ELSEVIER

Computer Physics Communications 117 (1999) 142–155

Computer Physics Communications

Scenario and parametric uncertainty in GESAMAC: A methodological study in nuclear waste disposal risk assessment

David Draper[a,1], Antonio Pereira[b], Pedro Prado[c], Andrea Saltelli[d], Ryan Cheal[a], Sonsoles Eguilior[c], Bruno Mendes[b], Stefano Tarantola[d]

[a] *University of Bath, UK*
[b] *University of Stockholm, Sweden*
[c] *CIEMAT, Madrid, Spain*
[d] *European Commission, Joint Research Centre, Ispra, Italy*

Received 30 April 1998

Abstract

We examine a conceptual framework for accounting for all sources of uncertainty in complex prediction problems, involving six ingredients: past data, future observables, and scenario, structural, parametric, and predictive uncertainty. We apply this framework to nuclear waste disposal using a computer simulation environment – GTMCHEM – which "deterministically" models the one-dimensional migration of radionuclides through the geosphere up to the biosphere. Focusing on scenario and parametric uncertainty, we show that mean predicted maximum doses to humans on the earth's surface due to I-129, and uncertainty bands around those predictions, are larger when scenario uncertainty is properly assessed and propagated. We also illustrate the value of a new method for global sensitivity analysis of model output called *extended FAST*. © 1999 Elsevier Science B.V.

Keywords: Bayesian prediction; Extended FAST; Level E/G test case; Parametric uncertainty; Scenario uncertainty; Sensitivity analysis

1. Introduction: the importance of prediction

It is arguable (e.g., [9]) that prediction of observable quantities is, or at least ought to be, the central activity in science and decision-making: bad models make bad predictions (that is one of the main ways we know they are bad). Prediction (almost) always involves a *model* embodying facts and assumptions about how past observables (data) will relate to future observables, and how future observables will relate to each other. Full deterministic understanding of such relationships is the causal goal, rarely achieved at fine levels of measurement. Thus we typically use models that blend determinism and chance,

$$\begin{aligned} y_i &= f(x_i) + e_i, \\ \text{observable} &= \begin{pmatrix}\text{"deterministic"}\\ \text{component}\end{pmatrix} + \begin{pmatrix}\text{"stochastic"}\\ \text{component}\end{pmatrix}, \end{aligned} \tag{1}$$

for outcome y and predictor(s) x (often a vector), as i ranges across the observations from 1 to n (say), and in which we may or may not pretend that f is known. The hope is that successive refinements of causal un-

[1] School of Mathematical Sciences, University of Bath, Claverton Down, Bath BA2 7AY, England; Tel: +44 1225 826222; Fax: +44 1225 826492; Email d.draper@maths.bath.ac.uk.

PII S0010-4655(98)00170-2

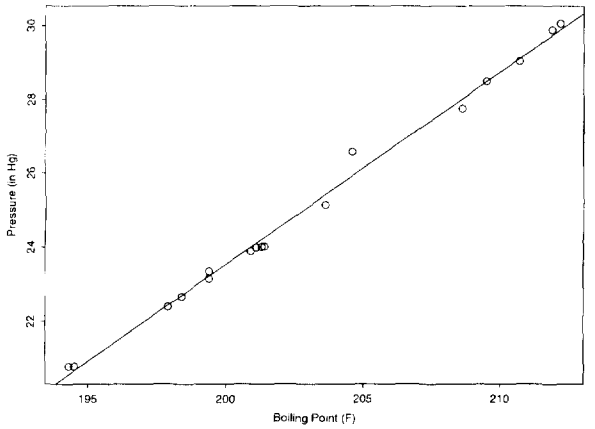

Fig. 1. Scatterplot of Forbes' data, from [29], relating barometric pressure to the boiling point of water, with linear fit superimposed.

derstanding over time will add more components to x, shifting the bulk of the variation in y from e to f, until (for the problem currently under study) the stochastic part of the model is no longer required.

In the period in which causal understanding is only partial, many uncertainties are recognizable in attempts to apply Eq. (1). An initial example – illustrating in a simple setting all the ingredients found in the more complicated case study in Sections 2 and 3 – is given by *Forbes' Law* [29], which quantifies the relationship between the boiling point of water and the ambient barometric pressure. Around 1750 the Scottish physicist Forbes collected data on this relationship, at 17 different points in space and time in the Swiss Alps and Scotland, obtaining the results plotted in Fig. 1.

In the problem of interest to Forbes, denoting pressure by y and boiling point by x, a deterministic model for y in terms of x would be of the form

$$y_i = f(x_i)\,, \quad i = 1, \ldots, n = 17\,, \tag{2}$$

for some function f whose choice plays the role of an assumption about the *structure* of the model. For simple f this model will not fit the data perfectly, e.g., because of imperfections in the measuring of x. A stochastic model describing the imperfections might look like

$$y_i = f(x_i) + e_{1i}\,,$$
$$e_{1i} \overset{\text{IID}}{\sim} N(0, \sigma_1^2)\,. \tag{3}$$

Here the e_{1i} represent *predictive uncertainty* – residual uncertainty after the deterministic structure has done its best to "explain" variations in y – and σ_1 describes the likely size of the e_{1i}.

To specify structure, an empiricist looking at Fig. 1 might well begin with a linear relationship between x and y,

$$S_1 : \; y_i = \beta_{10} + \beta_{11} x_i + e_{1i}\,, \; V(e_{1i}) = \sigma_1^2\,, \tag{4}$$

with β_{10} and β_{11} serving as *parameters* (physical constants) whose values are unknown before the data are gathered. However, nonlinear relationships with small curvature are also plausible given the Forbes data, e.g.,

$$S_2 : \; \log(y_i) = \beta_{20} + \beta_{21} x_i + e_{2i}\,, \; V(e_{2i}) = \sigma_2^2\,,$$
$$S_3 : \; y_i = \beta_{30} + \beta_{31} \log(x_i) + e_{3i}\,, \; V(e_{3i}) = \sigma_3^2\,,$$
$$S_4 : \; \log(y_i) = \beta_{40} + \beta_{41} \log(x_i) + e_{4i}\,,$$
$$V(e_{4i}) = \sigma_4^2\,. \tag{5}$$

Closer examination of Fig. 1 does in fact reveal a subtle upward curvature – Forbes himself theorised that the logarithm of pressure should be linear in boiling point, corresponding to structure S_2.

If you were proceeding empirically in this situation, your *structural* uncertainty might be encompassed, at least provisionally, by $\mathcal{S} = \{S_1, \ldots, S_4\}$, with each element in $\mathcal{S}$ corresponding to a different set of parameters, e.g., $\theta_{S_1} = (\beta_{10}, \beta_{11}, \sigma_1), \ldots, \theta_{S_4} = (\beta_{40}, \beta_{41}, \sigma_4)$. Note that changing from the raw scale to the log scale in x and y makes many components of the parameter vectors θ_{S_j} not directly comparable as j varies across structural alternatives.

Parametric uncertainty, conditional on structure, is the type of uncertainty most familiar to quantitative workers. With the data values in Fig. 1, inference about the β_{jk} and σ_j can proceed in standard Bayesian or frequentist ways, e.g., with little or no prior information the $n = 17$ Forbes observations – conditional on structure S_2 – yield $\hat{\beta}_{20} = -0.957 \pm 0.0793$, $\hat{\beta}_{21} = 0.0206 \pm 0.000391$, $\hat{\sigma}_2 = 0.00902 \pm 0.00155$ (although a glance at Fig. 1 reveals either a gross error in Forbes' recording of the pressure for one data point or a measurement taken under sharply different meteorological conditions).

The final form of potential uncertainty recognizable from the predictive viewpoint in Forbes' problem is *scenario* uncertainty, about the precise value x_i of fu-

ture relevant input(s) to the structure characterising how pressure relates to boiling point. For example, if in the future someone takes a reading of pressure at an altitude where the boiling point is $x^* = 200°$F, Forbes' structural choice (S_2) predicts that the corresponding log pressure $\log(y^*)$ will be around $\beta_{20} + \beta_{21}x^*$, give or take about σ_2. Here you may know the precise value of the future x^*, or you may not.

Thus (see [6] for more details) six ingredients summarise the four sources of uncertainty, arranged hierarchically, in Forbes' problem:

- *Past observable(s) D*, in this case the bivariate data in Fig. 1.
- *Future observable(s) y*, here the pressure y^* when the boiling point is x^*.
- Model *scenario* input(s) x, in this case future value(s) x^* of boiling point (about which there may or may not be uncertainty).
- Model *structure* $S \in \mathcal{S} = \{S_1, \ldots, S_k\}, k = 4$. In general structure could be conditional on scenario, although this is not the case here.
- Model *parameters* θ_{S_j}, conditional on structure (and therefore potentially on scenario); and
- Model *predictive* uncertainty, conditional on scenario, structure, and parameters, because even if these things were "known perfectly" the model predictions will still probably differ from the observed outcomes.

Each of the last four categories – scenario, structural, parametric, and predictive – is a source of uncertainty which potentially needs to be assessed and propagated if predictions are to be *well-calibrated* [4], e.g., in order that roughly 90% of your nominal 90% predictive intervals for future observables do in fact include the truth.

2. Nuclear waste disposal risk assessment: The case of GESAMAC

A key issue in the consolidation process of the nuclear fuel cycle is the safe disposal of radioactive waste. Although in the 1970s disposal of nuclear waste in deep sea sediments was considered [2], and despite some objections today [17], deep geological disposal based on a multibarrier concept is at present the most actively investigated option. Visualise a deep underground facility within which radioactive materials such as spent fuel rods or reprocessed waste, previously encapsulated, are emplaced, surrounded by other man-made barriers. While the safety of this concept ultimately relies on the safety of the mechanical, chemical and physical barriers offered by the geological formation itself, the physico-chemical behaviour of such a disposal system over geological time scales (hundreds or thousands of years) is far from known with certainty [21,24].

We have been involved since 1996 in a project for the European Commission, GESAMAC[2], which aims in part to capture all relevant sources of uncertainty in predicting what would happen if the disposal barriers were compromised in the future by processes such as *geological faulting, human intrusion,* and/or *climatic change*. One major goal of the project has been the development of a methodology to predict the radiologic dose for people in the biosphere as a function of time, how far the disposal facility and the other components of the multibarrier system are underground, and other factors likely to be strongly related to dose. The emphasis of the study is on the methodology, including the use of uncertainty and sensitivity analysis.

2.1. Description of the system model

The system model on which our work is based consists of a hypothetical underground radioactive waste disposal system represented by three coupled submodels: the *near field* (the source term), a *far field* (the geosphere), and a *biosphere*. The first submodel – the near field (the repository itself) – does not include any consideration of spatial structure or chemical complexities. It assumes an initial containment time for the wastes (only radioactive decay is considered), followed by a constant leaching rate of the inventory present at the time containment fails. The third submodel – the biosphere – is very simple and assumes that the radionuclides leaving the geosphere enter a stream of water from which a human population obtains drinking water, so that the dose received depends on the ratio of the drinking water consumption to the stream flow rate. This is clearly not a real,

[2] GEosphere modeling, geosphere Sensitivity Analysis, Model uncertainty in geosphere modeling, Advanced Computing in stochastic geosphere simulation: see http://www.ciemat.es/sweb/gesamac/.

site-specific safety study, but a simplified setup for the illustration of the methodology.

The main focus of our computer simulation work deals with the second submodel, the geosphere. The first version of our code, GTM-1 [22,24,25], was developed in 1989[3] and tested via extensive Monte Carlo simulations, through its inclusion in several versions of the *LISA* code and in the international *PSACOIN* benchmark exercises [19,20]. The new release of our code, GTMCHEM, developed in the framework of GESAMAC, expands on GTM-1 by incorporating other chemical phenomena than adsorption by linear isotherm (the only process included in the first release).

The geosphere submodel in GTMCHEM estimates the transport of radionuclides by groundwater through the geologic formations, represented by a one-dimensional column of porous material whose properties can change along the pathway and in which different chemical reactions (homogeneous or heterogeneous) can take place. The equation solved is

$$\frac{\partial C_i}{\partial t} = -V\frac{\partial C_i}{\partial X} + D\frac{\partial^2 C_i}{\partial X^2} + \mathrm{SoSi}\,, \tag{6}$$

where C represents concentration (mols/m^3), t is time (yr), X is the space coordinate (m), V is the groundwater velocity (m/yr), D is the hydrodynamic dispersion (m^2/yr), and SoSi is the source/sink term in which the chemical reactions are included.

Apart from the radioactive decay and the linear retention factor, the chemical phenomena modeled by GTMCHEM in the SoSi term are [7]: equilibrium complexation in solution; homogeneous first-order chemical kinetics in solution; slow reversible adsorption; and a sink associated with filtration or biodegradation. It is up to the user to choose which phenomena to be included in the simulation.

GTMCHEM solves Eq. (6) by means of a two-step procedure: first the advective and dispersive terms (and the equilibrium reaction in solution) are approximated by finite differences, following an implicit Crank–Nicolson scheme; and then the intermediate concentration values obtained in the first step are used to solve the physico-chemical reactions considered in the SoSi term. The sum of the values obtained by the two intermediate steps yields the final concentration values at each new time point. The output of GTMCHEM is "deterministic" in the sense that identical inputs will always lead to identical outputs.

For slow chemical reactions the error of decoupling the transport and chemical reaction parts is negligible if

$$T_C = \frac{\Delta X\, R}{V} < T_R \quad \text{and} \quad T_D = \frac{\Delta X^2\, R}{D} < T_R\,, \tag{7}$$

where T_R, T_C, and T_D are the time scales of the chemical reactions, convective transport, and dispersive transport, respectively. For heterogeneous reactions the two-step procedure without iteration leads to greater truncation errors than in the homogeneous case, but a good approximation still results if Eq. (7) is fulfilled. By definition this condition cannot be met for fast exchange.

The user, through logical options in the input file, selects the system model to be simulated (e.g., alternative source term models and analytical solutions) as well as the output desired. The code produces the peak fluxes at the end of each geosphere layer and the peak dose for each nuclide, as well as the associated time of the peak. The program may also be run with a fixed set of time points at which the fluxes and/or doses through space may be saved. Although this simplified mono-dimensional code is likely to be replaced by site specific 3D representations, GTMCHEM is good at modelling slow chemical processes involving the stationary phase, thus allowing the investigation of the relevance of these processes.

2.2. *The GESAMAC uncertainty framework*

The uncertainty framework developed above in Section 1 has a direct parallel in the GESAMAC context, as follows (Fig. 2):

- *Past data D*, if available, would consist of readings on radiologic dose under laboratory conditions relevant to those likely to be experienced in the geosphere or biosphere. (Fortunately for humankind, but unfortunately for the creation of a *predictive accuracy feedback loop* in our modeling – which would allow us to assess the most plausible structural and parametric possibilities – there have been

[3] Supported by a collaborative contract between the European Commission Joint Research Centre in Ispra and ENRESA, the Spanish radioactive waste management company, in Madrid.

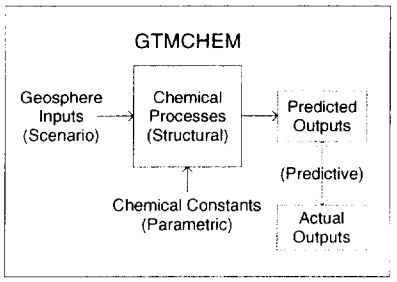

Fig. 2. Schematic illustration of the four sources of uncertainty in GESAMAC.

no accidents to date of the type whose probabilities we are assessing.)

- *Future observables* y^* consist of dose values at given locations L, t years from now, as L and t vary over interesting ranges.
- *Scenarios* $\mathcal{X}$ detail different sets of likely geosphere conditions at locations L and times t, as a result of human intrusion, faulting, and/or climate. We have found it useful to conceptualise scenario uncertainty in two parts:
  - *Macro-scenarios*, consisting of high-level statements of future geosphere conditions relevant to dose, such as climatic change; and
  - *Micro-scenarios*, which are low-level characterisations of how the macro-scenarios – e.g., how forces of climatic change such as erosion and deposition – would unfold chemically.
- *Structural possibilities* $\mathcal{S}$ include different combinations of chemical processes (e.g., sorption, equilibrium, and matrix diffusion) and different sets of partial differential equations (PDEs, such as (6)) to model them.
- *Parametric* uncertainty arises because the precise values of some of the relevant physical constants appearing in the PDEs are unknown. Note that parameters may be specific not only to structure but also to scenario (e.g., an early ice-age climatic scenario would have certain chemical constants driving it, whereas a worst-case geologic fracture scenario would be governed by different constants); and
- *Predictive* uncertainty is as speculative (at present) as past data in this project, and might be based on things like discrepancies between actual and predicted lab results, extrapolated to field conditions.

2.3. Uncertainty calculations

With the six ingredients above, the goal in uncertainty propagation is to produce two types of predictive distributions: *scenario-specific* and *composite*. The only hope of doing this in a way that captures all relevant sources of uncertainty appears to be a fully Bayesian analysis (e.g., [5]).

In the Bayesian approach past data D (if any) are known; future observable outcome(s) y^* are unknown, and to be predicted; and we must pretend that the sets $\mathcal{X}$ and $\mathcal{S}$ of possible scenarios and structures are known. Then the *scenario-specific* predictive distribution $p(y^*|\mathcal{S}, x, D)$ for y^* given $D, \mathcal{S}$, and a particular scenario x is given by

$$p(y^*|\mathcal{S}, x, D) = \int_{\mathcal{S}} \int_{\Theta} p(y^*|\theta_S, S, x)\, p(\theta_S|S, x, D) \times p(S|x, D)\, d\theta_S\, dS\,, \tag{8}$$

and the *composite* predictive distribution $p(y^*|\mathcal{S}, \mathcal{X}, D)$ for y^* given $D, \mathcal{S}$, and $\mathcal{X}$ is

$$p(y^*|\mathcal{S}, \mathcal{X}, D) = \int_{\mathcal{X}} p(y^*|\mathcal{S}, x, D)\, p(x|D)\, dx\,. \tag{9}$$

Here $p(y^*|\theta_S, S, x)$ is the *conditional predictive distribution* for y^* given specific choices for scenario, structure, and parameters, and $p(\theta_S|S, x, D)$, $p(S|x, D)$, and $p(x|D)$ are *posterior distributions* for the parameters, structure, and scenario (respectively) given the past data. Each of these posterior distributions depends on *prior distributions* in the usual Bayesian way, e.g., the posterior $p(S|x, D)$ for structure given the data and a particular scenario x is a multiplicative function of the prior $p(S|x)$ on structure and the likelihood $p(D|S, x)$ for the data given structure,

$$p(S|x, D) = c\, p(S|x)\, p(D|S, x)\,, \tag{10}$$

where c is a normalising constant.

2.4. Challenges to the Bayesian approach

This approach to full uncertainty propagation involves two major types of challenges: technical and substantive.

- *Technical challenge:* Computing with Eqs. (8)–(10) above requires evaluation of difficult, often high-dimensional integrals – for instance, the likelihood $p(D|S,x)$ in Eq. (10) is

$$p(D|S,x) = \int_{\Theta} p(D|\theta_S, S, x)\, p(\theta_S|S)\, d\theta_S\,, \quad (11)$$

and the parameter vector θ_S given a particular structure S may well be of length $l > 50$. The leading current technology for overcoming this challenge is *(Markov Chain) Monte Carlo integration* (e.g., [10,11]).
- *Substantive challenges:*
  - **Q:** How can you be sure that $\mathcal{X}$ contains all the relevant scenarios, and $\mathcal{S}$ contains all the plausible structural choices?
    A: You can't; in practice you try to be as exhaustive as possible given current understanding and resource limitations. There is no good way in this (or any other) approach to completely hedge[4] against unanticipated combinations of events that have never happened before.
  - **Q:** Where do the prior distributions $p(x)$ and $p(S)$ on scenarios and structures come from?
    A: One good approach [4,6] is to start with expert judgement, use sensitivity analysis (SA) to see how much the final answers depend on these priors, and tune them using *predictive calibration*: (1) compare the observed outcomes to their predictive distributions given past data – if the observed outcomes consistently fall in the tails, then the priors may have been inaccurately specified, so (2) respecify them and go back to (1), iterating until the predictions are well-calibrated.

[4] At the SAMO98 meeting at which this paper was presented, a booklet of promotional material featured a product for risk analysis. The advertising for this product included the section heading, in large, bold letters, "Account for all possible events – not just the most likely ones!" and went on to claim, "With [this product], your spreadsheet model goes from representing one possible scenario ... to several hundred ... to all possible scenarios – just by running a simulation!" This is arrant nonsense, of course; the result of any such exercise is always conditional on the set $\mathcal{X}$ upon which the simulations are based, and this set (almost) always falls far short of "all possible events".

3. GESAMAC preliminary results

We have used Monte Carlo methods to approximate the integrals in Eqs. (8), (9). For instance, to simulate in a way that fleshes out all four sources of uncertainty in Section 2, we would first draw a scenario at random according to an appropriate probability distribution, and then select one or more structural choices (e.g., chemical processes and/or PDEs to implement them) internal to our computer program GTMCHEM according to a second probability distribution specific to the chosen scenario. Parameters (chemical constants) specific to the chosen structure(s) would then be chosen according to a further set of appropriate probability distributions, yielding one or more GTMCHEM outputs, e.g., predicted maximum dose and dose values at location(s) L and time(s) t. We would then compare these with actual outputs (if available) to estimate the likely size of GTMCHEM's prediction errors; making a final series of draws (one for each value of L and t) from probability distributions to incorporate predictive uncertainty, which would then be added to GTMCHEM's predicted outcome values, would complete one iteration of the Monte Carlo. Repeating this simulation many times, making histograms or density traces of the resulting outputs, amounts to approximating the desired predictive distributions by Monte Carlo integration.

Since we have no past data D available on actual underground accidents (and we also have not made use of laboratory data in our work so far), there is no updating in the results given here from, e.g., a prior distribution $p(x)$ on scenarios to the corresponding posterior distribution $p(x|D)$. If we had past data available to support such updating, the Monte Carlo integration would become more complicated; Markov chain Monte Carlo (MCMC, e.g., [10]) would then become a natural simulation-based alternative.

To obtain the preliminary results reported here,

- We focused on the scenario and parametric inputs to GTMCHEM of greatest interest in the standard reference test case in the nuclear safety community, the *PSACOIN Level E Intercomparison* [19]. The Level E test case tracks the one-dimensional migration of four radionuclides – iodine (I-129), and a chain consisting of neptunium (Np-237), uranium (U-233), and thorium (Th-229) – through two geosphere layers characterised by different hydro-

Table 1
Example of parametric inputs to GTMCHEM in the simulation study: Fast Pathway scenario, iodine I-129 nuclide

Variable	Meaning	Distribution	Raw-scale	
			Min	Max
CONTIM	No-leakage containment time	uniform	100	1000
RLEACH	Leach rate after containment failure	log uniform	.001	.01
VREAL1	Geosphere water travel velocity in layer 1	log uniform	.1	1
XPATH1	Geosphere layer 1 length	uniform	200	500
RET1	Layer 1 retardation coefficient	uniform	1	2
STREAM	Stream flow rate	log uniform	1e4	1e6
C21F	Slow reversible adsorption forward rate	uniform	1e−9	1e−7
C21B	Slow reversible adsorption backward rate	uniform	1e−9	1e−7

geological properties (sorption, hydrodynamic dispersion, and ground water velocity);

- We developed a new test case called *Level E/G* ([23]) by modifying a reference scenario in five different ways, to create a total of six scenarios:
  - Reference (REF): Level E;
  - A fast pathway (FP) to the geosphere, corresponding to a geological fault passing directly through the containment chamber, or to the reduction of the geosphere pathway by erosion of the upper layer, or to the bypassing of the second layer through human activities. This scenario thus represents a reduction in radionuclide travel time through a reduction in the geosphere pathlength;
  - An additional geosphere layer (AG), the opposite situation from the previous scenario. This case arises, for instance, from a retreating glacier leaving behind another barrier layer between the repository and the biosphere, or when a geological event creates an alternative pathway that is longer than that in the reference case;
  - Glacial advance (GA), related to the AG scenario but arising from an advancing rather than retreating glacier;
  - Human disposal errors (HDE), corresponding to deficiencies in the construction of the repository and/or in waste disposal operations leading to premature failing of the near-field barriers; and
  - Environmentally induced changes (EIC), arising from human activities or geological events that indirectly are responsible for the modification of the disposal system conditions, such as the drilling of a pumping well or mining tunnel at a dangerously small distance from the containment chamber.

  These test cases were arrived at by creating a total of nine micro-scenarios – three in each of the categories *geological changes, climatic evolution*, and *human activities* – and merging similar micro-scenarios into the five non-Reference scenarios listed above; and
- We ignored structural and predictive uncertainty, focusing only on the scenario and parametric components of overall uncertainty.

3.1. Radionuclide migration calculations and the parallel Monte Carlo driver

The Monte Carlo simulations of iodine migration were made by coupling the transport code GTMCHEM to a program we have written called the Parallel Monte Carlo Driver (PMCD). PMCD is a software package for risk assessments developed for the GESAMAC project and written for a parallel computing environment. In developing this package, the aim was to offer to potential users of the code a high-performance computing tool that is user-friendly. The package is implemented using the Message Passing Interface (MPI) and the code presently runs on an IBM-SP2 parallel supercomputer.

Monte Carlo simulations are an ideal application for parallel processing in that every simulation can be made independently of the others. We have used the Single Program Multiple Data paradigm (SPMD) together with a master-slave approach. In brief, PMCD generates an input data matrix (master node), the rows of which are input parameters to the GTMCHEM code coupled to it. These parameters are sampled from

Table 2
Scenario-specific output summaries; note how much more I-129 gets into the biosphere with the Fast Pathway scenario

Scenario	Number of geosphere layers	Maximum dose		R^2	Number of model inputs
		Min	Max		
REF	2	6.65e−10	5.81e−6	.977	10
FP	1	2.12e−4	1.64e−1	.9998	8
AG	3	6.64e−9	5.89e−5	.982	15
GA	2	3.21e−11	6.35e−8	.961	11
HDE	2	5.06e−10	1.77e−5	.966	14
EIC	2	3.06e−9	2.69e−5	.984	12

probability density functions. Each slave node uses information from the input matrix to perform a set of GTMCHEM migration calculations. Whenever a slave node completes them, it sends a message to the master which, in turn, sends back more work to that slave.

We performed 1000 runs of GTMCHEM for each scenario. Because the set of input parameters specified by the Level E/G test case for the iodine nuclide results in relatively fast runs, the strength of PMCD is not fully appreciated in this study. The run times and number of nodes varied in accordance with which scenario was being simulated: e.g., for 1000 simulations the number of nodes varied between 4 and 25 and the run times between 10 and 96 minutes.

In our simulations the inputs to GTMCHEM were random draws from uniform or log-uniform distributions, with minima and maxima given (e.g.) in the last two columns of Table 1 for scenario FP for the iodine nuclide I-129. Full details on the Level E/G test case are available in [23].

3.2. A variance-based sensitivity analysis

In this section we present results for the maximum dose of the iodine nuclide I-129, which was close to lognormally distributed in all scenarios (we examine the total annual dose in a later section). We regressed a standardised version of log max dose on standardised versions of scenario-specific inputs, expressed on the raw or log scales as appropriate to produce approximate uniform distributions. All inputs were approximately pairwise uncorrelated. The regressions were well-behaved statistically; there was small (not practically significant) evidence of nonlinearity in only one input (VREAL1). As may be seen from the R^2 values in Table 2, nearly all of the variation in log max dose is "explained" by the simulation inputs entered linearly (with no interactions or nonlinear terms). Because the model is so additive in this case, a simple variance-based sensitivity analysis will suffice – more complicated methods are not needed (but see Section 3.4 below).

Table 3
Regression results for the Reference scenario; L at the beginning of a variable name means that variable entered the regression on the log scale

Variable	Standardised coefficient	Variance in log max dose "explained" by variable	
CONTIM	.00151	.000	
LRLEACH	−.00261	.000	
LVREAL1	.628	.394	←
XPATH1	−.113	.013	
RET1	−.200	.040	
LVREAL2	.0676	.005	
XPATH2	−.0147	.000	
RET2	−.0452	.002	
LSTREAM	−.724	.525	←
"error"	–	.023	

Table 3 gives an example of the regression results, in this case for the Reference scenario. When the regression is performed with standardised versions of the outcome y and all of the predictors x_j, the squares of the standardised coefficients may be used as a measure of the variation in y (on the variance scale) "explained" by each of the x_j, provided the predictors are (close to) independent. In this case we have dealt with sample correlations of small size between the x_j by averaging the squared standardised coefficients over all possible orderings in which the variables could be en-

Table 4
Summary of the important variables (those that "explain" 5% or more of the variance in log max dose), by scenario

Scenario	Variable	Standardised coefficient	Variance in log max dose "explained" by variable
REF	VREAL1	.628	.394
	STREAM	−.724	.525
FP	RLEACH	.419	.185
	STREAM	−.897	.815
AG	VREAL1	.593	.375
	STREAM	−.720	.552
GA	VREAL1	.800	.645
	RET1	−.252	.063
	STREAM	−.478	.229
HDE	VREAL1	.633	.392
	STREAM	−.719	.505
EIC	VREAL1	.633	.413
	STREAM	−.717	.531

tered sequentially into the regression equation. From Table 3 VREAL1 and STREAM are evidently the high-impact inputs for the Reference scenario.

Table 4 summarises the important variables for each scenario, by retaining only those inputs which "explain" 5% or more of the variance in log max dose. (The standard errors of the standardised coefficients ranged in size from .0041 to .0063; and VREAL1, RLEACH, and STREAM entered the regression on the log scale.) It is evident that, apart from occasional modest influence from other variables, the two inputs having to do with water travel velocity play by far the biggest role in the predicted variations in I-129, and in opposite directions: VREAL1 and max dose are positively related (large values of VREAL1 lead to faster travel times through the geosphere to the biosphere), whereas it is *small* values of STREAM that lead to large iodine doses (arising from less dilution of the fluxes coming from the geosphere to the biosphere).

3.3. A model uncertainty audit

How much of the overall uncertainty about maximum dose is attributable to scenario uncertainty, and how much to parametric uncertainty? To answer this question, following [5], we performed a kind of *model uncertainty audit*, in which we partitioned the total variance in max dose into two components, between scenarios and within scenarios, the second of which represents the component of uncertainty arising from lack of perfect knowledge of the scenario-specific parameters. The relevant calculations are based on the double-expectation theorem (e.g., [8]): with y as max dose, and scenario i occurring with probability p_i and leading to estimated mean and standard deviation (SD) of y of $\hat{\mu}_i$ and $\hat{\sigma}_i$, respectively (across the 1000 simulation replications),

$$\begin{aligned}\hat{V}(y) &= V_S[\hat{E}(y|S)] + E_S[\hat{V}(y|S)] = \hat{\sigma}^2 \\ &= \sum_{i=1}^{k} p_i(\hat{\mu}_i - \hat{\mu})^2 + \sum_{i=1}^{k} p_i\hat{\sigma}_i^2 \\ &= \begin{pmatrix}\text{between-}\\ \text{scenario}\\ \text{variance}\end{pmatrix} + \begin{pmatrix}\text{within-}\\ \text{scenario}\\ \text{variance}\end{pmatrix}, \end{aligned} \tag{12}$$

where

$$\hat{E}(y) = E_S[\hat{E}(y|S)] = \sum_{i=1}^{k} p_i\hat{\mu}_i = \hat{\mu}\,. \tag{13}$$

Table 5 presents the scenario-specific mean and SD estimates, together with three possible vectors of scenario probabilities. We obtained the first of these vectors by expert elicitation of the relative plausibility of the nine micro-scenarios described at the beginning of this section, and created the other two, for the purpose of sensitivity analysis, by doubling and halving the non-Reference-scenario probabilities in the first vector.

Table 6 then applies Eqs. (12), (13) using each of the three scenario probability vectors. It may be seen that the percentage of variance arising from scenario uncertainty is quite stable across the three specifications of scenario probabilities, at about 30% of the total variance[5].

Alternative approaches to the inclusion of scenario uncertainty in risk analysis have been put forward by

[5] In fact, taking the vector of scenario probabilities to be $p = (p_1, \lambda p_2)$, where p_2 is the last five values in the fourth column of Table 5 and $p_1 = 1 - \lambda\sum_{i=1}^{5} p_{2i}$, and letting λ vary all the way from 0.01 (corresponding to only 0.001 probability on the non-Reference scenarios) to 10 (at which point the Reference scenario has zero probability), the proportion of the variance in max dose arising from scenario uncertainty only drops from 30.9% to 25.7%.

Table 5
Scenario-specific estimated means and standard deviations of max dose, together with three possible sets of scenario probabilities

Scenario	Estimated		Scenario probabilities (p_i)		
	Mean ($\hat{\mu}_i$)	SD ($\hat{\sigma}_i$)	1	2	3
REF	2.92e−7	5.63e−7	.90	.80	.95
FP	1.49e−2	2.23e−2	.0225	.045	.01125
AG	2.47e−6	5.18e−6	.0125	.025	.00625
GA	4.11e−9	5.96e−9	.0125	.025	.00625
HDE	5.82e−7	1.24e−6	.02	.04	.01
EIC	1.17e−6	2.44e−6	.0325	.065	.01625

several investigators: see [1], as well as a recent special issue devoted to the treatment of aleatory and epistemic uncertainty [12]. There, predictive uncertainty is partitioned into stochastic (aleatory) and subjective (epistemic), the former corresponding to scenarios and the latter to model structures and parameters in our formulation. An older example of categorisation of uncertainty in risk analysis is [16]. Applications to risk assessment for nuclear waste disposal are in [12,14]. Such categorisations, as explained in [1], are for convenience only, as the two types of uncertainty are indistinguishable under the Bayesian framework. In the present work no distinction is made between scenarios and other factors as far as uncertainty is concerned.

Table 6 says that the mean maximum dose of I-129 is 600–2300 times larger when scenario uncertainty is acknowledged than its value under the Reference scenario, and the uncertainty about max dose on the SD scale is 5000–10000 times larger.

Fig. 3 presents scenario-specific estimated predictive distributions for log maximum dose, and also plots the composite predictive distribution with scenario probability vector 1. The Fast Pathway and Glacial Advance scenarios lead to max dose values which are noticeably higher and lower than the other four scenarios, respectively. Principally because of this, the composite distribution is considerably heavier-tailed than lognormal, in particular including a small but significant contribution of very high doses from scenario FP.

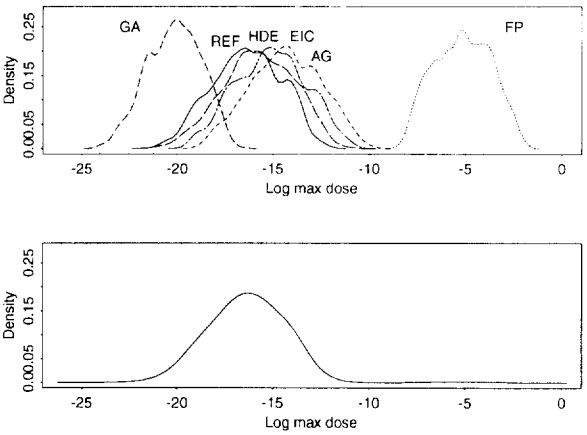

Fig. 3. Scenario-specific estimated predictive distributions (top panel) for max dose on the log scale, and composite predictive distribution (bottom panel) using scenario probability vector 1.

3.4. SA results for total annual dose in the REF scenario

Here we describe the results of a sensitivity analysis for total annual dose, arrived at by summing across all four nuclides monitored in Level E/G. It turned out that regression models relating this outcome to the inputs of our simulations were highly non-additive, meaning that simple variance-based methods of the type employed in Section 3.2 were insufficient as a basis for SA in this case.

To deal with such situations, we have developed a new method for global sensitivity analysis of model output [26] based on the Fourier Amplitude Sensitivity Test (FAST; [3]). We have named the new method *extended FAST* because of its ability to evaluate total effect indices for any uncertain factor involved in the model under analysis. The classic FAST is incomplete

Table 6
Sensitivity analysis of results as a function of scenario probabilities

Summary	Scenario probabilities		
	1	2	3
Overall mean max dose $\hat{\mu}$	3.36e−4	6.72e−4	1.68e−4
Overall SD $\hat{\sigma}$	4.01e−3	5.66e−3	2.84e−3
Overall variance $\hat{\sigma}^2$	1.61e−5	3.20e−5	8.08e−6
Between-scenario variance	4.90e−6	9.57e−6	2.48e−6
Within-scenario variance	1.12e−5	2.24e−5	5.61e−6
% of variance between scenarios	30.4	29.9	30.6
$\hat{\mu}/\hat{\mu}_{\mathrm{REF}}$	1153	2305	577
$\hat{\sigma}/\hat{\sigma}_{\mathrm{REF}}$	7128	10045	5049

in characterising the full model behaviour because it can only estimate the first-order (main) effects; the outcomes of the classic FAST are, however, acceptable for additive models.

The first-order index for a given factor measures the effect on the output uncertainty due to the variation of the factor itself, over its range of uncertainty, while all the others are allowed to vary simultaneously. The total effect index for a given factor includes the first order effect as well as all the interaction terms involving that factor, thus yielding an overall summary of the factor's influence on the output uncertainty. We argue that each factor has to be described by a pair of indices – first-order and total – and that this kind of representation allows an exhaustive and computationally inexpensive characterisation of the system under analysis.

Another method of SA currently employed by many analysts is based on the Sobol' measure [27]. Although computationally different from extended FAST, we have proved [26] that the Sobol' measure and FAST in fact lead to identical SA findings. In other words, the theory underlying FAST, Sobol' and another sensitivity measure, the correlation ratio [18], has been unified.

These approaches descend from a common root: the decomposition of the output variance. According to the analysis of variance, the total output variance D may be uniquely decomposed into orthogonal terms of increasing dimensionality,

$$D = \sum_{i=1}^{k} D_i + \sum_{1 \le i < j \le k} D_{ij} + \cdots + D_{12\ldots k}, \qquad (14)$$

where k indicates the number of factors and the terms $\{D_{i_1 i_2 \ldots i_s}, i_s \le k\}$ are called *partial variances*. By dividing these terms by D, *sensitivity indices* (which thus add up to one by definition) can be obtained: $S_{i_1 \ldots i_s} = D_{i_1 \ldots i_s}/D$. Formally, the *total effect index* for factor i is given by

$$S_{\mathrm{Ti}} = S_i + \sum_{j \ne i} S_{ij} + \sum_{j < l \ne i} S_{ijl} + \cdots + S_{12\ldots k}. \qquad (15)$$

Suitable summary measures can then be obtained by further normalisation: $S^*_{\mathrm{Ti}} = S_{\mathrm{Ti}} / \sum_{j=1}^{k} S_{Tj}$. The S^*_{Ti} are called the *total normalised sensitivity indices*.

The extended FAST is computationally more efficient than the Sobol' measure (see [28] for a comparison of the performance of the two methods). Indeed, the pair of first order/total effect indices for a given factor can be estimated via the same sample when FAST is used, while the Sobol' measure requires a different sample for each index to be computed. The number of samples required for computing the whole set of sensitivity indices via the Sobol' method is $2k$, whereas by using extended FAST only k samples are needed. The total number of model evaluations is obtained by multiplying the number of samples needed by the sample size, which is chosen as a function of the desired accuracy for the indices. The total number of model evaluations essentially determines the total cost of the analysis. Indeed, the computational cost of evaluating the sensitivity indices, given the set of model outputs, is nil.

An illustration of the first order and total sensitivity indices is given in Fig. 4 for the REF or base case scenario (Level E). The results are expressed as a function of time, from 10^4 to 10^7 years into the future.

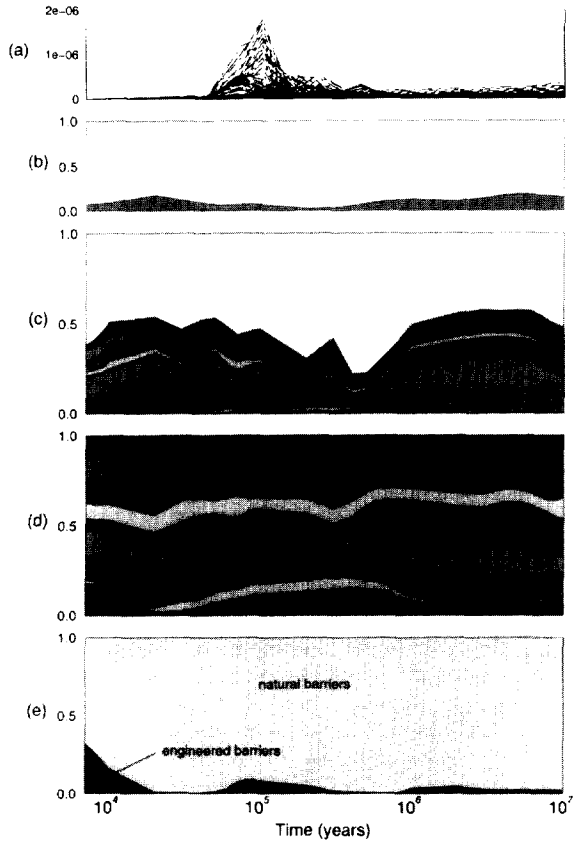

Fig. 4. Results from the base case (REF) scenario.

We chose a sample size of 257 to run the model. The curves displayed in panel (a) of the figure are the result of 3084 model evaluations. We used them both for estimating R^2 (b) and for evaluating first order (c) and total normalised indices (d) for all 12 factors of the underlying model ($257 \times 12 = 3084$). A much more restricted set of model outputs ($257 \times 2 = 514$) is sufficient to estimate the total normalised indices for engineered and natural barriers (e). Indeed, once the sample size is fixed, the computational effort is proportional to the number of factors or subgroups considered in the analysis.

Panel (b) of Fig. 4 shows that the underlying model is strongly nonlinear, given that R^2 is always below 0.2. A cumulative plot of first order indices is given in (c). The model under investigation is not additive because the shaded region is below 0.6 everywhere. More than 40% of the output uncertainty is due to interactions occurring among the factors.

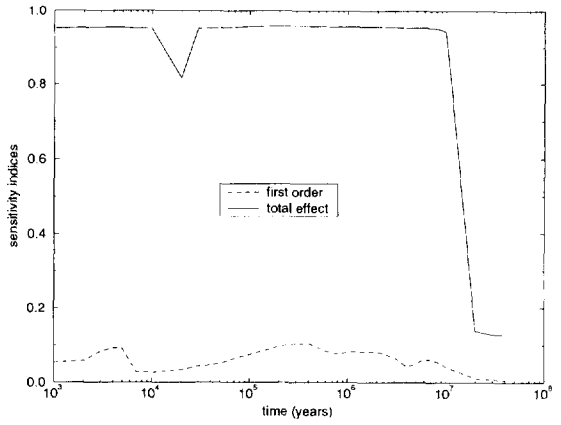

Fig. 5. Results from the Level E/G test case: first order and total sensitivity indices for the 'scenario' parameter.

A cumulative plot of the total indices for the 12 factors is given in panel (d). The most important factors can readily be identified:

- $v(1)$ = water velocity in the geosphere's first layer (VREAL1 in Section 3.2);
- $l(1)$ = length of the first geosphere layer (XPATH1);
- $Rc(1)$ = retention coefficient for neptunium (first layer; RET1 for Np-237) – note how the importance of this factor grows over time; and
- W = stream flow rate (STREAM).

In panel (e) the total normalised indices are displayed for the factors being grouped into two sub-sets (natural and engineered barriers). The modest role of engineered barriers is highlighted, as confirmed by risk assessment practitioners.

3.5. SA results for total annual dose in Level E/G

The results of a FAST computation for the total annual dose in Level E/G are described in this section. The first order and the total sensitivity indices for the parameter selecting among the various scenarios have been estimated via a set of 5763 model evaluations. The results, displayed in Fig. 5, are expressed as a function of time from 10^3 to $4 \cdot 10^7$ years into the future. It seems that one parameter has the function of triggering the development of different scenarios, and this parameter interacts with all the factors entering in each and every scenario.

In our example the first order index for the 'scenario' parameter is small but its total effect is close to 1! This can be considered as an obvious result, but

has implications for the theory of sensitivity analysis. It has been argued that when one (group of) factor(s) is important, its first order effect should be high. This could in some instances allow all the other factors to be fixed without loss of model accuracy [27]. On the other hand, Jansen [15] argues that the real reduction in model variability that one can achieve by eliminating a factor i (i.e., by considering the ith factor as a perfectly known quantity) is given by its S_{Ti}. In our example, if we were able to eliminate the 'scenario' (or 'triggering') parameter (i.e., by selecting the proper scenario), we would reduce the model variability around 95%, at most of the time points. This is indeed a measure of how much the 'scenario' parameter influences the output uncertainty.

Hence, in the problem setting where one seeks a group of factors accounting for most of the variance so that the others can be fixed, one should indeed focus on the total effect of the target group, and not on its first order effect.

4. Discussion

Ongoing GESAMAC work expands the modeling to two-dimensional chemical transport and incorporates structural and predictive uncertainty. Even before we reach that stage, however, it is clear (a) that propagation of uncertainty both within (parametric) and between scenarios in risk assessment work of this type can have a dramatic effect both on the predictive mean outcome values and the uncertainty bands; and (b) that total sensitivity indices are useful in quantifying SA results with nonlinear, non-additive models. Our goal, as we continue our study, is to contribute to the process of providing more realistic uncertainty assessment than ever previously performed in nuclear waste disposal risk management.

Acknowledgements

This work has been partially funded under contract FI4W-CT95-0017 (the GESAMAC project) of the European Commission through its R+D programme on 'Nuclear Fission Safety' (1994–1998). We are grateful to Henning von Maravic for comments and references, and to the European Commission for funding. Two of the authors have highly appreciated suggestions by Prof. George Apostolakis from the Massachusetts Institute of Technology.

References

[1] G. Apostolakis, The concept of probability in safety assessments of technological systems, Science 250 (1990) 1359–1364.
[2] W.P. Bishop, C.D. Hollister, Seabed disposal – where to look, Nucl. Technology 24 (1974) 425–443.
[3] R.I. Cukier, J.H. Schaibly, K.E. Shuler, Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. III. Analysis of the approximations, J. Chem. Phys. 63 (1975) 1140–1149.
[4] P. Dawid, Prequential analysis, stochastic complexity, and Bayesian inference (with discussion), in: Bayesian Statistics 4, JM Bernardo, JO Berger, AP Dawid, AFM Smith, eds. (Oxford Univ. Press, 1992) pp. 109–125.
[5] D. Draper, Assessment and propagation of model uncertainty (with discussion), J. Roy. Stat. Soc. Series B 57 (1995) 45–97.
[6] D. Draper, Model uncertainty in "stochastic" and "deterministic" systems, in: Proc. 12th Int. Workshop on Statistical Modeling, Vienna, C. Minder, H. Friedl, eds., Schriftenreihe der Österreichischen Statistichen Gesellschaft, Vol. 5 (1997) pp. 43–59.
[7] S. Eguilior, P. Prado, GTMCHEM computer code: Description and application to the Level E/G test case (GESAMAC Project) (CIEMAT, Madrid, 1998), in preparation.
[8] W. Feller, An Introduction to Probability Theory and its Applications, Vol. II, 2nd ed. (Wiley, New York, 1971).
[9] B. de Finetti, La prévision: ses lois logique, ses sources subjectives, Annales de l'Institut de H. Poincaré 7 (1937) 1–68; reprinted in 1980 as "Foresight: its logical laws, its subjective sources" in: Studies in Subjective Probability, H.E. Kyburg, H.E. Smokler, eds. (Dover, New York, 1980) pp. 93–158.
[10] W.R. Gilks, S. Richardson, D.J. Spiegelhalter, Markov Chain Monte Carlo in Practice (Chapman & Hall, London, 1996).
[11] J.M. Hammersley, D.C. Handscomb, Monte Carlo Methods (Chapman & Hall, London, 1979).
[12] J.C. Helton, D.E. Burmaster, Guest editorial: Treatment of aleatory and epistemic uncertainty in performance assessments for complex systems, Reliability Engineering and System Safety 54 (1996) 91–94.
[13] J.C. Helton, Uncertainty and sensitivity analysis in the presence of stochastic and subjective uncertainty, J. Stat. Comput. & Simul. 57 (1997) 3–76.
[14] J.C. Helton, Uncertainty and sensitivity analysis in performance assessment for the Waste Isolation Pilot Plant, Comput. Phys. Commun. 117 (1999) 156–180, this issue.
[15] M.J.W. Jansen, WINDINGS: uncertainty analysis with winding stairs samples, Poster; Workshop Assessment Environmental Impact: Reliability of Process Models (Edinburgh, 18–19 March 1996).

[16] S. Kaplan, B.J. Garrick,. On the quantitative definition of risk, Risk Analysis 1 (1981) 11–27.
[17] R.L. Keeney, D. von Winterfeldt, Managing nuclear waste from power plants, Risk Analysis 14 (1994) 107–130.
[18] M.D. McKay, Variance-based methods for assessing uncertainty importance in NUREG-1150 analyses, Los Alamos National Laboratory, LA-UR-96-2695 (1996).
[19] NEA PSAG User's Group, PSACOIN Level E Intercomparison, B.W. Goodwin, J.M. Laurens, J.E. Sinclair, D.A. Galson, E. Sartori, eds. (Nuclear Energy Agency, Organisation for Economic Cooperation and Development, Paris, 1989).
[20] NEA PSAG User's Group, PSACOIN Level 1A Intercomparison, A. Nies, J.M. Laurens, D.A. Galson, S. Webster, eds. (Nuclear Energy Agency, Organisation for Economic Cooperation and Development, Paris, 1990).
[21] A. de C. Pereira, Some developments of safety analysis with respect to the geological disposal of high-level nuclear waste, SKI technical report 89:06 (Stockholm, 1989).
[22] P. Prado, User's manual for the GTM--1 code, EUR 13925 EN (1992) ISSN 1018-5593.
[23] P. Prado, S. Eguilior, A. Saltelli, Level E/G test-case specifications (GESAMAC project), CIEMAT/DIAE/550/55900/04/08 (CIEMAT, Madrid, 1998).
[24] P. Prado, T. Homma, A. Saltelli, Radionuclide migration in the geosphere: A 1D advective and dispersive transport module for use in probabilistic system assessment codes, Radioactive Waste Management and Nuclear Fuel Cycle 16 (1991) 49.
[25] A. Saltelli, P. Prado, C. Torres, ENRESA/JRC cooperation agreement in the field of nuclear waste management, 3383-88-03 TG ISP E, Final Report: 1 Results, 2 Annexes, SPI89.36/I, II (1989).
[26] A. Saltelli, S. Tarantola, K. Chan, A quantitative, model-independent method for global sensitivity analysis of model output, Technometrics (1998), forthcoming.
[27] I.M. Sobol', On sensitivity estimation for nonlinear mathematical models, A free translation by the author from Mathematical Modelling, 2 (1990) 112–118 [in Russian].
[28] S. Tarantola, Analysing the efficiency of quantitative measures of importance: the improved FAST, in: Proc. SAMO 98: Second International Symposium on Sensitivity Analysis of Model Output, Venice, 19–22 April 1998, EUR report 17758 EN (1998).
[29] S. Weisberg, Applied Linear Regression, 2nd ed. (Wiley, New York, 1985).