Investigation of screening methods for sensitivity analysis and their application to a hydrogeological model

Thomas Henkel[a], Rene Blankenburg[b], Heike Wilson[a], Wilfried Krug[a], Holger Mansel[b]
t.henkel@dualis-it.de
[a] DUALIS GmbH IT Solution, Tiergartenstr. 32, 01219 Dresden
[b] Ingenieurbüro für Grundwasser, Nonnenstr. 9, 04229 Leipzig

Abstract

The performance of three different screening methods proposed by Morris [1] and Campolongo et al. [2,3] is investigated using analytical test functions. It is shown that one version of the third method, the radial design, demonstrates a superior performance in terms of computational efficiency and stability. This method is finally applied to a hydrogeological model that calculates water flow and solute transport in porous media [4]. Strengths and weaknesses of each method are discussed and conclusions are drawn with respect to efficient screening of model parameters.

1. Introduction

In many fields of environmental sciences mathematical models are used for simulation of different systems to understand and predict their potential behavior. However, the complexity of these models increased over the years. One consequence of model complexity is that uncertainty in model prediction increased too. Thus, to quantify where does the uncertainty come from (referred to as sensitivity analysis (SA)) is one essential part during model application.

Nowadays many SA techniques are available which can be classified into groups like screening methods, variance based methods etc. Screening techniques can be used to identify parameters that control most of the uncertainty of model output. Thus, a classification of input model parameters – hereinafter simply referred to as parameters – into important and non-important ones is possible with low computational effort. Therefore, these methods are frequently applied when the model is computationally expensive to evaluate and/or the model has a large number of parameters. Once the parameters have been classified in a first step, the model can be simplified by fixing non-important parameters to a constant value without significantly affecting the total output variance. Since screening techniques tend to provide qualitative sensitivity measures, i.e. a ranking of parameters in the order of importance, a second analysis step is usually done using e.g. a variance based method in order to study the uncertainty of the remaining important parameters in more detail.

One well-known screening method is the design proposed by Morris [1] which has been refined by Campolongo and co-workers [2]. Their improvement is based on a better exploration of the parameter space known as concept of „optimized trajectories“. Recently, Campolongo et al. investigated another design using a radial scheme as sampling strategy [3]. The authors concluded that this design has superior properties compared to other designs, and therefore suggested to use the radial design during a SA.

In this work, the performance of these screening methods is investigated using analytical test functions. Strengths and weaknesses of each method are discussed and conclusions are drawn with respect to efficient screening of parameters. Finally, the best method is applied to a hydrogeological model that calculates water flow and solute transport in porous media [4]. The paper is organized as follows: Notes on the screening methods applied are given in the next section. Section 3 briefly describes the numerical experiments. Results of the experiments are shown and discussed in section 4.

2. Screening methods

2.1. Elementary effect method (method A)

The elementary effect method as introduced by Morris [1] is based on a sample of r different trajectories where parameters are changed – one at a time of step Δ – on a grid of levels covering the parameter domain. The number of points of each trajectory is (k+1), where k is the number of parameters. Along a trajectory each parameter X_i is increased or decreased by the same step Δ assuming a uniform distribution in [0,1] for each X_i. The measure $EE_i = (Y^{(2)} - Y^{(1)}) / \Delta$, known as elementary effect (EE) of the parameter X_i, is calculated where $Y^{(1)}$ and $Y^{(2)}$ are the model output before and after the change of X_i, respectively. Now, r different EE's can be estimated by randomly sampling r different trajectories. The mean of these effects can be considered as a sensitivity measure. However, Campolongo et al. suggested calculating the following measure

$$\mu_i^* = \frac{1}{r}\sum_{i=1}^{r}|EE_i| \tag{1}$$

which uses the absolute value of the EE. This reduces computation errors in case of non-monotonic models since EE terms of opposite sign cancel out each other. Further details of the EE method can be found in Ref. [1][3][5].

2.2. Elementary effect method with optimized trajectories (method B)

The original EE method uses a random sampling procedure for the construction of trajectories. The design is based on generating a random starting point for each trajectory and completing it by changing one parameter at a time in a random order. This strategy may result in a poor coverage of the parameter space, especially for models with a large number of parameters. Therefore, Campolongo et al. [2] proposed an improvement of the sampling strategy which aims at a better scanning of the parameter domain using the same number of points. This is achieved by initially generating a high number of trajectories and then selecting r trajectories with the largest dispersion in the parameter space. In this context, dispersion is defined in terms of distances between couples of trajectories. The reader is referred to Ref. [2] for more details.

2.3. Radial sampling method (method C)

Recently, Campolongo et al. proposed a radial scheme as an alternative design to calculate EE's [3]. The radial scheme and the trajectory design described above are illustrated in Table 1, where a and b are two different k-dimensional random vectors. In radial sampling, points (1) and (2) of the parameter domain are used for the calculation of the EE for the first parameter, points (1) and (3) for the second parameter, points (1) and (4) for the third parameter etc. In trajectory sampling however, points (1) and (2) are used

for the calculation of the EE for the first parameter, points (2) and (3) for the second parameter, points (3) and (4) for the third parameter etc.

Point	Radial sampling	Trajectory sampling
(1)	$a_1, a_2, a_3, \ldots, a_k$	$a_1, a_2, a_3, \ldots, a_k$
(2)	$b_1, a_2, a_3, \ldots, a_k$	$b_1, a_2, a_3, \ldots, a_k$
(3)	$a_1, b_2, a_3, \ldots, a_k$	$b_1, b_2, a_3, \ldots, a_k$
(4)	$a_1, a_2, b_3, \ldots, a_k$	$b_1, b_2, b_3, \ldots, a_k$
...		...
(k+1)	$a_1, a_2, a_3, \ldots, b_k$	$b_1, b_2, b_3, \ldots, b_k$

Table 1: Comparison of different sampling schemes for k parameters

Such a block as illustrated on the left side of Table 1 corresponds to one trajectory. For proper screening however, more than one trajectory should be used. This means that several random vectors a and b are needed for the calculation. Campolongo et al. proposed to use Sobol's quasi-random numbers [6], known as LP_τ sequences, instead of plain random numbers for the generation of the vectors [3]. More details, especially of how to construct a setup with several trajectories using LP_τ sequences, can be found in Ref. [3].

3. Numerical experiments

3.1. Methodology

A series of computer experiments using different test functions was carried out in order to evaluate each screening method described above. The test functions applied are presented in the next section. For each function, the measure μ* according to formula (1) was calculated using the EE method (A), the EE method with optimized trajectories (B), and the radial sampling method (C). As for method A and B, a grid of 4 levels was used which results in a step $\Delta = 2/3$ [5]. In addition, 200 trajectories were generated first with method B. Then, r trajectories with the largest dispersion in the parameter space were selected for the test. As for method C however, two versions were applied: the original version by Campolongo et al. [3] and another version developed by the author. These versions, referred to as C_1 and C_2 respectively, differ in the way how LP_τ sequences are used to build a setup with several trajectories.

In order to evaluate the performance for both very low and typical sample sizes, the number of trajectories was varied between 2 and 10 for method B, and between 1 and 10 for all other methods. The reason is that method B requires at least 2 trajectories to enable a determination of distances between them. The μ* values computed were used to ascertain important parameters from non-important ones. According to Campolongo et al. [3], the following measure, referred to as score g, is calculated

$$g = \frac{card(\varphi \cap \xi)}{card(\varphi)} \tag{2}$$

which allows a comparison of results of the different screening methods. φ and ξ are both sets of important parameters, however, the measure which characterizes a parameter as being important or not is different. As for φ, a parameter is defined as an important one whose total order sensitivity index S_T is higher than 1/k, where k is the number of parameters. The index S_T, which can be estimated using a variance based SA method, characterizes the total effect of a parameter including its first order effect and interactions

with other parameters of any order. As for the test functions applied, analytical values for S_T can be computed. Thus, φ is considered to be a "true" set of important parameters. On the other hand, in screening experiments parameters are ranked according to their μ* value. Then, ξ defines the set of the first n important parameters identified by μ* where n is the cardinality of φ. The interested reader is referred to Ref. [5] for more details on variance based methods calculating sensitivity indices. Finally, the following experiment was performed:

- for each parameter calculate one EE per trajectory
- for each parameter calculate μ* over all trajectories r
- determine ξ and calculate g for this particular run

This procedure was repeated 1000 times for each screening method and test function in order to calculate mean and standard deviation of g. Experiments were performed using a notebook computer with Intel Core 2 Duo 2.1 GHz processor and 3.4 GB RAM.

3.2. Test functions

The screening methods described in section 2 were tested on the following analytical functions commonly used as benchmark in SA, where each parameter X_i was assumed to be uniformly distributed in [0,1].

(1) G-function as defined by Sobol [6] with k = 30 and the constants a_i :

$$G = \prod_{i=1}^{k} \frac{|4X_i - 2| + a_i}{1 + a_i} \tag{3}$$

X_i	a_i	S_T	X_i	a_i	S_T
24	0	0.4915	9	7	0.0102
8	0.5	0.2537	12	8	0.0081
15	0.5	0.2537	27	9	0.0065
19	1	0.1512	17	10	0.0054
3	3	0.0401	5	20	0.0015
1	4	0.0259	30	50	0.0003
20	5	0.0180	all other p.	99	0.0001
23	6	0.0133			

Table 2: G-function parameters for test case (1). S_T was calculated analytically [7].

The value of a particular a_i determines the relative importance of X_i, i.e. the smaller a_i the more important X_i. In this first test case, the constants a_i were set such that the 5 parameters X_{24}, X_8, X_{15}, X_{19}, and X_3 fulfill the criterion of an important parameter as described in section 3.1 because their index S_T is higher than the threshold 1/k = 0.033 (see Table 2). Thus, 5 out of 30 parameters have a significant impact on the uncertainty of the model. However, all parameters interact with one or more other parameters on a moderate level considering their first order sensitivity index (not shown here).

(2) A modified version of the G-function introduced as G4* in [3], i.e. 4 out of 20 parameters are identified as important parameters:

$$G^* = \prod_{i=1}^{k} \frac{(1+\alpha_i)|2(X_i+\delta_i - I(X_i+\delta_i)) - 1|^{\alpha_i} + a_i}{1+a_i} \tag{4}$$

The same parameter values were applied except for the parameter δ which has a very strong stochastic impact on the model output, i.e. it causes a shift of the function. Therefore, each random number generated for δ_i was multiplied by a factor 0.1. Thus, in contrast to test case (1) the function represents a stochastic component and a level of interaction among the parameters which is very strong.

(3) The K-function as applied with 20 parameters in [3]. Thus, 3 out of 20 parameters are identified as important parameters:

$$K = \sum_{j=1}^{k} (-1)^j \prod_{i=1}^{j} X_i \tag{5}$$

In contrast to the other test cases, the important parameters of this function exhibit a degree of interaction with other parameters which is on a low level. Thus, this test case may characterize a default case of SA which often occurred in practice.

3.3. Hydrogeological model

After having evaluated the screening methods, the best algorithm in terms of computational efficiency, reproducibility, and stability is applied to a hydrogeological model described in [4]. In general, this model allows the calculation of water flow and solute transport in porous media. In this work, the model is used to simulate the diffusion of a pollutant into an unsaturated soil zone typical for a groundwater risk assessment. In this case, the soil zone extends from the groundwater level to the soil surface. It is assumed that the groundwater level (lower boundary condition (LB)) and the groundwater recharge rate (upper boundary condition (UB)) do not change during simulation. Further, the pollutant dissolved in rain water diffuses into the soil for a period of 2000 days. Afterwards, the source of the pollutant is considered as fully exhausted. The model domain is divided into 5 horizontally extending soil layers where the top soil layer is considered to be identical to the bottom soil with respect to its hydraulic properties. Thus, 4 layers remain where the properties of a particular layer can be chosen independently of the other layers. In this application, the corresponding layer properties were adjusted such that the first (and fifth) layer represents coarse sand, the second layer middle sand, the third layer a medium-grained coarse sand, and the fourth layer sandy silt. As for the screening test, lower and upper boundaries of the soil parameters were set individually for each layer such that these main properties were preserved during simulation. Further, the concentration of the dissolved pollutant was calculated at a given level above the groundwater surface, i.e. in the capillary fringe, and considered as model output. Table 3 briefly describes all model parameters used. Finally, the test was carried out using 10 trajectories.

X_i	description	symbol	LB	UB	unit
1	residual water content of layer 1	thr_1	0.01	0.015	[-]
2	residual water content of layer 2	thr_2	0.02	0.045	[-]
3	residual water content of layer 3	thr_3	0.01	0.02	[-]
4	residual water content of layer 4	thr_4	0.07	0.16	[-]
5	porosity of layer 1	ths_1	0.31	0.34	[-]
6	porosity of layer 2	ths_2	0.35	0.41	[-]
7	porosity of layer 3	ths_3	0.31	0.35	[-]
8	porosity of layer 4	ths_4	0.41	0.49	[-]
9	van Genuchten parameter (scale) of layer 1	α_1	0.08	0.15	1/cm
10	van Genuchten parameter (scale) of layer 2	α_2	0.009	0.06	1/cm
11	van Genuchten parameter (scale) of layer 3	α_3	0.07	0.1	1/cm
12	van Genuchten parameter (scale) of layer 4	α_4	0.002	0.007	1/cm
13	van Genuchten parameter (slope) of layer 1	n_1	2.8	3.5	[-]
14	van Genuchten parameter (slope) of layer 2	n_2	1.8	2.5	[-]
15	van Genuchten parameter (slope) of layer 3	n_3	2.6	3.1	[-]
16	van Genuchten parameter (slope) of layer 4	n_4	1.3	1.5	[-]
17	saturated hydraulic conductivity of layer 1	Ks_1	500	800	cm/d
18	saturated hydraulic conductivity of layer 2	Ks_2	100	200	cm/d
19	saturated hydraulic conductivity of layer 3	Ks_3	400	700	cm/d
20	saturated hydraulic conductivity of layer 4	Ks_4	20	80	cm/d
21...24	longitudinal dispersivity of layer 1...4	$DispL_i$	15	100	cm
25...28	transversal dispersivity of layer 1...4	$DispT_i$	1	20	cm
29	pollutant diffusion coefficient in water	Difw	0	1.728	cm^2/d
30	pollutant diffusion coefficient in air phase	Difg	0	1.2E04	cm^2/d
31...34	adsorption coefficient in solid phase of layer 1...4	KS_i	0	20	l/kg
35...38	Henry distribut. coefficient betw. liquid and air phase of l. 1...4	H_i	0	0.01	1/d
39...42	first order break down rate of the pollut. in liquid phase of l. 1...4	$SnkL_i$	0	7E-06	1/d
43...46	first order break down rate of the pollutant in gas phase of l. 1...4	$SnkG_i$	0	7E-06	1/d

Table 3: Parameters of the hydrogeological model used in the simulation

4. Results and discussion

4.1. Test cases

Fig. 1 shows the mean, left side of Fig. 2 the standard deviation of the score g as a function of the number of trajectories for test case (1). As expected, the mean of g increases and the standard deviation decreases with increasing number of trajectories. As for one trajectory, a score of around 0.8 was obtained which means that 4 out of 5 parameters are identified as important ones. This result can be improved by increasing the sample size, i.e. using a higher number of trajectories. For example, 10 trajectories will lead to the identification of almost all important parameters in this test case. However, one can also see that g calculated with method A exhibits significant fluctuations. The reason for this behavior is not completely clear, however, the constant step width and the limited number of points in the parameter domain are limitations when scanning this domain. These constraints do not exist when applying method C, and therefore are believed to be responsible for the observed behavior. The score calculated with method B

also shows fewer fluctuations and is on average higher compared to method A. However, it should be mentioned that method B requires much larger computation time compared to the other methods. To provide a relative comparison, the computer run time for this test case was determined for one run with 10 trajectories using the hardware described in section 3.1. Computer run times were 4.5 ms, 6.2 ms, 3.7 ms, and 63 s for method A, C_1, C_2, and B, respectively. Obviously, the selection procedure of method B requires large computation resources. Fig. 1 also shows that both methods C_1 and C_2 demonstrate a good performance in terms of stability and convergence, whereas method C_2 delivers the highest score averaged across the trajectories of 6 to 10.

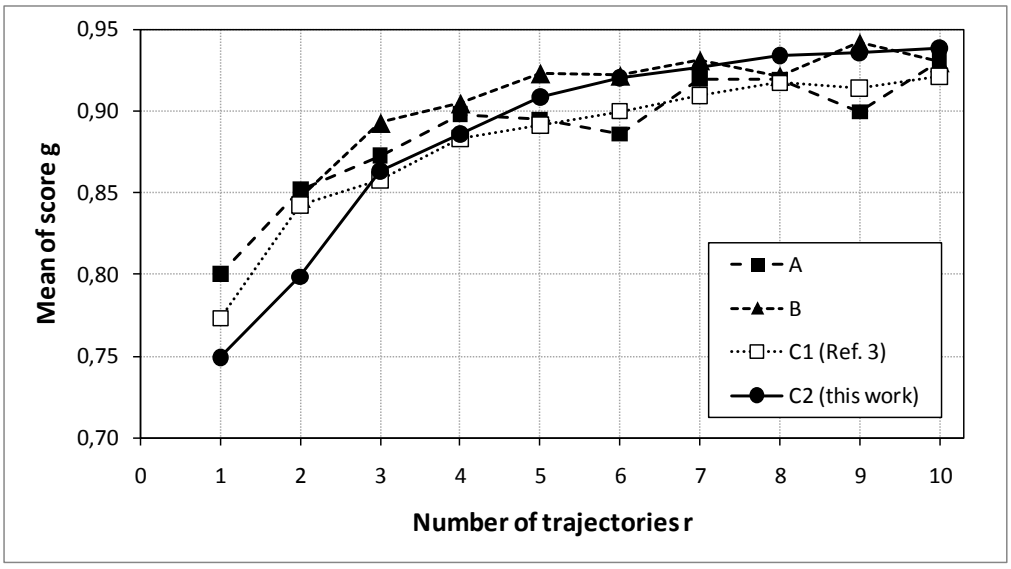

Fig. 1: Mean of score g as a function of the number of trajectories for test case (1)

Fig. 2 (right) and Fig. 3 illustrate results for test case (2). In general, the computed scores are lower and their standard deviations are higher due to the stochastic impact of the model parameter δ. The g values increase up to 0.9 if all stochastic parameters δ_i are set to zero (not shown here). Surprisingly, method C_1 delivers the lowest mean values and the highest standard deviations of g compared to the other screening methods. Further, the score does not improve with increasing sample size. The other methods exhibit more or less the same performance in terms of g except for small fluctuations seen with method A and B, respectively.

Results for test case (3) are depicted in Fig. 4 and 5. As can be clearly seen, method C_1 and C_2 demonstrate almost the same performance which is significantly higher compared to method A and B, respectively. For all trajectories applied, the mean of g is higher and the standard deviation of g is lower. In contrast to the methods A and B, fluctuations are almost absent and the convergence behavior is perfect. It should also be noted that the selection procedure inherent to method B cannot improve the performance compared to method A.

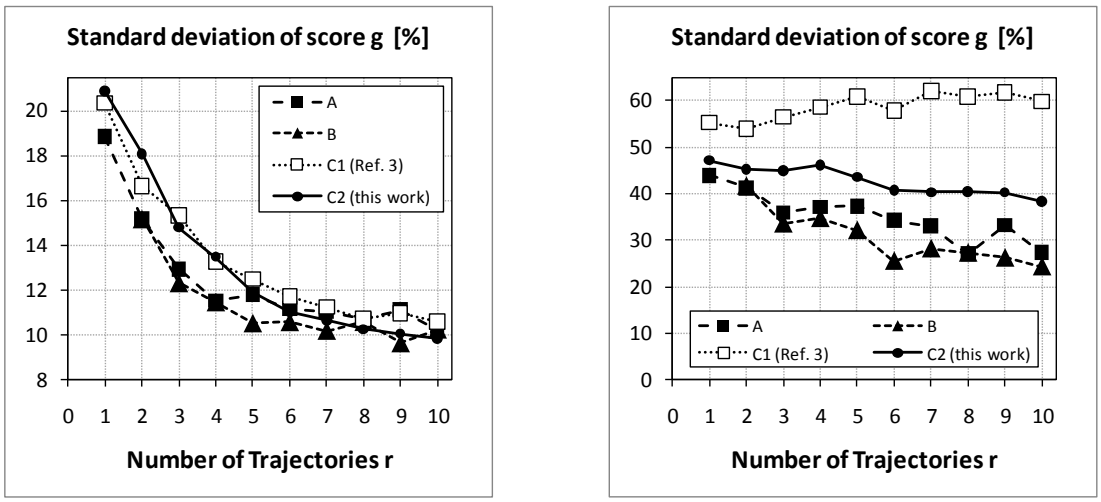

Fig. 2: Standard deviation of score g (related to the mean of g) as a function of the number of trajectories for test case 1 (left) and test case 2 (right)

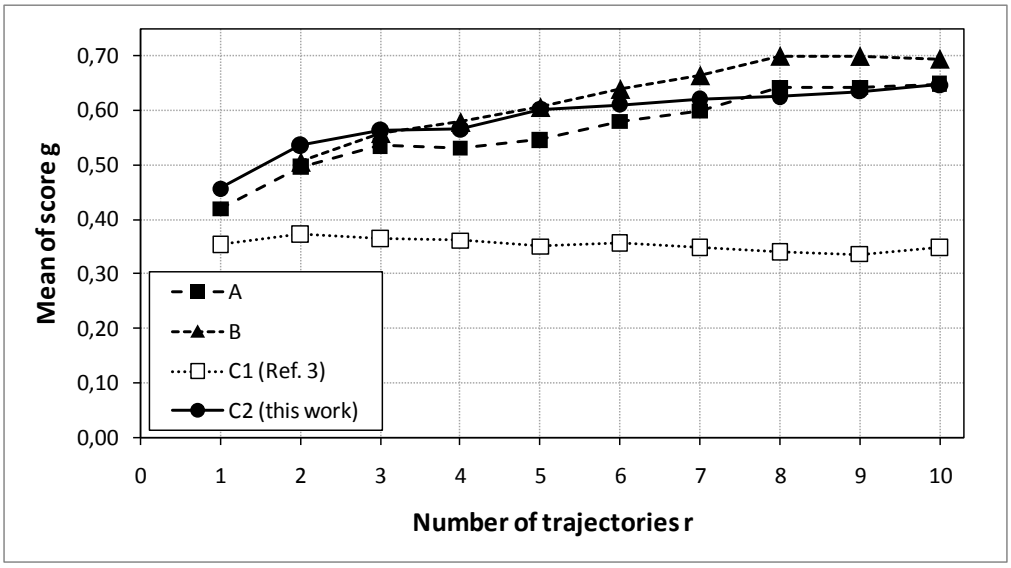

Fig. 3: Mean of score g as a function of the number of trajectories for test case (2)

To summarize these results, method C2 provided the best performance in terms of stability and convergence, and also on average the highest score and lowest standard deviation if 5 or more trajectories are considered. Therefore, this method is selected for a screening test on the model described in section 3.3.

Fig. 4: Mean of score g as a function of the number of trajectories for test case (3)

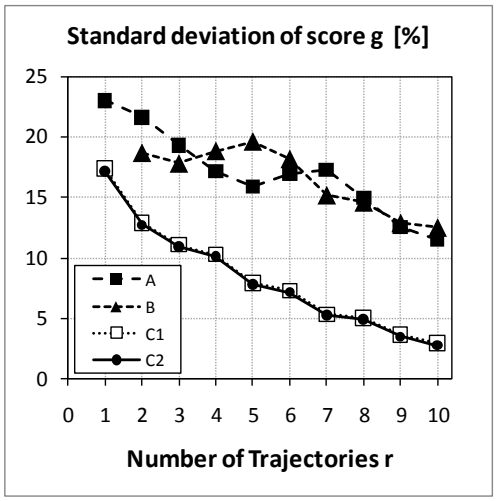

Fig. 5: Standard deviation of score g vs. number of trajectories for test case (3)

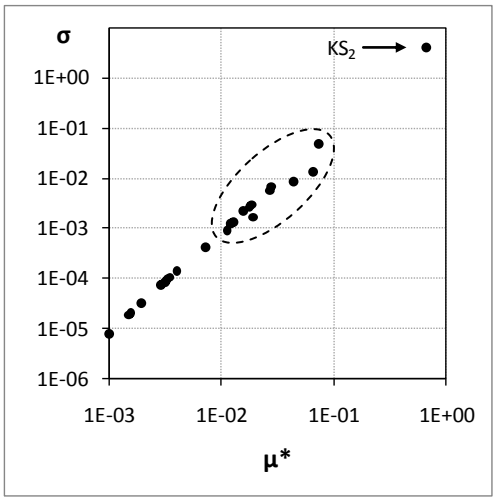

Fig. 6: σ versus μ* as calculated for the hydrogeological model

4.2. Hydrogeological model

Standard deviations (σ) and means of EE's (μ*) were calculated during a screening test on the model (see Fig. 6). The plot of σ as a function of μ* reveals that only one parameter (KS_2) has a strong influence on the model output due of its high μ* and σ. Following Morris, a high σ means that the corresponding parameter is characterized by a strong nonlinearity and/or a strong interaction with one or more parameters. Within the context of the application, this seems to be plausible because the adsorption process is

well known to have a large influence on the pollutant spreading. Further, this process is strongly assisted by the second layer because it has the largest volume in the model.

As can be seen in Fig. 6, the screening method detected a small subset of parameters which show a significant lower influence on the model than KS_2. Moreover, interaction and nonlinearity effects of these parameters are negligible due to their low σ. The following parameters, in descending order, belong to this subset: $DispL_2$, KS_1, KS_3, α_2, KS_4, Difg, H_2, $DispL_4$, n_2, H_3, H_1, $DispL_3$. Finally, the remaining parameters were identified as non-influential parameters because their μ* and σ values are far too low. This may open up the ground for a model simplification process.

5. Conclusions

In this work, the performance of three different screening methods has been investigated using analytical test functions. The well-known EE method by Morris, an improved version by Campolongo et al., and two different versions of a radial design have been tested in numerical experiments. It has been shown that the application of a radial design for screening of model parameters, if properly used, provides a better performance in terms of computational efficiency and stability compared to the other screening techniques. In particular, a more stable and convergent behavior of the algorithm has been demonstrated. In contrast, the classical EE method as well as its improved version using optimized trajectories tend to be affected by stability issues. Moreover, the latter algorithm requires far more computation resources than the other methods. Therefore, the algorithm developed in this work seems to be a good choice for an efficient SA, especially of deterministic models where only a small subset of parameters control the model uncertainty and parameters interact on a low level.

6. References

[1] *Morris, M.D.*: Factorial sampl. plans f. prel. comput. exp. Technometrics 33 (1991), 161-174.

[2] *Campolongo, F., Cariboni, J., Saltelli, A.*: An effective screening design for sensitivity analysis of large models. Environmental Modelling & Software 22 (2007), 1509-1518.

[3] *Campolongo, F., Saltelli, A., Cariboni, J.*: From screening to quantitative sensitivity analysis. A unified approach. Computer Physics Communications 182 (2011), 978-988.

[4] *Simunek, J., van Genuchten, M.T.:* The Chain_2D code for simulating 2-dim. movement of water, heat and multiple solutes in variably-saturated porous media [Report]. Riverside, California: U.S. Salinity Laboratory, U.S. Department of Agriculture (1994).

[5] *Saltelli, A.* Sensitivity Analysis. John Wiley & Sons, Ltd., 2000.

[6] *Sobol, I.M.:* Uniformly distributed sequences with an addition uniform property. USSR Computational Mathematics and Mathematical Physics 16 (1976) 236-242.

[7] *Saltelli, A., Annoni, P., Azini, I., Campolongo, F., Ratto, M., Tarantola, S.:* Variance based sensitivity analysis of model output. Computer Physics Commun. 181 (2010), 259-270.

[8] Simlab (2011) Software package for uncertainty and sensitivity analysis. Joint Research Centre of the European Commission. Downloadable for free at: http://simlab.jrc.ec.europa.eu

7. Acknowledgements

We gratefully acknowledge financial support from the German Federal Ministry of Education and Research (BMBF) under contract RESS-157-013 EGSIM. We also thank the Joint Research Centre (JRC) of the European Commission for providing the SimLab software package [8].