.

Two different procedures were used in the regression analysis. The first, hereafter called A, excluded all substances with a zero loss rate. The reason for this was that on plots of the data against pesticide properties (e.g. figure 3) one could notice that pesticides with zero-loss occurred throughout the whole range of DT_{50}, which indicates that other factors must affect the loss rate of these pesticides.
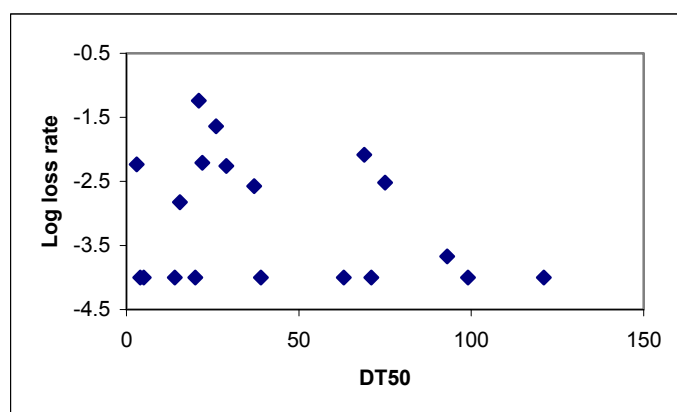

Figure 3. The distribution for 2000A with zero-losses put to 0.0001.

The B-procedure excluded pesticides that were applied in amounts of less than three kilograms for the entire area. The reason was that the amounts applied were thought to be so small that any leaching would probably be at concentrations below the detection limit in the water analysis. This approach includes in the analysis those pesticides that are applied in high amounts but still do not leach (table 1). The substances marked with one or two crosses were all excluded in the A-procedure. For the B-procedure, only those marked with one cross were excluded. An empty space in the table means that leaching was recorded for the substance in that year.

Table 1. Substances not detected are marked and a difference is made between those applied in amounts smaller or larger than three kilograms.

Substance	1997	1998	1999	2000	2001	2002	2003
Aclonifen	XX		X	XX	-	-	
Alpha-cypermethrin	-	-	X	X	X	X	X
Amidosulfuron	-	-	-	-	-	-	X
Azoxystrobin	-	-	-	XX			
Chloridazon	XX	XX	X	-	-	-	-
Cyanazine	-		X	-	-	-	-
Cyfluthrin	-	X	X	X	-	-	-
Cypermethrin	-	-	-	-	X	XX	X
Cyprodinil	-	-	-	-	-	-	XX
Deltamethrin	X	X	X	-	-	X	X
Esfenvalerat	X	XX	XX	XX	XX	X	X
Fenoxaprop-P	-	-	-	X		-	-
Fenpropimorph			XX				
Iprodione	-	-	-	-	XX	-	-
Lambda-cyhalothrin	X	-	-	-	-	-	-
Phenmedipham	XX	XX	XX	XX	XX		XX
Sulfosulfuron	-	-	-	-	-	-	X
Terbutryn	-	XX	-	-	-	-	-
Thifensulfuron-methyl	-	X	X	X	X	X	
Tribenuron-methyl				X		XX	X
Triflusulfuron-methyl	-	-	-	-	X	X	X

X = zero loss at an amount ≤3 kg
XX = zero loss despite applied amount >3 kg
- = not applied that year

The years were also clustered into groups. The applied amounts of a substance were added for all the years, then the amounts lost were added and finally the loss rate was calculated.

2.3. Catchment areas in the Swedish environmental monitoring programme
Four catchment areas are investigated within the Swedish environmental monitoring programme on pesticide leaching. One is Vemmenhög, as described above. The other catchment areas are in the counties of Östergötland, Halland and Västergötland. The area in Östergötland is 1681 ha with 89% consisting of arable land, in Halland 92% of the total 1460 ha is arable land and in Västergötland 91% of the 776 ha consists of arable land. Monitoring in these areas started in 2002, so there is only data available

for two years. Generally the same technique is used as for Vemmenhög but the sampling period is May-October (not November as is the case in Vemmenhög). Details on water sampling and pesticide analysis can be found in Kreuger at al. (2003 and 2004). The same pre-treatment was made as for the Vemmenhög data.

2.4. Lysimeter study
From May 1989 to February 1990 a large lysimeter study on leaching of dichlorprop was performed at Ultuna, Uppsala (Bergström & Jarvis, 1992 and 1993). The lysimeters were one metre deep and contained undisturbed soil from four different places. The soils were chosen to represent four different textures, sand, loam, clay and peat. There were four replicates for the sand and three each of the other soils. A more detailed description can be found in Bergström and Jarvis (1992 and 1993). In this analysis the loss rate, as percentage of applied amount, was used to investigate differences in leaching caused by texture.

2.5. Pesticide properties
Four pesticide properties are included in this study. The sorption coefficient, k_{oc}, describes the potential of the pesticide to bind to soil organic carbon. The higher it is the stronger the tendency for sorption and the smaller the risk for leaching. A high k_{oc} also leads to a longer residence time in the soil, since adsorption prevents the substance from being degradated. Roughly speaking a $k_{oc}<100$ means that the substance is mobile in the soil and a $k_{oc}>1000$ that it is immobile. The degradation half-life, DT_{50}, affects leaching such that the slower a substance is degraded the larger the risk. Both degradation and sorption are very sensitive to variations in soil and climate factors. Water solubility, S_w, describes the amount of pesticide that can be dissolved in a specific amount of water at a specific temperature. A pesticide with a high solubility is more likely to leach because it will be dissolved in the soil solution to a greater extent and can easily follow the percolating water through the profile. The octanol-water partition coefficient, $\log P_{ow}$, describes the extent to which the substance is dissolvable in organic solutes compared to water, a low $\log P_{ow}$ leads theoretically to high leaching.

Values for these properties were obtained from various sources and ranked according to their reliability (see appendix 1-4). For k_{oc} and DT_{50} two files consisting of information from EU monographs and a summary file created by N.J. Jarvis (pers. com.) were placed number one and two, followed by the pesticide properties used by Kreuger and Törnqvist (1998). Values from Linders et al. (1994), the ARS Pesticide Properties Database (www, Agricultural research service, 2004) and The Pesticide Manual (Tomlin, 2003) were used for calculating mean values when other sources were missing. For S_w and $\log P_{ow}$ the highest rank was given to values from the Swedish Chemicals Inspectorate (KemI, 2004) who have recently made a review of most of the substances. Thereafter, the rank followed the same pattern as for the other properties. Some exceptions were made from the ranking and these can be seen in appendix 4.

To avoid problems with autocorrelation between DT_{50} and k_{oc}, the ratio of these two were used in some cases. Also the GUS index, $GUS = \log_{10} DT_{50} * (4 - \log_{10} k_{oc})$, (Gustafson, 1989) was sometimes used for the same reason.

2.6. Multiple linear regression

The statistical analysis was performed in JMP IN version 5.1 (SAS Institute Inc., Cary, NC, USA, 1989-2003). Multiple linear regression (MLR) implies the prediction of a response variable from a linear combination of other (independent) variables. The best way to select the most important parameters is to make all possible comparisons (Quinn and Keough, 2002), but since that was regarded as too complicated, the analyses were made using a stepwise selection.

The stepwise procedure adds variables to the function depending on their significance level (P-value), in this case set at P≤0.25. In a second step, only variables with a significance of at least 0.10 are accepted in the function and the rest are removed. As a complement to the stepwise selection, some other combinations of variables were also tested. Since the r^2 value normally increases with an increasing number of variables, an adjusted r^2 was used for comparing the different functions obtained in MLR. This takes account of the fact that there are different numbers of variables in the functions.

The advantages of MLR are that it is a commonly used and well-investigated technique that is simple to interpret. The greatest disadvantages are that it assumes linear relationships and that the dataset should fulfil many criteria. For example, in regression analysis, the response variable should be normally distributed, have a constant variance and be independent, meaning in this case that x should not be a part of y (Grandin, 2004). Furthermore the predictor variables have to be uncorrelated. These conditions are hard to meet with data on pesticide leaching.

Correlation matrixes were constructed to study the strength of the correlations between different pesticide properties. Where the correlation was highest, these variables were avoided in the same function. Analyses where the variance of the loss rate varied in a systematic way with the predictor variables were excluded since these functions are invalid (Grandin, 2004). Outliers were identified using box plots in JMP. When points occurred outside the 1.5 interquartile range from the quartile (25th or 75th percentile), they were classified as outliers and therefore excluded. The loss rate was log-transformed to obtain an approximation of a normal distribution.

Cross-validation was used to test the derived prediction models. One year was excluded from the group of years forming the prediction function and then this function was tested on the year that was excluded. The procedure was repeated so that all years were excluded once. Model efficiency, ME, (Loague & Green, 1991) was used as a measure of the general agreement between the observed and the predicted values:

$$ME = 1 - \frac{\Sigma(obs_i - pred_i)^2}{\Sigma(obs_i - mean\_obs)^2}$$

where obs is the observed value, mean_obs the arithmetic mean of all the observed values and pred is the predicted value. Model efficiency always takes a value less than one, values larger than 0.5 are usually considered acceptable.

An extensive MLR analysis was performed on the Vemmenhög data. The data from the environmental monitoring catchment areas was used only to test if the prediction functions were generally applicable.

2.7. Other statistical analysis

ANOVA analysis and Tukey's LSD test were also performed in JMP. The methods were used on the lysimeter data to look for differences in leaching from different soils.

2.8. Artificial neural network (ANN)

Artificial neural networks imitate the function of a brain. A network can be trained on different events and "learn" to predict other events. The main advantage compared to statistical methods, such as MLR, is that ANN models non-linear data and there is no need to specify the form of the model in advance. Another advantage is that ANN does not have the same requirements on the dataset as regression. ANN also has a tolerance to noise and an ability to generalize. One disadvantage is that the results can be hard to interpret.

When ANN is compared to traditional techniques such as MLR, it is often concluded that ANN is the better method, especially for non-linear relationships and systems (Salt et al., 1992; Chen et al., 1999; Starrett et al., 1998). Starrett et al. (1998) compared ANN to first and second order regression and concluded that "ANN provided to be a feasible modelling technique for pesticide leaching".

The type of network that was used in this project was a feed-forward back-propagation network. A feed-forward ANN consists of several layers: one input layer, one output layer and one or more hidden layers (figure 4). To approximate continuous functions one hidden layer is usually enough (Masters, 1993; Basheer & Hajmeer, 2000). Each layer consists of a number of nodes. In a fully connected feed-forward ANN, each node in one layer is connected to all the nodes in the next layer and these connections have weights that are adjusted during training, so this is where the actual "learning" takes place.
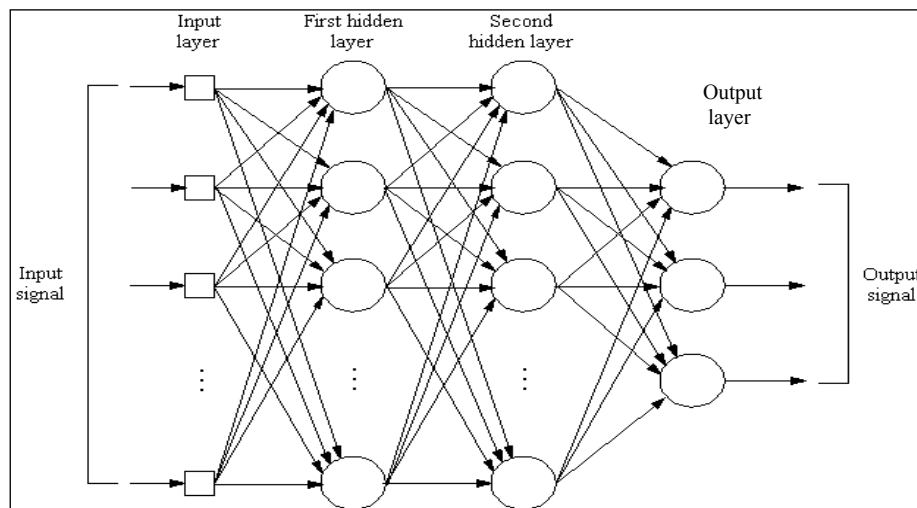

Figure 4. Schematic picture of connections in artificial neural network.

In order to compute an output signal from the ANN, an input pattern is presented to the nodes in the first layer. The input signal is fed forward through the nodes in the hidden layer to the output layer. At each node in the hidden layer, a weighted sum of the signal received from the input layer is computed:

$$v = \sum_{i=1..N}^{N} \left(w_i * I_i \right)$$

where w_i is the weight associated with the ith connection and I_i is the input signal from the ith node, and N is the number of input nodes. The output from each node is given by:

$$y_i = f(v_i)$$

where f(.) is the transfer function in use. The output from the nodes in the hidden layer is then transferred in the same way to the nodes in the output layer.

The second part of the name, back-propagation, describes the algorithm that is used to train the network, or the way the network learns. After the output value is calculated, this value is compared to the given (observed) value in the training pattern and an error is calculated. The error is then propagated backwards to the input layer and the weights in the network are adjusted to minimize the difference between the calculated and observed value. This process is repeated for each pattern. More details on how an ANN works mathematically can be found in Basheer & Hajmeer (2000) and Bishop (1994).

In this project, fully connected feed-forward back-propagation ANN's were constructed using the computer program PDP++ software (version 3,1; released Oct 2, 2003. Address to homepage in references). The back-propagation algorithm (Rummelhart et al., 1986) was chosen since it is the most widely used training algorithm and it is considered robust. The sigmoid transfer function was used.

One difficulty is to choose the number of nodes in the hidden layer (NHN). With too few hidden nodes, the network will tend to produce a linear estimation of the trend and with too many, the network will lose its ability to generalize. Several rules of thumb exist and as a starting point I followed Masters (1993), who suggested that:

$$NHN \approx \left(N_{in} * N_{out} \right)^{1/2}$$

where NHN is the number of hidden nodes, N_{in} is the number of input nodes and N_{out} is the number of nodes in the output layer.

The number of nodes also influences the number of training patterns required. It can be estimated using the ratio of the number of training patterns to the number of weights in the network. The ratio should be greater than four according to Masters (1993). The total dataset was divided into a training set and a test set with 80% and 20% of the dataset respectively, according to Basheer and Hajmeer (2000).

All ANN's were built with one node in the output layer representing the loss rate of the pesticide. The number of nodes in the input layer differed between networks (see table 2), but all networks had four input nodes representing the pesticide properties. It is desirable to have as few input parameters as possible (Bishop, 1994). To have only one node representing the weather, I chose to use the ratio of the amount of rain, summed for May-June, to the mean temperature for the same period. A higher value of the weather parameter implies a rainy and cold period while a warm dry period would result in a low value of the weather parameter. Theoretically, the lower the weather parameter the less leaching since a high temperature increases degradation and low rainfall reduces leaching. Using this weather parameter, it was possible to combine the pesticide loss for all individual years together in one large data set.

The data used were all pesticides applied in amounts of more than three kilograms (B-procedure in regression analysis). Measurements from 1997 to 2003 were combined in one data set. This was done for all networks except 9703S (table 2) where the same data as for the regression analysis was used. About 20% of the total data set was used for testing the performance of the built networks. The data used for testing was randomly chosen.

The pre-processing was considered by Bishop (1994), among others, to be one of the most important factors ensuring a good result. This is why different scaling intervals were adopted and also log-transformation in one case. However, there is no strict requirement for scaling the data. The training patterns were scaled to increase learning, using the formula:

$$x = \lambda_1 + (\lambda_2 - \lambda_1) * \frac{z - z_{min}}{z_{max} - z_{min}}$$

where x is the scaled number, λ_1 and λ_2 are the lower and upper limits of the interval to which the number will be scaled, z is the number itself, z_{min} is the minimum value and z_{max} is the maximum.

There are different ways of updating the weights during training, but in this project the weights were only updated after each 'epoch' (one epoch is when all the training patterns have been presented once). There are three different ways of presenting the data to the network: in the same order every time (sequential), random order within the epoch (permuted) or completely random. Another parameter that can be changed is the learning rate, which refers to how much the weights are adjusted each time. The momentum measures how much of the weight change is carried through to the next weight change and can therefore allow training to speed up when weights are adjusted in the same direction.

When to stop training the network is crucial because there is always a risk of over-training, meaning that the network loses its ability to generalize. A good way to avoid over-training is cross-validation (Bishop, 1994; Livingstone et al., 1997 and Masters, 1993). This involves training on one set of data and testing regularly on another and when the error on the test set is reduced to a minimum the training is stopped. Even with this method there is a risk of ending up in local minima. To avoid this, the network should be trained with different starting weights (Masters, 1993).

The different networks built in this study are presented in table 2. In all cases, trial and error was used to choose learning rate, momentum, weight updating and the order of pattern presentation to optimise the learning and performance of the ANN. To measure the performance, model efficiency (ME) was used (Loague and Green, 1991). The same measure was also used for MLR, which makes it possible to compare the performance of ANN with the regression equations.

Table 2. Details on the different properties of the neural networks.

	Nr of input	Nr of hidden nodes	Size of training set	Size of testing set	Transfor- mation	Scaling to interval	Learning rate	Momen- tum	Selection of pattern
9703K	5	3	73	18	-	0.1 - 0.9	0.2	0	sequential
9703L	5	3	70	16	-	0.1 – 0.9	0.25	0.5	permuted
9703M	5	5	69	16	-	0.1 – 0.9	0.15	0.7	permuted
9703N	5	30	69	16	-	0.1 – 0.9	0.15	0.3	permuted
9703P	5	50	69	16	-	0.1 – 0.9	0.15	0.3	permuted
9703Q	5	3	70	16	log	0.1 – 0.9	0.25	0.3	permuted
9703R	5	3	73	18	-	-1 to 1	0.2	0.3	permuted
9703S	4	3	19	4	-	0.1 – 0.9	0.15	0.7	permuted
9703T	5	10	68	15		0.1 – 0.9	0.25	0.3	permuted

K: no extra pre-processing

L: outliers excluded

M: like L but with more hidden nodes

N: like L but with more hidden nodes

P: like L but with more hidden nodes

Q: the response variable was log-transformed

R: variables scaled between -1 and 1

S: the same data set as for the MLR analysis 9703B

T: excluded more outliers than in L

3. Result and Discussion

3.1. Weather

For substances applied in spring pesticide loss is generally most affected by the weather in May and June. The long-term mean temperature (based on 30 years of measurements made at Vemmenhög by SMHI) for May is 10.8°C and 14.9°C in June, while the monthly precipitation is 39 mm and 48 mm respectively. The two years that differ the most in precipitation are 1997 and 2002. In those years, May was very rainy with twice the normal amounts. The temperatures did not differ more than one and a half degrees from the long-term mean in any year.

Stream runoff values for May and June at Vemmenhög can be seen in figures 5 and 6. There is generally more runoff in May than in June, which is due to higher temperatures and the presence of more fully developed crop cover in June, resulting in a higher evapotranspiration. 1999 stands out as a year with high runoff for both May and June. The precipitation was only slightly above the long-term average for May, but both Marsh and April had more rain than normal. In June it rained twice as much as normal explaining the high runoff. May 2002 was very rainy, which may explain the high runoff at this time.
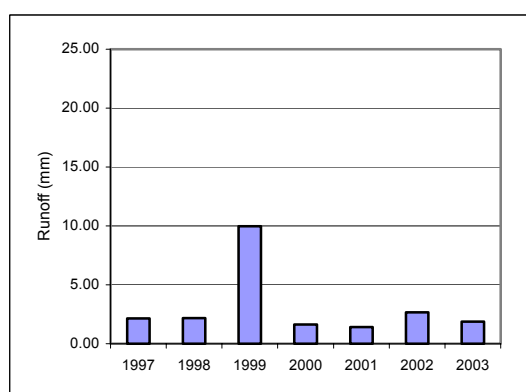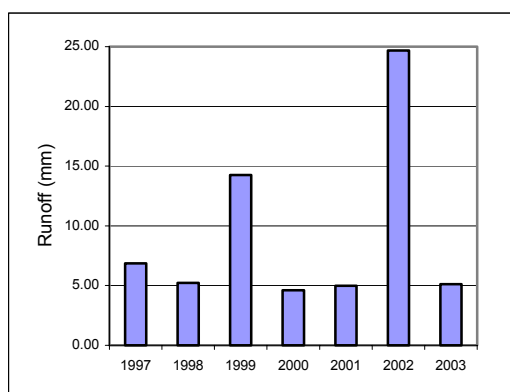
Figure 5. Runoff in Vemmenhög River for May. **Figure 6.** Runoff in Vemmenhög River for June.

3.2. Multiple linear regression

3.2.1. A-procedure

The results from the MLR analysis for the A-procedure (i.e. no zero-losses included) can be seen in table 3. Generally the stepwise procedure chose different combinations of variables with the highest degree of explanation in different years. One variable that occurs frequently is DT_{50}/k_{oc}. This seems to be a good choice of predictor variable if only one is to be selected. Sometimes the result improved with the addition of log S_w. The year with the highest degree of explanation was 1997, where r^2 equals 0.84 (P= 0.009). It was not possible to find any significant function for 2001 and 2003. Nor was it possible to find any significant regression equation for the year with sampling all winter (01-02A).

Table 3. Result from MLR analysis on the A-procedure, with the variables that were included in the best possible functions, their significance level and rate of explanation.

Year	Variables	P	r^2	r^2 ad	n
1997A	DT_{50}, k_{oc}, log P_{ow} [1,2]	0.009	0.84	0.76	10
	k_{oc}	0.005	0.65	0.60	10
1998A	k_{oc}, log P_{ow}, S_w [1]	0.031	0.61	0.48	13
	log P_{ow}, log S_w	0.005	0.66	0.59	13
	DT_{50}/K_{oc}, log S_w	0.034	0.49	0.39	13
1999A	GUS	0.002	0.71	0.68	10
	k_{oc}, DT_{50}	0.010	0.73	0.66	10
	DT_{50}/k_{oc}	0.023	0.49	0.43	10
2000A	k_{oc}, log P_{ow} [1]	0.031	0.62	0.52	9
	DT_{50}/k_{oc}	0.016	0.54	0.48	9
	DT_{50}/k_{oc}, log S_w	0.032	0.62	0.52	9
2001A	no significant function was found				13
2002A	DT_{50}, S_w [1]	0.002	0.75	0.69	12
	DT_{50}	0.002	0.63	0.60	12
	DT_{50}/k_{oc}	0.011	0.49	0.44	12
2003A	no significant function was found				13
01-02A	no significant function was found				17

A all pesticides with zero-loss were excluded

[1] stepwise selection with DT_{50}, k_{oc}, log P_{ow} and S_w included

[2] correlation between all variables on 0.7

3.2.2. B-procedure

The result changed when pesticides with applied amounts of less than three kilograms of active ingredient were excluded (B-procedure, table 4). It was possible to find significant functions at $P<0.05$ for all years, but for three of the years (1997, 1999 and 2001), the r^2 values were less than 45%. In these cases, it seems to be chance what variable is chosen. 2001 gave the worst result for both procedures. The annual runoff for this year is smaller than for the other years, but runoff in May and June did not differ greatly (figure 5 and 6). 1999 is the only year with a very high runoff in both May and June, which may be the reason for the poorer result in this year.

Table 4. Result from MLR analysis on the B-procedure, with the variables that were included in the best possible functions, their significance level and rate of explanation.

Year	Variables	P	r^2	r^2 ad	n
1997B	S_w [1]	0.025	0.41	0.35	12
1998B	DT_{50}/k_{oc}, log P_{ow}, GUS, log S_w [2]	0.0062	0.80	0.71	13
	DT_{50}/k_{oc}, log P_{ow}, S_w [3]	0.003	0.78	0.70	13
	DT_{50}/k_{oc}	0.0059	0.51	0.47	13
1999B	log k_{oc} [3]	0.030	0.45	0.40	10
2000B	k_{oc}, log P_{ow}, S_w, log DT_{50} [4]	0.021	0.82	0.69	11
	log S_w [5]	0.012	0.52	0.47	11
2001B	log P_{ow} [6]	0.036	0.37	0.30	13
2002B	DT_{50}/k_{oc}, S_w [3]	<0.0001	0.986	0.98	9
	DT_{50}/k_{oc}	<0.0001	0.95	0.95	9
	DT_{50}, S_w [1]	0.020	0.73	0.64	9
2003B	DT_{50}/k_{oc}, DT_{50} [3]	0.0004	0.76	0.71	14
	DT_{50}/k_{oc}, S_w	0.005	0.62	0.54	14

B: all pesticides with a dose < 3 kg were excluded and the losses were log (+1) transformed
[1] stepwise selection with DT_{50}, k_{oc}, log P_{ow} and S_w included
[2] stepwise selection with [1] except for S_w, but with log DT_{50}, log k_{oc}, log S_w, GUS and DT_{50}/k_{oc} included
[3] stepwise selection with [1] plus log DT_{50}, log k_{oc}, log S_w, GUS and DT_{50}/k_{oc} included
[4] stepwise selection with [1] plus log DT_{50}, log k_{oc}, GUS and DT_{50}/k_{oc} included
[5] stepwise selection with [1] plus log S_w, log DT_{50} and log k_{oc} included
[6] stepwise selection with log DT_{50}, log k_{oc}, log P_{ow} and log S_w included

For the years with high r^2 values (1998, 2000, 2002 and 2003), DT_{50} and k_{oc} are included in all the best functions. In three of these years adding S_w increases the degree of explanation. In two of these four years, log P_{ow} has the same effect. As in the A-procedure, DT_{50}/k_{oc} is the variable that is chosen most often. In three of the years it is included in the best function. The highest single r^2 value was found in 2002 where a function containing DT_{50}/k_{oc} and S_w had an r^2 value of 99% (P<0.0001). It can be noticed that this was a year where no significant function was found using the A-procedure. It was also a year with high runoff, especially in May.

The B-procedure was also applied to the other three catchment areas included in the environmental monitoring programme for 2002. The only site where it was possible to find any significant function was Östergötland. Here, 53% of the variation could be explained with a function containing DT_{50}/k_{oc} (P=0.0029, n=15). These catchment

areas are not as well investigated as Vemmenhög and all the farmers are not yet participating in all of the areas. The measurement period was also different from that in Vemmenhög, which may explain the slightly disappointing result. Another factor that can have affected the result is the threshold amount of three kilograms of pesticide spread within the area. With catchments smaller than Vemmenhög the threshold should probably be a smaller amount.

3.2.3. Grouped analysis

For the grouped periods 97-00A and 01-03A (table 5) DT_{50}/k_{oc} explains about half of the variation for both periods and also for the period as a whole. 01-03 is generally less well explained, which is not surprising since two of the included years do not show any significant relationships. For 97-03A, the variable DT_{50}/k_{oc} alone explains 56% of the variability (P=0.0008) but the highest degree of explanation is obtained with log P_{ow}, S_w and GUS (r^2=0.72; P=0.001).

Table 5. Result from MLR analysis on the grouped period, with the variables that were included in the best possible functions, their significance level and rate of explanation.

Year	Variables	P	r^2	r^2 ad	n
97-00A	log P_{ow}, S_w, GUS [1]	0.001	0.72	0.65	16
	DT_{50}/k_{oc}, log P_{ow}	0.002	0.63	0.57	16
	DT_{50}/k_{oc}	0.0008	0.54	0.53	16
01-03A	DT_{50}/k_{oc} [1]	0.003	0.46	0.42	18
	log DT_{50}, log k_{oc}	0.043	0.34	0.27	18
97-03A	DT_{50}/k_{oc}	<0.0001	0.60	0.57	19
	DT_{50}/k_{oc}, log P_{ow}	0.0003	0.60	0.55	19
97-03B	DT_{50}/k_{oc}, log P_{ow} [1]	<0.0001	0.69	0.66	20
	DT_{50}/k_{oc}	<0.0001	0.62	0.60	20

A all pesticides with zero-loss were excluded

B all pesticides with a dose < 3 kg were excluded and the losses were log (+1) transformed

[1] stepwise selection with DT_{50}, k_{oc}, log P_{ow}, S_w, log DT_{50}, log k_{oc}, log S_w, GUS and DT_{50}/k_{oc} included

For 97-03B 70% of the variation is explained by a function containing DT_{50}/k_{oc} and log P_{ow} (P<0.0001). These variables were also the most common ones chosen in the analysis for the individual years. The equation was the following: Log (loss+1) = $3.4*10^{-3} + 12.6*10^{-3} DT_{50}/k_{oc} - 0.7*10^{-3}$ log P_{ow}. One weakness is that since all the data was used for constructing the model, there is none left to evaluate it for the same catchment area. An attempt was made to predict the loss in the other three areas within the environmental monitoring programme, but negative model efficiencies were found for all the areas. The result was equally poor using the prediction formula from 97-03B containing only DT_{50}/k_{oc}. There are several possible explanations for this: first the predicted period consisted of only two years, 2002-2003. Secondly the range of substances used in the different areas varies considerably. As pointed out previously, the catchment areas within the environmental monitoring programme are not well defined, all farmers are not participating and the sampling periods are not the same. The areas do also differ in soil and weather.

Comparing the A- and B-procedure, the results obtained for the whole period are more similar than when comparing the years separately. It may be that grouping the years has the effect of evening out the differences between individual years.

3.2.4. Cross-validation

The best ME found when predicting pesticide loss in single years was 0.37 (table 6). The result could be improved by dividing the period into two groups and predicting for the other group. The functions containing only DT_{50}/k_{oc} from both 97-00A and 01-03A were acceptable for prediction, with ME values of 0.46 and 0.56 respectively (table 6). These functions were chosen since the variable DT_{50}/k_{oc} was often chosen in the previous regression analysis. Other functions did not prove to be as good for prediction. One example can be seen in table 6.

Table 6. Result of the cross-validation. The functions, how well they explained the variation for the analyzed years and the performance in prediction.

Period analysed	Function	r^2	P	Predicting on year	ME
without 97B*	$y=0.00058+0.01625*DT_{50}/k_{oc}$	0.711	<0.0001	1997B	0.12
without 98B*	$y=0.00118+0.01792*DT_{50}/k_{oc}$	0.566	0.0003	1998B	0.13
without 99B*	$y=0.00092+0.01558*DT_{50}/k_{oc}$	0.5699	0.0001	1999B	0.12
without 00B*	$y=0.00097+0.01770*DT_{50}/k_{oc}$	0.594	<0.0001	2000B	0.10
without 01B*	$y=0.00121+0.01417*DT_{50}/k_{oc}$	0.276	0.025	2001B	0.10
without 02B*	$y=0.00061+0.01864*DT_{50}/k_{oc}$	0.604	<0.0001	2002B	0.37
without 03B*	$y=0.00087+0.01855*DT_{50}/k_{oc}$	0.565	0.0002	2003B	0.24
97-00A	$y= -2.49385+2.00657*DT_{50}/k_{oc}$	0.56	0.001	01-03A	0.46
01-03A	$y= -2.42580+1.86766*DT_{50}/k_{oc}$	0.46	0.003	97-00A	0.56
97-00A	1)	0.68	0.001	01-03A	0.16

* The whole period except for one year was used for creating the function
1) $y= -1.17092+1.04465*\log DT_{50}+ -1.02290*\log k_{oc}$

It is remarkable that functions containing only DT_{50}/k_{oc} are able to explain and predict such a large part of the loss. Many parameters not included in the analysis will affect the pesticide loss, such as time of application, weather including both temperature and precipitation at the time of application and in the period immediately following, as well as mean temperature and precipitation during the whole year.

3.2.5. Sources of error

Data selection is crucial. Excluding all pesticides with a low application amount according to the B-procedure is the best way in principle and it also proved to give better estimations. The choice of threshold can be discussed, but I chose to follow Kreuger and Törnqvist (1998). For Vemmenhög this is probably a good choice, but for the other smaller areas it may be too large.

It would have been desirable with measurements of pesticide loss during the whole year. Kreuger (2003) showed that although the concentrations in the stream are low during the winter, the transported amounts can be large because of the large discharge. In my analysis I assumed that there is approximately the same relation between the loss in the measured period and the rest of the year. One weakness is that the partitioning between winter leaching and that in the other seasons is probably

different for different pesticides, mainly depending on their half-life. The substances fluroxipyr and mecoprop are applied in spring but according to the measurements made in winter 2000-2001 they had 74% and 79% of their yearly loss in the period that is not normally sampled (Kreuger, 2003). Consequently the loss is probably underestimated for some substances, while not for others. Another source of error is that some persistent pesticides can leach during long periods, so that "carry-over" can affect the measured values in the next year.

It is difficult to find representative values for some of the pesticide properties. The reported values depend strongly on how the analysis was made. Another important factor is that soil is not a homogeneous media, and conditions can differ between fields and also at smaller scales. Microbial activity for example differs greatly, as does water content, temperature and organic matter content. All these factors affect the partitioning between adsorbed and dissolved amounts of pesticides and also the rate of degradation and therefore also pesticide leaching. The weather differs between years and also affects, for example, the rate of degradation (the warmer and moister the soil, the faster the degradation).

The pesticide properties used in the analysis were the best ones that were available at the start of the project. Later, some updating of the information gathered from EU monographs (J. Asp pers. com.) became available as can be seen in appendix 5. A short comparison was made with the newest available pesticide properties, analysing for the grouped period of 1997-2003B. The result was that with only DT_{50}/k_{oc} the r^2 value increased from 0.62 to 0.68 (P<0.0001). After adding log P_{ow}, the r^2 value increased further to 0.72 (P<0.0001).

3.2.6. Comparison with previous analyses

In the previous analysis on data from Vemmenhög (Kreuger and Törnqvist, 1998), log P_{ow} was found to be the most significant variable explaining 26% of the loss rate. Now the degree of explanation has increased considerably and the variables chosen have changed. The conclusion is that DT_{50} and k_{oc} are the most significant properties for estimating pesticide loss.

The different results can be due to either the availability of better values describing the pesticide properties or to a decrease in point sources leading to a larger impact of the pesticide properties. A short comparison was made on data from 1992 and 1994 with the newer values of pesticide properties. The result did not differ much from what was found in the analysis by Kreuger and Törnqvist (1998). Therefore it is probably the decrease in point source losses (figures 5 and 6) that has improved the result. As discussed earlier, point sources are normally considered a large part of the loading to streams and pesticide properties can only be expected to influence diffuse leaching.

The result also contradicts the conclusions of Worrall and Kolpin (2004). In the area that they examined it was not possible to explain pesticide leaching from the pesticide properties. The study was made on leaching to ground water in the Mid-western United States. There are no reports on any actions against point sources and one can therefore expect that point sources contribute to leaching to a great extent and this may explain why there was no significant effect of pesticide properties. The soils

within the area also vary, which may affect the leaching more than pesticide properties.

In the future it would be interesting to look more carefully at the other areas within the Swedish environmental monitoring programme. When all the areas are well defined, measured at the same period and when point sources have decreased they will be comparable. It would be interesting to see if the same pesticide properties would be chosen for all the areas. It may also be possible to investigate what difference the soil and the weather makes. For future studies it would be valuable to have analysed water samples representing the whole year.

3.3. Lysimeter study

The data from the lysimeter study was not appropriate to use for an MLR analysis. Instead ANOVA and Tukey's test were performed. The ANOVA analysis showed that there was a significant difference ($P=0.023$) between the pesticide losses from the different soils. Tukey's test showed that the peat leached more than both the sand and the loam. No significant differences were found between the other soils. The most likely explanation for the fact that the peat showed such high leaching despite the high content of organic matter (which would theoretically strongly adsorb pesticide, leading to low leaching) is that it was a very dry summer and large cracks had developed, favouring rapid macropore flow.

3.4. ANN

The ANN with the best performance was 9703T (for construction details see table 2), which had an ME value of 0.35 (table 7). This network was based on data that had undergone a rigorous cleansing of outliers. It may indicate that one of the problems with the data is that there are a few large values of loss rate and the rest are small. The network could have had difficulties separating the low values. Even after log-transformation, the distribution was still uneven and the result was not improved much (compare 9703L to 9703Q).

Table 7. How well the ANN models performed in prediction of pesticide loss rate.

	9703K	9703L	9703M	9703N	9703P	9703Q	9703R	9703S	9703T
ME	-0.01	0.06	0.17	0.30	0.22	0.11	-0.94	-0.16	0.35

For explanation on the networks, see table 2.

The results of ANN can be compared to the regression analysis for 9703S, which was trained on exactly the same data as was used in the regression for 9703B (table 5). The conclusion is clear, that the network did not function, not being able to identify the trend that was found in the regression analysis.

One of the main problems was that the network got stuck in local minima, which prevented it from learning properly. The results, on many occasions, seemed to be based on pure chance, since they differed considerably for runs using different starting weights. These initial weights are usually randomized, and the network trained properly only with certain combinations of weights.

The number of hidden nodes that were needed proved to be higher than suggested by the rule of thumb. The number of hidden nodes could be increased without causing problems since a cross-validation procedure was used to stop the training. If the error is checked against a different data set, the risk of losing the ability to generalise is small. Another factor that can have had an effect on the poor performance of the ANN is the relative small number of data patterns. Maybe there were too little data to obtain adequate training of the network. When the number of hidden nodes increases, the demand for training data also increases.

The weather parameter was chosen partly on the basis of what was possible to calculate. It may have been better to use the number of times with more rain than say ten millimetres per day for a period following the application. This was not possible however, since there are many farmers in the area applying the same substance on different occasions.

There are preliminary results from other researchers working on ANN, indicating that it may well be possible to built good networks with the kind of data that I used. In these cases, a different computer program was used as well as a different learning algorithm (the Levenberg-Marquardt method, Stenemo pers. com., 2004). This indicates that one of the main problem of the ANN used in this study was related to the learning algorithm used (back-propagation). ANN analysis is an interesting approach worth looking more at in the future since it has some advantages compared to traditional regression techniques.

4. Conclusions

Multiple linear regression proved in this study to be a better analysis tool than artificial neural networks. The results of MLR are also easier to interpret. All the following conclusions are from the MLR-analysis.

Using loss rate as response variable, it was possible to find significant functions explaining up to 37-99% of the variation in the individual years at Vemmenhög in the period 1997 to 2003. Excluding pesticides that were applied in low amounts gave better results than excluding those with zero-loss. The combination of variables in the functions with the highest rate of explanation differed for different years, but DT_{50}/k_{oc} was the most frequently occurring variable.

Grouping the years, the best significant function for the period 1997-2003 (excluding pesticides used in low amounts) contained DT_{50}/k_{oc} and log P_{ow}. It explained 70% of the variation in pesticide loss rate ($P<0.0001$). In Vemmenhög point source losses have decreased considerably since the beginning of the nineties. Pesticide properties are of great importance for the amounts leached when losses occur mainly through diffuse leaching, explaining the improved result compared to previous studies.

It proved generally not very successful to use the formulas for predicting pesticide loss for individual years, but it worked better on the grouped years. The highest model efficiency found in this case was 0.56.

5. References

5.1. Literature references

Basheer, I.A. & Hajmeer, M., 2000. Artificial neural networks: fundamentals, computing, design, and application. Journal of microbiological methods 43, 3-31.

Bergström, L. & Jarvis, N., 1992. Evaluation and comparison of pesticide leaching models for registration purposes. Background information, experimental methods, soil properties, pesticide data and model driving variables. Workshop held at Ekenäs, Stockholm, Sweden, 8th-11th of December, 1992. Division of water quality management, Department of soil science, Swedish University of Agricultural Science. Uppsala, Sweden.

Bergström, L. & Jarvis, N., 1993. Leaching of Dichlorprop, Bentazone, and ^{36}Cl in undisturbed field lysimeters of different agricultural soils. Weed science, 41, 251-261.

Bishop, C.M., 1994. Neural networks and their applications. Review of scientific instruments, 65, 1803-1832.

Chen, Y.Q., Chen, D.Z., He, C.Y. & Hu, S.X., 1999. Quantitative structure-activity relationships study of herbicides using neural networks and different statistical methods. Chemometrics and intelligent laboratory systems, 45, 267-276.

Grandin, U., 2004. Dataanalys och hypotesprövning för statistikanvändare. Institutionen för miljöanalys, Sveriges Lantbruksuniversitet. På uppdrag av svenska naturvårdsverket.

Gustafson, D.I., 1989. Groundwater ubiquity score: A simple method for assessing pesticide leachability. Environmental toxicology and chemistry, 8, 339-357.

Kreuger, J. 1996. Pesticides in stream water within a small catchment in southern Sweden, Report from the 'Vemmenhög-project' 1990-1994. Swedish University of Agricultural Science, Division of water quality management, Uppsala. Technical report 32.

Kreuger, J. 1997., Report from the 'Vemmenhög-project' 1995-1996, Pesticide concentrations and transport in water from a small agricultural catchment in southern Sweden. Swedish University of Agricultural Science, Division of water quality management, Uppsala. Technical report 40.

Kreuger, J., 1998. Pesticides in stream water within an agricultural catchment in southern Sweden, 1990-1996. The Science of the total environment, 216, 227-251.

Kreuger, J., 2000. Övervakning av bekämpningsmedel i vatten från ett avrinningsområde i Skåne, Årsredovisning för Vemmenhögsprojektet 1998 samt en kortfattad långtidsöversikt. Swedish University of Agricultural Science, Division of water quality management, Uppsala. Ekohydrologi 54.

Kreuger, J., 2002a. Övervakning av bekämpningsmedel i vatten från ett avrinningsområde i Skåne, Rapport för Vemmenhögsprojektet 2000. Swedish University of Agricultural Science, Division of water quality management, Uppsala. Technical report 65.

Kreuger, J., 2002b. Övervakning av bekämpningsmedel i vatten från ett avrinningsområde i Skåne, Årsredovisning för Vemmenhögsprojektet 2001. Swedish University of Agricultural Science, Division of water quality management, Uppsala. Ekohydrologi 69.

Kreuger, J., 2003. Vinterprovtagning av bekämpningsmedel I Vemmenhögsån 2001/2002, Redovisning av specialprojekt inom miljöövervakningen. Swedish University of Agricultural Science, Division of water quality management, Uppsala. Technical report 69.

Kreuger, J. Övervakning av bekämpningsmedel i vatten från ett avrinningsområde i Skåne, Årsredovisning från Vemmenhögsprojektet 1999. Manuskript 2002.

Kreuger, J. & Hessel, K., 1998. Report from the 'Vemmenhög-project' 1997, Pesticide concentrations and transport in water from a small agricultural catchment in southern Sweden. Swedish University of Agricultural Science, Division of water quality management, Uppsala. Technical report 44.

Kreuger, J. & Törnqvist, L., 1998. Multiple regression analysis of pesticide occurrence in streamflow related to pesticide properties and quantities applied. Chemosphere, 37, 189-207.

Kreuger, J., Holmberg, H., Kylin, H. & Ulén, B., 2003. Bekämpningsmedel i vatten från typområden, åar och i nederbörd under 2002, Årsrapport till det nationella programmet för miljöövervakning av jordbruksmark, delprogram pesticider. Swedish University of Agricultural Science, Division of water quality management, Uppsala. Ekohydrologi 77.

Kreuger, J., Törnquist, M. & Kylin, H., 2004. Bekämpningsmedel i vatten och sediment från typområden och åar samt i nederbörd under 2003. Swedish University of Agricultural Science, Division of water quality management, Uppsala. Ekohydrologi 81.

Linders, J.B.H.J., Jansma, J.W., Mensink, B.J.W.G. & Otermann, K., 1994. Pesticides: Benefaction or Pandora's box? A synopsis of the environmental aspects of 234 pesticides. RIVM-report 679101014.

Livingstone, D.J., Manallack, D.T. & Tetko, I.V., 1997. Data modelling with neural networks: advantages and limitations. Journal of computer-aided molecular design, 11, 135-142.

Loague, K. & Green, R.E., 1991. Statistical and graphical methods for evaluating solute transport models: overview and application. Journal of contaminant hydrology, 7, 51-73.

Masters, T., 1993. Practical neural network Recipes in C++. Academic Press, London.

Quinn, G. & Keough, M., 2002. Experimental design and data analysis for biologists. Cambridge University Press, Cambridge.

Rummelhart, D.E., Hilton, G.E. & Williams, R.J., 1986. Learning internal representations by error propagation. Parallel distributed processing, 1, chapter 8, Rummelhart, D.E. & McCleland, J.L. (Eds.). Cambridge, MA:MIT Press.

Salt, D.W., Yildiz, N., Livingstone, D.J. & Tinsley, C.J., 1992. The use of artificial neural networks in QSAR. Pesticide science, 36, 161-170.

Sorensen, P.B., Bruggemann, R., Carlsen, L., Morgensen, B.B., Kreuger, J. & Pudenz, S., 2003. Analysis of monitoring data of pesticide residues in surface waters using partial order ranking theory. Environmental toxicology and chemistry, 22, 661-670.

Starett, S.K., Najjar, Y., Adams, G. & Hill, J., 1998. Modelling pesticide leaching from golf courses using artificial neural networks. Communications in soil science and plant analysis, 29, 3093-3106.

Tomlin, C.D.S. (Ed), 2003. The pesticide manual. 13th edition. British Crop Protection Council, Hamshire.

Worrall, F. & Kolpin, D.W., 2004. Aquifer vulnerability to pesticide pollution-combining soil, land-use and aquifer properties with molecular descriptors. Journal of Hydrology, 293, 191-204.

5.2. Internet references

Agricultural Research Service, The ARS Pesticide Properties Database, September 2004. www.arsusda.gov/acsl/services/ppdb/

PDP++ Software Home Page. September 2004. www.cnbc.cmu.edu/Resources/PDP++/PDP++.html

Swedish environmental protection agency (SEPA), December 2004. www.naturvardsverket.se

5.3. Personal contacts

Asp, Jeanette, 2004. Review of EU-monographs for the Department of Soil Science, Division of environmental physics, Swedish University of Agriculture. Uppsala

Jarvis, Nicholas, 2004. Department of soil science, Division of environmental physics, Swedish University of Agriculture. Uppsala

KemI. Swedish chemical inspection, extract from pesticide database, via Christina Snöbohm Rudin. 20040915.

Kreuger, Jenny, 2004. Department of Soil Science, Division of water quality management, Swedish University of Agriculture. Uppsala

Stenemo, Fredrik, 2004. Department of Soil Science, Division of environmental physics, Swedish University of Agriculture. Uppsala

6. Appendix

Appendix 1. Table containing the soil sorption coefficient, k_{oc}, (l/kg) from different sources.

Substanser	k_{oc}	k_{oc}	mean	k_{oc}	k_{oc}	k_{oc}	chosen
Aclonifen				5992.48	5318-1264		5992.48
Alpha-cypermethrin			73				73
Amidosulfuron							
Azoxystrobin		423			Approx. 500		423
Bentazone	115		33.7	0.688		34	33.7
Bromoxinil						109-1079	594
Chloridazon	129			110.08			129
Clopyralid	5	36				36	36
Cyanazine	94			94.6			94
Cyfluthrin	33800			58050		64125	33800
Cypermethrin					26492-144652	84000	84786
Cyprodinil		2972					2972
Deltamethrin				818.72	4.6E5-1.63E7		8380000
Dichlorprop-P	26		22				22
Diflufenican				1978			1978
Esfenvalerate	395				5300	5273	395
Ethofumesate	166	144	147	144.48			147
Fenoxaprop-P				14.964			15
Fenpropidin				1978			1978
Fenpropimorph	4382	3962		3569			3962
Fluroxypyr	68	84	68	60.2			68
Glyphosate		884		11236.76			884
Ioxinil	170						170
Iprodione				483.32	373-1551		483.32
Isoproturon	123	122	122				122
Karbendazim					200-250		225
Lambda-cyhalothrin				309600	330000		319800
MCPA	31	62	74	49.88			74
Mecoprop-P	17		28.7				28.7
Metalaxyl			55.5	46.44			55.5
Metamitron	76	141	110	172			110
Metazachlor	80	139		139.32			139
Methabenzthiazuron	689			696.6			686
Metribuzin					3.14		3.14
Metsulfuron-methyl			39.5	48.16			39.5
Pendimethalin	18000			190.92		12000	18000
Phenmedipham		625		584.8	2400		625
Primicarb	1387			792.92			1387
Prochloraz	3331						3331
Propiconazole	1428	1229		1233.24			1229
Sulfosulfuron			33.2				33.2
Terbutryn				670.8	2000	2863	1844
Terbuthylazine					162-278		220
Thifensulfuron-methyl			28	32.68			28
Triadimenol	355			230.48			355
Tribenuron-methyl		43	30.9				30.9
Triflusulfuron-methyl						32	32
Referenser	**1**	**2**	**3**	**4**	**6**	**7**	

1 Kreuger & Törnqvist, 1998
2 Jarvis pers com, summary file
3 Jarvis pers com, EU monographies
4 Linders et al., 1994
5 KemI, 2004
6 Tomlin, 2003
7 ARS Pesticide properties database, 2004

Appendix 2. The octanol-water partition coefficient, log P_{ow}, from different sources. The values for pH 7 were chosen when several alternatives were available.

Substances	log P_{ow}	log P_{ow}	log P_{ow}	log P_{ow}	log P_{ow}	log P_{ow}	chosen
Aclonifen			4.17	4.37	4.37		4.37
Alpha-cypermethrin		5.5		5.5	6.94		5.5
Amidosulfuron				1.63	1.63		1.63
Azoxystrobin				2.5	2.5		2.5
Bentazone	-0.5	-0.46		-0.46			-0.46
Bromoxinil				2.8	5.06		3.93
Chloridazon	2.2			1.19			1.19
Clopyralid	-0.2			-2.6	-2.63		-2.6
Cyanazine	2.2			2.1			2.1
Cyfluthrin	6.2		0.002			5.95	6.1
Cypermethrin			6.32	6.6	6.6		6.6
Cyprodinil				3.9	4		3.9
Deltamethrin			5.4	4.6	4.6		4.6
Dichlorprop-P	2	1.77		-0.25	-0.25		-0.25
Diflufenican			4.9	4.9	4.9		4.9
Esfenvalerate	4.1			6.24	6.22		6.24
Ethofumesate	2.7	2.7	2.7	2.7			2.7
Fenoxaprop-P				4.28	4.58		4.28
Fenpropidin			2.59	2.59	2.9		2.59
Fenpropimorph	4.1		2.48	4.1			4.1
Fluroxypyr	3.8	1.7	1.74	1.7			1.7
Glyphosate		-3.2		-3.2	<-3.2		-3.2
Ioxinil	3.4						3.4
Iprodione				3.3	3		3.3
Isoproturon	2.5	2.5		2.5			2.5
Karbendazim				1.51	1.5		1.5
Lambda-cyhalothrin			7	7	7		7
MCPA	2.1	-0.81		-0.81			-0.81
Mecoprop-P	0.1	0.02	1.6 – 2.2	0.02			0.02
Metalaxyl		1.71	1.65	1.75			1.75
Metamitron	0.8	0.83	0.83	0.83			0.83
Metazachlor	2.1		2.13	2.13			2.13
Methabenzthiazuron	2.6			2.64			2.64
Metribuzin				1.6	1.6		1.6
Metsulfuron-methyl		-1.7	-1.7	-1.7			-1.7
Pendimethalin	5.2			5.2			5.2
Phenmedipham				3.6	3.59		3.6
Primicarb	1.7			1.7			1.7
Prochloraz	4.1						4.1
Propiconazole	3.7		3.65	3.72			3.72
Sulfosulfuron		-0.77		-0.77	<1		-0.77
Terbutryn			3.5		3.65		3.6
Terbuthylazine				3.2	3.21		3.2
Thifensulfuron-methyl		0.021	-1.7	1.7	0.02		1.7
Triadimenol	3.1					3.08	3.1
Tribenuron-methyl		0.78		0.78			0.78
Triflusulfuron-methyl				3	0.96		3
Referenser	**1**	**2**	**3**	**4**	**5**	**6**	

1 Kreuger & Törnqvist, 1998

2 Jarvis pers com, EU monographies

31

3 Linders et al., 1994
4 KemI, 2004
5 Tomlin, 2003
6 ARS Pesticide properties database, 2004

Appendix 3. The solubility in water, S_w (mg/l), at pH 7 where several alternatives were available.

Substances	S_w	S_w	S_w	S_w	S_w	S_w	Chosen
Aclonifen			1.5-2.5	1.4	1.4		1.4
Alpha-cypermethrin				0.00397	3.97 ug/l		0.00397
Amidosulfuron				3.3			3.3
Azoxystrobin				6.7	6		6.7
Bentazone	500	570	500	570			570
Bromoxinil					89	27	53.5
Chloridazon	400		400	340	0.005		340
Clopyralid	8400			143000	7850*	9000	8700
Cyanazine	171		171	171			171
Cyfluthrin	0.002		5.62			0.002	0.002
Cypermethrin			0.004	0.004	0.004		0.004
Cyprodinil				16	13		16
Deltamethrin			<0.002	0.0002	<0.2 ug/l		0.0002
Dichlorprop-P	590	*0.59*	350	590	590	350	590
Diflufenican			<0.05	<0.05	<0.05		0.05
Esfenvalerate	1			0.002	0.002		0.002
Ethofumesate	50	50	110	50			50
Fenoxaprop-P				0.7	0.7		0.7
Fenpropidin			350	707	530 g/m3		707
Fenpropimorph	4.3		6.8	4.32			4.32
Fluroxypyr	91	6500	91	5700			5700
Glyphosate		995		10500	10.5 g/l		10500
Ioxinil	1.8						1.8
Iprodione				12.2	13		12.2
Isoproturon	65	70.2		70.2			70.2
Karbendazim					8	8	8
Lambda-cyhalothrin			6	0.005			0.005
MCPA	825	293.9	1500	294			294
Mecoprop-P	860	250	620	860			860
Metalaxyl		26	7100	8.4			8.4
Metamitron	1700	1700	1820	1700			1700
Metazachlor	430		30	4.3			4.3
Methabenzthiazuron	60		60	59			59
Metribuzin				1050	1.05 g/l		1050
Metsulfuron-methyl		2790	109-9500	2790			2790
Pendimethalin	0.3		0.3	0.33			0.33
Phenmedipham				6	4.7		6
Primicarb	3060		2700	3060			3060
Prochloraz	34						34
Propiconazole	110		110	125			125
Sulfosulfuron		1626.8		1627	1627		1627
Terbutryn			22-25		22		23
Terbuthylazine				8.5	8.5		8.5
Thifensulfuron-methyl		2.24		2240	2240		2240
Triadimenol	62					62	62
Tribenuron-methyl		2.04		2040			2040
Triflusulfuron-methyl				110	260		110
Referenser	**1**	**2**	**3**	**4**	**5**	**6**	

1 Kreuger & Törnqvist, 1998   * in dest.water

2 Jarvis pers com, EU monographies

3 Linders et al., 1994

4 KemI, 2004
5 Tomlin, 2003
6 ARS Pesticide properties database, 2004

Appendix 4. The degradation half-life, DT_{50}, given in days, from different sources. pH 7 was chosen in pesticide manual.

Substances	DT_{50}	DT_{50}	mean	Mean	DT_{50}	DT_{50}	chosen
Aclonifen				71	68-80		71
Alpha-cypermethrin			121		21		121
Amidosulfuron					3-29		13
Azoxystrobin		39			14		39
Bentazone	44		21	48			21
Bromoxinil					<1	0.08	0.08
Chloridazon	115			31	42-56	18	24.5
Clopyralid	20	26					26
Cyanazine	16			16			16
Cyfluthrin	99			116			99
Cypermethrin					60	63	61.5
Cyprodinil		31			20-60		31
Deltamethrin				25	<2-3		12
Dichlorprop-P	20		16	15			16
Diflufenican				192	15-30weeks		235
Esfenvalerate	63				287		63
Ethofumesate	45	37	80	37			37
Fenoxaprop-P					1--10		5
Fenpropidin				111	58-95		94
Fenpropimorph	139	93		67			93
Fluroxypyr	52	24	29	27			29
Glyphosate		16			1-130		16
Ioxinil	8						8
Iprodione				41	20-160		41
Isoproturon	28	13	13				13
Karbendazim					8-32		20
Lambda-cyhalothrin				41	6-40		32**
MCPA	22	22	29	15			22
Mecoprop-P	12		6	11			6*
Metalaxyl			40	42			40
Metamitron	35	23	15.5	30			15.5
Metazachlor	72	18		18			18
Methabenzthiazuron	135			135			135
Metribuzin					30-60		45
Metsulfuron-methyl			24	31			24
Pendimethalin	123			171			123
Phenmedipham		20			ca 25		20*
Primicarb	69			108			69
Prochloraz	132						132
Propiconazole	59	75		96			75
Sulfosulfuron			106		ca 32		106
Terbutryn				74	14-50		53
Terbuthylazine					30-60		45
Thifensulfuron-methyl			4	6			4
Triadimenol	255						255
Tribenuron-methyl		12	14				14
Triflusulfuron-methyl					3	5.5-14.4	3
Referenser	**1**	**2**	**3**	**4**	**5**	**6**	

1 Kreuger & Törnqvist, 1998   * Jarvis pers com
2 Jarvis pers com, summary file   ** Kreuger pers com

3 Jarvis pers com, EU monographies
4 Linders et al., 1994
5 Tomlin, 2003
6 ARS Pesticide properties database, 2004

Appendix 5. The substances used in this analysis and the newest available from EU monographies.

Substances	DT_{50}	DT_{50}	k_{oc}	k_{oc}	log P_{ow}	log P_{ow}	S_w	S_w
Aclonifen	71		5992		4.4		1	
Alpha-cypermethrin	121	121	73	76344	5.5	5.5	0.00397	0.00297
Azoxystrobin	39	78	423	423	2.5	2.5	7	7
Bentazone	21	21	34	33.7	-0.5	-0.5	570	570
Chloridazon	25		129		1.2		340	
Clopyralid	26	30.6	36	5	-2.6	-2.6	8700	143000
Cyanazine	16		94		2.1		171	
Cyfluthrin	99		33800		6.1		0.002	
Cypermethrin	62	60	84786	763444	6.6	4	0.004	0.1
Cyprodinil	31		2972		3.9		16	
Deltamethrin	12	32	8380000	10270000	4.6	4.6	0.0002	0.0002
Dichlorprop-P	16	16	22	22	-0.3	1.8	590	0.59
Esfenvalerate	63	59	395	630000	6.2	6.2	0.002	0.001
Ethofumesate	37	80	147	147	2.7	2.7	50	50
Fenoxaprop-P	5	1	15	8800	4.3	3.8	0.7	0.9
Fenpropimorph	93		3962		4.1		4	
Fluroxypyr	29	29	68	68	1.7	1.7	5700	6500
Iprodione	41	32	483.32	381	3.3	3	12	12
Lambda-cyhalothrin	32	56	319800	157000	7	7	0.005	0.005
MCPA	22	29	74	74	-0.8	-0.8	294	294
Mecoprop-P	6	6	29	28.7	0.02	0.02	860	250
Metamitron	15.5	15.5	110	110	0.8	0.8	1700	1700
Methabenzthiazuron	135		686		2.6		59	
Phenmedipham	20	26	625	888	3.6	3.6	6	2
Primicarb	69	67	1387	1387	1.7	1.7	3060	3100
Propiconazole	75	85	1229	648	3.7	3.7	125	150
Sulfosulfuron	106	103	33.2	33.2	-0.8	-0.8	1627	1627
Terbutryn	53		1844		3.6		23	
Terbuthylazine	45		220		3.2		9	
Thifensulfuron-methyl	4	4	28	28	1.7	0.02	2240	2
Tribenuron-methyl	14	11	31	13	0.8	0.8	2040	2040
Triflusulfuron-methyl	3		32		3		110	
Reference	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**

1 Values used in the analysis

2 The newest available values from EU (Asp pers.com.)

List of publications in Emergo

2003:1   Holmberg, H. Metodutveckling för utvärdering av simuleringsmodeller med hjälp av fluorescerande ämnen (Development of methods to evaluate simulation models using fluorescent dye tracers). M.Sc. thesis. 50 pages.

2003:2   Olsson, C. Översvämningsåtgärder i Emån – simulering i Mike 11 modellen (Measures for flood control in the river Emån – simulation in the Mike 11 model). M.Sc. thesis. 34 pages.

2003:3   Gärdenäs, A. Eckersten, H. & Lillemägi, M. Modeling long-term effects of N fertilization and N deposition on the N balance of forest stands in Sweden. 30 pages.

2003:4   Jarvis, N. Hanze, K. Larsbo, M. Stenemo, F. Persson, L. Roulier, S. Alavi, G. Gärdenäs, A. & Rönngren, J. Scenario development and parameterization for pesticide exposure assessments for Swedish groundwater. 26 pages. ISBN: 91-576-6588-5

2003:5   Eckersten, H., Gärdenäs, A. & Lewan, E. (Eds.) Biogeofysik – en introduktion (Environmental physics – an introduction). 141 pages. ISBN: 91-576-6591-5

2003:6   Larsbo, M & Jarvis, N. MACRO5.0. A model of water flow and solute transport in macroporous soil. Technical description. 47 pages. ISBN: 91-576-6592-3

2003:7   Nylund, E. Cadmium uptake in willow (*Salix viminalis* L.) and spring wheat (*Triticum aestivum* L.) in relation to plant growth and Cd concentration in soil solution. M.Sc. thesis. 33 pages.

2003:8   Strömqvist, J. Leaching of fungicides from golf greens: Simulation and risk assessment. M.Sc. thesis. 41 pages.

2003:9   Blombäck, K, Strandberg, M. & Lundström, L. Det organiska materialets betydelse för markens biologiska aktivitet och gräsets etablering och tillväxt i en golfgreen (The influence of soil organic matter on soil microbial activity and grass establishment and growth in a putting green).

2003:10   Stenemo, F & Jarvis, N. Users guide to MACRO5.0, a model of water flow and solute transport in macroporous soil. ISBN: 91-576-6610-5.

2004:1   Årevall, H. Mark- och lokaliseringsaspekter i miljökonsekvensbeskrivning – fem fallstudier av vindkraftsprojekt (Aspects of land and localization in environmental impact assessment – five case studies of wind power projects). M.Sc. thesis. 62 pages.

2004:2   Almqvist, S. Simulering av bekämpningsmedelstransport i banvall – utveckling av ett prognosverktyg (Simulation of pesticide transport in railway embankments – development of a predictive tool). M.Sc. thesis. 47 pages.

2004:3   Eckersten, H., Gärdenäs, A. & Lewan, E. (Eds.) Biogeofysik – en introduktion (Environmental physics – an introduction). 143 pages. ISBN: 91-576-6591-5

2005:1   Ortiz, C.Calibration of GenRiver with GLUE for Northern Vietnamese conditions. M.Sc. thesis 27 pp.

2005:2   Adielsson, S. Statistical and neural network analysis of pesticide losses to surface water in small agricultural catchments in Sweden. M.Sc. thesis. 28 pages.