International Model Validation

Mathematical models are often used for estimation of environmental transport of radionuclides and other contaminants and for assessing the resulting exposures, doses, and risks that may occur or may have occurred. Generally, models rather than measurements are used because the relevant measurements were not made or cannot be made. Because model predictions may be used as the basis for important decisions, the consequences of which may be significant in terms of human, ecological, or economic costs, it is essential to evaluate the reliability of the model predictions. One of the major approaches used in recent years to address the reliability of model performance is to test model predictions against field measurements, an approach sometimes referred to as “model validation.”

For models of radionuclide transport in the environment, some of the most valuable model testing opportunities have come about as a result of the Chernobyl accident in April 1986. Cesium-137 and other radionuclides have been measured in various media (air, soil, water, foodstuffs, animal feed, animal products, human bodies) at various times and places, and the location and time of the release are known. Thus it has been possible to compare predicted concentrations of radionuclides at specified locations and times with actual measurements. Several test exercises have been carried out over the last 15+ years under international programs sponsored by the International Atomic Energy Agency, the Swedish Radiation Protection Institute, and other agencies: the Biospheric Model Validation Study, Phases I and II (BIOMOV and BIOMOVS II), the Co-ordinated Research Programme on Validation of Environmental Model Predictions (VAMP), and most recently the Biosphere Modelling and Assessment Methods programme (BIOMASS).

Dr. F. Owen Hoffman, President of SENES Oak Ridge, Inc., organized and served as general chairman of the first international workshop dedicated to the evaluation of mathematical models used to assess the environmental transport and dose of radioactivity. In 1985, he consulted with the Swedish Radiation Protection Institute in cofounding the BIOMOVS program. In 1988, Dr. Hoffman became Chief Scientist to the International Atomic Energy Agency for the VAMP program. He also headed the Multiple Pathways Assessment Working Group of the VAMP program and the Post Chernobyl Data Working Group of the BIOMOVS II program. Dr. Kathleen M. Thiessen, Senior Scientist for SENES Oak Ridge, Inc., chaired the Dose Reconstruction Working Group of the BIOMASS program, one of two working groups in BIOMASS Theme 2, Environmental Releases. Dr. Thiessen was also instrumental in completing the efforts of the Multiple Pathways Assessment Working Group of the VAMP program and the Post Chernobyl Data Working Group of the BIOMOVS II program. Drs. B. Gordon Blaylock and A. Iulian Apostoaei of SENES Oak Ridge, Inc., have also been active participants in some of the international model validation programs.

The table below provides a summary of the four international programs and their major sponsors. References describing the programs and the individual model testing exercises are also provided below. A number of major conclusions about model performance and the modeling process have been drawn to date from these model testing exercises.

- First, intercomparison of predictions from several modelers provides a valuable opportunity to compare modeling approaches and results (Fig. 1). Differences in the results are often due to differences in interpretation of the input information or to errors.
in input or coding, and such situations can easily go unrecognized apart from such an intercomparison. However, blind testing of model predictions against field data permits evaluation of the accuracy of the model predictions and uncertainty statements (Figs. 1-3). Testing of individual pathways and midpoints (model components or submodels) is essential, both to identify any compensatory effects in a model (Fig. 4) and to identify areas (e.g., critical parameter values) where the state of knowledge should be improved. Model testing has also shown that consensus among model predictions should not be confused with accuracy of the predictions (Fig. 5).

- Differences in model predictions for the same endpoint may be seen even when two assessors use the same computer code (compare Models D1 and D2 in Fig. 5). In general, although computer codes should be flexible to permit adaptation for site-specific situations, the experience and judgment of the assessor are much more important than the computer code used, due to differences in interpretation of incomplete input information and in selection of appropriate parameter values. An experienced assessor who is familiar with the processes being modeled and with site-specific data on parameters (e.g., transfer coefficients) may achieve model predictions within a factor of 2 of the test data. This level of accuracy in assessment models cannot generally be guaranteed; use of generic models and transfer coefficients, especially by an inexperienced assessor, may result in errors exceeding a factor of 10.

- Estimates of uncertainty are also dependent on the judgment of the assessor and therefore will differ among different analysts for the same assessment endpoint (Figs. 2-3, 5). Uncertainty estimates are frequently incomplete, in that not all sources of uncertainty have been included. Therefore, the uncertainty estimates on a model prediction often do not encompass the measured value; this constitutes a statement of overconfidence.
Table 1. Summary of major international model-testing studies.

<table>
<thead>
<tr>
<th>Study and sponsoring organizations</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BIOMOVS</strong> (Biospheric Model Validation Study)</td>
<td>1986-1990</td>
</tr>
</tbody>
</table>
| Swedish Radiation Protection Institute  
S-171 16 Stockholm, Sweden |
| **BIOMOVS II** (Biospheric Model Validation Study, Phase II) | 1990-1996 |
| Canadian Nuclear Safety Commission  
(formerly The Atomic Energy Control Board of Canada)  
P.O. Box 1046, Station B  
Ottawa, Ontario, K1P 5S9 Canada  
Atomic Energy of Canada Limited  
Chalk River Laboratories  
Chalk River, Ontario, K0J 1J0 Canada  
Centro de Investigaciones Energéticas Medioambientales y Tecnológicas  
Avenida Complutense 22  
28040 Madrid, Spain  
Empresa Nacional de Residuos Radiactivos, S.A.  
Emilio Vargas, 7  
28043 Madrid, Spain  
Swedish Radiation Protection Institute  
S-171 16 Stockholm, Sweden |
| **VAMP** (Validation of Environmental Model Predictions) | 1988-1994 |
| International Atomic Energy Agency  
Wagramer Strasse 5  
P.O. Box 100  
A-1400 Vienna, Austria |
| **BIOMASS** (Biosphere Modelling and Assessment Methods) | 1996-2000 |
| International Atomic Energy Agency  
Wagramer Strasse 5  
P.O. Box 100  
A-1400 Vienna, Austria |
Figure 1. Comparison of model predictions with measurements for the vertical distribution of Cs-137 in the soil of an experimental plot near the Chernobyl Nuclear Power Plant in October 1986. The thin lines with the predictions from Models 1 and 2 represent the 95% subjective confidence intervals on the model predictions; the vertical lines represent the 95% confidence intervals on the measurements.
Figure 2. Comparison of model predictions with observations for the deposition density of Cs-137 in southern Finland following the Chernobyl accident in 1986. The vertical lines on most of the predictions indicate the 95% subjective confidence intervals on the model predictions; the thin horizontal lines indicate the 95% confidence intervals on the observations.
Figure 3. Comparisons of model predictions with measurements for the time-varying concentration of Cs-137 in human bodies in Central Bohemia (Czech Republic). The solid and dashed lines represent the best estimates and 95% subjective confidence limits, respectively, for the model predictions. The dark circles and vertical lines represent the means and 95% confidence intervals, respectively, of the measurements.
Figure 4. Example of compensatory error. The graph indicates the ratio of the predicted value to the observed value (P/O ratio) for I-131 deposition at two sites (Tokai, Japan, and Neuherberg, Germany) and for I-131 concentrations in forage and milk at the same sites. Model R gave good predictions (P/O ratio close to 1) for both the deposition and the concentration in milk. However, the high P/O values for forage indicate that the model overpredicted the transfer of deposited I-131 into forage and underpredicted the transfer of I-131 from forage to milk.
Figure 5. Comparison of model predictions with an estimate made from measurements for the internal dose to humans from ingestion of Cs-137 in Central Bohemia (Czech Republic) from 1986 to 1989. The vertical lines on some of the predictions indicate the 95% subjective confidence intervals on the model predictions, and the thin horizontal lines indicate the 95% confidence intervals on the estimated value. Models D1 and D2 refer to predictions made by two assessors with the same computer code.
References for International Model Validation Studies

General


Biospheric Model Validation Study (BIOMOVS)


BIOMOVS (Biospheric Model Validation Study). 1989. Irrigation with Contaminated Groundwater (H.A. Grogan, ed.). Swedish Radiation Protection Institute, Stockholm,


Biospheric Model Validation Study, Phase II (BIOMOVS II)


Radiation Protection Institute, BIOMOVS II Technical Report No. 10.


**Validation of Environmental Model Predictions (VAMP)**


The IAEA Programme on Biosphere Modelling and Assessment Methods (BIOMASS)


Final versions of the BIOMASS reports are expected to be published in 2002.

Updated December 19, 2001