International Congress on Environmental Modelling and Software

3rd International Congress on Environmental Modelling and Software - Burlington, Vermont, USA - July 2006

Jul 1st, 12:00 AM

Current Trends In Environmental Modelling With Uncertainty.

Jiří Hřebíček

Jan Pesl

Current Trends in Environmental Modelling with Uncertainty

J. Hrebícek[a], **M. Hejc**[b] **and I. Holoubek**[c]

[a] *Centre of Biostatistics and Analyses, Masaryk University, Brno, Czech Republic*
[b] *Faculty of Informatics, Masaryk University, Czech Republic*
[c] *Research Centre for Environmental Chemistry and Ecotoxicology, Masaryk University, Brno, Czech Republic*

Abstract: The important characteristic feature of environmental modelling is the complexity and uncertainty of its mathematical representation (uncertainty of formula). Imprecision of its input data is another characteristic feature, where it is not possible to omit influences of primary monitoring (e.g. gaps of data, errors of measuring facilities, human factor, etc). Many parameters in algorithms and their mathematical formulations are substituted by empirical constants in praxis, although it is well known that their values are very volatile and input data are not validated. Nowadays, information and communication technology (ICT) capabilities are growing rapidly and applied mathematical software (e.g. computer algebra systems, statistical packages, etc) becomes more powerful to overcome problems with formula complexity and uncertainty. The basic methods how to deal with the data uncertainties are well known and standardized from the last century, but some of their comparisons and recommendations for environmental modelling are not known enough. Paper presents generalized approach and shows universal methodology how to use current ICT tools for the implementation of mathematical models with formula and data uncertainties. The Checkland's soft system methodology is modified for its use by current ICT in environmental modelling with uncertainties. Further, results of the case study for the transport influence and all the related air pollution in the Czech Republic are presented. Various approaches for solving uncertainty with the computer algebra system Maple are simulated. The modification of the model COPERT III developed in Maple is almost free of any guessed emp irical values, but the results are still crisp enough and they are as useful (or more useful because of solving input volatility problem) as the original ones to analyze the situation and allow thinking about improvements of this specific environmental model**.**

Keywords: Modelling; Uncertainty analysis; Symbolic Algebraic Computation; Air pollution; Transport

1. INTRODUCTION

In practice, deeper knowledge about development of environmental models is gained through e.g. *uncertainty*, and *sensitivity analysis*; see [USEPA, 2003]. We remember about basic terminology of these analyses.

Uncertainties in the scientific sense are the component of all aspects of the environmental modelling process. They describe *lack of knowledge* about models, their parameters, constants, data, and beliefs. There are many sources of uncertainty, including: the science underlying a model, uncertainty in model parameters, scientific constants and input data, observation error, and implementation uncertainty. However, identifying the types of uncertainty that significantly influence environmental model outcomes (qualitatively or quantitatively) is the key to successfully integrating the solution of model into the knowledge about solved environmental problem. *Uncertainty analysis* investigates the effects of lack of knowledge or potential errors of model inputs (e.g. the "uncertainty" associated with its parameter values) and together in combination with sensitivity analysis informed about the confidence that can be placed in model results. Uncertainties can be divided into three interrelated categories [USEPA, 2003]:

Model framework uncertainty, i.e. uncertainty in the underlying science, the system of governing equations that make up the mathematical model and developed algorithms of this model which are the result of incomplete scientific data or lack of knowledge about the factors that control the behaviour of the system being modelled.

Data uncertainty is caused by *measurement errors*, analytical imprecision and limited sample sizes during the monitoring or collection and treatment of data.

Application niche uncertainty, i.e. uncertainty regarding the appropriate application of a model, e.g. using certain ICT tools. This is therefore a function of the appropriateness of a model for use under a specific set of conditions.

Sensitivity represents the degree to which the environmental model outputs are affected by changes in selected input parameters. *Sensitivity analysis* measures the effect of changes in input values or assumptions (including boundaries and model functional form) on the outputs [Morgan and Henrion, 1990]. It studies how uncertainty in a model output can be systematically apportioned to different sources of uncertainty in the model input [Saltelli, Tarantola, and Campolongo, 2000].

2. MODIFIED UNCERTAINTY ANALYSIS OF MODELS, ALGORITHMS AND DATA WITH USING ICT

2. 1 Introduction

The uncertainty analysis of environmental models and their algorithm implementation using ICT tools consists of following stages [Castrup, H., 2004]:

- *characterization of input uncertainties*, i.e. estimation of uncertainties in algorithm inputs and parameters;
- *uncertainty propagation*, i.e. estimation of the uncertainty in algorithm outputs resulting from the input uncertainties, see [Dong, 2002];
- *characterization of model uncertainty*, i.e. characterization of the uncertainties associated with different algorithm structures and model formulations, and
- *characterization of the uncertainties in algorithm predictions* resulting from uncertainties in the evaluation data.

Although the mathematical background for various methods of uncertainty analysis is quite well known for a long time, most of the environmental models are not working with the uncertainty at all. Instead, the scientific constant values or parameters of model are taken either from tables in books [Mohr and Taylor, 2000] or from some repositories managed by the responsible institutions (e.g. the EPA, EEA, etc.). If the above data are not available, than the measuring model parameters must be initiated (e.g. in appropriate laboratories) including data uncertainty analysis (EPA, GUM) and the values coming from here are used. In opposite case it is necessary either to give up or to adapt the parameters as the last step of the model development in such way that the results are matching reality.

2.2 New approach to uncertainty analysis

Proposed uncertainty analysis of environmental models with using ICT tools issues from following approaches:

Interval arithmetic is used to address data uncertainty that arises either due to imprecise measurements or due to the existence of several alternative methods, techniques of theories to estimate parameters [Kearfott and Kreinovich, 1996]. Especially when the probability structure of inputs is known, the application of interval analysis would in fact ignore the available information, and hence is not recommended.

Fuzzy theory is a method that facilitates uncertainty analysis of systems where uncertainty arises due to vagueness or fuzziness rather than due to randomness alone, [Uncertainty in Engineering, 2006]. Fuzzy theory appears to be more suitable for qualitative reasoning, and classification of elements into a fuzzy set, than for quantitative estimation of an uncertainty. The formal description of fuzzy randomness chosen by these authors is not suitable for formulating the uncertainty encountered in nonlinear structural analysis, e.g. in the nonlinear environmental models.

Probabilistic analysis is the most widely used method for characterizing uncertainty in environmental models, especially when estimates of the probability distributions of uncertain parameters are available [Helton and Davis, 2002]. The uncertainties are characterized as probabilities associated with events.

Methodology of Checkland. Peter Checkland stated in his book *Systems Thinking, Systems Practice* (published in 1981), that the complexity of the universe is beyond expression in any possible notation. His new Soft Systems Methodology (SSM) was an attempt to apply science to human activity systems. By examining the ecological systems in this manner, we can draw some knowledge about interaction and perception. This knowledge will help us in understanding and improving mathematical model of these systems. The SSM iterative approach is divided into seven distinct stages, forming a life cycle of mathematical model [Checkland, 1999], [Checkland and Poulter, 2006], e.g. ecological system:

1. *Finding out about the environmental problem situation*. This is basic research into the problem area. Who are the key players? How does the process work now?
2. *Expressing the environmental problem situation through so-called "Rich Picture"*. As with any type of diagram, more knowledge can be communicated visually. A picture is worth a thousand words.
3. *Selecting how to view the situation and producing root definitions*. From what different perspectives can we look at this environmental problem situation?
4. *Building conceptual environmental models of what the system must do for each root definitions*. We have basic "Whats" from the root definitions. Now we begin to define "Hows".
5. *Comparison of the conceptual environmental models with the real world*. We compare the results from steps 4 and 2 and see where they digger and are similar.
6. *Identify feasible and desirable changes*. Are there ways of improving the situation?
7. *Recommendations for taking action to improve the environmental problem situation*. How would we implement the changes from step 6.

Because each of above approaches needs different algorithm and data representation, many authors of the environmental models usually choose one of the approaches and keep it from the beginning to the end solving. Otherwise it will be necessary to create all algorithms and data structure sets more times in different forms. Moreover, the approaches are living their own independent lives and no uncertainty analysis of the model is available. *Therefore we present new trends in environmental modelling with uncertainties to show that the current ICT tools have no performance problems with uncertainty approach. They automate the algorithms developing and processing, and data obtaining and processing using the Internet so that it will be possible to try all the approaches with one data structure (possibly also combinations, not only one uncertainty representation in the whole model) and choose the proper uncertainty representation at the end of the modelling, not at the beginning*. These above four approaches were discussed and compared [Pešl, 2005] and some of them were already used in practice.

2.3 Appropriate ICT tools: Computer algebra based systems

Computer algebra based systems (CAS) involve the direct symbolic and algebraic computation (SAC) of the governing equations of mathematical models of environmental problem and also the estimation of the sensitivity and uncertainty of model outputs with respect to model inputs. The symbolic technology allows CAS to maintain all of the essential mathematical knowledge and structure inherent in a formula, equation, model, or program. Consequently, SAC can apply rules of mathematics to environmental problems and quickly produce answers that are much more meaningful than just numbers or graphs. For example, often, the actual solution of the problem is not the final step as many applications require further mathematical processing after a given solution computation. SAC approach provides greater flexibility for post-processing because the system maintains the history of its computation and is not just a black box. CAS maintains all of the underlying mathematical structure including those of previous expressions that were used to create an expression. By applying the associated packages of mathematical and graphical operations, one can analyze sensitivities of parameters, convergence of solutions, parametric dependencies, and much more.

The process of environmental modelling using CAS consists of the spiral cycle IDENTIFY – DEVELOP – IMPLEMENT – SOLVE – ANALYZE – MODIFY (Figure 1), which shows the way how complex CAS automate all phases of environmental modelling.

Today there are various CAS, from the simple utilities to complex systems [Gander and Hrebícek, 2004]. The known SAC systems are e.g. Yacas, HartMath, The OpenXM project, Prologie, GiNaC, ArtLandia, Axiom, CoCoA, Derive, Algebra Domain Constructor, Fermat, GAP, GANITH, GRG, GRTensor, LiDIA, GNU DOE Maxima, Magma, Maple, Mathematica, Mathomatic, MathSoft, MATLAB, MathTensor, Milo, MP, MuPAD, NTL, Pari, Reduce, Schur, Singular, SymbMath, TI-92 Calculator, and TI-92 Plus. Further, we will concentrate on Maple - one of the most used complexes CAS. It has own programming language and exports its worksheets into MathML, LaTeX, RTF, HTML and XML files or Java, C#, Fortran and MS Visual Basic languages. Its suitable tools for network communication enable connecting Maple to processes on remote hosts on a network (such as an Intranet or the Internet) and exchange data or program codes with these processes.

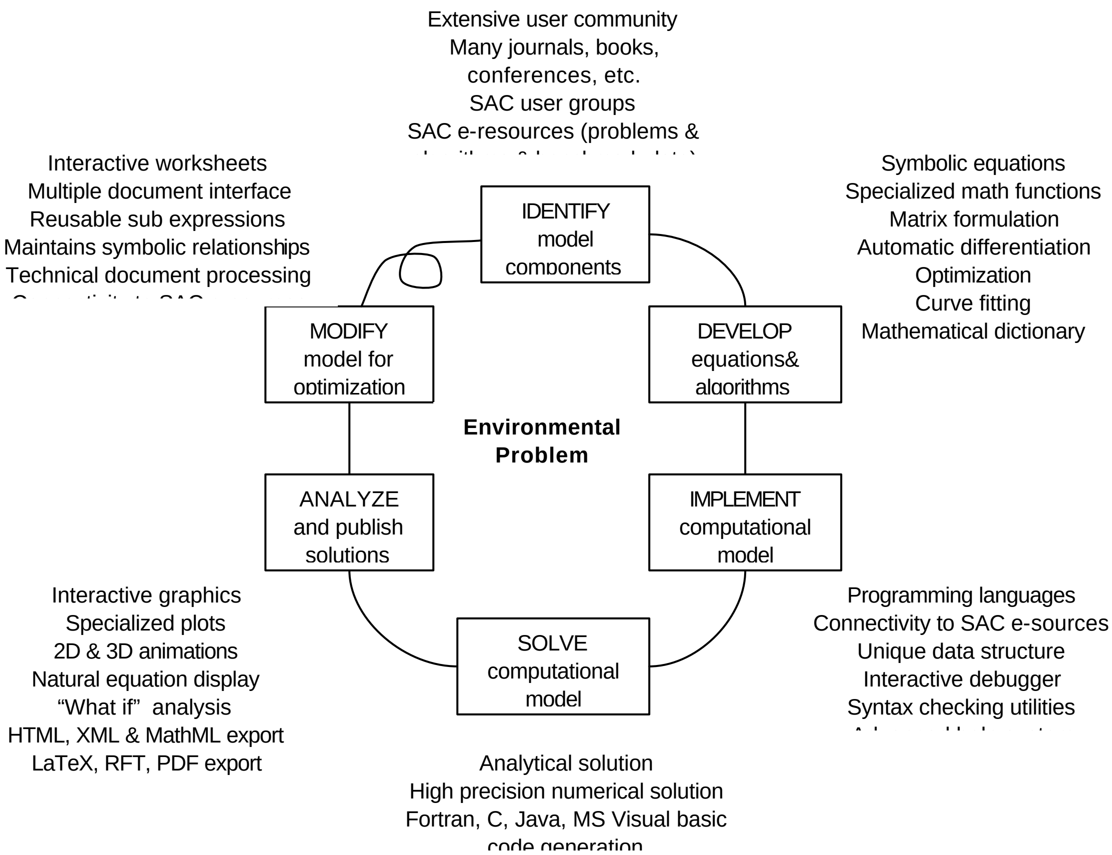

Figure 1. Life cycle (spiral) of environmental modelling using SAC

3. UNCERTAINTY ANALYSIS SUPPORT USING MAPLE

At first, Maple tools dealing with the uncertainty are introduced. Of course, the Checkland's SSM is not automatically implemented in Maple. Its usage in our case study will be described in the following chapter.

Interval arithmetic implementation. The package *intpakX* of Maple provides basic data types and operations for interval arithmetic as well as additional features for further interval computation. It contains the type checking functions, all arithmetic functions including powers, trigonometric and hyperbolic ones, set operations on the interval, range operations for a given function, complex number support and some basic numeric methods as the Newton's method for finding a root of an uncertain function.

Fuzzy theory implementation. The *Fuzzy Sets* toolbox of Maple allows constructing and working with fuzzy subsets of both the real line and of user-defined finite sets. Its modules automatically generate fuzzy controllers from a collection of user-defined rules. This allows modelling, testing, and modifying fuzzy systems in the interactive Maple worksheet environment.

Probabilistic analysis. The *ScientificErrorAnalysis* package of Maple provides representation and construction of numerical quantities in Maple that has a central value and associated uncertainty or error, which is some measure of the degree of precision to which the quantity's value is known. The associated uncertainty can be specified in absolute, relative, or units in the least digit form. In the returned object, the uncertainty is quantified in absolute form.

Getting the online data and program codes. The *Sockets* package of Maple allows getting data and program codes for the computation online from the web. In particular, it enables two independent Maple processes running on different computers on a network to communicate with one another.

4. CASE STUDY: AIR POLLUTION BY THE TRANSPORT IN CZECH REPUBLIC

The emissions from transport in the Czech Republic has been analyzed with respect to uncertainties using ICT tools of Maple [Hrebícek, Holoubek and Pešl, 2005], [Pešl, 2005], where the implemented mathematical model of transport air emissions in Maple has issued from the well-known mathematical model COPERT III [Ntziachristos and Samaras, 2000]. Therefore, we will not describe its mathematical equations here. The COPERT III methodology is assumed to reflect real world conditions, but it is not fully clear from its documentation to what extent fuel consumption estimates have been based on official test cycle results and to what extent they are based on measurement of real world cycles. Currently, the EU funded projects ARTEMIS and PARTICULATES are further extending the knowledge on emission factors for all transport modes and all pollutants and we have taken into account their results and recommendations. We have introduced another set of emission factors for the computations at the local level. These factors do not represent the pollutant emission from one kilogram of fuel, but from travelling one kilometre. It looks more suitable to make comparison of these two emission factor sets, but the final answer is still opened. Therefore we try to rearrange the model COPERT III in the following way:

- Treating the emission factors, fuel consumptions and transport powers as uncertain.
- Unifying the formulas for various pollutants.
- Unifying the formulas for various transport types.

Of course, not all of these changes must be desirable, but using the Checkland's SSM iterative approach has allowed us to change the model afterwards, taking into account its bad properties which were not corresponded to the situation in the real world.

The original COPERT III methodology has been improved with respect to possibilities of uncertainty analysis in Maple, for the calculation of emissions, which are measured and statistically estimated.

Selected emission factors, which are based on measured values, used probabilistic approach, and further the direct dependence of the relationship of transport performances given in passenger kilometres or ton-kilometres were eliminated.

The results presented on Figures 2 and Figure 3 of Carbon dioxide (CO_2) and VOC emissions were obtained after two iterations of the SSM.

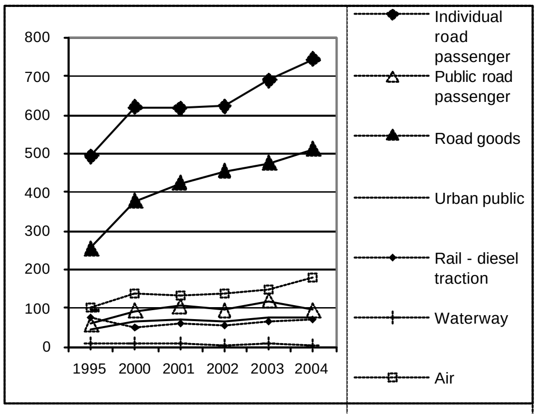

Figure 2. Carbon dioxide (CO_2) emissions (kg/inhabitant) generated in the Czech Republic by all types of transport.

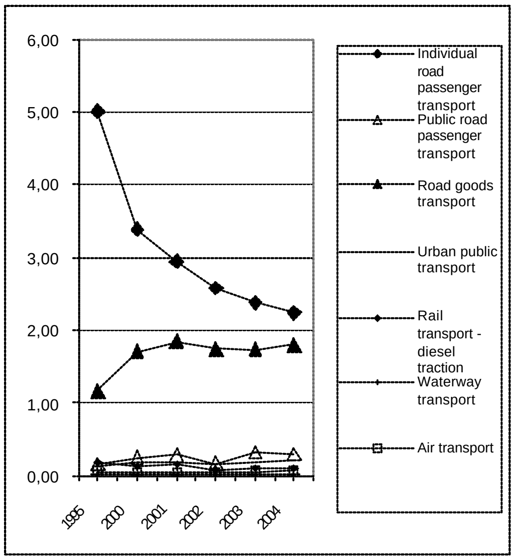

Figure 3. VOC emissions (kg/inhabitant) generated in the Czech Republic by all types of transport.

5. CONCLUSIONS

We can see that the uncertainty handling is not further problematic in environmental modelling using current ICT tools. Our several years' research at Masaryk University [Hrebícek, Pešl, 2003, 2005], [Pešl, 2005] has shown that introducing uncertainty into environmental modelling is suitable. Deeper knowledge of the mathematical model and the data together with uncertainty and sensitivity analysis can show how much the input uncertainty influence the outcome of the model. Classification of the parameters and the data into clusters (where some of them are sufficient to be known roughly and some of them more accurately) can divide the problem of uncertainty into parts, solved by different approaches (interval arithmetic, fuzzy and probabilistic theory).

The further opened question is the sense of using the SSM because now it seems that we need at least two iterations to get as good results as the original model. Such approach makes environmental modelling little bit slower, but the iteration guarantees that there are no useless formulas and keeps the model complexity at the lower bound corresponding to the results we have.

We can conclude that the uncertainty analysis has been done without big problems, including the theoretical and practical comparison of the approaches and redesigning the case study model in this way. The model simplification of case study has been considered as partially problematic, some formulas have been successfully eliminated, but other simplifications have to be revised. This revision demand corresponds to the SSM and this philosophical approach seems to be very suitable in the field of environmental modelling. Obtaining the data from the Internet and using one set for various uncertainty handling approaches has been done at the basic level, showing the possibility well, but with omitting the perfect error-prone interface.

There is still a lot of research in the area of environmental modelling and uncertainty analysis, but we hope this presented investigation helped a little at least to give the future research the proper direction - to have environmental models with acceptable complexity, covering the influence factors and using current ICT.

6. REFERENCES

Castrup, H., Estimating and combining uncertainty Analysis Tools, Proc. 8th Annual ITEA Instrumentation Workshop, Lancester, Canada, May 2004.

Checkland, P., *Soft Systems Methodology in Action*, John Wiley & Sons, 418 pp., Chichester, 1999.

Checkland, P., and J. Poulter, *Learning For Action: A Short Definitive Account of Soft Systems Methodology, and its use Practitioners, Teachers and Students*. John Wiley & Sons, 192 pp., Chichester, 2006.

Dong, Q., Reliability of Uncertainty Assignments and Its Impact on Generating Recommended Data. Proc. 18th CODATA Conference, Montreal, Canada, September 2002

Gander, W., Hrebicek, J., *Solving Problems in Scientific Computing Using Maple and Matlab*. 4th. exp. and rev. ed. Springer, 476, Berlin, 2004.

Helton, J.C., and F.J., Davis, Illustration of Sampling-Based Methods for Uncertainty and Sensitivity Analysis. *Risk Analysis*, 22(3), 591-622, 2002

Hrebícek, J., and J., Pešl, Soft Systems Methodology Applied to Environmental Modelling. *Environmental Informatics Archives*, 1(2003), 261-266, 2003.

Hrebícek, J., I., Holoubek, and J., Pešl, Environmental Modelling with Uncertainness – Case Study of Air. In The International Environmetrics Society TIES2005 Being, China, August 2005.

Hrebícek, J., Pešl, J., Dealing with estimated parameters in the environmental models. In Proc. ITEE 2005 Second International ICSC Symposium on Information Technologies in Environment Engineering,. Magdeburg, Germany, p. 655-664, October 2005.

Kearfott R.B., Kreinovich V., *Applications of Interval Computations*, Kluwer, 444 pp., Boston, 1996.

Mohr, P., J., and B., N., Taylor, CODATA recommended values of the fundamental physical constants: 1998. *Rev Mod Phys*, 72(2), 351-495, 2000.

Morgan, G., and M. Henrion, The Nature and Sources of Uncertainty. In: Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge: Cambridge University Press. pp. 47-72. 1990.

Ntziachristos L. and Samaras Z., COPERT III Computer Programme to Calculate Emissions from Road Transport, Methodology and Emission Factors, European Environment Agency, Copenhagen, 2000.

Pešl, J., Uncertainty handling in the environmental modelling using computer algebra systems with online data manipulation, Ph. D. Thesis, Faculty of Infromatics, Masaryk University, Brno, 2005

Saltelli, A., S. Tarantola, and F. Campolongo, Sensitivity analysis as an ingredient of modelling. *Statistical Science*, 15(377-395), 2000.

USEPA, Draft Guidance on the Development, Evaluation, and Application of Regulatory Environmental Models (CREM), 2003. http://cfpub.epa.gov/crem/cremlib.cfm

ARTEMIS: http://www.trl.co.uk/artemis/

Maple: http://www.maplesoft.com

PARTICULATES: http://lat.eng.auth.gr/particulates/

Uncertainty in Engineering: http://www.uncertainty-in-engineering.net/