Los Alamos Environmental Restoration Records Processing Facility

ER Record I.D.# 63521

LOS ALAMOS NATIONAL LABORATORY
ENVIRONMENTAL RESTORATION
Records Processing Facility
ER Records Index Form

ER ID NO. 63521 *Date Received:* 8/5/99 *Processor:* YCA *Page Count:* 14

Privileged: (Y/N) N *Record Category:* P *Record Package No:* 0

FileFolder: N/A

Correction: (Y/N) N *Corrected No.* 0 *Corrected By Number:* 0

Administrative Record: (Y/N) Y

Refilmed: (Y/N) N *Old ER ID Number:* 0 *New ER ID Number:* 0

Miscellaneous Comments:

REFERENCE CITED

14042

63521

(14)

ELSEVIER

Journal of Environmental Radioactivity 42 (1999) 117–130

JOURNAL OF ENVIRONMENTAL RADIOACTIVITY

BIOMOVS II: An international test of the performance of environmental transfer models

P.A. Davis[a,*], M.R. Avadhanula[b], D. Cancio[c], P. Carboneras[d], P. Coughtrey[e], G. Johansson[f], R.H. Little[g], G.M. Smith[g], B.M. Watkins[g]

[a]AECL, Chalk River Laboratories, Chalk River, Ontario, Canada K0J 1J0
[b]AECB, Box 1046, Station B, Ottawa, Ontario, Canada K1P 5S9,
[c]CIEMAT/IMA, Avenida Complutense 22, 28040 Madrid, Spain
[d]ENRESA, Emilio Vargas 7, 28043 Madrid, Spain
[e]Mouchel Consulting Ltd, West Hall, Parvis Road, West Byfleet, Weybridge, Surrey KT14 6EZ, UK
[f]Swedish Radiation Protection Institute, S-171 16 Stockholm, Sweden
[g]QuantiSci Ltd, Chiltern House, 45 Station Road, Henley on Thames, Oxfordshire RG9 1AT, UK

Received 21 April 1997; accepted 2 March 1998

Abstract

The Biospheric Model Validation Study-Phase II (BIOMOVS II) was an international cooperative program that tested the accuracy of predictions of environmental assessment models. Model evaluation was based on calculations made by individual participants for 10 test scenarios that addressed both short- and long-term releases of radioactivity from facilities such as power reactors, solid waste disposal repositories and uranium mill tailings. Model predictions were compared with each other and, where possible, with independent field observations, and reasons were sought for any differences that arose. Qualitative topics were also considered, including development of systematic methodologies for radiological assessments. This paper addresses conclusions arising from the study as a whole. Confidence intervals on predictions and differences between predictions and observations were often less than a factor of 10, although there was much variability among models and scenarios. Model performance depended critically, not only on the formulation and parameter values of the model itself, but also on the experience and assumptions made by the user. The study demonstrated the need to better explain and justify all aspects of model structure and application and to assess all sources of uncertainty. A key recommendation was that assessments should not be undertaken in isolation by one individual using one model.

*Corresponding author. Tel.: 001 613 584 3311; fax: 001 613 584 1221; e-mail: davisp@aecl.ca

0265-931X/98/$19.00
PII: S0265-931X(98)00049-6

1. Introduction

Mathematical models are commonly used to simulate contaminant transport through the environment following real or hypothetical releases. Invariably, these models are simplifications of complex natural processes and predictions derived from them are approximations. Model reliability must therefore be evaluated and understood before predictions can play a part in decision making. The comparison of model predictions with observations collected following actual releases provides the most rigorous test of model credibility. Where such data are not available (for example, in scenarios involving hypothetical releases from deep geological waste repositories), confidence can be established in a number of other ways: by intercomparing predictions from several models applied to a common scenario and identifying and explaining any discrepancies; by demonstrating that the assumptions on which the model is based are reasonable in light of current scientific practice; by quantifying the uncertainty in model predictions; by investigating model behaviour through sensitivity analysis and by verifying the computer code of the model. In this paper, the full suite of activities undertaken to establish model credibility is termed model evaluation.

The need to test assessment models for radioactive contaminants has recently become more urgent given the increasingly prominent role that models play in assessing real and postulated releases from the growing number of nuclear facilities worldwide. In response to this need, several international cooperative studies were set up in the 1980s. Among these were programs that focussed on the geosphere (SKI, 1987; SKI, 1990), on probabilistic safety assessment (NEA, 1991) and on the behaviour of radioactivity in the environment following the Chernobyl accident (IAEA, 1993). Also established was the Biospheric Model Validation Study (BIOMOVS, 1993), which dealt with models of contaminant transfer through the biosphere and intake by man. The release scenarios in BIOMOVS encompassed a broad range of assessment issues and included some relatively novel modelling problems, particularly for long-term releases from solid waste disposal facilities. BIOMOVS ran from 1986 to 1990 under the auspices of the Swedish Radiation Protection Institute (SSI).

Some of the issues addressed in BIOMOVS were not fully resolved by the end of the study, and not all potentially important pathways, radionuclides and scenarios had been considered. Accordingly, a second phase of the program, BIOMOVS II, was established in 1991 with funding from five organisations: Atomic Energy Control Board, Canada; Atomic Energy of Canada Limited; Centro de Investigaciones Energeticas Medioambientales y Tecnologicas, Spain; Empresa Nacional de Residuos Radiactivos SA, Spain; and SSI. BIOMOVS II had three main objectives:

- to test the accuracy of the predictions of environmental assessment models for selected contaminants and exposure scenarios;
- to explain differences in predictions among models; and
- to recommend priorities for future research to improve model performance.

The study was also designed to act as a forum for the exchange of ideas, experience and information to help improve confidence in the models. Scientific direction was provided by a Coordinating Committee made up of one member of each of the 26 organizations that formally joined BIOMOVS II. A Steering Committee composed of

representatives from the funding agencies organized and managed the administrative aspects of the project. Both committees received support from the Scientific and Technical Secretariat, QuantiSci Ltd. Ultimately, the study involved 160 organisations from 31 countries.

BIOMOVS II activities were carried out through Working Groups (WGs). The first task of each WG was to define clearly the questions it sought to address. Most WGs then developed test scenarios to answer those questions through a series of model calculations. The scenarios included all of the information needed to carry out the calculations: source terms, transfer pathways, properties of the environmental system and end points. Wherever possible, the scenarios were based on field data. Individual members of the WGs submitted predictions for the end points using the models and resources at their disposal. The predictions were intercompared and discussed at semi-annual meetings to identify, explain and resolve any differences among them in terms of differences in assumptions, model formulation and/or parameter values. In cases where data were available, predictions were also compared with observations, which were withheld from the modellers until calculations were complete. Participants were requested to quantify uncertainties in their predictions. Interaction among WGs was encouraged.

BIOMOVS II ended in 1996. This paper briefly reviews its activities as a whole, presents major overall conclusions and makes recommendations for further work. Detailed discussions of the results of individual WGs can be found in the papers that follow in this issue and in the BIOMOVS II Technical Reports published by SSI on behalf of the Steering Committee (Table 1).

Table 1
Scenarios comprising the BIOMOVS II program

Scenario	Technical Report[a]	Reference in this issue
Guidelines for uncertainty analysis	1	b
Guidelines for comparison of model predictions	3	b
Reference biospheres for radioactive waste disposal	2 and 6	van Dorp et al.
Contaminant migration and impacts from uranium mill tailings	4 and 5	Camus et al.
Effect of user interpretation on model predictions	7	Kirchner et al.
Tritium transfer in the foodchain	8 and 13	Barry et al.
Chernobyl data — washoff from experimental plots	9	Konoplev et al.
Chernobyl data — consequences of contamination of aquatic media	10	Kryshev et al.
Chernobyl data — atmospheric resuspension of radionuclides	11	Garger et al.
Complementary studies — biosphere modelling for assessment of radioactive waste disposal	12	Klos et al.
^{14}C migration and accumulation in a lake	14	Bird et al.
Water flow and radionuclide transport in lysimeters	15	Butler et al.
Effect of model complexity on uncertainty estimates	16	Elert et al.
Overview of the BIOMOVS II program and its results	17	This paper

[a] All technical reports are available from the Swedish Radiation Protection Institute, 171 16 Stockholm, Sweden.

[b] There is no paper for this scenario in this issue.

The concept of validation has long been controversial. Philosophers such as Popper (1959) have argued that scientific theory cannot be validated, only invalidated, and that past demonstrations of accuracy do not guarantee similar performance in the future. Other scientists (e.g. McCombie and McKinley, 1993) have claimed that a model can be considered validated if it achieves an acceptable level of predictive accuracy given its history of testing, the complexity of the system in question and the future application of the model. In this paper, the term validation is restricted to past demonstration of accuracy without implications as to future merit. Most BIOMOVS II participants, however, took a more pragmatic view: given the need for quantitative radiological assessments, a model with a good track record is preferable to no model at all, even if it has not been rigorously validated. It was also generally recognized that, regardless of the meaning given to the term validation, the process of model evaluation leads to a better understanding of the transport mechanisms in question and to improved confidence in the use of model results.

2. Scenarios

The scenarios undertaken in BIOMOVS II reflected the interests of participants and issues left unresolved in the first phase of BIOMOVS. Account was also taken of the activities of other international programs such as VAMP (Validation of Model Predictions; IAEA, 1993) and the Nuclear Energy Agency's Performance Assessment Advisory Group. The final program included scenarios that addressed a broad range of assessment issues (Table 1). Scenarios based on field data were developed for ^{14}C transport in lakes, tritium fluxes from soil and vegetation, formation of organically bound tritium in spring wheat and the upward transport of radionuclides in soils in lysimeters. Data collected following the Chernobyl accident were used to develop scenarios for atmospheric resuspension, wash-off from soils to water bodies, and transport and accumulation in the Chernobyl cooling pond. Scenarios involving releases from solid radioactive waste repositories (Complementary Studies) and uranium mill tailings were developed as model intercomparison exercises since test data were not available. The mill tailings scenario considered the transport of stable elements as well as a number of radionuclides. The end points for most scenarios were radionuclide concentrations in various environmental compartments, but some WGs calculated doses and risks as well. Two WGs used the scenario approach to investigate specific aspects of uncertainty analysis, namely the effects of model complexity and user assumptions on uncertainty estimates. A WG on uncertainty and validation provided general advice on these issues to other participants and also developed a glossary and guidelines for uncertainty analysis and comparison of model predictions.

The Reference Biospheres WG (van Dorp et al., this issue) undertook the more qualitative task of developing a methodology for constructing defensible biosphere models for specific applications. The work was motivated by the recognition that the underlying premises of a biosphere assessment are often taken for granted in the early stages of model development, with the result that the model often cannot be

demonstrated to fit its intended purpose. The reference biospheres methodology overcomes this by providing a systematic approach to the definition and justification of the composition and structure of the model. It takes into account model objectives, assumptions made about enviromental conditions and human behaviour, and features, events and processes (FEPs) that may need to be included in the model. Application of the methodology results in a conceptual model of the biosphere (a reference biosphere) for the assessment in question, which can subsequently be translated into a mathematical description and computer code. The WG illustrated the methodology for solid radioactive waste disposal. The resulting reference biosphere reflected consensus among the WG members, who represented eight different countries, and provided a consistent framework for comparing performance assessments for alternative radioactive waste disposal facility designs and locations. The methodology was also applied to the waste management scenario developed by the Complementary Studies WG.

For scenarios based on observations, the amount of background information included in the scenario description was governed by what the supplier of the data could provide. Information was often incomplete, forcing participants to make many assumptions to adapt the data to their models or their models to the data. For scenarios without observations, the level of detail could be set arbitrarily. A small amount of data allowed more leeway in interpreting the scenario and encouraged discussion on a conceptual level. However, experience from the first phase of BIOMOVS suggested that, with this approach, differences in predictions due to scenario interpretation could overwhelm differences due to model formulation or parameter values. In BIOMOVS II, the attempt was made to define all scenarios tightly to reduce this problem.

3. Models and parameter values

Many of the models used for BIOMOVS II calculations were compartment models, reflecting the traditional approach used in environmental transport simulation. In a compartment model, each of the major environmental media (soil, surface waters, animals and so on) is represented by one or more compartments which are spatially homogeneous in their properties. Transfers between compartments are usually described using rate constants. Radionuclides entering a compartment are assumed to be uniformly and instantaneously mixed within its volume. Most of the differences between compartment models used for a given scenario related to the number of compartments and pathways considered and not to the basic formulation. But compartment models were not used exclusively. Radionuclide transfer in soils was often treated by solving the advection–diffusion equation and most participants employed simple empirical models in the Resuspension scenario. A few of the models used in BIOMOVS II were specially coded for the study but most were pre-existing models that were changed as necessary to accommodate a particular scenario.

The predictions of a model depend not only on its mathematical formulation and structure but also on the numerical values of the transfer parameters used in it. Some

values were defined as part of the scenario descriptions but in other cases participants were free to choose and practice varied considerably. Some relied entirely on generic values selected from the general literature or taken from one or the other of several compilations of recommended default values (e.g., Baes et al., 1984; Posten and Klopfer, 1986; IAEA, 1994). Others chose values that they believed reflected the information given in the scenario description. The result was frequently a substantial difference among the values adopted for a given parameter.

4. Uncertainty

In recognition of the vital role that uncertainty estimates play in establishing model credibility, a main focus of BIOMOVS II was to quantify uncertainties for the models and scenarios considered in the study. Guidelines for performing an uncertainty analysis were published (BIOMOVS II, 1993a) and advice on specific issues related to uncertainty was available throughout the program. Most participants carried out a parametric uncertainty analysis and included a quantitative uncertainty statement, such as the 95% confidence interval, with their predictions. Unfortunately, the analyses were sometimes inconsistent with regard to the parameters included, their probability density functions (pdfs), the sampling scheme, and the method used to propagate uncertainties through to model endpoints. The result was substantial variability in the magnitude of the uncertainty estimates. The smallest 95% confidence interval estimated by an individual participant in a given WG was usually about a factor of 2, but intervals up to 10 times larger were not uncommon (Fig. 1). To

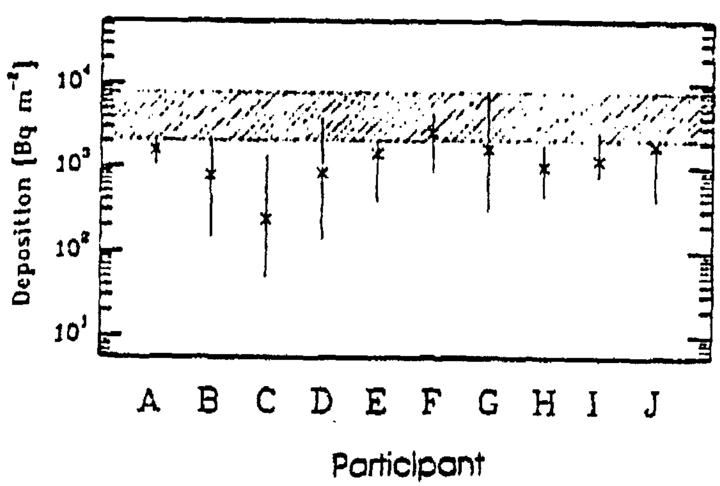

Fig. 1. Uncertainties associated with predictions for the user interpretation scenario (Kirchner et al., this issue). The end point of the calculation was the total amount of ^{137}Cs deposited from the atmosphere at Bremen following the Chernobyl accident. The vertical lines are the 95% confidence intervals calculated by each participant, with best estimates shown as crosses. The hatched area denotes the 95% confidence interval of the field measurements.

some extent this is to be expected given the many subjective judgments that are inherent in uncertainty analysis (IAEA, 1989). But the variability makes it difficult to draw general conclusions regarding the magnitude of parametric uncertainty for the models and applications considered here. It would be prudent to assume that the magnitude for each scenario corresponds to the largest uncertainty estimated by any of the participants. This means that the parametric uncertainty for most scenarios was about a factor of 10 and closer to a factor of 100 for the Resuspension, Complementary Studies and Model Complexity scenarios. In general, uncertainties tended to be relatively small in the scenarios that dealt with releases directly to the atmosphere or to surface water bodies and larger for those that treated releases from underground waste repositories or movement through soils.

Overall, the uncertainties estimated in BIOMOVS II were substantially less than those in the first phase of BIOMOVS. There were several reasons for this:

- Over time, participants gained greater understanding of their models and the problems to which they are applied. Uncertainties, which are subjective estimates of the confidence the participants have in the accuracy of their predictions, have therefore decreased.
- The WGs spent considerable time defining the scenarios to ensure they were complete, consistent and well understood. This allowed participants to choose parameter values with some confidence and assign relatively narrow distributions to them.
- With time, participants became more familiar with uncertainty analysis tools and were able to apply them appropriately.
- Many WGs used a consensus approach to specify pdfs for the distributed parameters in the analysis. This led to greater confidence in the parameter values and narrower distributions. Informal expert elicitation of this sort helps to overcome the bias that a single individual inevitably brings to the process and avoids the need for formal elicitation (Hofer, 1986), which is expensive and time-consuming.

Sources of uncertainty other than those due to parameter values were also investigated in BIOMOVS II. The user interpretation WG showed that uncertainties arising from assumptions made by the model user can easily exceed an order of magnitude for scenarios involving radionuclide transport in terrestrial food chains. Similarly, the model complexity WG showed that variability of this size must also be expected in soil transport scenarios due to uncertainties in model formulation. The work of the reference biospheres WG suggested that a systematic approach to the composition and structure of a model at the start of its development could help to reduce model uncertainty. Various WGs noted that simple human errors such as miskeying of numbers, misplaced decimal points and incorrect units appeared occasionally in the results and must be guarded against. When all sources are taken into account, the total uncertainty in model predictions for each scenario considered in BIOMOVS II was likely considerably larger than parametric uncertainty alone.

5. Model evaluation

For scenarios involving experimental data, the primary tool used by most WGs to evaluate model performance was graphical comparison of predictions and

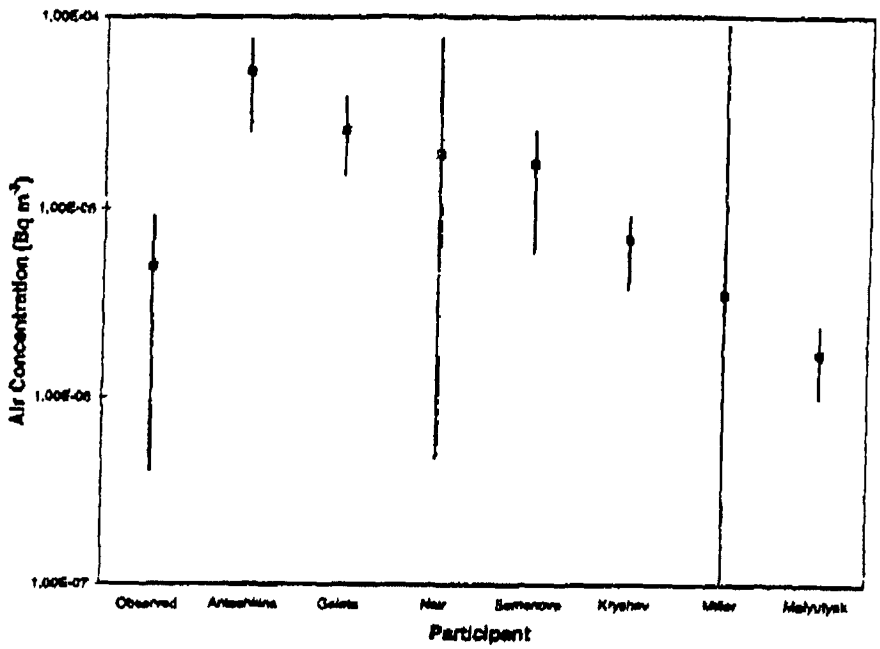

Fig. 2. Observations and predictions for the resuspension scenario (Garger et al., this issue). The endpoint of the calculation was air concentration in Kiev in December 1991 due to resuspension of ^{137}Cs previously deposited from the airborne Chernobyl plume. The bars on observations and predictions are 95% confidence intervals.

observations (Fig. 2). Conclusions regarding model validity were based on a qualitative assessment of the extent to which confidence intervals overlapped. A similar approach was taken in those scenarios for which no test data were available, but in these cases the comparisons involved only the predictions and uncertainties of the various models. For a given scenario and endpoint, the best estimate predictions of the various models tended to scatter around the observations. Highest and lowest predictions often lay within about two orders of magnitude and it was not uncommon for individual predictions to differ from the observations by less than a factor of 10. Confidence intervals on predictions and observations overlapped for some models but not for others, suggesting that uncertainties were underestimated in some cases.

Statistical methods for making quantitative statements of model performance were presented in a BIOMOVS II Technical Report (BIOMOVS II, 1995) but found little scope for application in the WGs. There were three main reasons for this. First, the model predictions were usually not independent, normally distributed or drawn randomly from populations and so violated the fundamental assumptions of many statistical tests. Moreover, the uncertainty estimates in most scenarios may not have been consistent enough to provide the basis for quantitative comparison. Second, although statistical tests are quantitative, their output cannot be used without making subjective judgments. When applied to a given set of predictions, a particular test might return a value α where perfect agreement would be signified by β. There are no

general criteria that indicate how close α must be to β for the agreement to be considered acceptable in some way. Thus statistical measures are of little use for validation purposes, although they can indicate which of a group of models performs best with respect to a given set of observations.

The third reason lies in the recognition that other tools are available for establishing the credibility of model predictions. The most important of these is a physically based evaluation of all parts of the model, justifying the processes included in it, how they are formulated mathematically, the choice of parameter values and the assumptions made. From this comes an understanding of why the model behaves as it does, the ability to explain differences between its predictions and observations, and the insight to improve it. Most WG efforts went into such evaluations to the benefit of the participating models. Many of the conclusions are scenario specific and can be found in the reports of the individual WGs, but a few are generally valid. For the most part, participants in a given scenario considered the same processes and pathways and simulated them in a similar manner. This suggests that there is consensus on what processes and pathways are important and how they are best modelled. Some of the differences in predictions could be ascribed to the complexity of the model used for a given process. But in many cases the explanation for differences lay in the values chosen for the model parameters. Many processes are described by bulk transfer parameters that depend strongly on the environmental conditions under which they were measured. Values reported in the literature vary widely, as did the choices made by the participants. In general, predictions made using site-specific parameter values agreed more closely with observations than did those using generic values.

6. Conclusions

Some of the more important overall conclusions to emerge from BIOMOVS II are summarized below:

1. Collaborative, multi-disciplinary efforts such as BIOMOVS II play a vital part in improving environmental transport models. This applies as much in terms of broad methodology as in details of data interpretation and process understanding. International efforts are cost effective and provide a mechanism for identifying and resolving points of difference. Consensus on individual discrepancies represents a collective view of the state-of-the-art; if no consensus can be reached, an area of future research has been identified. The participation in BIOMOVS II of a wide range of people with differing expertise (field experimentalists and mathematical modellers) and perspectives (researchers, regulators and operators) helped contribute to the development of consensus. The study also provided participants with the opportunity to explain and justify their models to a diverse audience, a procedure that often identified significant issues of concern and mechanisms for dealing with those concerns. The open framework for discussion and debate was useful for achieving these goals.

2. At the beginning of BIOMOVS II a glossary was developed and published for use by participants (BIOMOVS II, 1993b). However, experience within the study and elsewhere has shown that, for a variety of reasons, consistent international use of

terminology is hard to achieve. This contributes to difficulties in interpreting the assessment problem, the information and data available for solving the problem and the presentation and explanation of results.

3. In most multi-pathway, multi-contaminant scenarios (e.g. Uranium Mill Tailings and Complementary Studies), exposures were not dominated by a single pathway or radionuclide. Multiple pathways and radionuclides must be considered, at least at the screening stage of assessment calculations.

4. Complex models do not necessarily provide results in better agreement with observations than simple models, nor do their predictions necessarily have smaller uncertainties. On the other hand, complex codes can incorporate explicitly a larger range of processes and may be required to model time-dependent behaviour.

5. Environmental transport models can rarely predict the actual course of future events in long-term assessments because of difficulties in accounting for changes in the physical environment and human lifestyles over very long periods of time. Model outputs for these types of applications should be interpreted, not as predictions of reality, but as indications of what might happen based on best available knowledge.

6. Environmental transfer models are effective tools which can be employed as one of several inputs to the decision making process. They should not be considered in isolation but should be combined with other inputs such as reasoned arguments, bounding calculations and public opinion.

7. Blind testing of model predictions against field data plays a key role in establishing model credibility. Confidence in the model will increase as more supporting data become available. Quantitative statements regarding the performance of the model can be made only for the conditions under which it was tested, although the model may perform well outside those conditions.

8. Testing model predictions against field data is not sufficient to establish model validity. Confidence building should also involve sensitivity and uncertainty analyses and code verification. Most importantly, the model must be shown to be fit-for-purpose by demonstrating rigor in defining the assessment context, deriving the model, justifying the assumptions made, selecting the parameter values and interpreting the results. It must be shown that each of these activities is treated appropriately given prevailing levels of understanding. These methods provide the only approach to establishing confidence in models for which experimental validation is not possible, such as models that predict the long-term impact of radionuclides released from deep repositories.

9. As many intermediate results and endpoints as possible should be tested to identify compensatory errors in the models and areas where the state of knowledge should be improved.

10. Evaluation of uncertainties involves a significant element of subjective judgment. Despite this, uncertainty analysis can be undertaken in a way that will provide meaningful and useful results. Modellers should be familiar with the techniques that are available and choose an approach that is consistent with the purpose of the model, the quality of the data and the nature of the application. The pdfs used in parameter uncertainty analyses should be set by a process of consensus. The methods used and assumptions made in carrying out the analysis must be documented in detail.

11. The uncertainty estimates provided by participants with their model predictions almost invariably reflected parameter uncertainty only. Better methods for estimating uncertainties due to conceptual model formulation and user interpretation should be developed and applied routinely.

12. A variety of steps can be taken to reduce uncertainties:

- Model calculations must be based on a clear understanding of the assessment context. The nature of the endpoints and properties of the system being addressed must be clearly defined. Calculations should be undertaken in an iterative fashion to provide opportunities to remove ambiguities and inconsistencies. This will reduce uncertainties due to user interpretation.
- Initial construction of the model should be based upon a rigorous and systematic approach to ensure that all relevant FEPs have been included. This will decrease uncertainties due to model formulation.
- The use of site-specific values for the most sensitive parameters of the system will reduce parameter uncertainties.
- Greater clarity and consistency in model requirements and assessment guidelines will lead to models with greater similarity in goals and endpoints, making comparison and verification easier.

13. The level of agreement between predictions and observations depended on scenario and endpoint but differences in many cases were less than a factor of 10. Uncertainties were of the same magnitude but would likely have been larger had participants included uncertainties due to model formulation and user interpretation.

14. Much of the difference in model predictions could often be traced to the use of different parameter values. It is important to evaluate data critically in terms of their source, quality and intended application before values are chosen. Modellers who fail to do this run the risk of making meaningless calculations. The best model is often the one that is best supported by data.

15. A given modelling project should not be carried out in isolation by one individual using one model. This will likely introduce bias and limit the credibility of the results. The independent use of two or more models and rationalization of any differences in their predictions will lend more confidence to the results. Such problems can be further reduced by ensuring that a multi-disciplinary team is used, including those providing information on the physical system, those responsible for model development and those using the model.

7. Suggestions for the future work

1. More work is required to improve and implement methods for evaluating model performance and uncertainties to gain a better understanding of the true level of confidence that can be placed in the predictions of environmental transport models. In particular, continuing emphasis should be given to

- explaining and justifying assumptions made in the models;
- methods to assess sources of uncertainty other than parameter uncertainty;

- greater understanding of the effects of user interpretation on model predictions;
- protocols for replicating assessments to reduce the subjectivity introduced by individuals;
- development of improved physically based methods for deriving values for key transfer parameters (sorption coefficients, concentration ratios and so on) from site-specific data; and
- presentation and interpretation of results so that a wide range of audiences (e.g. scientists, regulators, operators, members of the public) can better understand the implications.

2. The reference biospheres methodology was successfully developed and partially tested in cooperation with the Complementary Studies WG, and an important degree of international consensus was reached. This work should be extended to:

- develop further the principles for defining critical groups relevant to long-term radiological assessments;
- apply the methodology to a range of basic systems and alternative assessment contexts; and
- use the methodology to develop more fully and formally a set of conceptual models, including clarification of the ways in which FEPs are represented and the corresponding databases defined.

3. A variety of specific environmental transport and accumulation processes were identified during BIOMOVS II for which further experimental and modelling studies are warranted. These include:

- lysimeter studies to clarify the role of various soil and plant processes in the upward migration of contaminants through vegetated soil profiles;
- behaviour of contaminants in vegetation following irrigation;
- long-term behaviour of tritium following continuous releases as well as specific processes such as re-emission from plants and soil and the formation of organically bound tritium;
- contaminant migration from uranium mill tailings and consequent behaviour in aquatic ecosystems;
- chemical transformations of contaminants in environmental media and an understanding of species-dependent behaviour;
- behaviour of ^{14}C in the terrestrial environment;
- collation of data to facilitate the comparison of human health risks from radioactive and non-radioactive contaminants;
- assessment of exposures and risks to non-human biota; and
- the transport of radionuclides in solid form by processes such as erosion and bioturbation.

4. The increased understanding of transport processes gained in BIOMOVS II and other programs should be used to help update transfer parameter values published in various handbooks.

5. Simplified versions of some BIOMOVS II scenarios have been used as exercises in student training. More scenarios could be published to further help educate students of radioecology.

6. More effort should be put into preparing historical data or generating new data for use in model testing to help establish the credibility of environmental transport models in different applications.

7. Data sets and predictions from BIOMOVS II are available for use by modelling groups that were unable to participate formally in the study. The Technical Reports provide sufficient information to enable other groups to calculate predictions and compare them with the original results and with the observations, where they exist.

There is a continuing need to provide a forum to bring together the different types of expertise and disciplines involved in understanding and improving the performance of environmental transfer models. The BIOMASS (Biosphere Model Assessment) program set up by the IAEA in 1996 (IAEA, 1996) is filling this need, vigorously continuing and extending the work described here.

Acknowledgements

The ideas for this paper were drawn from conclusions synthesized for each scenario by the various Working Group leaders: Peter Barry, Ulf Baverstam, Glen Bird, Adrian Butler, Henri Camus, Mark Elert, Evgenii Garger, Owen Hoffman, Gerald Kirchner, Tom Kirchner, Ryk Klos, Alexei Konoplev, Ivan Kryshev, Marion Scott, Siegfried Strack and Frits van Dorp. But most credit must go to the individual participants of BIOMOVS II, whose hard work and insight provided the basis for the material presented here.

References

Baes, C. F. II, Sharp, R. D., & Shor, R. W. (1984). A review and analysis of parameters for assessing transport of environmentally released radionuclides through agriculture. Oak Ridge National Laboratory Report ORNL-5786, Oak Ridge, TN, USA.

BIOMOVS (1993). Final report. BIOMOVS Technical Report 15, Swedish Radiation Protection Institute, 171 16 Stockholm, Sweden.

BIOMOVS II (1995). Qualitative and quantitative guidelines for the comparison of environmental model predictions. BIOMOVS II Technical Report No. 3, Swedish Radiation Protection Institute, 171 16 Stockholm, Sweden.

BIOMOVS II (1993a). Guidelines for uncertainty analysis. BIOMOVS II Technical Report No. 1, Swedish Radiation Protection Institute, 171 16 Stockholm, Sweden.

BIOMOVS II (1993b). Glossary developed for the participants in the BIOMOVS II study. Swedish Radiation Protection Institute, 171 16 Stockholm, Sweden.

Hofer, E. (1986). On surveys of expert opinion. *Nuclear Engineering and Design*, *93*, 153.

IAEA (1996). International programme on biosphere modelling and assessment methods (BIOMASS). Themes for a new coordinated research programme on environmental model testing and improvement. Theme 1: Radioactive Waste Disposal, Theme 2: Environmental Releases, Theme 3: Biospheric Processes. International Atomic Energy Agency, Vienna, Austria.

IAEA (1994). Handbook of parameter values for the prediction of radionuclide transfer in temperate environments. International Atomic Energy Agency, Technical Report Series No. 364, Vienna, Austria.

IAEA (1993). Validation of environmental model predictions. International Atomic Energy Agency, STI/PUB/932, Vienna, Austria.

IAEA (1989). Evaluating the reliability of predictions made using environmental transfer models. International Atomic Energy Agency, Safety Series No. 100, Vienna, Austria.

McCombie, C. & McKinley, I. (1993). Validation – another perspective. *Groundwater*, *31*, 530–531.

NEA (1991). The international probabilistic system assessment group: background and results. Nuclear Energy Agency, OECD, Paris, France.

Popper K. (1959). *The Logic of Scientific Discovery*. New York: Harper and Row.

Posten, T. M., & Klopfer, D. C. (1986). A literature review of the concentration ratios of selected radionuclides in freshwater and marine fish. Pacific Northwest Laboratories Report PNL-54084, Richland, WA, USA.

SKI (1990). The international INTRAVAL project — background and results. Swedish Nuclear Power Inspectorate, OECD, Paris, France.

SKI (1987). The international HYDROCOIN project — background and results. Swedish Nuclear Power Inspectorate, OECD, Paris, France.