

Application of data assimilation techniques to post accident management, assessment and presentation of uncertain environmental and radionuclide data

N.A. Higgins and N. Pérot

*National Radiological Protection Board, Environmental Assessments Department,
Chilton, Didcot OX11 0RQ, U.K.*

Abstract. A major collaboration between IPSN and NRPB has recently commenced with the intention of developing and implementing practical data assimilation techniques. This joint work builds on the CITRAME and SECTAR projects of IPSN and NRPB respectively. The tools developed through the collaboration will assist decision-makers after an accident to protect the public more effectively. They will clarify the state of knowledge available through the display of map based estimates of the likely environmental concentrations of radioactivity and the uncertainties in those concentrations. Considerable attention will be paid to the development of risk maps that show clearly the current uncertainties in the estimated concentrations in a range of foods and the predicted uncertainties at different times arising from the migration of radioactivity through the food chain. The estimates of radioactivity concentration and its temporal development in different foods, together with the associated uncertainties, will change as more measurement information becomes available. The maps will therefore help decision-makers to allocate resources more effectively by concentrating measurement effort in areas of high uncertainty. A discussion of the role these techniques might play in the management and selection of intervention strategies is given together with an initial progress report on their development.

1. INTRODUCTION

The assessment of the radiological and environmental consequences of an accidental release of radioactivity is particularly difficult in the period immediately following the release. At this time actions are being implemented to protect the local population and food production, and longer-term measures are being considered. These considerations are of necessity occurring under conditions of relative ignorance. Large measurement programmes will be underway following the accidental release but these inevitably take time to build up a comprehensive picture of the contamination that has arisen from the release. It is therefore important to make the best use of the information that is available and to recognise the uncertainty in any predictions made on the basis of that information. To this end the collaboration will explore three related aspects of the problem of presenting decision-makers with the information they require in an effective way.

Thus the collaboration aims to:

- Establish a realistic estimate of the spatial distribution and variability of the contamination based on the measurement information available and any supporting information that may be applicable.
- Project the contamination estimate and the uncertainty associated with it into the future and into other foods, through the use of environmental transfer modelling.
- Develop graphical displays that allow the spatial distribution of contamination and uncertainty in the exact level of contamination in a range of foods and environmental materials to be effectively presented to decision-makers.

The culmination of these three tasks will allow decision-makers to make difficult decisions more effectively. They will be able to take the most appropriate action to mitigate the consequences of an

accident in the knowledge that they have as comprehensive a view as possible under the circumstances of both the contamination and the associated uncertainties.

The work to be undertaken to realise these three objectives is discussed in turn below.

2 SPATIAL PREDICTION

Data assimilation is the term generally adopted to describe a merging of measurement data and model prediction through the use of statistical techniques. The result is a prediction that is likely to be a more accurate representation of the distribution of contamination in the environment after an accident than could be achieved by the use of either limited monitoring results or a model alone. This task will bring together work carried out by NRPB on data assimilation under the SECTAR programme [1] with future developments by both IPSN and NRPB.

The main impetus for this work is to develop practical tools, based on the experience gained of data assimilation in this and previous work. Additionally, new techniques will be investigated and, as more data becomes available, previous techniques revisited to check on the suitability of different techniques under a variety of conditions.

2.1 Geostatistics

Geostatistics is the analysis of measurement data taken at different locations whose values are likely to be correlated by an amount that is related to their proximity to one another. At its simplest the technique does not use any other information than the measured values themselves. It can therefore be considered as an approach that is entirely complementary to the use of a simple model that takes no or little account of measured values. However, the techniques of geostatistics also offer ways of combining other information, including information from models, with direct measurements of the quantity of interest.

The main contributor to improved deposition prediction is the ability to use the information gained from measurements effectively. At the very early stages of an accidental release simple geostatistics will have to work with very little data. This is not practical, but as shown in [1], when other information is used to aid an analysis then the techniques of geostatistics may be applied in circumstances when they would be inappropriate in isolation.

The techniques range from optimal linear interpolation to the Monte Carlo simulation of unknown values that are consistent with the spatial correlation structure of the known data. This latter option can, for example, be used to report the estimate value at each location that has a given probability of being greater than the true value. This technique can be used to provide some of the display information discussed in Section 3.

2.2 Bayesian Methods

An alternative technique, using Bayesian analysis, that will work from the start of an event has been formulated under SECTAR [1]. In this case the prior expectations of critical model parameters are updated through the influence of measurements. However, the methods applied in the SECTAR project go further than adjusting parameter values in a calibration process and include consideration of model inadequacy [2]. Model inadequacy is defined to be the difference between the true values at a particular location and the prediction of the process model (in this case an atmospheric dispersion model) at that location when the best parameter values are used to evaluate the process model. The best values of the model parameters are initially unknown but may be solved for approximately by calibrating the process model. The inadequacy function is modelled as a Gaussian process with the correlation between the inadequacy function at different locations assumed to be represented by one of several simple alternative forms. The choice of the covariance function is important, as this represents beliefs about the smoothness properties of the deposition. It defines the nature of the local influence on the observed data points and acts in a way analogous to the semi-variogram used in geostatistics. In the case of the *product Gaussian* correlation form, for example, the constant coefficients of the correlation function are parameters of the Bayesian model, whose values are estimated from the data. The product form allows distinct correlations

in the x and y directions to be represented, for example, to characterise differences in the extent of correlation in the downwind and crosswind directions following deposition from the atmosphere. If the measurements indicate that the process model provides a very poor representation of the phenomena under study its contribution to the final estimate will be very small and the technique will produce a form of kriging estimate through the action of the inadequacy term [3]. The ability of a technique to handle inadequacy is particularly important when models are used in the initial assessment of accident consequences. In this case only simple models that require only the small amount of input data available can be used. These models reproduce, for example, the general trend of atmospheric dispersion over a short distance but quickly run into problems with changes in wind direction, rainfall, complex terrain etc.

3 FOOD CHAIN MODELLING WITH UNCERTAINTY

Uncertainty analysis studies have been undertaken in many areas of radiation protection. They have considered the range of outcomes possible from a model by varying model parameters, or more strictly the measurable quantities used in the derivation of model parameters, over their feasible ranges. These studies have been applied as part of a probabilistic risk assessment of accident consequences [4], in dose reconstruction studies [5] or to assess the uncertainty in doses from routine releases [6]. These analyses are difficult and time consuming to undertake. Particular problems arise when parameter ranges are obtained through the inherently difficult procedure of expert elicitation when appropriate or sufficient experimental data are not available. These difficulties are likely to be compounded after an accident when any change of the prearranged default local parameter ranges will need to be decided quickly. Even if only a few changes were required to parameter ranges etc. to prepare for an event specific uncertainty analysis, the analysis could still take considerable computing effort and hence time to complete. Such time is simply not available shortly after an accident.

The potential benefits of undertaking uncertainty analysis after a nuclear accident are considerable. It immediately recasts the decision required of a decision-maker from deciding the best course of action to take based on a single best estimate of the consequences to one that recognises that the best estimate might not be the realised result. The best response in the light of uncertainty will be one where the decision-maker takes account of the likelihood and consequences of the situation being much better or worse than the best estimate would indicate. It may even enable the decision-maker to wait until more information is obtained and the uncertainty reduced before introducing disruptive countermeasures.

Some uncertainty information will be available to the decision-maker through the use of the statistical techniques of SECTAR. This information is of immediate value but when used in conjunction with a food-chain model will only provide a partial measure of the total uncertainty in the predicted radioactivity concentration in food from the affected area.

This project will investigate techniques to provide the decision-maker with a more complete view of the uncertainty. This will enable a prediction of the radioactivity concentration in food to be produced that includes both the uncertainty in the amounts deposited and intercepted and the modelling uncertainty that results from estimating the migration and uptake of radioactivity as it moves through the environment. A programme of work is already established to look at estimating the uncertainty in the radioactive input [1], what must be established are the potential ways of advancing the uncertainty modelling of the food chain. There are several aspects and approaches that can be considered. Food chain models are generally represented as a series of coupled first order linear differential equations describing the fluxes of radioactivity concentration between the different parts of the environment. The simpler this representation is the fewer parameters there will be to consider uncertain and the greater the likelihood that the problem will be tractable in a short period of time. However, a simple process model will generally lead to a poorer representation of the radionuclide transfers taking place. This route has been applied in the past and has the 'advantage' of removing parameter correlations that are very difficult to handle in the more complete model by amalgamating processes and introducing parameters that are less physically based. However, it is likely to be more difficult to estimate the uncertainty of the hybrid parameters in the simpler model. An alternative route to simplification is through simulation or emulation where a complex model is evaluated at a range of points in the parameter space and an interpolating function fitted to enable estimates to be made at any intermediate selection of parameter values. This function is likely to be much quicker to run on a computer but the approach has the added uncertainty inherent in results

estimated by interpolation. This process is analogous to the geostatistical simulation discussed in section 2, except that in the geostatistical case reality, in the form of measurements, has taken the place of a complex model.

Statisticians have recently referred to this general area of study as the Statistical Analysis of Computer Code Outputs (SACCO). It is work arising out of SACCO that offers yet a further opportunity to improve the speed and efficiency of uncertainty analysis by replacing the computationally time consuming Monte Carlo analysis with a much less time consuming Bayesian approach [2]. The Bayesian approach is faster because many fewer function evaluations are required to produce a cumulative distribution function e.g. approximately 10 in comparison to the several hundred required by a standard Monte Carlo approach [7].

4 DISPLAY OF UNCERTAIN INFORMATION

It has already been indicated that providing a decision-maker with information on the likely uncertainty of estimates of current contamination and future radionuclide concentrations in crops and the environment could be of great help. However, it is clear that if care is not exercised in the way that information is presented to a decision-maker then the result could be increased confusion and a loss of effectiveness in the decision making process. This project will therefore consider what aids can be used to provide a clear impression of the situation as currently viewed and the range of forecasts that stem from that incomplete picture.

Work is at a preliminary stage in the design of graphic representations of uncertainty. Generally, when trying to communicate uncertainty, there are two aspects that the decision-makers need to be able to visualise simultaneously. Firstly, some indication of where, over the possible range of the variable, representative values are located, typified by the mean, the median, the mode or some other measure that represents the central tendency or location of the likely values. The second is the spread of the distribution, indicating the likelihood of values that are much greater or less than the central tendency, represented by the standard deviation, variance or inter-quartile range for example. Both aspects vary spatially, and it is notoriously difficult to display any two spatially varying attributes on the same map. There are other related attributes that may be of interest, for example, the degree of uncertainty about particular percentiles, the uncertainty about the temporal extent of any imposed restrictions and how the forecast uncertainty in say, activity concentration, is predicted to change with time.

There are five general approaches that could be adopted to display the information.

- Different symbology could be used on the same map. There are many combinations that could be used. For example: two different line styles for two sets of contours, a set of contours overlaying a varying colour set, a varying colour set mixed with a varying cross-hatching style, spot values of one field overlaying a contour set of the other etc. Whichever is chosen, it is difficult to avoid being confusing, particularly if the displays are generated automatically.
- Orthographic display of one variable draped over the other. Again this can be confusing and hard to interpret.
- The simplest approach is probably the display of two maps next to each other representing, for example, the spatial variation in the mean and variance respectively. However, in this case there is perhaps a greater potential for the degree of uncertainty to be ignored by the decision-maker. It is also potentially difficult to marry corresponding features on two maps, that to avoid clutter display little other information to guide the eye.
- Computer animation would be complex to implement, possibly confusing and not useful for inclusion in reports, communicating remotely etc.
- The combination of the two fields into a single meaningful value.

The last approach requires more discussion. Uncertainty was described above as having two aspects; location or central tendency and spread. This is the classical way of describing distributions and is a convenient shorthand that captures the essence of a distribution without enumerating each possible outcome and its associated probability. However, this approach may not provide the best way of communicating information to decision-makers on aspects of the distribution. The most important aspects of a contaminant distribution from the perspective of a decision-maker may be the extreme percentiles. If the decision-maker's aim is to be conservative then rather than displaying the central tendency it may be more appropriate to display estimates at each location that are only likely to be exceeded 5% of the time, i.e. the 95-percentile of the distribution. Alternatively, a particular value could be selected and the probability of exceeding that value could be displayed as contours or an overlay of different coloured regions. The second approach is particularly useful when there are action levels, e.g. the degree of contamination that triggers certain countermeasures, such as food restrictions. Figure 1 shows an example of this form of display produced by statistical simulation. The calculation used a sample of 40 measurements of Chernobyl data in Nobozibkov district of Bryansk in Russia to indicate likely areas for restriction that are broadly consistent with the restriction areas implied by the use of all the data and the same notional criterion.

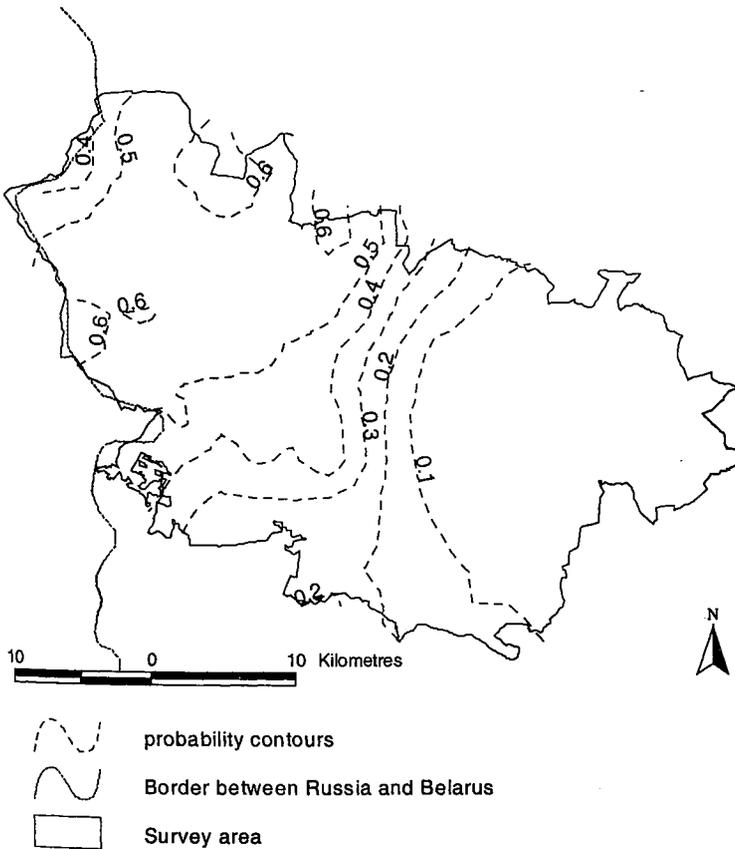


Figure 1 Probability of exceeding a notional restriction threshold

The simple and more complex displays of uncertainty discussed above will help decision-makers. Additional, yet more complex, graphical representations may also be of use to decision-makers, and will certainly help technical managers provide the decision-makers with more reliable results. For example, the uncertainty displayed could be partitioned into its various contributory parts [8]. This sensitivity analysis perspective would offer the decision-maker or technical manager an insight into the source of the uncertainty. It would also highlight instances where it was possible to reduce the uncertainty before a

decision was required. Lack of sample measurements is the most likely cause of uncertainty that could potentially be remedied quickly, but supplementary information on e.g. where it had rained could also fall into this category. As the emergency operation wound down and longer-term remediation and recovery operations took over, sensitivity analysis would be an important aid to the fine-tuning of models and the assessment of the effectiveness of any long-term countermeasures that might be considered.

Another aspect of decision aiding that could be explored is the use of *cost functions* to weight uncertainty displays. Uncertainty in the level of contamination in particular areas will be less important than in others. The map displayed could therefore be adjusted to show information that is weighted to the importance of the uncertainty in different regions.

5 DISCUSSION

Space precludes the discussion of many aspects of the joint work between NRPB and IPSN. However, it is clear from the above discussion that this project should enable decision-makers to gain a much better understanding of the uncertainties in the deposition after an accident and how that uncertainty propagates into estimates of the contamination likely to be found in food.

Acknowledgements

The authors would like to thank Dr Boris Yatsalo of RRC KI (Kurchatov Institute), Moscow for the Chernobyl data and T. Charnock (NRPB) and F. Burge (NRPB) for helping to prepare the Figure.

References

- [1] Higgins N. A., Charnock T. W. and Smith T. J. SECTAR: Statistical Estimation and Characterisation Techniques for Accident Response ECORAD 2001 *ibid*.
- [2] O'Hagan A., Kennedy M. C., Oakley E. (1999) Uncertainty Analysis and other Inference Tools for Complex Computer Codes. In: Bayesian Statistics 6. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds. (Oxford University Press, Oxford 1999) pp. 503-524 (with discussion).
- [3] Kennedy M. C. O'Hagan A. and Higgins N., Bayesian analysis of computer code outputs. In Quantitative Methods for Current Environmental Issues. C. Anderson, V. Barnett, P. Chatwin and A. El-Shaarawi Eds. (Springer-Verlag, in Press 2001).
- [4] Expert Judgement and Accident Consequence Uncertainty analysis. L. H. J. Goossens and G.N. Kelly Eds. Special Issue Radiation Protection Dosimetry **90**(3) (2000).
- [5] Hoffman, F. O. Peer review of HEDR uncertainty and sensitivity analyses plan. Hanford environmental dose reconstruction project. Battelle, Pacific Northwest Laboratories Richland, WA PNWD-2162 HEDR (1993).
- [6] Smith K. R., Brown J., Jones J. A., Mansfield P., Smith J. G. and Haywood S. M., Uncertainties in the Assessment of Terrestrial Foodchain Doses NRPB-M922 (NRPB, Chilton, 1998).
- [7] Haylock, R. and O'Hagan, A. (1996) On inference for outputs of computationally expensive algorithms with uncertainty on the inputs. In Bayesian Statistics 5, J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds. (Oxford University Press, Oxford, 1996) pp. 629-637.
- [8] Saltelli, A., Chan, K. and Scott, E. M. eds. Sensitivity Analysis. (Wiley, New York, 2000).