t2.11

Risk-based prioritisation and categorisation of potentially contaminated land using GIS

Stuart D Gillies GIS Team Manager, Babtie Group Ltd.
Steve Morrison Senior Environmental Scientist, Babtie Group Ltd.
Colin Harding Divisional Director, Babtie Group Ltd.
Michael Ayres Senior Applications Developer, ESiT Ltd

Summary bullet points

- Source-Pathway-Receptor risk prioritisation and categorisation of potentially contaminated land
- Spatial analysis
- Metadatabase
- Part IIA Environmental Protection Act 1990

Abstract

Legislation enacted under Part IIA of the Environmental Protection Act 1990 introduced a statutory obligation on all Local Authorities (LAs) to prepare, adopt and publish a written strategy for the identification and remediation of contaminated land by July 2001 (October 2001 for Scottish Councils).

Babtie Group Limited have developed a methodology for the prioritisation and categorisation of potentially contaminated sub-areas, which integrates established Source-Pathway-Receptor risk assessment principles with the spatial analysis capability of GIS. A model has been developed using standard GIS-based overlay and proximity techniques, to assess the spatial relationship between potentially contaminated sub-areas of land and statutorily recognised human and environmental receptors, taking cognisance of environmental pathways. In essence, the model outputs a risk-ranked list of potentially contaminated sub-areas to provide LAs with a rationale for focusing resources on detailed inspection of higher-risk sites. Automating the risk assessment process using GIS provides consistency in data interrogation and will realise a valuable time saving for LAs, reducing to minutes a manual process that may otherwise take several months to complete.

A Land Quality Management Database (LQMD), which is linked seamlessly to the GIS, has been developed for recording sub-area specific data including reconnaissance reports, ground investigation data, photographic evidence, archived reports, records of communications and other information relevant to the determination of potentially contaminated land. Additional GIS tools have been developed to facilitate the automation of the risk prioritisation and categorisation process and to provide auditable and user-friendly data management. The application necessitates the creation of a standard metadatabase to manage the large number of datasets used in the risk-ranking stages of the assessment exercise.

The database provides an auditable record of all datasets, which is an essential requirement since the methodology must stand up to public scrutiny. Map production and reporting tools have been developed which are linked to the LQMD to enable users to output standardised sub-area specific reports and maps.

1 Introduction

Legislation enacted during 2000, under Part IIA of the Environmental Protection Act 1990, introduced a statutory obligation on all Local Authorities (LAs) to prepare, adopt and publish a written inspection strategy for the identification and remediation of contaminated land by July 2001 for English LAs and October 2001 for Scottish LAs.

The Contaminated Land (England) Regulations 2000 (S.I. 2000/227) require that the approach taken by LAs in implementing the inspection strategy should *be "rational, ordered and efficient"* and one which attaches the greatest priority to those potentially contaminated areas of land considered to have the greatest potential to cause "*significant harm"* or *"pollution to controlled waters"*. This requires an initial area wide "strategic" review followed by more detailed specific area inspections. Further, the legislation also requires that the data used in the strategic assessment process is publicly available, auditable and that the assessment process as a whole is rational and transparent. The Environment Agency, in their guidance document "Contaminated Land Inspection Strategies – Technical Advice For Local Authorities", state that LAs should employ a methodology which begins "by comparing the location of areas of potential contamination with areas where there are sensitive receptors…" and having established their geographical coincidence and therefore two parts of a potential pollutant linkage this will allow LAs to define "inspection areas or sub-areas".

The inspection approach methodology developed by Babtie Group has been aligned to these fundamental principles to enable LAs to move efficiently from the strategic consideration of the whole authority area to a situation where smaller "sub-areas" of land and eventually individual "sites" within these sub-areas can be appraised. In the context of the Babtie methodology, "sub-areas" are taken to represent the potentially contaminating land-use areas identified from the source information provided by, amongst others, the Landmark Information Group, the EA (or SEPA) and the LA, which can, in some cases, extend over several hectares. Individual sub-areas may be made up of many smaller "sites", which represent those areas of land under separate ownership or occupation.

The large quantity of geospatial data coupled with the need to store, retrieve, analyse and report this data on a regular basis lends itself to the use of Geographical Information Systems.

This paper describes the GIS-based methodology developed by Babtie Group to support LAs in fulfilling their statutory obligations under this recently enacted legislation.

2 Summary Requirements of Section 57 Risk Assessment Methodology

Historical records indicate that many LAs, particularly those with large tracts of industrial and urban land, may have hundreds, and in some cases thousands, of potentially contaminated sub-areas to assess within the new regulatory framework for Contaminated Land. Ensuring that such a large number of sub-areas are assessed in a consistent and transparent manner would be a time-consuming and arduous process if undertaken manually. The methodology developed by Babtie Group automates the assessment methodology by integrating standard GIS proximity analysis techniques into prioritisation and risk assessment algorithms. As well as significantly speeding up the process, the approach enables LAs to optimise the assessment parameters to suit their specific concerns and requirements. Further, the Babtie approach allows LAs to easily re-evaluate the prioritisation and categorisation of potentially contaminated land, following the identification of new source or receptor data.

The spatial analysis process is built around the well established and widely adopted Source-Pathway-Receptor risk assessment principles. Initial prioritisation of potentially contaminated sub-areas is based on the Part I assessment procedures of DoE Contaminated Land Research Report No. 6, *"Prioritisation and Categorisation Procedure for Sub-areas Which May Be Contaminated"* (DoE, 1995), with initial priority groupings determined on the basis of sub-area proximity to identified target receptors, including development, surface waters and groundwaters.

Categorisation, as defined under Part II of CLR 6, is used to further assess sub-areas for their probable impact on recognised receptors and to place them into one of four Priority Categories (PC), relating to the seriousness of the potential impact and the urgency of action required. The process relies on the availability

of detailed, reliable and statistically valid data to answer the queries in the CLR 6 logic flowcharts and has significant limitations when applied directly to multiple sub-area categorisation at a strategic level, due to the fundamental requirement for extensive quantitative data on ground conditions for each sub-area. If the CLR 6 Part II process is strictly adhered to from the outset for such large numbers of sub-areas there would be no way of selecting in a “rational, ordered and efficient” manner which sub-areas to consider first. Further, the assignment of Priority Categories under CLR 6 Part II assumes that both a desktop assessment and walkover are completed beforehand; a process which would require years to complete for a LA area, which is clearly not intended.

The Babtie Preliminary Risk Ranking System (BPRRS) approach was developed as an intermediate step to overcome the limitations of CLR 6 Part II and is one that provides a rank-ordered assessment of the relative risks presented by potentially contaminative sub-areas.

The BPRRS provides a scored hierarchy of sub-areas each with a Preliminary Priority Category (PPC) to identify the worst-case areas for more detailed desktop assessment and walkover survey. It is a semi-quantitative ranking methodology, which makes the best use of available digital data to enable sub-areas to be ranked quickly and efficiently in an objective and consistent manner. The application of the BPRRS eliminates much of the front end "subjective" judgement typically associated with a purely qualitative approach to risk assessment through the inclusion of weighted scores for each factor used in the characterisation of the potential pollutant linkages.

Scores or ‘weightings’ are assigned to potential sources according to their assessed ‘severity', pathways according to their likely ‘efficiency’ in enabling contaminants to migrate and receptors according to their ‘sensitivity’. These scores are then used in the risk assessment algorithm, scripted into Avenue, to derive an overall score for each potentially contaminated sub-area. The scores themselves are not an absolute measure of the risk presented by a sub-area, and are intended merely to provide a measure of the relative risk to statutory receptors in comparison to other sub-areas.

There are two fundamental principles behind the risk ranking system. The first is that each element in the potential pollutant linkage will have equal importance. The second is that the relative score allocated to the various factors within each element of the linkage are proportionate to their relative importance. To prevent any element of the linkage dominating the final risk ranking score a standard technique of normalisation is applied.

Given the large number and variety of geo-datasets required to fulfil the requirements of the methodology, robust data management is an important and fundamental part of the process. To achieve this Babtie used a core set of tools – "BabtieCoreTools" - created for Arcview GIS to facilitate data and metadata management, based on the national standard for metadata. "Section57CoreTools" (S57CT) is an extension of BabtieCoreTools, which includes the prioritisation and risk assessment tools for Part IIA work and other associated Arcview functionality, as well as a transparently linked Land Quality Management Database.

3 Input Data

The site prioritisation algorithm is based around the established Source-Pathway-Receptor risk assessment principle. Data is therefore required for each of these components, as discussed below:

Source Datasets Identifying Sub-areas of Potentially Contaminated Land

The methodology adopted for the assessment of potentially contaminated land is dependent upon the historical information regarding the previous uses of each sub-area in a LA area. In general, source information may be considered as discrete sub-areas, which, at some point in their history, have been occupied by a potentially contaminating land-use. For most areas of the UK there are a number of potential suppliers of such information, including each LAs own archive, data from the Environment Agency (or SEPA) and, crucially, data gathered from historical topographical mapping. The inherent diversity in the nature of these datasets – the way it is collected, the way it is recorded, the way it is stored – creates a problem in that it must all be integrated within the model in a manageable way. In order to do this a standard and consistent data structure was adopted for all source information regardless of where it was generated.

The data structure of the Landmark historical land-use datasets was seen as a suitable reference for this standard as this data is available for a large proportion of the UK, it has been collected using a standard, consistent and robust methodology and is well documented and understood. It is not subject to local variability and is consistent for all parts of the country for which it has been collected. As the data structure is straightforward and well documented, it is also relatively easy to take data from other suppliers and modify it to the format of the Landmark historical land-use data, while ensuring that the conversion process is documented and auditable.

The Landmark datasets have been captured and collated in a standard manner based on historical Ordnance Survey mapping dating back to around the 1850's. The data was collected as three distinct datasets according to the geographic characteristics of the sub-area they represent:

- Points – sub-areas with an area less than 100m^2
- Lines – linear sub-areas (such as railways)
- Polygons – sub areas with an area greater than or equal to 100m^2

In addition to the spatial component of the source data, there is a temporal characteristic. This recognises that the aerial extent over which potentially contaminative activities occur may change through time. The concept adopted to deal with this is to associate an epoch with each sub-area recorded.

The epoch represents the time period during which base Ordnance Survey information was collected and provides an indication of when a potentially contaminative land-use existed. Historic Land-use data is usually supplied for 6 epochs that generally commence in the mid-1800's.

The Landmark methodology assigns a GIS Unique Reference Number (GISURN) to each sub-area recorded. This number persists through epochs for the same sub-area if the land-use remains constant. This means that it is possible to merge all of the geo-datasets for each dataset type and remove all occurrences of the object with the GISURN except the most recent, which gives an indication of the latest period when the potentially polluting activity was being undertaken. If the same area is subsequently used for some other activity, it will be represented as a sub-area with a new GISURN. The result of this is that there can be more than one object for the same geographic location. The above approach is straightforward for points and lines, however for polygons it is slightly more complex as the final extent of a land-use may not be its maximum extent. This is accommodated by identifying the epoch with the maximum polygon size and associating it with the attributes of the polygon from the most recent epoch.

Ddata from other suppliers is modified into the same format as the Landmark historic land-use dataset. All of the individual source datasets – point, line and polygon for each epoch - are merged into a single polygon, point and line dataset. Using the above criteria the number of objects is significantly reduced, but without affecting the integrity of the information. The term 'Special Three' has been adopted to differentiate between the original source datasets and the merged datasets that are used by the risk assessment algorithm. As new data becomes available the 'Special Three' can be re-created ensuring that the model is always up to date.

As such, the 'Special Three' represents the dataset for potential contaminant sources used in the model. To achieve this S57CT provides tools that treat the "Special Three" as a single data source.

Environmental Pathway Datasets

The following pathway datasets are used by the S57CT model:

- BGS 1:150,000 artificial geology dataset (i.e. made ground);
- BGS 1:50,000 drift geology dataset (natural soils);
- BGS 1:50,000 solid geology dataset (natural rocks); and
- Coal Authority mine entries.

The efficiency of the pathway is an estimation of the ability of liquid and gaseous to migrate and disperse in the underlying soils and strata. The scoring of the pathway factors was therefore based upon the transmission properties of the artificial, drift and solid geology characterised by the BGS datasets. The ability of mine shafts to transport gas and vapours effectively from mined seams was also recognised in the model and the proximity, or spatial coincidence, of such artificial migration pathways with potentially contaminative sub-areas was taken account of in the scoring of pathway efficiency.

The overall assessment of pathway efficiency was based on all four factors. Where any single component of the overall pathway acts as an effective barrier to contaminant transport this must be reflected in the total score. On this basis, the pathway scores are multiplied together, prior to normalisation, which allows a low score to be calculated to reflect the presence of an ineffective component in the overall geological pathway. Mine entries are considered as an additional form of artificial pathway and the scores for these and artificial geology were summed prior to the multiplication operation across the total pathway. The model does not take account of other localised pathways (e.g. culverts, pipes, etc) at the strategic inspection level. These features will be considered during the detailed inspection stage of the assessment.

Receptor Datasets

The following receptor datasets are used by the S57CT model:

- Human receptors
- Groundwater (vulnerability) receptors
- Controlled surface water receptors
- Ecological receptors, and
- Heritage receptors.

The Receptor Sensitivity score is formulated on the basis of an assessment of the spatial coincidence of the potentially contaminative sub-areas with the recognised receptor datasets previously described, as follows: (1) the most-sensitive human receptor type; (2) the most sensitive/vulnerable form of Groundwater Source Protection Zone (SPZ I, II or III) or aquifer (major, minor, non-aquifer); (3) the highest quality controlled surface watercourse (within 50m of the sub-area); (4) the most important designated ecological receptor; and (5) the presence of a Scheduled Ancient Monument. The overall Receptor Sensitivity score associated with a particular sub-area is calculated by a root mean square approach, which allows the total receptor score to be relatively high when only a single receptor has a high score weighting.

4 The Model

In order to ensure compliance with the Contaminated Land (England) Regulations 2000 (S.I. 2000/227) it was necessary to ensure that the assessment process is carried out consistently for the whole LA area – fulfilling the strategic review requirement.

The model itself is extremely flexible in terms of the datasets that can be utilised. The parameters of the model are recorded in three files, which store information on data requirements and behaviour associated with prioritisation, risk ranking and final numerical score assessments and PPC assignment. These also include the score and weightings associated with proximity and coincidence and the type of proximity analysis to carry out. The parameter files can be modified to suit the particular requirements of any LA. However, for the risk assessment to be reliable and consistent it is necessary to ensure that all datasets utilised are 'fit for purpose' and that the LA is content that they are an accurate representation of the required source, pathway or receptor information.

The model considers the spatial relationships between each sub-area in the ‘Special Three’ and all available pathway and receptor information. The spatial relationships considered usually involve either coincidence with, or proximity to, the pathway or receptor and the model carries out a sequential set of activities in a specified order – to systematically assess the potential Source-Pathway-Receptor linkages – and each activity results in a scoring or grouping value being assigned to each sub-area. Each of these activities causes an attribute in the ‘Special Three’ to be updated and these values are used to calculate the final

score for each sub-area. The overall list of scores for each sub-area can then be sorted to prioritise the follow up sub-area inspection strategy. During the modelling process each activity is carried out as a series of actions. Each action determines which of a specified range of values is applied. Actions are performed in a specific order beginning with the least sensitive and progressing to the most sensitive assessment , which will assign the highest score.

The result is that the 'Special Three' are given numerous attributes (one for each assessment activity) that represent all the stages of the modelling process. The final two attributes contain the final normalised risk-ranked score and the Preliminary Prioritisation Category which is based on the normalised risk-ranked score.

5 Using the Output from the Model - the LQMD

The primary objective of the BPRRS is to provide LAs with an internally consistent risk-ranked list of potentially contaminated sub-areas of land, which allows the LA to undertake detailed inspections in a order which is proportionate to the estimated level of risk. Initially, it is likely that LAs will need to collate all available information from planning and environmental records for each sub-area. To provide a repository for this information the S57CT package includes an MS Access database (The Land Quality Management Database – or LQMD), which is linked seamlessly to Arcview GIS. The LQMD stores a diverse range of information relating to each sub-area. An integral part of the LQMD is a suite of spreadsheets that, based on DoE Industry Profiles, provide a checklist for each of the contaminants that may be present for a particular industrial use and should be investigated. It is anticipated that LAs will use the LQMD to collate and reference information until sufficient data has been obtained on which to undertake a quantitative risk assessment to decide whether the sub-area should be determined as "Contaminated Land", as per the statutory definition.

6 Data Management Issues

A fundamental requisite of the legislation is that the data used in the strategic assessment is transparent and auditable. It is therefore necessary to record detailed metadata for each of the datasets used in the assessment process, such that at any time in the future it will be possible to identify the source, nature, accuracy, extent, intended use, copyright issues etc. of all data used by the model. The metadatabase serves this fundamental purpose and must be populated before the model, or any of the associated tools, can function. The risk assessment algorithm has been designed such that it will not function properly unless the appropriate metadata fields have been populated. The metadatabase is linked with the GIS, such that the user is able to query the metadatabase from the GIS for any selected dataset at any. The metadatabase is also linked to an automated data loader in S57CT, which enables datasets to be represented by a windows-style browser list.

7 Implementation Approach

Babtie work closely with LAs providing on-the-job training for LA staff such that on completion of the project (i.e. once the risk ranked table has been created and populating the LQMD has commenced) the LAs have the capability to assume long-term responsibility for management of the system.

8 Summary

The new Contaminated Land Regulations require every Local Authority to inspect their areas to determine whether any contaminated land exists. This must be undertaken in a logical, ordered and efficient manner using appropriate scientific and technical techniques. The Local Authority must not simply review those areas that it knows or currently suspects are contaminated, but it must strategically review the whole area to assess from historic uses and factual data whether it is likely that a contaminating use may previously have been present. Having identified potentially contaminated land it must then assess whether such land is causing, or has a significant potential to cause, significant harm to specified receptors; or that pollution of controlled waters is occurring, or is likely to occur.

The strategic review requires the assessment of a large volume of geospatial datasets on potential sources of contamination, pathways by which contaminants can be moved from the sources and the locations and

sensitivity of receptors. Geographical Information Systems provide a rapid, auditable and cost-efficient method for managing and analysing this data. Furthermore, following the strategic review the Local Authority will carry out detailed assessments of the high risk areas which will entail collecting more data on physical conditions, ownership, inspections, investigations as well tracking communications and regulatory actions concerned with individual sub-areas or sites. A Local Authority will find the stored data of use in the performance of its other functions, such as planning and asset management, hence the reporting functions of GIS will be invaluable.

The Babtie Section57CoreTools provides a system that simplifies and extends the functionality of ArcView GIS particularly to the assessment and management of potentially contaminated land. It has built-in tools that help overcome many of the difficulties in carrying out the strategic identification of land that is potentially contaminated and which may require further detailed inspection.