Mapping nitrate leaching to upper groundwater in the sandy regions of The Netherlands, using conceptual knowledge

Leo Boumans · Dico Fraters · Gerard van Drecht

Received: 6 October 2006 / Accepted: 6 April 2007 / Published online: 23 August 2007
© Springer Science + Business Media B.V. 2007

Abstract The European Community asks its Member States to provide a comprehensive and coherent overview of their groundwater chemical status. It is stated that simple conceptual models are necessary to allow assessments of the risks of failing to meet quality objectives. In The Netherlands two monitoring networks (one for agriculture and one for nature) are operational, providing results which can be used for an overview. Two regression models, based upon simple conceptual models, link measured nitrate concentrations to data from remote sensing images of land use, national forest inventory, national cattle inventory, fertiliser use statistics, atmospheric N deposition, soil maps and weather monitoring. The models are used to draw a nitrate leaching map and to estimate the size of the area exceeding the EU limit value in the early 1990s. The 95% confidence interval for the fraction nature and agricultural areas where the EU limit value for nitrate (50 mg/l) was exceeded amounted to 0.77–0.85 while the lower 97.5% confidence limit for the fraction agricultural area where the EU limit value was exceeded amounted to 0.94. Although the two conceptual models can be regarded as simple, the use of the models to give an overview was experienced as complex.

Keywords Water Framework Directive · Nitrate Directive · Monitoring · Field survey · Regression analysis · GIS

Introduction

The European Commission has stipulated a limit value for nitrate concentration in groundwater of 50 mg/l (Nitrates Directive, EC 1991). Groundwater is defined as all water in the water saturated zone. Member States are obliged to monitor and to control nitrate concentrations in water by means of legislation. In the “Draft guidelines for the monitoring required under the Nitrates Directive” (EC 2003) it is stated:

Member States should choose their groundwater sampling points so as to get a representative picture of nitrate concentrations in their groundwaters. The selection of sampling points will depend on land use and hydro-geological conditions. Both shallow and deep groundwater should be included in the monitoring network. However, shallow groundwater is more susceptible for influences at soil surface than deeper groundwaters.

Shallow groundwater is defined to be in the first 5 m of the saturated zone. The European Commission also asks its members to design a monitoring network to provide a coherent and comprehensive overview

L. Boumans (✉) · D. Fraters · G. van Drecht
National Institute of Public Health
and the Environment (RIVM),
1, 3720 BA Bilthoven, The Netherlands
e-mail: LJM.Boumans@rivm.nl

of groundwater chemical status within each river basin (Water Framework Directive, EC 2000). In Guidance Document No. 7, “Monitoring under the Water Framework Directive,” p. 105 (EC 2003) it is stated that conceptual models are necessary to allow assessments of the risks of failing to meet the Directive’s environmental objectives to be made. The term conceptual model is used as shorthand for the understanding, or working description, of the real hydro-geological system that is needed to design effective groundwater monitoring programmes. It is stressed that the term implies that a mathematical model is required for all bodies of groundwater.

The Dutch government has decided to monitor the quality of the uppermost metre of the saturated zone for evaluating the effects of legislation and for evaluating the exceedings of the limit value. This uppermost metre is most susceptible to influences and furthermore, sampling attenuates seasonal variations in groundwater quality (Boumans et al. 2005). There are two monitoring programmes for the uppermost metre of groundwater in The Netherlands, one for agriculture and one for nature. We focus on the sandy regions of The Netherlands (see Fig. 1). These regions deserve special attention because:

- Nitrogen surpluses are greater than in other European countries, exceeding 200 kg/(ha.a) (de Walle and Sevenster 1998);
- Highly intensive animal farming occurs more frequently here than in other parts of The Netherlands, causing higher nitrogen surpluses (Fraters et al. 1998) and nitrate concentrations are higher in these regions (Fraters et al. 2004);
- Sandy soils are more vulnerable to nitrate leaching than clay and peat soils (van Drecht 1993);
- Groundwater in the sandy region is used for drinking water, while in other regions surface waters are mainly used.

Two conceptual regression equations arising from the monitoring programmes for agriculture and for nature (Boumans et al. 2004, 2005), will be used for drawing a map and the estimation of the total area exceeding the limit value. Publicly available data on land use, forest inventory, fertiliser use, atmospheric N deposition, soil conditions and climate are used to interpolate to non-sampled areas. Well known statis-

Fig. 1 Exceedance of the EU limit value for nitrate in groundwater by nature and agriculture in the sandy areas of The Netherlands in the early 1990s

tical methods for mapping, kriging and semi-parametric regression, are not used because the two regression models available exhibit just a small spatial trend and it is very easy to ‘over fit’ features of the data using semi-parametric regression or kriging (Venables and Ripley 1994, p. 266; de Kwaadsteniet 1990). For example, Pebesma and de Kwaadsteniet (1997) map groundwater quality using block kriging, whereas Tiktak et al. (1999) use a generalised additive regression model with a locally weighted smoother. With kriging the model for the expectation value is maintained and small-scale variation is modelled by modelling the residual part. Semi-parametric regression methods skip the rigid form for the expectation value and use a flexible approach to estimate the expectation value. The decision between modelling an expectation and modelling residuals can be arbitrary. If the expectation value is modelled, as was done by the regression models for nature and agriculture, the result is an estimated median for units with the same values for the independent variables. If the residuals are modelled too, the result is an estimated median value for each unit. As more spatial information, based upon conceptual knowledge, is incorporated into the model, spatial autocorrelation of model residuals will

decrease and accordingly the necessity of the use of the two statistical mapping methods as well.

Our regression models for nature and agriculture are based upon data from national databases and therefore knowledge about nitrogen inputs is less precise. Andersen et al. (1999) used an empirical model to estimate nitrate leaching from the root zone for about 1,200 individual fields in six catchments. The empirical model of Anderson is based upon a data set derived from a large number of controlled field and lysimeter experiments where nitrogen inputs are well known. As a consequence, the model of Anderson should relate nitrate leaching better to nitrogen input. However the estimations of Anderson for the individual fields may show a bias as experimental circumstances will be different from agriculture in practice.

An advantage of deterministic to statistical methods is that physico-chemical knowledge can be better incorporated in the model, making extrapolating estimations more reliable. Wolf et al. (2005) describe a mechanistic deterministic model, STONE, which is meant to be applicable for The Netherlands at national and regional scales. Nitrate leaching is only one characteristic of STONE. A disadvantage is that uncertainty of input data for deterministic models can be large and uncertainty for the estimations cannot be estimated. Wolf et al. (2005), conclude that testing of a large-scale model, like STONE, on measured data from field experiments can hardly be expected to be satisfactory. Furthermore calibration of a large-scale model on well-managed experiments may be wrong for practical applications. According to van der Molen and Boers (2002) uncertainties in input parameters probably overrule uncertainties in the model structure and parameters, and may limit the need for the complexity of models.

Conceptual regression models for nitrate in groundwater wells using GIS variables, which are comparable with the variables of our regression models (nitrogen loading from atmospheric deposition, animal manure, commercial fertiliser, soil use, climate and soil drainage characteristics), are made by Nolan (2001) and Burkart et al. (1999). However their regression models are not used for mapping. In an earlier paper Nolan et al. (1997) produced a map for the risk of nitrate occurrence in groundwater wells less than 30 m (100 ft) deep in the United States. They did not use regression models as they considered their data set

was not sufficiently uniform and consistent. In our case the construction and location of the groundwater wells was specially adapted for monitoring purposes and regression models could be derived.

Our goal is to use statistical models, containing as much conceptual knowledge as possible, for mapping and estimating. This allows an optimum combination of the advantages of statistical and deterministic models.

Material and methods

Mapping

For mapping nitrate concentrations in the uppermost metre of groundwater in nature and agricultural areas, two regression models are available that relate groundwater nitrate concentrations to GIS data. A regression model for natural areas was developed with data gathered during 1989–1990 (Boumans et al. 2004). The regression model for agricultural areas was developed with data gathered during 1992–1995 (Boumans et al. 2005).

Nitrate limit value exceedance was estimated with the two regression models for grid units of 500 m×500 m. Statistics Netherlands (CBS) is responsible for collecting, processing and publishing statistics to be used in practice, by policymakers and for scientific research. Statistics Netherlands divides The Netherlands in grid units of 500 m×500 m and gives a census of the area soil use in each grid unit (CBS 1987). The area of interest consists of 65,013 units (representing 950,000 ha) containing nature and or agricultural area. There are 26,467 units for nature areas, (representing 287,000 ha) and 152 of these were sampled. Agriculture was not sampled in the grid units but at 99 farms participating in the monitoring programme for agriculture (Fraters et al. 1998). The farm area was digitised and overlays were made with remote sensing images for crop types and the soil map. N inputs were available per soil type-crop combination for each municipality and an N input was attributed to each farm (Boumans et al. 2004). In the same way the N input data were also superimposed on the agricultural land of each grid unit. To estimate nitrate concentrations for the grid units, the measured nitrate concentrations of the farms are treated as measured nitrate concentrations for grid units with the same fractions soil and crop types.

That means that we interpret our sample of 99 farms as a sample of 99 grid units.

For each unit j the 97.5% lower confidence limit for the median for nature, $LC_{j,nat}$ and for agriculture, $LC_{j,agr}$ was determined. A mean lower confidence limit for unit j LC_j was calculated weighted for the extent of area's nature and agriculture of the unit. Next we calculated the ratio Q_j of the nitrate limit value to the mean lower confidence limit $Q_j = LC_j / C_{EU}$ and assigned each unit to one of five classes 0, 1×, 2×, 3×, or 4×, according to whether $Q_j < 1$, $1 < Q_j < 2$, $2 < Q_j < 4$ or $Q_j > 4$, respectively.

Estimating the extent of the area exceeding the EU limit value

The interval estimate for the total extent of areas of units exceeding the limit value A is derived by simulating the distribution of $\hat{A}$,

$$\hat{A} = \sum_{j=1}^m \sum_{u=1}^2 \hat{p}_{ju} a_{ju}$$

A = total nature and agricultural area of units exceeding the limit value

p_{ju} = probability for land use, u , in unit j , to exceed the limit value

a_{ju} = area land use, u , in unit, j (CBS 1987)

for p_j is substituted:

$$p_j = 1 - \Phi \left[\left(C_{EU} - \mu_j \right) / \sigma \right],$$

$\mu_j = \beta X_j$ is the expected nitrate concentration value for that unit, with β the vector of regression coefficients, and X_j the values of the independent variables.

σ^2 is the variance of the residuals which is assumed to be Gaussian

$\Phi[]$ is the standard Gaussian distribution.

C_{EU} is the nitrate limit value.

$\hat{p}_j$ is derived by substituting for μ_j , $\hat{\mu}_j = \hat{\beta} X_j$, with the estimated β coefficients and $\hat{\sigma}$ for σ

Realisations for β are derived by sampling the joint probability distribution for the regression coefficients of the regression models and trend surface models,

using their variance-covariance matrix (Haining 1990; p.116–117). A realisation for σ is achieved by sampling from $\hat{\sigma}^2 \chi^2$.

The distribution of $\hat{A}$ was simulated by 1000 realisations of A . The 97.5% lower confidence limit for the extent of areas of units exceeding the limit value is calculated as the 2.5 percentile of the realisations of A .

Results

The two regression models are used to construct a map providing a coherent overview of limit value exceedance in the sandy region of The Netherlands during the early 1990s, see Fig. 1. For units in class 1X the weighted median nitrate concentration exceeds the EU limit with at least 97.5% confidence, which may be interpreted to mean that for all these units the probability of exceedance of the EU limit is at least 50%. For units in classes 2× and 3× similar statements can be made concerning exceedance of respectively two, three and four times the EU limit value. Exceedance can be found in the south-east of The Netherlands, known for its intensive animal farming. Areas without exceedance consist mainly of nature areas with little agricultural land use within 1 km. The estimated total unit area nature and agriculture exceeding the limit value is 77–85%. The estimated total area agriculture is 94–99% and the estimated total area nature is 21–30%.

Discussion

Mapping method

Guidance Document 7, “Monitoring under the Water Framework Directive” (EC 2003) suggests simple conceptual models for assessing the chemical status of groundwater in a river basin. The (simple) regression models for nitrate in the uppermost groundwater of agricultural and nature areas are based on simple conceptual models. The construction of the independent variables and the use of the regression models can be considered complex, as complex software was used to calculate the precipitation excess, and complex GIS manipulations and trial and error to minimise residual variance. Our method to estimate the fractions

limit value exceedance using the models, was not found in textbooks or in the literature and can therefore be considered complex as well. However developing such a method was not our goal.

The WFD requires its members to give the chemical status for each groundwater body. Currently the shapes of these bodies are not exactly known. The same technique presented above to estimate limit value exceedance for the sandy regions of The Netherlands can also be used for the uppermost groundwater of separate regions being groundwater bodies. Yet this will require investigating whether or not the residuals of the two regression models are different for different regions.

Although we incorporated as much conceptual knowledge as possible, we did not use a deterministic model. Our method uses the same GIS databases from which the models are derived, to estimate groundwater nitrate for non-sampled units. If the GIS data are more unreliable, this will result in a larger residual variance and more uncertain estimates. However this uncertainty can be quantified.

Estimating a map for limit value exceedance

For organisational reasons (see “[Material and methods](#)”), nature and agricultural areas were sampled at different times and in two different types of unit. Moreover for both areas different variables are supposed to be of influence on the groundwater nitrate concentration. Therefore two separate regression models have to be used for mapping. For each grid unit we defined the nitrate concentration to be the area-weighted average of the concentrations for nature and agriculture within the unit. Geographic information for farm areas is not available but the area of a sampled farm is in the same order of magnitude as the area of a grid unit and we assumed a farm area to be representative for agriculture in a grid unit with the same geographic information. Using separate models for nature and agriculture implies that the predicted unit concentration is the area-weighted average of the predicted values for nature and agriculture. For both models, nature and agriculture, the residual variance is rather large, only 35 and 50% of the variance is explained respectively. So model based predictions for the unit concentration will show large prediction intervals, making a map of these predictions rather non-informative. The standard deviation of the nature and agriculture estimator for the

expected value is related to the number of observed responses and is in our case small compared to the standard deviation of the distribution of the nature and agricultural model residuals. Moreover for a symmetrical distribution, the expected value equals the median. This means that the mean concentration will be greater than the expectation (median of unit values) for 50% of the units. Therefore we decided to classify all units according to the estimated probability of EU limit value exceedance, instead of predicting the concentrations for each unit. This can be done with greater precision. For this classification we relied upon the estimation of the expected value for the concentration per unit and the symmetry of the distribution of residuals in our models. The response of the regression model for nature is a transformed variable, to make this distribution symmetrical. Back transformation results in a median value. To indicate which units have an evident probability to exceed the limit value, an area-weighted average of the two 97.5% lower confidence limits was calculated. This average is a lower bound for the median of the weighted medians.

Estimating the extent of the area with limit value exceedance

The problem of local variability was solved in the former paragraph by estimating local probabilities. A map with probabilities can be difficult to interpret by policymakers. Nolan et al. (1997) also recognise the problem that local predictions for nitrate show high variability. Their solution is to estimate mean values for regions of the USA. Our solution is to estimate the area exceeding the limit value. The total area (of a region) exceeding the limit value is calculated as the sum of all areas of units where the concentration exceeds the limit value. Again the concentration of a unit area is calculated as the weighted average of the concentrations for the nature and the agricultural areas within the unit. If our sampling units had been selected randomly with both nature and agriculture sampled in the same units, the fraction units in the sample exceeding the limit value would be an unbiased estimate for the fraction of all units. A simple interval estimate for a fraction in case of a random sample can be made with the cumulative binomial probability distribution. This distribution gives the probability P for k occurrences or more in case of n trials and probability p per trial (Press et al. 1988). The k is

interpreted as the realised number of units with limit value exceedance and p is interpreted as the unknown fraction units with limit value exceedances. The probability p can be varied until P equals 2.5 and 97.5%. The two p values found are interpreted as the lower and upper boundary of the estimated 95% interval. The number of farms exceeding the limit value was 94 of 99 (0.96). The 95% interval estimate for this fraction, calculated in the simple way, is 0.90–0.99. For nature areas the fraction found and estimated 95% interval is 0.18 and 0.13–0.26 respectively. However, we are interested in the total area, nature and agriculture, exceeding the limit value and our sample was not completely random. Therefore we had to use a different method, leaning on the models for agriculture and nature and the availability for all units of the values of the independent variables. For agriculture the difference between the binomial interval and our interval is negligible. However for nature our interval (0.21–0.30) does not even contain the fraction found (0.18) but lies above this. As stated in Boumans et al. (2004) the sample for nature is underrepresented with respect to vulnerable situations for nitrate leaching, as situations with low groundwater tables (>5 m below soil surface level) could not be sampled. This has caused the difference between our interval and the interval calculated with the cumulative binomial probability distribution.

Nolan (2001) and references cited herein, classify measured concentrations because many observations are below the limit of detection. A logistic regression model is used to predict classes. Nitrate concentrations could have been classified into above and below the EU limit value. Instead of classifying the dependant variable we classified the predictions because there were only few measured values below the limit of detection. The parameters of a logistic regression model will depend upon the classification of the dependant variable and information is neglected.

Conclusions

Two simple statistical models, one for agricultural areas and one for nature areas, that contained as much conceptual knowledge as possible, were used to map The Netherlands' rural area in the sandy region with different nitrate EC limit value exceedance probab-

ilities and to estimate the area limit value exceedance. This allowed an optimal combination of the advantages of deterministic and statistical models.

Data on groundwater quality were gathered, modelled and mapped using simple conceptual models as suggested by Guidance Document 7, "Monitoring under the Water Framework Directive." However, the use of the simple conceptual models to map groundwater quality is a complex process.

References

- Andersen, H. E., Kronvang, B., & Larsen, S. E. (1999). Agricultural practices and diffuse nitrogen pollution in Denmark: Empirical leaching and catchment models. *Water Science and Technology*, 39, 257–264.
- Boumans, L. J. M., Fraters, D., & Van Drecht, G. (2004). Nitrate leaching by atmospheric N deposition to upper groundwater in the sandy regions of The Netherlands in 1990. *Environmental Monitoring and Assessment*, 93, 1–15.
- Boumans, L. J. M., Fraters, D., & Van Drecht, G. (2005). Nitrate leaching in agriculture to upper groundwater in the sandy regions of The Netherlands during the 1992–1995 period. *Environmental Monitoring and Assessment*, 102 (1–3), 225–241.
- Burkart, M. R., Kolpin, D. W., Jaquis, R. J., & Cole, K. J. (1999). Agrichemicals in ground water of the Midwestern USA: Relations to soil characteristics. *Journal of Environmental Quality*, 28, 1908–1915.
- CBS (1987). Soil use statistics. Voorburg, The Netherlands: Statistics Netherlands (CBS), (in Dutch).
- de Kwaadsteniet, J. W. (1990). On some fundamental weak spots of kriging technique and their consequences. *Journal of Hydrology*, 114, 277–284.
- de Walle, F. B., & Sevenster, J. (1998). *Agriculture and the environment*. Dordrecht: Kluwer.
- EC (1991). *Directive of the Council of 12 December 1991, concerning the protection of waters against pollution caused by nitrates from agricultural sources*, 91/676/EEC. Brussels: European Community.
- EC (2000). *Directive of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy*, 2000/60/EC. Brussels: European Community.
- EC (2003). *Monitoring under the Water Framework Directive. Common implementation Strategy for The Water Framework Directive*, Guidance document no. 7.
- Fraters, D., Boumans, L. J. M., van Drecht, G., de Haan, T., & de Hoop, W. (1998). Nitrogen monitoring in ground water in the sandy regions of the Netherlands. *Environmental Pollution*, 102(Suppl 1), 479–485.
- Fraters, B., Hotsma, P. H., Langenberg, V. T., Leeuwen, T. C., Van Mol, A. P. A., Olsthoorn, C. S. M., Schotten, C. G. J., & Willems, W. J. (2004). Agricultural practice and water quality in the Netherlands in the 1992–2002 period. Background information for the third EU Nitrate

- Directive Member States report. RIVM report 500003002 (In English).
- Haining, R. (1990). *Spatial data analysis in the social and environmental sciences*. UK: Cambridge University Press.
- Nolan, B. T. (2001). Relating nitrogen sources and aquifer susceptibility to nitrate in shallow ground waters of the United States. *Ground Water*, 39, 290–299.
- Nolan, B. T., Ruddy, B. C., Hitt, K. J., & Helsel, R. (1997). Risk of nitrate in groundwaters of the United States – A national perspective. *Environmental Science and Technology*, 31, 2229–2236.
- Pebesma, E. J., & de Kwaadsteniet, J. W. (1997). Mapping groundwater quality in The Netherlands. *Journal of Hydrology*, 200, 364–386.
- Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1988). *Numerical recipes*. UK: Cambridge University Press.
- Tiktak, A., Leijnse, A., & Vissenberg, H. (1999). Uncertainty in a regional-scale assessment of cadmium accumulation in The Netherlands. *Journal of Environmental Quality*, 28, 461–470.
- van der Molen, D. T., & Boers, P. C. M. (2002). Credibility and acceptability of mathematical models of environmental impacts of agriculture in The Netherlands. Agricultural effects on ground and surface waters, research at the edge of science and society. (Proceedings of a symposium held at Wageningen, October 2000.) *LAHS Publication*, 203, 303–310.
- van Drecht, G. (1993). Modeling of regional scale nitrate leaching from agricultural soils, The Netherlands. *Applied Geochemistry* (Suppl 2), 175–178.
- Venables, W. N., & Ripley, B. D. (1994). *Modern applied statistics*. Berlin Heidelberg New York: Springer.
- Wolf, J., Hack-ten Broeke, M. D., & Rötter, R. (2005). Simulation of nitrogen leaching in sandy soils in The Netherlands with the ANIMO model and the integrated modeling system STONE. *Agriculture, Ecosystems & Environment*, 105, 523–540.