EXPERT SUPPORT SYSTEM (ESS) FOR DATA ACQUISITION IN ENVIRONMENTAL RISK ASSESSMENT OF CONTAMINATED SITES

Ammar Abbachi[1]

1: Assistant Professor ABHA College of Technology

abbachi@yahoo.com

ABSTRACT

Environmental Impact Assessment of suspected contaminated sites requires the use of quatitative methods based on factual evidences and scientific methods to derive robust conclusions and to plan effective metigation and risk management solutions.Therefore, designing sampling strategies for site investigation is an important input to the process of conducting a quantitative risk assessment of contaminated sites. These sampling strategies require all available information to be taken into account including prior information derived from desk studies and walkover surveys.

This paper describes an expert support system for helping site assessors to compile this prior information and to develop an initial hypothesis on the likely locations of hotspots (as source of contamination) on contaminated sites. Using a knowledge-base approach the system converts desk study information into a score of indicators in order to build an a prior probability map of hotspots locations. The total number of sample locations is then computed and distributed over the site to reflect the prior information and the likelihood of hotspots locations.

A case study is presented in which a desk study report and site walkover survey were used as source of prior information. Using the expert support system, a series of spatial sampling strategies were worked out for different hotspots sizes to assist in designing subsequent site investigation phases as part of an environmental and human health risk assessment process.

Keywords: *Contaminated land, Risk assessment, Site investigation, Sampling, Expert Judgement.*

الملخص

إن التقييم البيئي للمساحات التابعة للمناطق الصناعية والمعتقد أنها مصدر تلاوث بيئي تكون في حاجة إلى طـرق تقيـيم خطر هذا التلاوث يعتمد فيها على بيانات مادية حقيقية و على منهجية علمية ترسم معالم حلول موضوعية، تساعد علـى وضع خطط سليمة لمعالجة خطر هذا التلاوث على البيئة في هذه الناطق. لذلك يعد وضع استراتيجية واضحة للحصول على عيينات ميدانية من أهم العناصر في عملية التقييم الكمي لخطر هذا التلاوث. و تكمن قوة هذه الاستراتيجية في جمع

كل المعلومات و البيانات المتاحة و ذات الصلة بالنشاط الصناعي في هذه المناطق و منها خاصة التقارير المعدة مسبقا و الزيارات الميدانية و محاولة استغلالها في تكوين فكرة عن مواقع انتشار الملوثات في المنطقة. يقدم هذا البحث نظام حاسب آلي لمساعدة و تدعيم خبراء التقييم لهذه المساحات الملوثة لجمع معلومات أولية وبناء فكرة عن الأماكن التابعـة لها و الأكثر احتمالا أن تكون مواقع لمصادر التلوث. وذلك بتطوير قاعدة بيانات تشمل المعلومات الأوليـة و تجربـة خبراء التقييم و ملاحظاتهم (Knowledge-Base) حول المنطقة وتحويل هذه البيانات إلى مؤشرات تساعد على رسم خريطة احتمالات Probability map يشار فيها إلى الأماكن التي من المحتمل أن تكون مصدر تلوث. فيقـوم فريـق التقييم باستعمال هذا النظام لحساب عدد العينات اللازمة وتوزيعها لتشمل الأماكن المشار إليها في خريطة الاحتمـالات التي تم تطويرها مسبقا. كما يقدم البحث عرضا تطبيقيا لهذه الطريقة، بحيث تمت دراسة تقارير أولية و زيارات ميدانية لمنطقة صناعية نتج عنها تطوير خريطة احتمالات أولية عن مواقع يعتقد أنها مصادر تلوث بيئي. واسـتعمل النظـام المقدم في هذا البحث لرسم عدة استراتيجيات توضح عدد العينات و مواقعها في المنطقة. و تختلف هذه الاستراتيجيات باختلاف حجم مصدر التلوث Hotspot Size المحدد من طرف خبراء التقييم وذلك فـي المراحـل الأوليـة لجمـع المعلومات اللازمة و الكافية لتقييم خطر التلوث في هذه المنطقة على البيئة و صحة الإنسان.

1. INTRODUCTION

Pressure to develop new industrial facilities and to assess the environmental risk of existing ones has increased the need for better site investigation tools. An important component of environmental risk assessment is determining the nature, extent and significance of air, soil and groundwater contamination, and how to manage it. Sufficient information of acceptable quality is required for defensible decisions based on the likely position of contamination hotspots and the estimated spatial distribution of contaminants.

The use of an Expert Support System (ESS), developed to design and streamline site investigation planning and increase the likelihood of collecting appropriate data cost-effectively, is an important step in environmental risk assessment of suspected contaminated land. The ESS presented here is planned to be developed on several modules to design sampling strategies for detection of soil, air and groundwater hotspots as source(s) of contamination on industrial sites. In this paper a module related to soil contamination is presented in details.

The system aggregates information relevant to possible soil contamination and thus indicates the areas most likely to contain hotspots. The indicators are then used to calculate prior probabilities and hence to design spatial sampling strategies.

A case study illustrating the use of ESS to design a first stage sampling strategy is presented where a Desk Study Report [ICC, 1995] and site walkover survey were used to compile a list of weighted indicators of likely contamination. The user enters this preliminary information in

Top

to ESS in order to help develop an initial hypothesis on the location of suspected hotspot. Then, a cost effective spatial sampling strategy is designed using this initial hypothesis. The number of samples required is expressed as a function of hotspot size, while sampling locations reflect the spatial distribution of the weighted indicators.

2. USING PRIOR INFORMATION

Regular sampling strategies designs, which assume that all parts of a site have an equal probability of containing a hotspot can lead to large numbers of sampling points [Ferguson, 1992]. More cost-effective sampling strategies can be devised when there are grounds for relaxing the equiprobability assumption concentrating investigative efforts in areas suspected of being contaminated. The two main approaches for improving the efficiency of sampling designs are:

* Variable density sampling using weight-of-evidence scores
* Multi-stage sampling

Combinations of both approaches may be used when appropriate. In certain cases on-site analytical methods require multi-stage sampling.

Experienced assessors often suspect that some parts of a site are more likely to contain hotspots than others. They may then wish to design a sampling strategy that reflects their degree of suspicion or strength of belief about the targets likely location. One approach to this type of sampling design is to partition the site into subareas (say on a 1 to 10 scale) to reflect strength of belief as to where are targets likely to be.

Figure 1 shows an example of this type of scoring , and corresponding sampling plan . The scoring scale is arbitrary, where the highest score being used as a normalising density. The most favoured subarea is calculated to give 0.95 probability of hitting a target if it exists in that subarea [Ferguson and Abbachi, 1993].

If the assessor's judgment was correct the 0.95 probability of success is thus achieved with fewer sampling points and lower cost. But if the assessor is wrong the penalty is a reduced probability of success in locating the target (see probability values adjacent to figure 1B).

Top

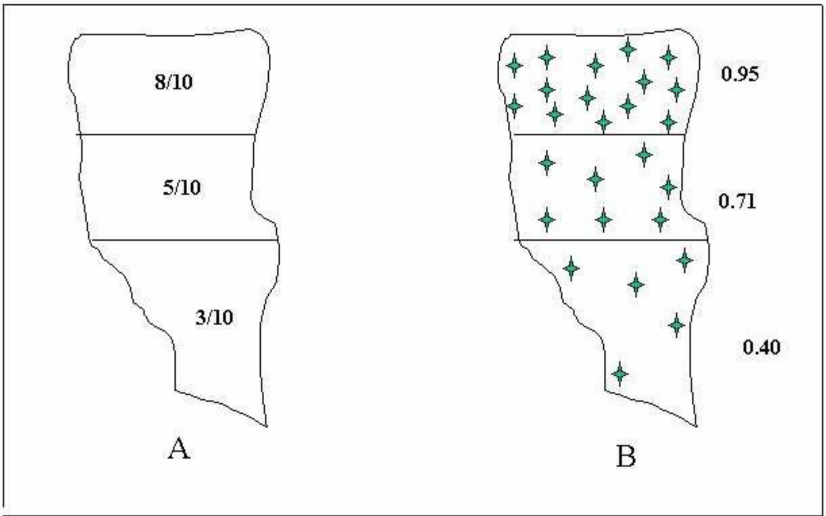


Figure 1 Variable density sampling for different strengths of belief.

The problem with this approach is that it requires the site investigator to convert a variety of disparate information into a score reflecting his or her strength of belief about hotspot location. To overcome this difficulty, a computer-bas Expert Support System (ESS) has been designed so that specific items of information derived from a review of site history and from a preliminary walkover survey can be used directly to optimise sampling designs.

3. THE EXPERT SUPPORT SYSTEM

ESS is a decision support system for the design of sampling strategies for contaminated sites. Eventually, the system will comprise the following linked modules:

* Likely contaminants
* Location of contaminant hotspots
* Soil gas survey
* Groundwater sampling
* Data analysis of first-stage sampling results
* Design of second-stage sampling

In the current version only the Hotspot location module is fully developed for first stage sampling of a site investigation. ESS Hotspot module is intended to help site investigators to design the initial, or first stage, sampling by developing an optimised sampling pattern for detecting (though not delineating) hotspots, and provide a statistical justification for the chosen strategy.

The extent of site coverage by sampling will depend on the density of sample locations. A compromise between total coverage and cost must therefore be achieved, with sampling becoming a statistical exercise to maximise useful information at a cost as low as possible consistent with the objectives of site investigation. When the objective is to locate hotspots, it is sufficient to place just one sampling point on the area covered by a hotspot. Therefore, if a circle equal to the radius of the putative hotspot is drawn around a sampling point, the circle can be thought of as the zone of coverage for that sampling point.

The total sampling coverage is the sum of all zones of coverage excluding overlaps. The ratio of the total coverage to the total site area can also be thought of as the probability of hitting a hotspot if it exists. ESS is developed around this concept of coverage and hit probability. The computational part calculates an optimum sampling coverage, covering areas with the strongest prior evidence for the existence of a hotspot.

Before running ESS Hotspot, the user should have already completed a preliminary survey of the site and have compiled:

(i) data on the historic and current use of the site

(ii) a record of any visual, or other indications of potential contamination gained during a site walkover [Ferguson and Abbachi, 1994].

The user may additionally have some preliminary chemical analysis data from previous investigations or from ad hoc sampling of the site.

Archive data on previous site use (chemicals and processes) should when possible, relate to the length of time in use time since last use, chemicals present, quantities of chemicals handled and the properties of these chemicals, specifically toxicity, biodegradation potential and leaching potential. It is recognised that these data are often very difficult to obtain precisely. In ESS high precision is not required and the user may simply make a best judgment according to a 3-point scale high, low or unknown.

It is helpful when conducting a walkover survey to subdivide the sit into a number of square cells to compile the visible or other indicators for each cell. These cells represent an information grid (see Figures 3 and 4) into which all prior information is aggregated. Users must specify the size of the information cells, balancing spatial resolution with the time required to input all the information cell by cell. It is also very prudent to avoid to specify large number of grid cells as this rarely leads to a significantly improved solution but substantially increases data entry and computation time.

These total scores can be viewed on the screen (Figure 6) in the form of a map. A high score indicates a high a-priori probability (i.e. strong evidence that a hotspot exists) and therefore the need for a relatively high sampling density in order to locate it with confidence. Attribute scores are converted to a priori probabilities by the user specifying the probabilities he or she thinks most appropriate for the highest and lowest scoring information cells; intermediate scoring cells are scored proportionately. Default values are provided in ESS as initial guesses.

3.1. Estimating Local Sampling Densities.

Analysis of the prior information provides a spatially distributed set of a priori probabilities that a hotspot exits within specified subareas of the site defined by the information grid cells. The following approach is used to convert the a priori probability assigned to each information cell (grid square) into a target number of samples for that grid square.

The primary motivation for a sampling scheme is to ask; "what is the probability of locating a hotspot if it exists?" if, however, the sampling scheme fails to locate a hotspot, the question then becomes what is now the probability that a hotspot exists given that the sampling scheme has failed to find one? This (after the event) probability is termed the a posteriori probability. The probability of locating the hotspot, if it exists can be considered as the hit probability, which is equal to the sampling coverage as discussed above.

A bayesian probability approach has been advocated to relate the above probabilities [Gilbert, 1987] and [Ferguson and Abbachi, 1993]

$$PrH_i = 1 - PrA_i\,(1 - P_i)/(P_i(1 - PrA_i))$$

Where PrH_i , is the hit probability for grid square i, P_i the a priori probability and PrA_i the a posteriori probability. By setting one of the sampling objectives (i.e. PrH_i or PrA_i) to a fixed target value, the above equation can be used to compute the other probability . We generally set all PrA_i to 0.05, i.e. if, after sampling, a hotspot has not been located in any given square, there is 95% confidence that a hotspot does not exist within that square. The value 0.05 is not prescriptive although it provides a confidence level with which most users seem to feel comfortable.

The hit probabilities PrH_i for each grid square are used to calculate the target number of samples $N_T=\Sigma N_i$ for the whole site, where N_i is the nominal number of samples allocated for grid square i. The problem is how to distribute the total number of samples N_T over the site such that each individual requirement on N_i is satisfied.

3.2 Optimisation of Sample Locations

The given number of samples, N_T needs to be placed such that their aggregate weighted coverage is maximised. Therefore, the problem is to minimise the aggregate a priori probability score of all parts of the site that falls outside the zones of coverage of the sampling points . This is a well defined optimisation problem that can be solved using the Quasi-Newton method [Fletcher, 1987]. However, this is a local optimisation procedure and in practice, a good initial estimate was found to be essential.

The initial estimate is found using an approximate sequential placement algorithm. For computational convenience, each information cell is subdivided into a fine grid typically comprising 3*3 or 4*4 smaller grid squares. The centres of the fine grid squares define the set of possible sample locations in the approximation. The a priori probability scores of all the fine grid squares are first placed in rank order. When fine grid squares have the same score, a possible sample location whose zone of coverage lies wholly in a high probability subarea is ranked higher than one whose zone of coverage overlaps into an adjacent lower probability subarea. More generally the rank order is based on the average a priori probability over the whole zone of coverage rather than the probability at the sample placement point. Any remaining ties in rank order are broken arbitrarily. In practice, the sequential solution usually performs almost as well as the optimised solution, which typically improves the overall weighted sampling coverage by less than 5%.

4. CASE STUDY

The initial objective in the case study was to design a preliminary sampling strategy to locate suspected hotspot and to provide an overall picture of the spatial distribution of soil contaminants within a former gas work planned to be developed for other end uses.

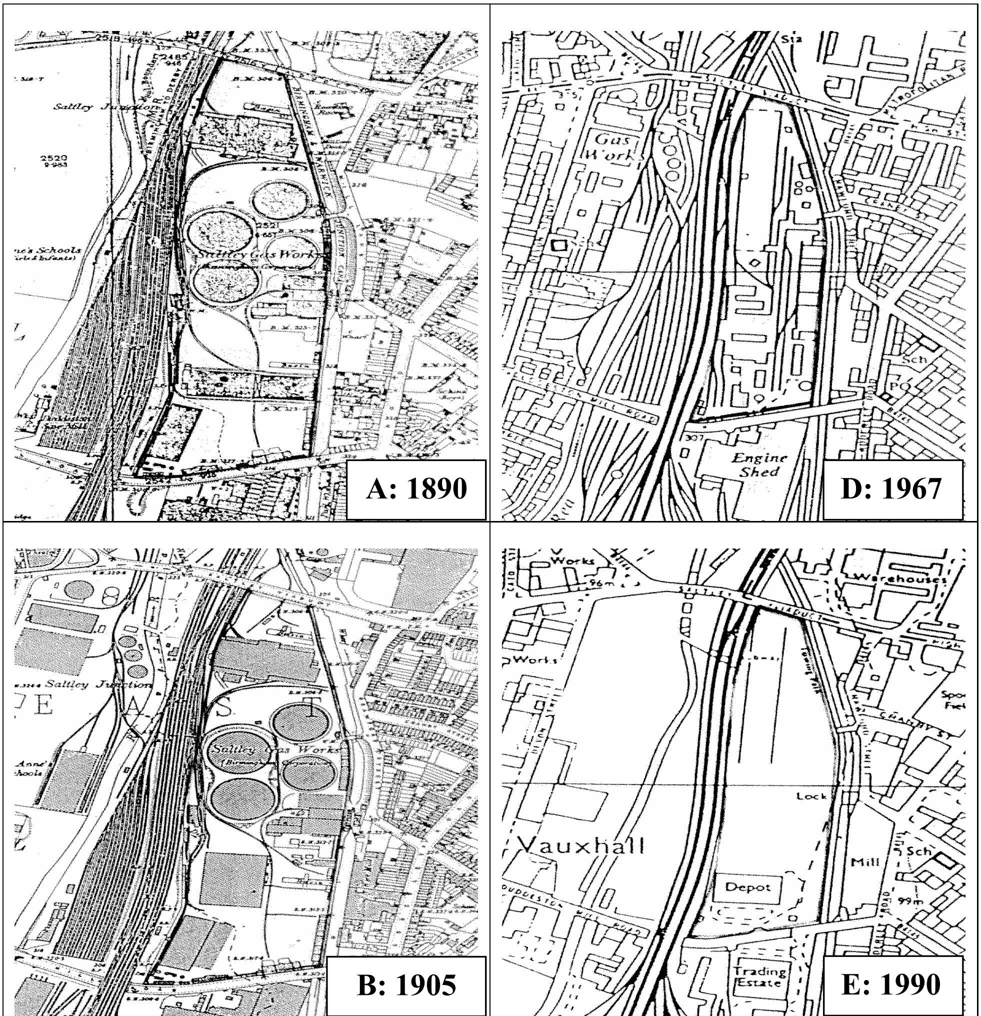
The site used to illustrate ESS extends over approximately 4ha and has been the location of several past industrial activities. Figure 2 shows survey maps of the site at different periods.

The presence of a gas works in the central area (from 1890 onwards) regarded as a subarea with high potential for containing contaminant hotspots. Loading and off loading areas are not accurately recorded and could be anywhere next to the railway tracks to the west or the road to the south of the site. The canal junction to the east of the site may also have been used as a loading and off loading area; waterways were in use until the middle of this century.

The following are the historical attributes used:

- Process areas
- Storage areas of raw materials
- Waste disposal areas
- Loading and off-loading areas
- Filled areas

Top



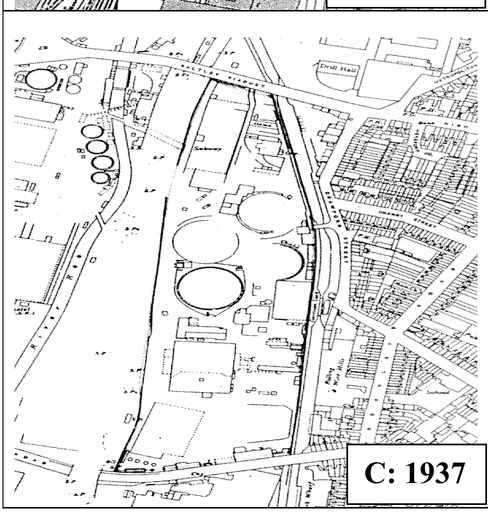


Figure 2 Different Ordnance Survey maps of the site with gas works in central area (A: 1890 – B: 1905), the site abandoned after closure of gas works (C: 1937) and gas works replaced with new infrastructures (D: 1967), the infrastructures cleared later (E:1990)

Top

However, a site walkover indicated the existence of the following attributes:

- Irregular surface
- Poor drainage
- Anomalous soil type
- Oily patches
- Bare areas with sparse vegetation
- Remains of site infrastructure
- Waste tips

The scanned map of the site was imported in to the ESS and divided in to sixty-six 25* 25m information cells (Figure3). Cells outside the site have been eliminated as shown in Figure 4, allowing the sampling plan to cover the site area only. Although a greater number of information cells will result in higher resolution. It will considerably increase the amount of time required to input all the information.

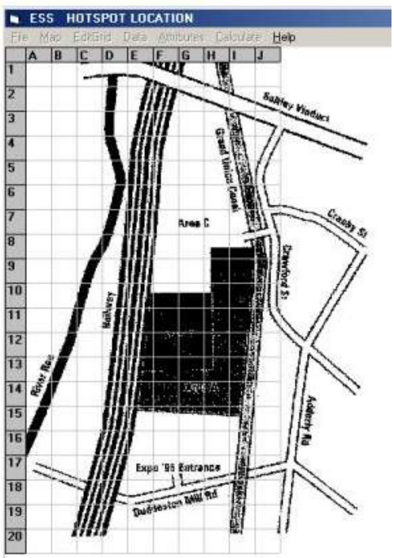


Figure 3: Scanned map of the site

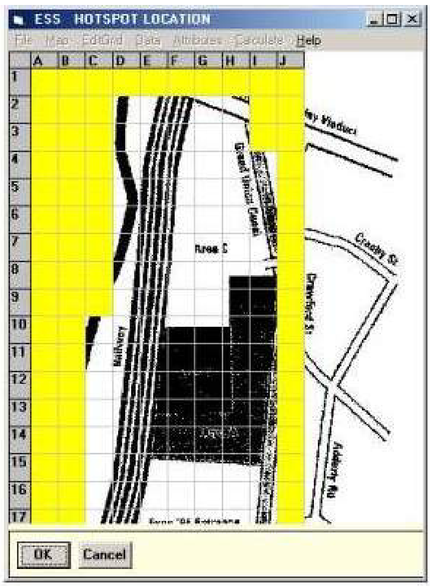
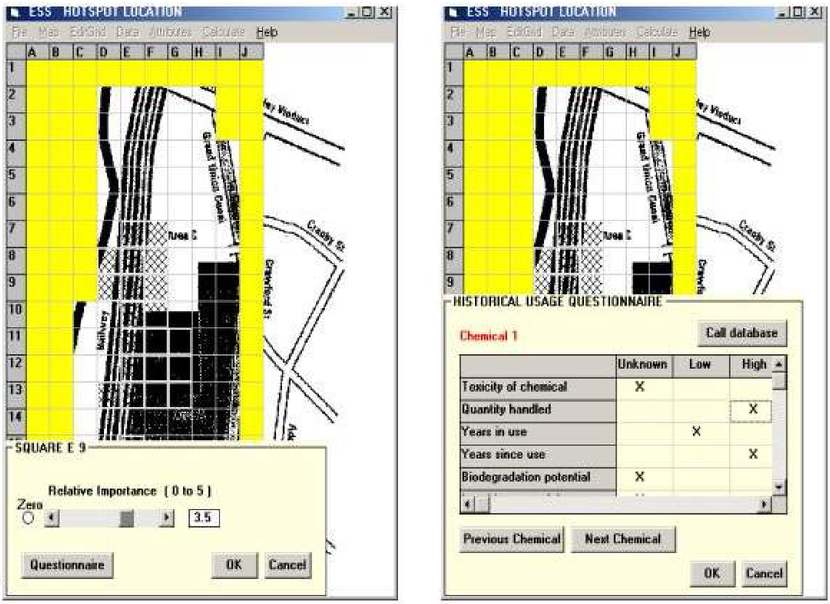


Figure 4: Information cells outside the site eliminated

Figure 5 shows two ways of allocating attributes to information cells and working out their influence on the final score. Figure 5A shows that user judgement on the importance of an attribute can be set using a sliding scale to specify a score in the nominal range 0-5.

Top

Alternatively, Figure 5B shows a menu which allows site assessors to respond to simple questions; the answers are then used to work out a score for the designation information cell. The user can overrule the knowledge base if specific data on items such as quantity handled, years in use, years since last use and leaching potential, are available.



A B

Figure 5 Attributes allocated and different scoring systems (A: Expert Judgement. B: Using Knowledge Base)

The hotspot size is made variable to study its impact on the number of sample locations required. The hotspot size assumed of circular shape is expressed as a percentage of the total site area. The score shown in form of a map in Figure 6 are then reviewed and crosschecked with the assessor's strength of belief on level of contamination in different part of the site. The central area is most suspected of being contaminated and was the location of a gas works for many years (see Figure 2). The far north east part of the site is also of high a priori probability; it is suspected to be a site where waste material was deposited.

The sampling strategies worked out using ESS allow the number of sample locations to be expressed as a function of hotspot size as shown in Figure 7. An optimum sampling strategy could be selected if the target hotspot size and other economic parameters of the site investigation were known.

The spatial distribution of the samples is characterised by higher sampling density in the central areas and the north east part of the site (Figure 8) reflecting the higher priori probabilities as described above.

Using ESS, the assessor is able to compare sampling strategies designed on the basis of prior information with strategies based on the equiprobable assumption [Ferguson, 1992]. The reduction in total number of sample locations is apparent especially at smaller hotspot sizes (Figure 7) for this site roughly a 30% reduction in total number of sample locations is achieved. Of course this reduction is obtained at the expense of relaxing the hit probability in some areas of the site. Site investigators should make there own judgement as to whether priori information gathered on a site is robust enough to give confidence in the sampling strategy adopted.

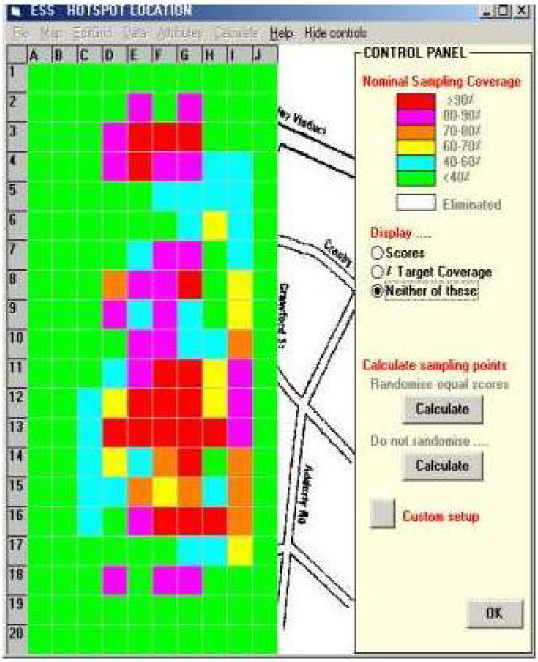


Figure 6: Nominal Scoring Coverage
(high scores in central areas)

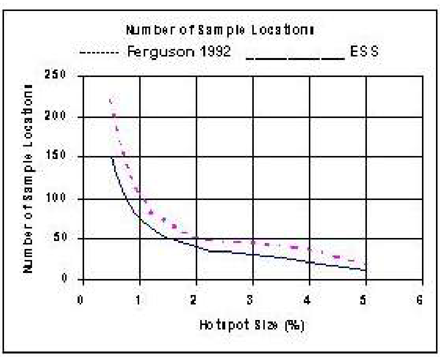


Figure 7 Number of sample locations as a function of hotspot size, expressed as a percentage of the total site area

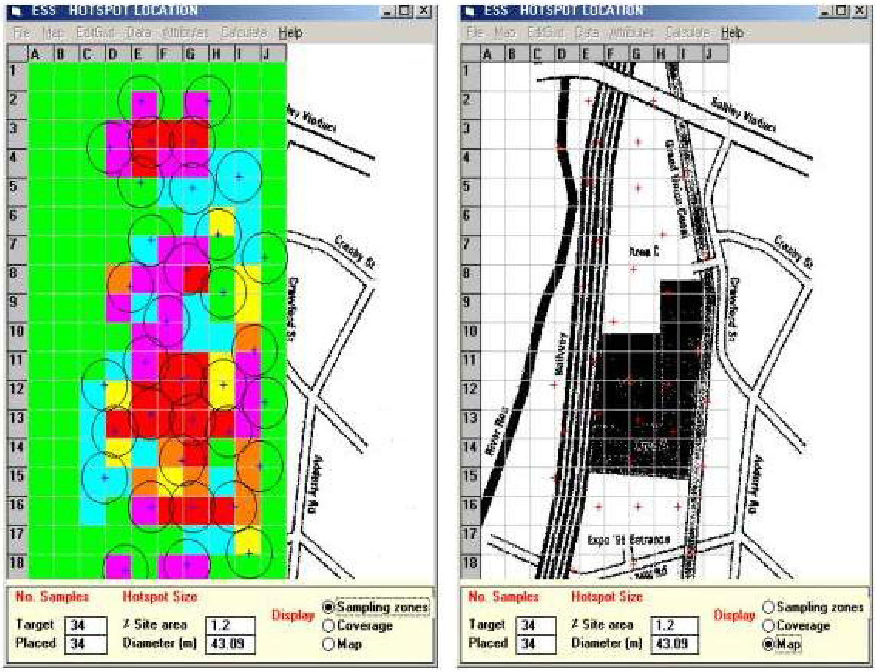


Figure 8 First Stage sampling strategy with clustering in central area (A: sampling coverage B: Sampling locations on the map)

5. CONCLUSION

The system developed here provides a standardised framework for sampling design. It uses readily available prior information acquired during a desk study and walkover survey. It should be noted that many practitioners fail to use data in sampling design (although the data have been collected at some expense) and recourse to a regular grid pattern is the norm for site sampling. Sampling and analysis costs may amount to high figures for each sample taken. So, minimisation of samples numbers is important, the subsequent costs of missing a contaminant hotspot through insufficient sampling could however be orders of magnitude higher. It is necessary, therefore, to find a balance between minimising sample numbers and maintaining an acceptable statistical confidence. The system provides a pioneering methodology for helping to achieve that balance.

The system is centred on expert knowledge. The knowledge has been structured as a series of numerical coefficients. As knowledge increases through a better understanding of contamination indicators, it is expected that these coefficients can be improved.

An overall evaluation of system performance is difficult to undertake. One criterion would be to compare system predictions with those made by experts. An objective criterion could be the relative success in the long term at hitting hotspot targets found on investigated sites. It must be borne in mind, however, that there is no definitive answer to the sampling design problem .

ACKNOWLEDGEMENT

Paper based on the research project funded by the Department of Environment UK. The author would like to thank the ICC Group for making the ordnance maps public.

REFERENCES

1. Department of the Environment, 1993, "Guidance on Preliminary Site Inspection of Contaminated Land". CLR report No2. Department of the Environment London.
2. Ferguson, C. C., 1992, "The Statistical Basis for Spatial Sampling of Contaminated Land". *Ground Engineering* 25/5, 34-38
3. Ferguson, C.C and Abbachi, A., 1993, "Incorporating Expert judgement into Statistical Sampling Designs for Contaminated Sites". *Land Contamination & Reclamation* 1, 135-142
4. Ferguson, C.C. and Abbachi, A., 1994, Department of the Environment: "Sampling Strategies for Contaminated Land". Report No4, Department of the Environment, London
5. Fletcher, R.O. 1987, Practical Methods of Optimisation. Academic Press London.
6. Gilbert, R.O. 1987, Statistical Methods for Environmental Pollution Monitoring. Van Nostrand Reinhold, New York .
7. ICC [ICC information Group ltd], 1995, Search for Contaminative Uses Report at EXPO 95 SITE. Available from ICC information Group ltd, London.

Top