Error propagation in groundwater pesticide vulnerability modelling

Paulette Posen¹, Andrew Lovett¹, Kevin Hiscock¹, Brian Reid¹, Sarah Evers² and Rob Ward²

¹ School of Environmental Sciences
University of East Anglia, Norwich NR4 7TJ, UK
Tel.: +44 1603 592547; Fax: +44 1603 591327
p.posen@uea.ac.uk

² Environment Agency
Olton Court, 10 Warwick Road, Olton, Solihull B92 7HX, UK

Abstract

Although environmental modelling is increasingly performed within a GIS framework, analysis of the associated error is far from routine, and rarely presented with the results. An important benefit of performing error analysis is its value in determining which elements of a vulnerability assessment framework need improving. With this in mind, it was decided to examine the extent to which error might propagate through a model of groundwater vulnerability to pesticide contamination. A pesticide leaching model was developed and incorporated into an assessment of groundwater contamination risk from normal agricultural use of the herbicide isoproturon, in a 30 km x 37 km area of river catchment to the north-west of London, England. The model, which comprised two main components accounting for (i) degradation and (ii) attenuation of the pesticide, was based on conventional contaminant transport calculations, combined with existing soil, rainfall, hydrogeological and depth to water table data. The results of an error analysis on the model were used to assign confidence limits to the resulting risk maps. In this instance, correlation of model variables led to a reduction of error in the final output. However, the results of the analysis showed how inclusion of low quality input data can lead to a large increase in output uncertainty. It is suggested that error propagation analysis should be routinely included in groundwater vulnerability assessment.

Keywords: error propagation, GIS, groundwater, pesticide, leaching model

1 Introduction

Concerns over the risk of contamination by diffuse agricultural pollution (notably nitrates) have prompted the development of groundwater vulnerability assessment tools since the 1980s. More recently, the widespread use of agricultural pesticide compounds has been of growing concern in the area of groundwater protection. Consequently an upper limit of 0.1 µg l⁻¹ for individual concentrations of pesticides in drinking water has been set and EU Member States have adopted various programs of measures to assess the likelihood of groundwater meeting the Environmental Objectives of the Water Framework Directive (WFD) (CEC, 2000). The WFD calls for water management to be based on the river basin, being the natural geographic and hydrological unit, hence vulnerability assessment is required at a regional scale.

The cost of continuous widespread national monitoring and sampling of groundwater bodies is often prohibitive, so it is necessary to use models to delineate areas of greatest vulnerability in order to target resources in the most efficient manner. Against this background, a number of
different parametric and mathematical models have been incorporated into groundwater vulnerability systems with the aim of assessing risk at the regional scale (Aller et al., 1985; Herbst et al., 2005). Many such systems combine contaminant transport models with information relating to soil and contaminant properties, rainfall, depth to water table, etc, to derive risk maps. Although this type of modelling is increasingly performed within a GIS framework (e.g. Hollis et al., 2000; Lovett et al., 2001), error analysis is far from routine and rarely presented with the results. Only a few studies have tackled the matter of uncertainty in the modelling of risk from non-point source pollution (Li et al., 1998; Worrall et al., 2000) and there is a clear need for reliability indicators to be included with model output.

Identification of the elements to which a model is most sensitive, in terms of error propagation, can be useful in assessing which input parameters need to be measured most rigorously (Loague et al., 1989). Another benefit of undertaking such error analysis is its value in determining which elements of a vulnerability assessment framework need improving. It was therefore decided to examine the extent to which error might propagate through a regional-scale groundwater pesticide risk model, and to use the results to apply confidence limits to the model output. The aims of the study were (i) to establish the importance of accuracy in model input parameters; (ii) to determine the difference, in terms of error propagation, between a pesticide ‘attenuation-only’ model and an ‘attenuation-with-degradation’ model; and (iii) to ascertain how correlation between model variables affects error propagation.

2 Background
For the purposes of the current study, a model was developed to assess the relative risk of groundwater contamination by the widely-used herbicide isoproturon in a 30 km x 37 km area to the north-west of London, England. The study area, which straddles the Colne and Lea river basins (Figure 1), includes a mixture of arable and non-arable land interspersed with small conurbations. The area has a wide variety of soil cover and overlies a Major Cretaceous Chalk Aquifer, which is confined in the south-eastern corner by the Tertiary London Clay. Consequently, this location provided a suitably varied context for a groundwater vulnerability assessment to be made at a regional scale.

Figure 1 Location and model boundary of the study area, in the Thames region to the northwest of London, England.
3 Methods

3.1 GIS layers used in the pesticide risk model

A simple pesticide leaching model, the structure of which is shown in Figure 2, was developed. This model combined contaminant transport equations with depth to water table data and the results of a pesticide degradation experiment.

An ‘attenuation factor’ was used to describe pesticide attenuation in the topsoil and unsaturated zones. It combined parameters relating to soil texture and organic carbon content with pesticide partitioning and decay values, effective rainfall and depth to water table data (Figure 2). Current estimates of unsaturated zone thickness are based on relatively limited data, so an updated depth to water table data set was incorporated in the model for the purposes of the current study. This data set was derived by interpolating observed water levels from 94 Environment Agency monitoring boreholes within the study area, measured during a period of high rainfall in October 2001, and subtracting the interpolated surface from a digital layer of surface topography (Ordnance Survey, Land-Form PANORAMA digital terrain model, 50 m grid resolution). The pesticide degradation experiment (Posen et al., in press) assessed differences in the topsoil degradation potential of eleven dissimilar soil types identified in the study area, with respect to isoproturon. The results of this experiment were used to assign a ‘degradation factor’ to each soil type, to describe microbial degradation processes in the topsoil, which are not routinely considered in such models.

Each of the model input parameters (Figure 2) was converted to a 1 km grid layer for computation in the PCRaster program (Van Deursen and Wesseling, 1995).
3.2 Calculation of the attenuation factor

Attenuation of pesticides in the topsoil and unsaturated substrate layers can be described by the equation:

\[ A = A_0 e^{-0.693t/t_{1/2}} \]  

(1)

where:
- \( A \) – is the pesticide concentration at the base of each respective layer,
- \( A_0 \) – is the pesticide concentration at the top of each respective layer,
- \( t \) (days) – is the time taken for the pesticide to travel through each layer,
- \( t_{1/2} \) (days) – is the half-life of the pesticide in each layer.

In the current study, \( A_0 \) (topsoil) was assigned a value of 1, and \( A_0 \) (substrate) was assigned the calculated value of \( A \) at the base of the topsoil layer. Thus the calculated value of \( A \) at the base of the unsaturated substrate (i.e. at the water table) represented the fraction of isoproturon remaining after leaching through both layers, hence describing an attenuation factor, \( A_f \), where:

\[ A_f = \frac{A_{\text{substrate}}}{A_{\text{topsoil}}} \]  

(2)

Published half-life values for isoproturon were taken from Mackay et al. (1997) and were set at 21 days and 250 days for \( t_{1/2} \) topsoil and \( t_{1/2} \) substrate, respectively, to account for changes in pesticide decay rates at different depths in the soil column.

Calculation of \( A_f \) had to be performed in several steps, for (i) the topsoil, and (ii) the unsaturated substrate layer, by combining the various model parameters identified in Figure 2. To this effect, Equation 1 was broken down as follows:

\[ t = \frac{L}{v_c} \]  

(3)

where:
- \( L \) (m) – is the depth of the layer,
- \( v_c \) – is the velocity of the contaminant.

A constant depth of 0.2 m was assigned to the topsoil layer, except at locations where groundwater was in contact with the topographic surface, in which case the value was set at zero. For the unsaturated substrate layer, the value of \( L \) was described by:

\[ L = (\text{surface topography} - \text{topsoil depth}) - \text{water table elevation} \]  

(4)

Contaminant velocity, \( v_c \) in Equation 3, is described by

\[ v_c = \frac{i}{\theta \cdot 1/R_f} \]  

(5)

where:
- \( i \) (m day\(^{-1}\)) – is the depth of the layer, is the infiltration rate, and is taken to be the equivalent of effective rainfall (from the UK Meteorological Office Standard Average Annual Rainfall records for the period 1961-1990),
- \( \theta \) – is soil porosity (expressed as a decimal fraction),
- \( R_f \) – is a retardation factor for contaminant transport through the soil layer.
The retardation factor, $R_f$, is dependent on factors related to the hydrophobicity of an organic compound (Freeze and Cherry, 1979) and the organic carbon content, bulk density and saturated porosity of the soil. It can be described by the equation:

$$R_f = 1 + \left( \frac{K_{oc} f_{oc} \rho_b}{\theta} \right)$$

(6)

where:
- $K_{oc}$ – is the organic carbon partitioning coefficient of the pesticide,
- $f_{oc}$ – is the fraction of organic carbon in the soil,
- $\rho_b$ – is soil bulk density,
- $\theta$ – is soil porosity.

Values for the soil parameters ($f_{oc}$, $\rho_b$, $\theta$) were obtained from the NSRI-derived database (Williamson et al., 2001) and the $K_{oc}$ value for isoproturon (which is variable, but, for the purposes of this study, was applied as a constant) was taken from Mackay et al. (1997).

Once all of the input parameters had been converted to 1 km grid layers in PCRaster, the attenuation factor, $A_f$, was calculated according to Equations 1 to 6.

### 3.3 Inclusion of the degradation factor

Finally, the product of the degradation factor and $A_f$ layers was calculated to give an ‘attenuation-with-degradation’ factor, $A_f$-deg, representing the fraction of isoproturon that could impact on the water table after leaching through the entire unsaturated soil column, including the potential for microbial degradation in the topsoil.

The respective $A_f$ and $A_f$-deg layers were exported to ArcView GIS (Version 3.2, ESRI, http://www.esri.com) and overlaid onto a digital map of the major Chalk aquifer in the study area (based on British Geological Survey, Solid Geology, 1:50,000). The resulting maps showed the relative risk, in each case, of groundwater contamination by isoproturon.

### 3.4 Error propagation analysis

A mean value and standard deviation were calculated for each of the soil physical attributes. The degradation factor was calculated according to soil type, using the range of values obtained in the pesticide degradation experiment. As this model parameter was derived experimentally from a small number of random samples and used to represent the degradation potential for each soil type, it was subject to sampling error and, therefore, better expressed by a mean value with associated standard error. Relative error values were calculated for the soil parameters and degradation factor to facilitate comparison between these variables.

The water table and rainfall input layers were derived by kriging interpolation in the GS+ program (Version 5.1.1, Gamma Design Software, http://www.gammadesign.com), the output from which included the respective kriging standard deviation values. The standard deviation value for the surface topography layer was obtained from the UK Ordnance Survey website (http://www.ordnancesurvey.co.uk/oswebsite/products/landformprofile/techinfo.html). For the purposes of this study, pesticide parameters ($K_{oc}$ and half-life) were treated as constants.

Error propagation was computed for both the attenuation-only and attenuation-with-degradation models, using the method described in Burrough and McDonnell (1998), p. 247-250. The error analysis was performed twice: the first time it was assumed that all model variables were independent, and the second time intercorrelation of variables was accounted...
for. Values of correlation coefficients, $r$, associated with the second analysis are given in Table 1.

Table 1 Values of the correlation coefficient, $r$, between variables in the pesticide leaching model.

<table>
<thead>
<tr>
<th></th>
<th>Water table</th>
<th>Porosity</th>
<th>Bulk Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topography</td>
<td>0.833</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Organic Carbon Content (topsoil)</td>
<td>-</td>
<td>-0.050</td>
<td>-</td>
</tr>
<tr>
<td>Organic Carbon Content (substrate)</td>
<td>-</td>
<td>0.910</td>
<td>-</td>
</tr>
<tr>
<td>Organic Carbon Content/Porosity (topsoil)</td>
<td>-</td>
<td>-</td>
<td>-0.586</td>
</tr>
<tr>
<td>Organic Carbon Content/Porosity (substrate)</td>
<td>-</td>
<td>-</td>
<td>-0.766</td>
</tr>
</tbody>
</table>

4 Results

4.1 Error in the input data

Error values for the model input parameters are given in Table 2. The relative error values indicate that soil organic carbon content and the degradation factor introduce greatest variability into the model. In the case of the latter parameter, the high variability reflects the small number of experimental results used to derive this value.

Table 2 Error values for the pesticide leaching model input parameters.

<table>
<thead>
<tr>
<th>Input parameter</th>
<th>Maximum Relative Error (%) $^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bulk density (g/cm$^3$)</td>
<td>21.99</td>
</tr>
<tr>
<td>Fraction of organic carbon</td>
<td>129.73</td>
</tr>
<tr>
<td>Porosity (fraction of total volume)</td>
<td>12.05</td>
</tr>
<tr>
<td>Degradation factor</td>
<td>147.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface topography</td>
<td>100.29 m</td>
</tr>
<tr>
<td>Water table elevation</td>
<td>72.00 m</td>
</tr>
<tr>
<td>Rainfall</td>
<td>214.02 mm</td>
</tr>
</tbody>
</table>

$^a$ Maximum relative error, rather than mean and standard deviation/standard error, have been displayed here, due to the large number of soil types and their respective values.

4.2 Error propagation through intermediate steps of the model

Error analysis indicated that that the greatest error in calculation of the retardation factor, $R_f$, was associated with soil organic carbon content and variability, and for the calculation of depth to water table, $L$, error was almost entirely attributable to interpolation of the water table surface. Error associated with the calculation of contaminant travel time, $t$, was negligible in the topsoil compared with that in the substrate, due to the extremely low velocities involved and the deep thickness of the substrate at some locations. Consequently the lowest error values for $t$ occurred where the water table lay closest to the ground surface.
4.3 Error in the final models

The resulting error maps (for the non-correlated analysis) indicated that error associated with
the final step of the ‘attenuation-only’ (Af) model was negligible over most of the study area
(Figure 3(a)), whereas inclusion of a degradation factor (Af-deg) led to a marked increase in
model error (Figure 3(b)), reflecting the high values of standard error associated with the
calculation of this factor. Lowest error in the Af model, and highest error in the Af-deg model,
(Figures 3(a) and (b), respectively) occurred at locations where groundwater was in close
contact with the topographic surface.

When correlated variables were accounted for in the analysis, the error was reduced to values
close to zero in both models, so the error maps have not been displayed here. This reduction in
error in the latter analysis can be attributed to the way in which errors associated with various
parameters cancelled each other out at certain model stages.

![Figure 3](image)

Figure 3 Maps showing the spatial distribution of error (without correlation) in the calculation of (a) the
‘attenuation-only’, Af, risk factor, and (b) the ‘attenuation-with-degradation’, Af-deg, risk factor.

4.4 Applying the error to the final models

The derived error values were applied to the Af and Af-deg models, respectively, in order to
assign confidence intervals to the mean values. Figure 4 shows the mean risk factor for the Af
model. Both the lower and upper 95% confidence limits for this model produced risk maps
identical to the mean risk factor map, indicating that error propagation was negligible.
Figure 4 Map showing mean risk factor values (on an arbitrary scale where 0 and 1 represent lowest and highest risk, respectively) for the ‘attenuation-only’, Af, model (without correlation). Maps representing the values of lower and upper 95% confidence limits for the Af model were identical to this map.

Figure 5 (a) Mean, (b) minimum, and (c) maximum risk factor values for the ‘attenuation-with-degradation’, Af-deg, model (without correlation).
However, application of the error to the mean Af-deg model (Figure 5(a)) effectively reduced the minimum risk values to zero over much of the study area (Figure 5(b)), but produced a twelve-fold increase in maximum risk value for the high-risk cells (Figure 5(c)).

5 Discussion

The largest contributors of input error (as indicated in Table 2) came from the soil organic carbon, interpolated water table and degradation factor layers, suggesting that these parameters may have the greatest influence on model error. As the standard deviation for the topography layer was constant over the entire study area, error in the depth to water table calculation, \( L \), was almost entirely attributable to the water table kriging standard deviation. The highest error values for \( L \) occurred in areas that were furthest from groundwater observation borehole locations, highlighting the importance of having good spatial distribution of data when interpolating surfaces that are to be used as input to other models.

It should be noted here that error at the intermediate model steps gives an indication as to which elements of the model are particularly susceptible to error propagation. Highest model error occurred in the calculation of \( t \) for the substrate, a parameter that could be used to assess the likelihood of a contaminant reaching a groundwater body within a certain number of the compound’s half-lives. If the value of \( t \) was to be used for such a purpose, particular attention should be paid to the minimising of error values in this calculation.

In the final models, error in the attenuation factor, \( A_f \), was negligible throughout the study area (Figure 3(a)), but increased by several orders of magnitude at some locations when the degradation factor was included in the calculation (Figure 3(b)). The most significant error increases in the attenuation-with-degradation model, \( A_f \)-deg, were seen at locations where the water table is in close contact with the topographic surface and the role of degradation (with its associated high error values) assumes greater importance.

Application of the error calculations to the mean \( A_f \) and \( A_f \)-deg models (Figures 4 and 5(a), respectively) showed that error propagation only assumed importance when the degradation factor was included in the model. There was a marked difference between minimum and maximum risk maps for the \( A_f \)-deg model, as shown in Figures 5(b) and (c), respectively, the minimum risk map showing only six high-risk cells in the eastern half of the study area. The range of values (-11 to 12) for the \( A_f \)-deg risk factor gave a clear indication that inclusion of the degradation factor greatly reduced the reliability of the model, leading to a large increase in output uncertainty.

From a modelling perspective, the \( A_f \)-deg model is more informative than the \( A_f \) model, as it has a clearly-defined mean value, with upper and lower confidence limits. In contrast to this, the \( A_f \) model shows no differentiation between mean, upper and lower limits. This indicates that the \( A_f \) model is more precise than the \( A_f \)-deg model, in terms of minimising the associated error, but greater precision does not necessarily correspond to greater accuracy. Therefore, from a risk assessment perspective, the \( A_f \)-deg model (with its associated confidence limits) gives a clearer indication of the areas of greatest risk uncertainty.

The analysis highlights the importance of minimising model input error. Regarding model error, it was not possible, in this particular study, to avoid using correlated variables, or certain mathematical functions. However, despite the existence of strong correlations between some of the model parameters, the mathematics acted in such a way as to reduce the propagation of error through the model, and taking account of correlation improved the model precision.
6 Conclusions

The investigation presented in this paper has emphasized the importance of considering, and including, error propagation analysis as a routine part of any modelling exercise. In particular, it highlighted input parameters that introduced the most uncertainty into the model, as well as indicating the geographical locations of greatest output uncertainty. Error is inherent in all models and, with respect to the study presented here, its analysis should be incorporated in all groundwater vulnerability assessment exercises. Every effort should be made to obtain good quality input data and associated metadata for all model parameters, and confidence limits should be routinely presented with modelling and mapping output. It is suggested that output should also include information for subsequent users, explaining any limitations and giving instructions for appropriate use.

Acknowledgements

This study was conducted as part of a Natural Environment Research Council PhD studentship (NER/S/A/2001/06/48), with CASE partner support from the UK Environment Agency.

References


