EnviroInfo 2000: Umweltinformatik ’00 Umweltinformation für Planung, Politik und Öffentlichkeit
Copyright 2000 Metropolis Verlag, Marburg, ISBN: 3-89518-307-5

# Protecting Water Supply Quality – Decision Support Using Geographical Information Systems (GIS)

Jim Foster[1] and Adrian McDonald[2]

**Abstract**

In order to optimise the quality and use of abstracted waters it is imperative to assess the risk of pollution from the catchment area upstream of the supply intake. This process can be greatly facilitated through the use of environmental spatial data and decision support software. This paper describes the development of a number of tools to support environmental risk assessment and operational decision making using a variety of environmental spatial data.

The process of risk assessment for pollution prediction is a novel application of geo-environmental information. This process requires the assimilation of data which are spatially variable in nature, making geographical information systems (GIS) an ideal tool for such assessments. PC-based geographical information systems (WINGS and MapInfo Professional) are used in the evolution of a risk assessment methodology to determine catchment risk.

Examples are given showing how raster and vector data are used within a GIS framework to produce maps indicating areas of potential hazard to water quality within the River Wharfe catchment of North Yorkshire (UK). Data are also coupled with known modelling techniques to predict and quantify risk frequency and impact.

The work illustrates the potential of GIS to encourage the predictive management of water supply intakes through the assessments of hazard and risk and the modelling of management strategies such as specified grazing areas and the selective use of supply sources. The information science aspects of this development work are described and a number of example applications are illustrated that are of potential interest to end users.

[1]School of Geography, University of Leeds, Leeds, LS2 9JT, UK,
Email: mail@jimfoster.fsnet.co.uk
[2] School of Geography, University of Leeds, Leeds, LS2 9JT, UK,
Email: eedean@geography.leeds.ac.uk.

## 1. Problem Orientation

Water supply companies operating within the European Union (EU) have a statutory duty to provide water that is wholesome to drink. Over half of the drinking water supplied in the UK is obtained from surface water abstractions, i.e. where resources are abstracted from rivers, streams or impounding reservoirs.

The quality of these waters is a product of the rainfall chemistry, surface characteristics, land-use, soils and geology of the river catchment areas. Several water quality issues have been found to be prevalent in upland catchments, and therefore influence the activities of a water supply company abstracting drinking water from such sources. These include agricultural pollution, discoloration, microbial contamination and trace metal mobilisation. In addition to these problems, river intakes throughout Britain have been subject to pollution from industrial spills, road traffic accidents and untreated sewage effluent on numerous occasions (Foster 1998).

In order to comply with drinking water standards, and optimise treatment costs and efficiency, supply companies need to be aware of the characteristics of their water gathering grounds and how these characteristics, combined with the activities taking place, can affect raw water quality. Knowledge of potential contamination sources in a river catchment is often limited to local knowledge of the site operators, with no procedure for the collection or analysis of this knowledge. One approach to assessing potential impacts is through the application of an environmental risk assessment. The assessment of pollution risks to a surface water intake involves the assimilation of large volumes of spatially variable data. The intrinsic capacity of geographical information systems (GIS) to store, analyse and display such data makes them ideal tools for assisting risk assessments (Rejeski 1993).

The process of risk assessment for pollution prediction is a novel application of geo-environmental information and as such there are very few existing applications in the contemporary literature. Those applications which have been documented either do not utilise digital data sources (Cole/Lacey, 1995) or refer specifically to groundwater resources (Barrocu/Biallo 1993), the transport of hazardous goods (Brainard/Lovett/Parfitt 1993) or specific contamination issues such as individual pesticide products (Battaglin/Goolsby 1996). Only one application of GIS-based techniques to the protection of water supplies has been carried out (Breach/Porter/Court/Hollis/Keay/Hallett 1994), but again this only considered one aspect of raw water quality (pesticide contamination). Foster (1998) produced the first attempt the develop a generic risk assessment procedure utilising the potential of GIS techniques. This paper describes the information science aspects of this development work and describes a number of example applications of interest to end users.

## 2. System Specification

### 2.1 User Requirements

The water company involved in this research (Yorkshire Water Services Ltd.) required a decision support system that would act as a store for digital geo-referenced information, provide predictive assessments of water quality risk and act as a planning tool to model potential scenarios such as spill incidents.

The system was developed in such a way that it could be used throughout a large water utility. Thus it was necessary to use both hardware and software that facilitated a cost-effective rollout of the system across a broad spectrum of end users. The MapInfo-based system was therefore developed to enable risk assessments and hazard maps to be displayed on the corporate desktop mapping software (WINGS geographical information system). This allows over 1000 end users to access the information across the company's sites and at remote operational locations (Mitchell 1994).

### 2.2 Technical Description of System

The system was developed on a medium specification (at the time of development) desktop personal computer incorporating a Pentium 133Mhz processor, 32 megabytes of RAM and 2 gigabytes of hard disk storage. This was done to enable the cost effective use of the solution in a commercial environment.

A number of software elements were utilised in the development of the system. The need for two specific geographical information system (GIS) packages came to light as much of the source data were made available through the WINGS-based GISLAB initiative at Yorkshire Water. However, the WINGS software (distributed by Systems Options Ltd) package was known to have limited data manipulation capabilities, as the spatial analysis tools were only in the early stages of development. MapInfo (Version 4.0) was therefore selected to carry out the spatial analysis as this represented a leading PC-based desktop GIS software at the time of the study.

In general, it was found that data could be displayed equally well in either software package, although the creation of original data layers (or maps) was often easier in MapInfo due to its simple tabular data structure and the ease of interaction with other packages such as Microsoft Excel.

The use of "Mifmerge" software that enables data transfer from MapInfo to the WINGS data format was a significant advantage. However, the reverse of this, when data were transformed from WINGS format for analysis in MapInfo was not totally automated and involved considerable manual conversion using alternative software such as text editors and spreadsheets. The text files associated with the WINGS data structure had to be manually converted into the correct format for import to

MapInfo. Some of this was automated by the creation of text editing and data sorting macros in Microsoft Word and Excel, but considerable time was spent manually transforming and especially verifying the data. However, once these transformations had been carried out, interaction between the MapInfo-based GIS analysis and the WINGS-based desktop mapping “front-end” was not computationally intensive due to the use of the Mifmerge software.

Whilst the display and interrogation of the resultant hazard maps was possible in both WINGS and MapInfo, the data processing capabilities of MapInfo were found to be much more suitable for the analysis of digital data despite the use of experimental (beta release) polygon algebra functions within WINGS. The successful creation of new data layers through the combination of existing data was found only to be possible using MapInfo. For example, the 25m ITE Landcover data could be used to identify in which part of the catchment animals were likely to be grazing. Using MapInfo’s SQL selection tools, the cells could then be assigned animal grazing intensity ranks in accordance with the rank of the 2km agricultural census data cell that they were geographically within. In all such cases cell rankings were utilised so as not to misrepresent absolute values when combining data of different resolutions. This was seen as a suitable method of addressing many of the problems associated with approximating information from values transferred from data of several scales.

The table-based format of maps in MapInfo allowed the combination of the different data layers based on either attribute data or geographical location, thus allowing true GIS analysis to be carried out to create the hazard maps. Data created in MapInfo could be easily returned to WINGS for display purposes. The resultant hazard maps can therefore potentially be displayed across the corporate-wide mapping system in use at Yorkshire Water Services Ltd.

The automated searching of digital data to identify potential point sources of pollution did make extensive use of WINGS’ capabilities. The Ordnance Survey Landline data were stored within WINGS in such a manner that text strings could be easily searched to identify features of interest using the WINGS “dump” feature. In addition, the 1:10,000 raster background maps were visually searched in WINGS for those areas where the larger scale vector maps were not available. This was also carried in WINGS to avoid the need for the large picture files containing the mapped information to be repeated in a format readable by MapInfo.

The amalgamation of the GISLAB data and the improved functionality are seen as key elements in the development of sophisticated GIS-based tools to assist in the integrated management of the company’s assets within a complex geographic context such as groundwater, reservoirs and river catchments. The research described in this paper is closely linked to the utilisation of these data to develop pro-active approaches to raw water management.

The techniques described above were utilised in the development of a GIS-based risk assessment system to aid decision making in the catchment areas. Several examples of the system's use are described below.

## 3. Pollution Hazard Mapping

The first stage of any assessment of risk is to identify the potential hazard presented by the existing conditions (or potential changes to these conditions). In a river catchment this involves identifying the location of the factors influencing raw water quality upstream of the supply intake site.

### 3.1 Identifying Pollution Sources Using Logic Trees

Logic trees are used to combine the physical characteristics that result in potential pollution sources and therefore predict their spatial distribution using GIS (i.e. they form the basis of the creation of hazard maps). Fault trees were used in the identification of causal activities in the catchment that may result in an undesirable concentration of a particular substance reaching the intake. Event trees can then be used as a method of assessing the significance of the consequences of such concentrations at the supply intake. The terms *Source Trees* and *Consequence Trees* respectively, are more applicable to their use here.

Using the terminology commonly applied to these tree structures, the *root* of the diagram is representative of the riverine concentration in each case and the *leaves* determine the catchment information or GIS data needs for the Source Trees and possible outcomes for the Consequence Trees. The schematic basis of these tree structures is shown in Figure 1. The use of such Source Trees combined with hazard mapping is the primary method of hazard identification in the methodology.

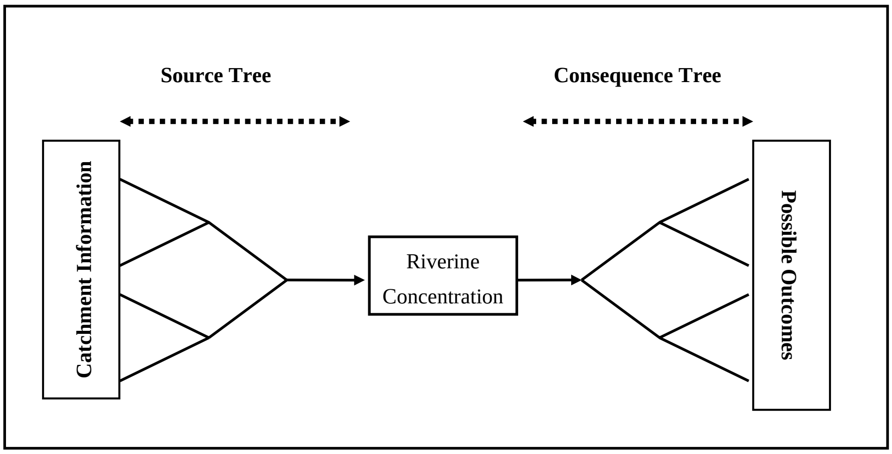


Figure 1
Schematic logic tree for risk assessments of potable water supply intakes

The tree structures are inherently flexible and can be modified according to the level of understanding of the processes involved. For example, the configuration of the structure will vary between the categories of point, quantifiable non-point and unquantifiable non-point pollution sources (these may in turn lead to quantitative, semi-quantitative and qualitative results). Careful construction of the logical structure describing a process allows common-cause elements to be highlighted. The preparation of such trees for hazard identification stage is a fundamentally important part of the whole risk assessment as it necessitates an open-minded view of the potential pollution hazard scenarios in the catchment and may therefore highlight problems previously unidentified.

The output from the creation of the Source Trees is used to model catchment relationships in the GIS. Hazard maps will be created, identifying potential water quality hazards in the intake catchment.

Source trees are used to relate the characteristics of the sub-catchment areas to potential water quality hazards. In this way the areas of land that are of particular concern can be identified and mapped using the GIS. This allows risk mitigation strategies to be targeted at the key areas. The source trees form the basis of the GIS data models, identifying the necessary data layers and how they need to be combined. Where generic, published values (such as infection or application rates) exist these are used to help quantify the relative importance of contributory factors and hence sources areas. GIS overlay and map algebra techniques are used to combine these data to produce the hazard maps. The aim of this approach is not to quantify exact amounts reaching the river or model flows to the intake site, but to look at the relative contribution of potential source areas in terms of their level of hazard.

### 3.2 Example Application: C*ryptosporidium* Hazard Mapping

The significance of all pollution sources will be strongly influenced by the catchment hydrology. The runoff of water over the land surface after a rainfall event may transport pollutants quickly and efficiently into the stream network. It is therefore extremely important to assess the potential for runoff occurring over a given area of land when considering pollution sources on the land surface. In order to assess the significance of such areas, a Runoff Potential Index (RPI) was calculated in the GIS to assist in hazard identification.

The RPI, displayed as a ranking from 1 (very low) to 5 (very high), was created by combining the annual effective rainfall, land slope, soil hydrology and proximity to a watercourse. These ordinal ranked data were combined using geographical overlay, with all calculations being based on the smallest spatial unit (i.e. the slope – created from a triangular irregular network (TIN) in ArcInfo). This resulted in some spatial approximation due to the varying resolution of the data being used, but provided an indication of the most important pollutant source areas. One example application of the RPI is in the identification of *cryptosporidium* source areas.

*Cryptosporidium parvum* is a protozoan parasite that can cause the disease cryptosporidiosis in human and animal populations. Fatalities may occur in immuno-suppressed populations as a result. The protozoa are transported in the raw waters via microscopic oocysts, which are extremely difficult to remove by conventional drinking water treatment. Large numbers of oocysts (>$10^{10}$ animal$^{-1}$ day$^{-1}$) may be deposited on the ground in animal faeces and subsequently washed off the land surface by overland flow events. *Cryptosporidium* represents a major potential hazard to public health through the contamination of water supplies obtained from upland catchments and it is important to identify sources areas of this harmful pathogen.

Agricultural census data from annual farm surveys have been converted from paper records based on English parishes to 2km grid square digital data by the Edinburgh Data Library. Animal stock numbers stored in these tables were multiplied by average animal infection and faecal production rates to determine the predicted *cryptosporidium* loading for each grid cell. In order to reduce spatial approximations, the Institute of Terrestrial Ecology Landcover data map was introduced to identify where stock were located. Using the "contains within" command within MapInfo's SQL menus, all of the 25m Landcover cells on which agricultural animals may be found were assigned an average oocyst production value for *cryptosporidium* from the 2km data. This meant that only those areas where the source could be found (i.e. where animals were present) were displayed on the resultant hazard map.In order to identify which areas of the land were likely to represent the most significant sources, the RPI cells calculated previously were combined with the oocyst loading. This was again done by joining tables in MapInfo using the ‘SQL Select’ command to chose all RPI values (geographically) contained with the load ranked cells. The resultant table contained attribute data describing the runoff potential and *cryptosporidium* load (both expressed as a rank). These cells were then displayed on a thematic map to illustrate the areas of greatest risk to water quality from *cryptosporidium*. In addition to non-point sources (i.e. faecal deposition), potential point sources were identified from the OS vector maps. Locations such as stock holding areas, farm yards and buildings were selected from the maps using one of two techniques. For much of the catchment data could be obtained using the automated techniques described below. The remainder were identified from visual inspection of the 1:10,000 OS map layer held in GISLAB.

Figure 2 shows the *cryptosporidium* hazard map for the upper River Wharfe catchment area (within the GIS this can be interrogated down to field scale).

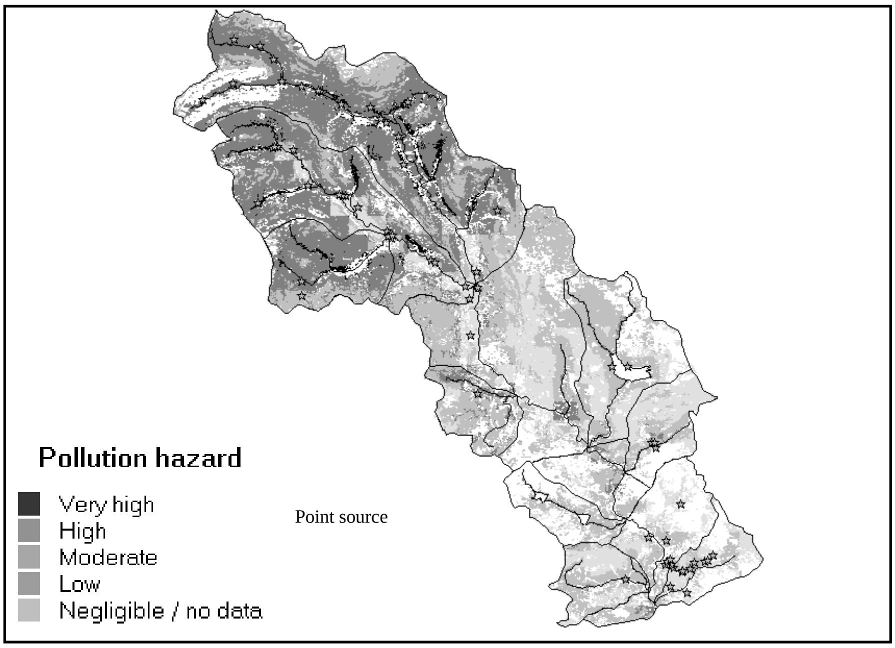


Figure 2.
*Cryptosporidium* hazard map of the Upper Wharfe catchment

High risk (dark ) areas represent land where high animal densities promote *cryptosporidium* deposition and the environmental conditions are most suitable for oocyst transport. From maps such as these, high risk areas can be easily targeted for remedial action. Similar maps have been created for contamination from pesticides, oils & greases, naturally occurring colour, trace metals (iron, manganese & aluminium), faecal bacteria, lead, nutrients and phenolic compounds.

Pro-active catchment management such as the use of buffer zones, grazing restrictions and the promotion of alternative land uses, can thus be used to reduce the risk to water quality at the intake site.

## 4. Coupling geo-environmental data to predictive modelling techniques

Thus far only the spatial categorisation of hazard has been considered. This has allowed (i) the identification of sensitive areas, and thus the targetting of remedial strategies within a catchment and (ii) the ranking of catchments to target better monitoring and investment programmes. However, the decision support system outlined here also has the capacity to predict some of the hazard components. Two examples, having very different prediction time horizons, are presented which relate to colour and to road spills.

Colour is a significant problem for water suppliers because (i) it is easily seen and gives rise to customer complaint, (ii) it is expensive to remove and (iii) when chlorinated, produces potentially carcinogenic trihalomethanes.

Past research (e.g. Mitchell and McDonald 1992 & 1995) has yielded a number of predictive models describing the relationship between catchment characteristics and stream water quality. Here the ability of the GIS to analyse several large files of spatially referenced data is used to quantify water discoloration and aluminium concentrations at a sub-catchment outlet using such models. The equations, based on step-wise multiple regression, are as follows:

$$\log_{10} \text{Colour} = 0.00512\ (\%\text{TCLA}_5^0) - 0.609\ (\text{MSS}) + 0.00368\ (\%1011\text{b}) + 0.21435$$
$$\log_{10} \text{Al} = 0.034\ (\%\text{TCLA}_5^0) - 10.0019\ (\text{RR}) - 0.653$$

*Where:*$\%TCLA_5^0$ *= percentage of total channel length in areas of less than or equal to* $5^o$ *slope, MSS = main stream slope (slope between 10 and 85 percentiles of main stream), %1011b = percentage of catchment area of soil type 1011b (Winter Hill Peat), RR = Relief ratio (Basin relief / basin length)*

These models, developed by Mitchell & McDonald (1992), have been found to explain 82% and 68% of variations in colour and aluminium respectively in upland catchments similar to the one reported here. Considering a small sub-catchment as an example, Table 1 shows predictions of colour and how these compare to values obtained by on-site sampling. In general the mean values were over predicted for this sub-catchment, with the predictions for colour actually being closer to the maximum values obtained from the sampling. Predicted values can potentially be compared to the water quality regulations to determine frequency estimates.

| | **Colour (Hazen)** |
|---|---|
| Predicted value (mean) | 46.54 |
| Maximum (from sampling) | 45.70 |
| Mean (from sampling) | 19.16 |
| Minimum (from sampling) | 0.01 |

Table 1.
Predicted and actual values of Colour in Littondale

In addition Naden and McDonald (1989) have shown that risk of polluting events may be forecast from the moisture deficits 3 months and 14 months previous to the pollution event. For those areas identified as potentially generating colour, the size of the likely event can be forcast as an envelope curve up to a year in advance and as a detailed forecast three months in advance. Since moisture deficit information is held as Morecs data on a 40km grid it is posiible to provide a spatially based gross forecast of colour although at this grid size the prediction is less likely to have practical value to a water company.

The system has a predictive capacity for short term response events. Pollution from road spills is a serious issue for the management of river intakes. The relatively small low discharge rivers of the UK have limited dilution potential but are relatively steepslope and have rapid travel times from spill source to abstraction site. (Spill probabilities for catchments will be covered in a later section.) The decision support system allows a reported spill to be located either by direct operator action or by grid reference. Travel path is determined from digital elevation data and travel time is determined from the sums of the overland flow (to a mapped channel feature) and channel flow components. For both elements the worst case, maximum velocity, minimum time value is used. A more sophisticed approach could use accumulated dead zone or kinematic wave models to identify, more precisely, travel times and dilution but in practise water companies are risk-averse in such situations. They will close down intakes at risk and only reopen them when physical sampling has indicated that the hazard has passed. The predictive element in the systems presented here are held at a fit-for-purpose level.

## 5. Combining Geo-environmental Data with Macro-scale Predictions

Road traffic accidents are potentially a major source of pollution to water supply intakes. The main hazards exist from the bulk transport of liquid substances in large tankers. Information about the road network was combined with accident and spill statistics to give a probabilistic assessment of a spill occurring in the catchment. Equation 1 is used to calculate the total road spill risk in a catchment area.

$$Total\_Road\_Spill\_Risk = \sum_{i>0}^{i=n} \left[ L \times F(0.1) \times A \times P_i \right]$$

Where: L = total length of road of a particular class in the catchment,

F = average flow of Heavy Goods Vehicles (HGVs) per year,

0.1= proportion of HGVs that are bulk tankers,

A = accident involvement rate per million vehicle kilometres travelled,

Pi = probability of a spill of size i occurring as a result of an accident.

Equation 1.
Calculation of road spill risk (adapted from Cole and Lacey, 1995)

A MapBasic program was written in MapInfo to extract the relevant data from the attribute tables and calculate the equation. An information box was then displayed on-screen to show the result. (The attribute extraction can also be done easily in WINGS by creating a macro to extract the relevant information from the selected data layers). Expected probabilities were calculated for two classes of road, Primary routes and Minor routes and are shown in Table 2.

| Spill size (kg) | Primary Routes | Minor Routes | Total |
|---|---|---|---|
| **Less than 150** | 0.0360 | 0.0048 | 0.0408 |
| **150 – 1500** | 0.0090 | 0.0010 | 0.0100 |
| **1500 +** | 0.0200 | 0.0030 | 0.0230 |
| **Total** | 0.0650 | 0.0088 | **0.0738** |

Table 2
Probability per annum of a tanker spill occurring in the River Wharfe Catchment
(Source: Foster 1998)

Primary routes represent a major potential hazard in the Wharfe catchment due to their extended proximity to the watercourse. Table 1 shows that a spillage from a primary route is predicted once every 15.5 years of operation (p=0.0650), compared to once every 188 years (p=0.0088) for minor roads. This approach can be enhanced by using link-specific traffic flow data and accident rates, mapping the extent of highway drainage and local traffic surveys but provides a coarse identification of the major risk areas upon which the supply company‘s resources can be focussed. In addition, such an assessment can also form the basis of emergency procedures in response to information regarding road traffic accidents in the catchment.

## 6. Conclusions / Future Research Directions

Several regulators control the water industry in the UK. In the last year the regulator has required that water companies evaluate the risk involved in their catchments. In particular they are required to monitor for *cryptosporidium* if a catchment is deemed high risk. This externally driven requirement has raised the value of the generic risk assessment systems reported here. Such systems also have potential to be used by the Environment Agency to inform their Catchment Abstraction Management Strategy (CAMS). This is currently in an embryonic state but will surely exhibit rapid growth and become a fundamental tool in environmental management. Many of the proposed elements of CAMS can be derived directly from this information system. Similarly the results of such initiatives as the Upper Wharfedale best practice project could well be demonstrated to have much wider potential by applying the results to the information system in order to display the regional effects of best practice. Finally there is a growing competitive element in the UK water industry through common carriage agreements and through the proposed 'mutualisation' of companies such as Yorkshire Water (although many others will follow this route driven by commercial pressures and a domino effect). Companies wishing to supply waters in the area of another company would be informed by knowledge of the catchment water supply risk faced by a competitor. Companies wishing to tender to be the managers of the assets of a mutual would wish to know the risks involved should they be successful and become the managers of a mutual's assets

Three areas of future work appear important. The first to simplify further the interface to the operator. The second to consider whether the integration of the catchment risk software with the water resource allocation models or intelligent network models would be advantageous particularly for scenario planning for climate change. Finally to consider whether the continued expansion and integration is the best route to a useful tool. In practice the different element of the tool are required by different parts of the business for which the presence of other tools, menus and windows can be a distraction. A question for social science research might focus on why operators appear to need a simple unitary objective system (say to show only flows from a spill) and for software scientist to provide what *appears* to be a dedicated single objective system.

## 7. Acknowledgements

The authors gratefully acknowledge the assistance of Yorkshire Water Services Ltd (part of the Kelda Group) in carrying out this work. The views expressed are those of the authors alone and are not representative of Yorkshire Water Services Ltd, or the authors current employers. Part of this work was supported by a research grant from the EPSRC to whom we are grateful.

The trademarks of all software packages are acknoweldged and the authors are grateful to the Soil Survey and Land Research Centre and Institute of Hydrology for permission to use the digital data (under licence to Yorkshire Water) in this research. Agricultural census data were provided by Edinburgh University Data Library. Ordnance Survey map data are reproduced with the permission of the controller of Her Majesty's Stationary Office, Crown Copyright. Licence No. WU298581.

## 8. Bibliography

Barrocu, G., Biallo, G. (1993): Application of GIS for aquifer vulnerability evaluation. In: Kovak, K., Nachtnebel, H.P. (Eds): HydroGIS 93: Application of Geographic Information Systems in Hydrology and Water Resources. Proceedings of the Vienna Conference, April 1993. IAHS Publication No. 211, pp571-589.

Battaglin, W.A., Goolsby, D.A. (1996): Using GIS and regression to estimate annual herbicide concentrations in outflow from reservoirs in the Midwestern USA 1992-93. Proceedings of the AWRA Symposium on GIS and water resources. American Water Resources Association, USA.

Brainard, J.S., Lovett, A.A., Parfitt, J.P. (1993): GIS as a risk assessment tool for aqueous waste transport. In: EGIS '93: Proceedings of the Fourth European Conference and Exhibition on Geographical Information Systems. Utrecht EGIS Foundation, pp. 391-400.

Breach, R.A., Porter, M.J., Court, A.C., Hollis, J.M., Keay, C.A., Hallett, S.H. (1994): CATCHIS – a new computer based catchment planning and information system to assess the vulnerability of surface and groundwater catchments to contamination. In: Proceedings of the American Water Works Association Annual Conference (Water Quality). AWWA, pp 545-562.

Cole, J.A. and Lacey, R.F. (1995): Estimation of the frequency of spills of liquids from road tankers involved in traffic accidents. CIWEM seminar on intake protection and pollution risk assessment. London, 03/05/95.

Foster, J.A. (1998): Development of a catchment risk assessment procedure for the protection of potable water supply intakes. Unpublished Ph.D thesis, School of Geography, University of Leeds, UK.

Mitchell, G.N., McDonald, A.T. (1992): Predicting the spatial distribution of coloured water in North Yorkshire. School of Geography Working Paper 92/9. School of Geography, University of Leeds, Leeds, UK.

Mitchell, G.N., McDonald, A.T. (1995): Catchment characterisation as a tool for upland water quality management. Journal of Environmental Management. Vol.43. pp. 83-95.

Naden, P.S., McDonald, A.T. (1989): Statistical modelling of water colour in the upland: the Upper Nidd catchment 1979-1987. Environmental Pollution 60, 141-163