VOL. 13, NO. 12, JUNE 2018 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

GLOBAL SENSITIVITY ANALYSIS, INVERSE MODELLINON SOIL WATER INFILTRATION

Goh Eng Giap[1], Kosuke Noborio[2] and Asmadi Ali[1]
[1]School of Ocean Engineering, Universiti Malaysia Terengganu, Kuala Terengganu, Terengganu, Malaysia
[2]School of Agriculture, Meiji University, Higashimita, Tamaku, Kawasaki, Japan
E-Mail: sunnygoh@gmail.com

ABSTRACT

Richards' equation was used to govern water infiltration into the unsaturated soil. Haverkamp's hydraulic properties were used to govern the matric pressure head and hydraulic conductivity relations. Sobol' variance-based method as a global sensitivity analysis tool was coded to study sensitivity and uncertainty values of input parameters. The same input parameters uncertainty values used in Sobol' variance-based method could be subjected to an inverse method to search for a set of globally optimized input parameters. Therefore, a simple Genetic Algorithm, i.e. an inverse method, was coded and used to improve the current framework for the uncertainty analysis. A simple significant digit approximation method was developed, applied on the input parameters, and the generated uncertainty parameter values were found to cause 15.4 % of error, which could be reduced if the number of input parameter significant digit used in the simulation were increased. The simple Genetic Algorithm was able to search for a set of globally optimized input parameters at a low population size, i.e. 50, on three cases tested and also was found able to handle noise data. Data points at the lower and upper plains of the water infiltration front are necessary for the inverse method.

Keywords: Richards' equation sobol', variance-based method, simple genetic algorithm, significant digit approximation.

1. INTRODUCTION

Traditionally, a sensitivity analysis was carried out by changing a single input parameter while keeping other parameters at constant values. It is better known to the statistical community as one-at-a-time (OAT) procedure. A better method is to simultaneously changing more than one parameter at a time known as global sensitivity analysis (GSA). It is because the simultaneous variation of few input parameters at a time could result in sensitivity coefficient that is different from adding up the entire varied individual OAT from base case sensitivity coefficients.

GSA tools can be ranged from qualitative screening method to quantitative technique (Sobol' 1990; Morris 1991; Campolongo *et al*. 2007).Sobol' variance-based method is one of the GSA tools. It is useful for sensitivity analysis and also for uncertainty analysis. In the former, it is used to identify the most and the least sensitive parameters. While in the latter, it is used to determine the greatest influence parameter for parameter prioritization and the least influential parameter for parameter fixing (Saltelli *et al*. 2008). Some models were subjected to this method are SWAT (Nossent *et al*. 2011), inhalation dose model (Avagliano and Parrella 2009), CERES-EGC (Drouet *et al*. 2011) and SLAMM 5 (Chu-Agor *et al*. 2011). Moreover, the variance-based method has been listed in the US Environmental Protection Agency list of attributes as preferred sensitivity analysis method (EPA 2009).

The range of input parameters uncertainty used in Sobol' variance-based analysis could be further processed by an inverse method to determine an optimized set of input parameters. In order to obtain globally optimized set of input parameters, a simple Genetic Algorithm (SGA) was used for this purpose. SGA was introduced by Holland (1975). It is a class of evolutionary algorithms that is a nature inspired computational method. Varieties of GA methods were used in estimating hydraulic properties in unsaturated soils. For instance, SGA used by Schneider *et al*. (2013) and micro-GA used by Ines and Droogers (2002) and Shin *et al.* (2012). Inverse method other than GA, e.g. Levenberg-Marquardt (L-M) algorithm, was used by Kelleners *et al*. (2005) to predict hydraulic properties. Similarly, the L-M algorithm was also used by Mollerup *et al*. (2008). SGA is a global search method, whereas L-M is a local search method. It means that L-M could potentially found a local minimum that is not a global minimum found by SGA. Global minimum means the prediction is closer to experimental or benchmark data than local minimum prediction. Therefore, in the current work, it was limited to SGA.

Numerical simulation of water infiltration is not only allowed prediction of water movement in the unsaturated zone, it has significant implication for contaminant movement (Šimůnek and Bradford 2008). Generally, Richards' equation (Richards 1931)is used to governed water infiltration in variably saturated soils. It has been applied in various disciplines such as hydrology, meteorology, agronomy, and environmental protection (Pachepsky *et al.* 2003). Mollerup *et al.* (2008) coupled Philip's semi-analytical solution of Richards' equation with an inverse method, i.e. L-M algorithm. In the current studies, Philip's semi-analytical solution from Haverkamp *et al*. (1977) was used as benchmark data to verify model simulation for SGA method. The published data from Haverkamp is one of the resources available for modelers to verify their newly developed simulation source code, e.g. Noborio *et al*. (1996), Fayer and Jones (1990), Fayer (2000) and Goh and Noborio (2013).Therefore, the current studies are built on the research for the chosen case that has been a common case study for modelers, and hence, the finding from the present work would be a useful

reference for other researchers with interest in developing model and couple with GSA and SGA methods on Richards' equation.

The aim of these studies is to manage uncertainty input parameter values. Firstly, the water infiltration model was subjected to Sobol' variance-based method to study global sensitivity analysis. Secondly, the model was run through SGA. While the former manages to determine important parameters and also to identify influential parameter for prioritization, the latter extends uncertainty analysis to search for a globally optimum set of input parameters that is best described the water infiltration front. The first objective was achieved by three cases. Case A was based on significant digit approximation that was discussed in Goh and Noborio (2013), and Cases B and C were based on 10 and 20 % variation of input parameters, respectively. The significant digit approximation was to determine the effect of uncertainty in model output due to the uncertainty values of input parameters established from the published data. The model output in all case studies is volumetric water content value.In the SGA analysis, the Gaussian noise was imposed on the benchmark data and the inverse method was used to predict the volumetric water content, under the uncertainty values of input parameters in Cases A, B, and C. The extreme variation of input parameter uncertainty values was also examined.

2. MATERIALS AND METHODS

2.1 The governing equation for water flow in unsaturated soil

The θ_L −based form used in the simulation model is as following:

$$\frac{\partial \theta_L}{\partial t} = \frac{\partial}{\partial z}\left[\left(K\frac{\partial \psi_m}{\partial \theta_L}\right)\frac{\partial \theta_L}{\partial z} - K\vec{k}\right] \quad (1)$$

where: θ_L is the volumetric water content (m^3/m^3); t is the time of simulation (s); z indicates the vertical distance of simulation (m); K is hydraulic conductivity of the medium (m/s);$\psi_m$is the matric pressure head (m);$\vec{k}$is the vector unit with a value of positive one when it is vertically downwards. The $\partial\theta_L/\partial t$ term indicates partial derivative of θ_L with respect to time, while $\partial\psi_m/\partial\theta_L$ term is partial derivative of ψ_m with respect to θ_L. The $\partial\theta_L/\partial z$ is the partial derivative of θ_L with respect to vertical distance. The initial condition volumetric water content value, in Table 1, indicates the initial spatial distribution of water content. The value was used along with upper and lower boundary conditions for simulation. The upper boundary was fixed to a constant volumetric water content given in Table-1. The lower volumetric water content was set equal to adjacent cell. The Equation.1, in general, is known as Richards' equation (Richards 1931). The equation was numerically approximated using finite-difference method, and its algebra was coded in FORTRAN 2008.

2.2 The constitutive functions of matric pressure head and hydraulic conductivity

In this study, the constitutive functions of Haverkamp *et al.* (1977) were used:

$$\psi_m = -10^{-2}\exp\left[\frac{\alpha(\theta_s-\theta_r)}{\theta_L-\theta_r} - \alpha\right]^{\frac{1}{\beta}} \quad (2)$$

$$K = K_s\frac{A}{A+(-100\psi_m)^B} \quad (3)$$

where: α, β, A and B are fitting parameters; θ_r (m^3/m^3) is residual volumetric water content; θ_s (m^3/m^3) is saturated volumetric water content; and K_s (m/s) is saturated hydraulic conductivity.

2.3 Numerical experiment, initial and boundary conditions

Infiltration of Yolo light clay was used as our numerical experiment, and the coefficient values (base case) were listed in Table-1. Initial condition for $\theta_L$was 0.2376 m^3/m^3. The Lower boundary was set permeable to water. The upper boundary was set at 0.495 m^3/m^3. After considering the mass balance ratio (MBR) and the number of iteration, the simulation parameters such as time-step, simulation time, convergence value (CV), and spatial discretization size were set as 500 s, 10^5 s, 10^{-12} m^3/m^3 and 1 cm, respectively.

Table-1. Goes approximately here.

Parameter	Base Case	Case A	Case B	Case C
α	739	738.5 – 739.499	739 – 812.9	739 – 886.8
θ_r (m^3/m^3)	0.124	0.1235 – 0.124499	0.124 – 0.1364	0.124 – 0.1488
θ_s (m^3/m^3)	0.495	0.495 – 0.495499	0.495 – 0.5445	0.495 – 0.594
β	4	3.5 – 4.499	4 – 4.4	4.0 – 4.8
A	124.6	124.55 – 124.6499	124.6 – 137.06	124.6 – 149.52
B	1.77	1.765 – 1.77499	1.77 – 1.947	1.77 – 2.124
K_s (cm/hr)	$4.428\text{x}10^{-2}$	$4.4275\text{x}10^{-2}$ – $4.428499\text{x}10^{-2}$	$4.428\text{x}10^{-2}$ – $4.8708\text{x}10^{-2}$	$4.428\text{x}10^{-2}$ – $5.3136\text{x}10^{-2}$
θ_L(initial cond.) (m^3/m^3)	0.2376	0.23755 – 0.2376499	0.2376 – 0.26136	0.2376 – 0.28512
θ_L(upper bound.) (m^3/m^3)	0.495	0.495	0.495	0.495
Δt (s)	500	500	500	500
Δz (cm)	1	1	1	1

2.4 Sobol' variance-based method as global sensitivity analysis

Model output (Y) is a function of input parameters $(X_1, X_2, \ldots, X_q)$. In a variance-based method, total unconditional variance, V(Y), is given by Sobol' (1990):

$$V(Y) = \sum_i^q V_i + \sum_i^q \sum_{j>i}^q V_{ij} + \cdots + V_{12\ldots q} \quad (4)$$

where $\sum_i^q V_i$ is the sum of partial variances that include individual effects of each input parameter, $\sum_i^q \sum_{j>i}^q V_{ij}$, all the partial variances of two input parameters interaction, and so on.

Dividing equation. 4 by total unconditional variance, V(Y), the equation becomes,

$$\sum_i^q S_i + \sum_i^q \sum_{j>i}^q S_{ij} + \cdots + S_{12\ldots q} = 1 \quad (5)$$

where the first term on the left side of the equation $S_i = V_i/V(Y)$ is first order indices, $S_{ij} = V_{ij}/V(Y)$ is second order indices for the second term, and so on. The indices are scaled between 0 and 1. If there is no interaction between input parameters, $\sum_i^q S_i$ would be equal to unity, and if $\sum_i^q S_i$ is less than unity, it indicates the presence of interaction effects, which could be any combination of indices that is greater than first order.

The quasi-Monte Carlo estimators compiled by Saltelli *et al*. (2010) allowed estimation of first order sensitivity index (S_i) and total effect index (S_{T_i}) for input parameter i. As an example, three input parameters model has a total sensitivity index on input parameter that is given by:

$$S_{T_1} = S_1 + S_{12} + S_{13} + S_{123} \quad (6)$$

$$S_{T_2} = S_2 + S_{12} + S_{23} + S_{123} \quad (7)$$

$$S_{T_3} = S_3 + S_{13} + S_{23} + S_{123} \quad (8)$$

Since the quasi-Monte Carlo estimators can estimates S_i and S_{T_i}, the difference between the two estimators, for instance $(S_{T_1} - S_1)$ would remove single effect value and only higher than first order sensitivity indices are left, which is usually used to indicates interaction effects between input parameters.

The first order sensitivity index and total effect index were estimated by the following quasi-Monte Carlo estimators (Saltelli *et al*. 2010):

$$S_i = 1 - \frac{(1/2N)\sum_{m=1}^N \left(y_B^{(m)} - y_{C_i}^{(m)}\right)^2}{(1/N)\sum_{m=1}^N \left(y_A^{(m)}\right)^2 - f_0^2} \quad (9)$$

$$S_{T_i} = \frac{(1/2N)\sum_{m=1}^N \left(y_A^{(m)} - y_{C_i}^{(m)}\right)^2}{(1/N)\sum_{m=1}^N \left(y_A^{(m)}\right)^2 - f_0^2} \quad (10)$$

where $y_A^{(m)}$, $y_B^{(m)}$, and $y_{C_i}^{(m)}$ are model outputs. Each of these model outputs was given by a set of input parameter values by matrix of N × k. The k columns were referred to the number of input parameters. Sobol' quasi-random sequences were used to generate quasi-random numbers for k-dimensional input parameters, where the k value used in this study was eight. The quasi-random number generated was ranged between zero and unity, and it was scaled with respect to the uncertainty values in Table-1 for Cases A, B, and C. For each row of the matrix, e.g. matrix input parameter values of $y_A^{(m)}$, was a complete set of input parameter values for an execution of a model simulation. The number of model execution depended on the N rows, where the N value used in this study was fifteen thousands. Nossent *et al*. (2011) had shown that an N value of twelve thousands for twenty-six input parameters was sufficient to obtain reliable

estimation of sensitivity indices. Hence, the matrix input parameter values of $y_A^{(m)}$ and $y_B^{(m)}$would require 2N model executions. The matrix input parameter values of $y_{C_i}^{(m)}$was referred to parameteri, and it has N rows for each parameter. Since there were k input parameters that must be tested, and the model executions would be kN. The matrix input parameter values of $y_{C_i}^{(m)}$for parameter i was obtained by using all the columns of matrix input parameter values of$y_A^{(m)}$, except the column of parameter iwas taken from the matrix input parameter values of $y_B^{(m)}$. This step was repeated based on the number of input parameter as indicated by the value k. Therefore, in the current studies, the total model executions that were carried out for each case (Cases A, B, and C) was given by N(2+k)=15,000(8+2)=150,000. All the cases combined would require 450,000 model executions. The f_0^2 was given by $\left((1/N)\sum_{m=1}^{N} y_A^{(m)}\right)^2$.

2.5 Simple genetic algorithm (SGA) as inverse method

The uncertainty values of input parameters were used to generate a random population of initial solutions. The solution is referring to a set of input parameter values for use in model simulation. The random population was generated using Sobol' sequence to spread widely between individual solutions in order to encompass every possible combination of input parameter values that could best describe the model output. Vrugt *et al*. (2008) used Latin hypercube sampling to generate the initial population. Each solution of the population is known as chromosome, which is consisting of input parameter values that were encoded into binary string of 1 and 0. Two initial chromosomes (parent) were randomly selected to go through crossover process in order to generate two new chromosomes (offspring). In this study, a single crossover point was used (Figure-1).

1	0	1	1	0	0	parent 1
0	1	1	0	0	1	parent 2
1	0	1	0	0	1	offspring 1
0	1	1	1	0	0	offspring 2

Figure-1. Goes approximately here.

The offspring was decoded into a real value of each input parameter values, and it was then used as input for the simulation and evaluated with an objective function. The objective function values from the parent and offspring were compared, and if offspring objective function value is better than the parent, the offspring genes will overwrite the parent genes. Similarly, under the process of mutation, two parents were selected, and random genes in the parent were selected for the mutation (Figure-2). They were then decoded and used as input for the simulation and then evaluated with the objective function. Genes from offspring with better objective function value would replace the parent genetic values.

1	0	1	1	0	0	parent 1
0	1	0	0	0	1	parent 2
1	1	1	1	1	0	offspring 1
0	1	1	0	0	0	offspring 2

Figure-2. Goes approximately here.

The crossover and mutation processes would continue based upon the preset number of population size. A complete iteration of crossover and mutation would signify a complete creation of a new generation where the size of the initial population and the new generation are the same.

3 RESULTS AND DISCUSSIONS

3.1 Simulation outcomes of Sobol' variance-based method

There are various forms of an uncertainty of input parameter distribution, for instance, normal, uniform, triangular and lognormal. In this study, the uniform distribution was used; the distribution was based on the significant digit approximation method (Case A), 10 and 20 % increment for Cases B and C, respectively, as in Table-1.The uniform distribution used in the current studies was similar to those used in Saltelli *et al*. (2004), Vrugt *et al*. (2008), Drouet *et al.* (2011), Yang (2011), and Younes *et al.* (2013). The significant digit approximation for Case A was developed from the base case in Table-1. For instance, the residual volumetric water content, i.e. 0.124 m^3/m^3 as the base case, could be approximated by lower and upper bounds of 0.1235 and 0.124499 m^3/m^3, respectively. Vice versa, the lower and upper bounds could be reduced to three significant digits; it would return to the original value as before. Hence, the lower and upper bounds were the minimum and maximum uncertainty limits resulted from the given significant digit of an input parameter. If the base case was to increase its significant digit to 0.1240 m^3/m^3 the uncertainty range for the parameter would be bounded between 0.12395 and 0.1240499 m^3/m^3, respectively, which has a lower uncertainty range than the former example. All the parameters in Case A were developed from the base case values with the same method. Hence, the uncertainty values of input parameters distribution on model output due to the effect of the significant digit approximation was able to be investigated.

Table-2. Goes approximately here.

Parameter	Case A		Case B		Case C	
	S_i	S_{T_i}	S_i	S_{T_i}	S_i	S_{T_i}
α	0	0.00001	0.00089	0.00130	0	0.00143
θ_r	0	0.00002	0	0.00005	0	0.00006
θ_s	0.22980	0.26600	0.37800	0.46200	0.15290	0.27914
β	0.51770	0.58400	0.01630	0.02380	0.00904	0.02684
A	0	0.00010	0.00322	0.00768	0.00027	0.00855
B	0.18580	0.21300	0.30900	0.38900	0.20441	0.33170
K_s	0	0.00002	0.00978	0.01460	0.00578	0.01649
θ_L(initial cond.)	0.00017	0.00019	0.19100	0.19400	0.48030	0.48563

Using the quasi-Monte Carlo estimators from Equations. 9 and 10, the first order sensitivity index (S_i) and total effect index (S_{T_i}) were estimated for every input parameter (Table-2). The model variability was calculated by simply multiplying the first order sensitivity index and total effect index with one hundred. The results for Case A showed that β had the highest $S_i$that 51.77 % of the model variability was due to this factor alone.It was followed by $\theta_s$andB at 22.98 and 18.58 %, respectively. A minor percentage was accounted by θ_L(initialcond.) at 0.017 %. Insignificant contributions were found from parameters α, θ_r, A and K_s. The combined model parameters β, $\theta_s$andBwere found responsible for 93.33 % of all the model variability. By taking $1-\sum_{i=1}^{8} S_i$, only 6.67 % was due to the interaction between parameters for a combination of indices higher than first order. The specific contribution of each input parameter on interaction effect was investigated through the use of $S_{T_i}-S_i$. The interaction effect calculation was found similarly and mainly contributed by β and then followed by $\theta_s$andB at 6.63, 3.62 and 2.72 %, respectively. It is important to note that the combined value of the three parameters was greater than 6.67 %, as given by $1-\sum_{i=1}^{8} S_i$, that was found in the previous calculation. It can be justified by that some second and higher order sensitivity indices were repeatedly calculated and accounted for as shown in the example of Equations. 6-8. In addition, similar to S_i, the S_{T_i} for parameters α, θ_r, A, K_s and θ_L(initialcond.) were found to be insignificant.

While some input parameters were found to be more significant than the others, it should be noted that this was basedon the test case (base case) of the uncertainty values of input parameters that were given in Table 1. Under different uncertainty values of input parameters, the results of S_i and S_{T_i}would be different. To demonstrate our claim, the base case input parameters in Table-1 were modified to have an increment of 10 % to form a new uniform distribution for each parameter, i.e. Case B. In Table-2, the parameter β was no longer providing dominant influence on model variability.Parameters demonstrated significant influences wereθ_s, B and θ_L(initialcond.)at 37.8, 30.9 and 19.1 % of the model variability, respectively. Some characteristics were remain unchanged, for instance, S_{T_i} was greater than S_i, and also parameters with greater S_i resulted in greater S_{T_i}. Since the results in Case B, as in Table-2, were based on 10 % increase in value for all input parameters, in short, it could be termed as a sensitivity analysis. However, the results in Case A were based upon different percentage variations on each input parameter and therefore, it should be termed as uncertainty analysis. The word uncertainty signifies that each time the sensitivity indices are to be calculatedin different cases with different uncertainty values for each input parameter. Hence, in uncertainty analysis, the importance of parameters would be differed from one case to the others. The consistency of sensitivity analysis results was further tested in Case C that all input parameter values were increased by 20 %. The results of Case C showed that the model variability was dominated by θ_L(initialcond.), B, and $\theta_s$in decreasing order. The parameter importance ranking appeared to be different from those from Case B. The discrepancy could be explained, refer to Figures 3(b) and 3(c), that the interception point of three S_i input parameters (θ_s, θ_L(initialcond.), B) was located below and above the horizontal line of 0.185 m depth corresponding to Cases B and C. A general comparison between first order sensitivities of Cases A, B, and C (Figures 3(a)-(c)) showed that there were similarities of sensitivity index curvature for Cases B and C and irrespective of percentage increment applied, while Case A was distinctively different. In a general perspective, sensitivity analysis is a particular case of uncertainty analysis, and it happened only when all parameters were varied with an equal percentage.

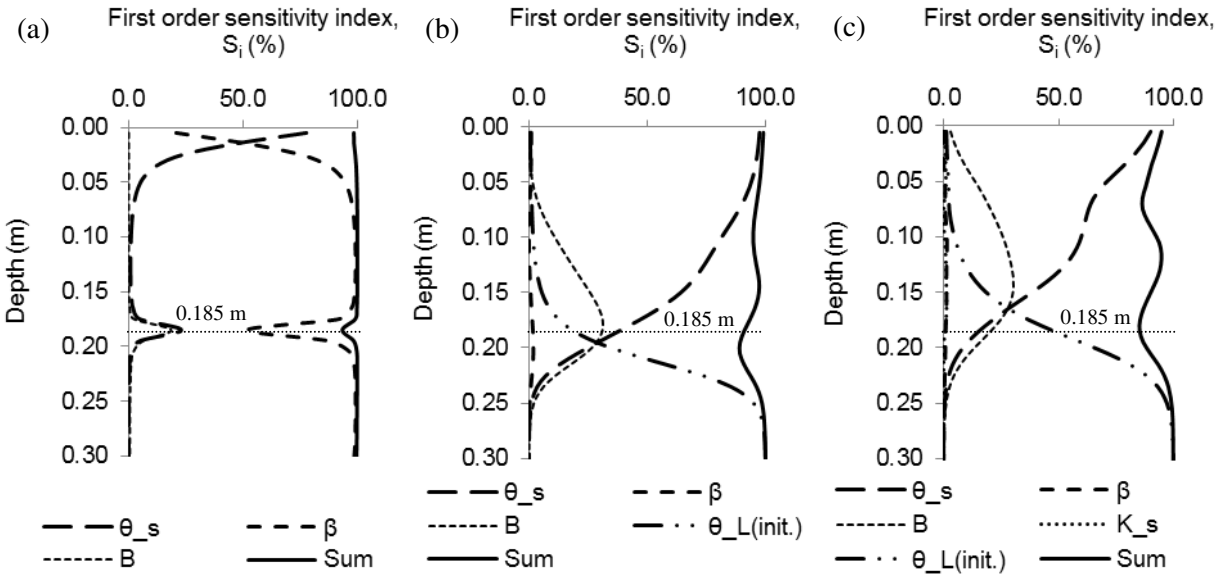

Figure-3(a)-(c). Go approximately here.

In Case A, refer to Figure-3(a), the parameter β was the most important parameter where almost 100 % of model variability could be explained by the parameter alone, except at the near surface layer and the depth of 0.185 m. In Case B, Figure-3(b), $\theta_s$was the single most important parameter at near surface depth. It was gradually decreased and followed by a steady increase of B and also a slow minor increase of θ_L(initialcond.)at adeeper depth between 0.05 and 0.15 m. After 0.15 m depth, the increase of θ_L(initialcond.)was exponential and the decrease of$\theta_s$was steep with increasing depth, while B gradually increase to a peak before it decreases to zero after water infiltration front. The upper depth volumetric water content was largely influenced by θ_s, and at the lower depth it was entirely governed by θ_L(initialcond.). A similar observation was found in Case C and only that the increased and decreased of parameters were found to take place earlier than Case B.

The simulation output was also processed for other statistical measures, Figures 4-6. The figure that indicate the difference between mean and median (Figure-4-6(d)) was found in close trend to that skewness in Figures 4-6(c). The mean value in Figures 4-6(a) appeared to have surrounded by boundaries from 95 percentile confident intervals, as inner boundaries, and Minimum, Maximum values as outer boundaries. Case C has greater variability of volumetric water content than Case B, and also the variability of Case B was found higher than Case A; as shown in Figures 4-6(b). The figure for standard deviation has a similar trend as those observed in Max-Min and standard error of the mean. The standard error of mean data was not shown here. At 0.185 m depth of Case A, there was sudden decreased of uncertainty as indicated in standard deviation, Max-Min and the 95 percentile confident intervals. It was rather an uncommon event as it was not evident in Cases B and C. At a distance before the end of water infiltration front, a peak of kurtosis and skewness was observed in all cases, except in Case A the peak of skewness was rather subtle.At deeper depth of water infiltration front, i.e. some distances somewhere in the middle before the end of water infiltration front and after the rest of 0.4 m depth, kurtosis remained subtle, i.e. approximate -1.2, signifying low and flat peak as shown in the distribution curve in Figure-4(e). For Cases B and C, the low and flat peak happened after the depth of 0.3 m. Skewness and kurtosis were found to have reached its peak after the peak of standard deviation, which suggest that the distribution curves were initially horizontally spread (as indicated by standard deviation) before skewed right (or positive), and raised high frequency (as indicated by kurtosis). The positive (or right) and negative (or left) skewness values appeared to be rather intermittent, but it became zero (no effect of skewness) at deeper depth after the water infiltration front. A normal distribution was rarely found, because the skewness and kurtosis, in Figures4-6(c), were consistently deviated from zero (or perfect normal distribution curve) as shown in Figures4-6(e). In addition, an apparent relation between kurtosis and sensitivity index was observed. In regions, i.e. depths, where only single input parameter has dominant influenced on simulation output, kurtosis value tended to approximate equal or less than -1.0. Since the negative kurtosis value indicates the appearance of small and flat peak, a normal distribution with a unique peak could not be found in these regions. Such observation was predominantly observed in all depths of Case A, except at near surface and areas surrounding 0.185 m depth; whereas in Cases B and C, predominant regions were similarly observed at near surface and depth after water infiltration front. The generated low and flat peak distributions should be defined as uniform distribution where β was found primarily responsible for Case A, while θ_s and θ_L(initialcond.) were found responsible for near surface and depth after water infiltration front, respectively, for corresponding to Cases B and C.

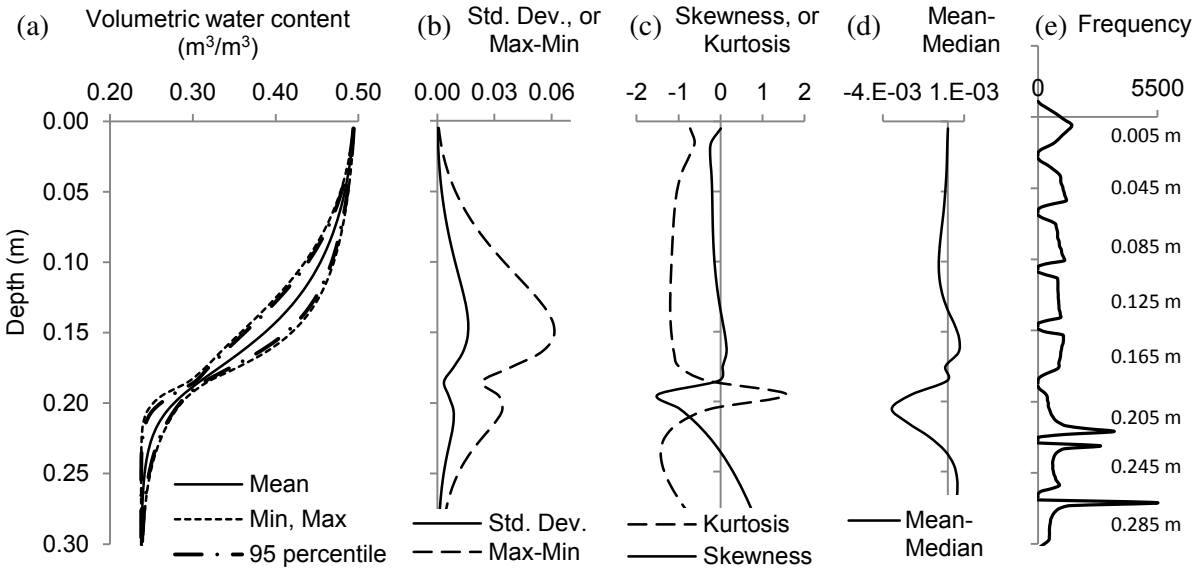

Figure-4(a-e). Go approximately here.

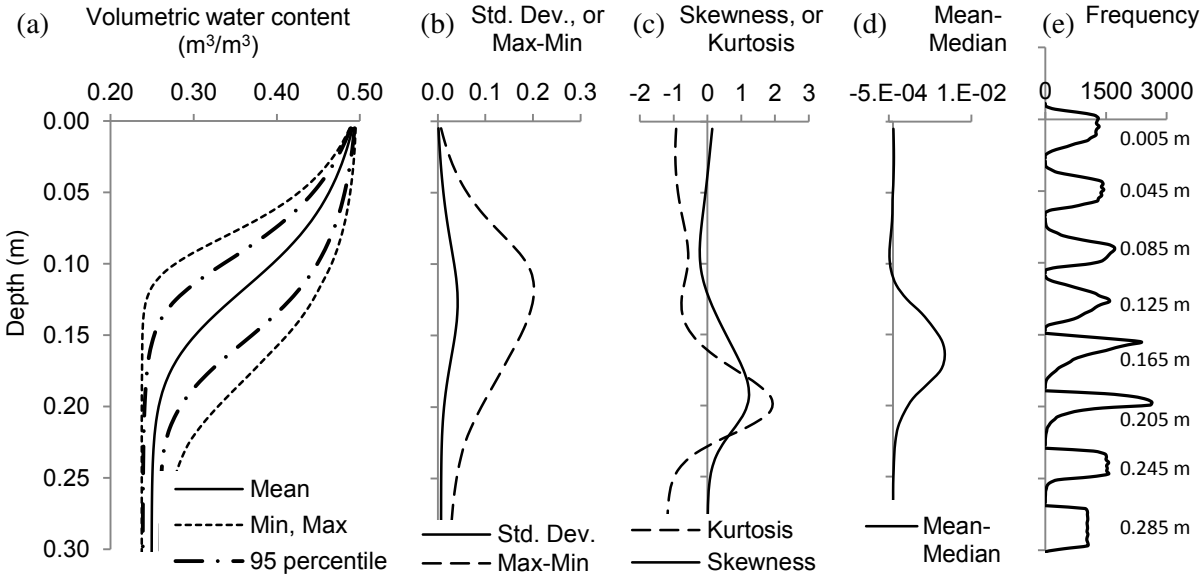

Figure-5. Go approximately here.

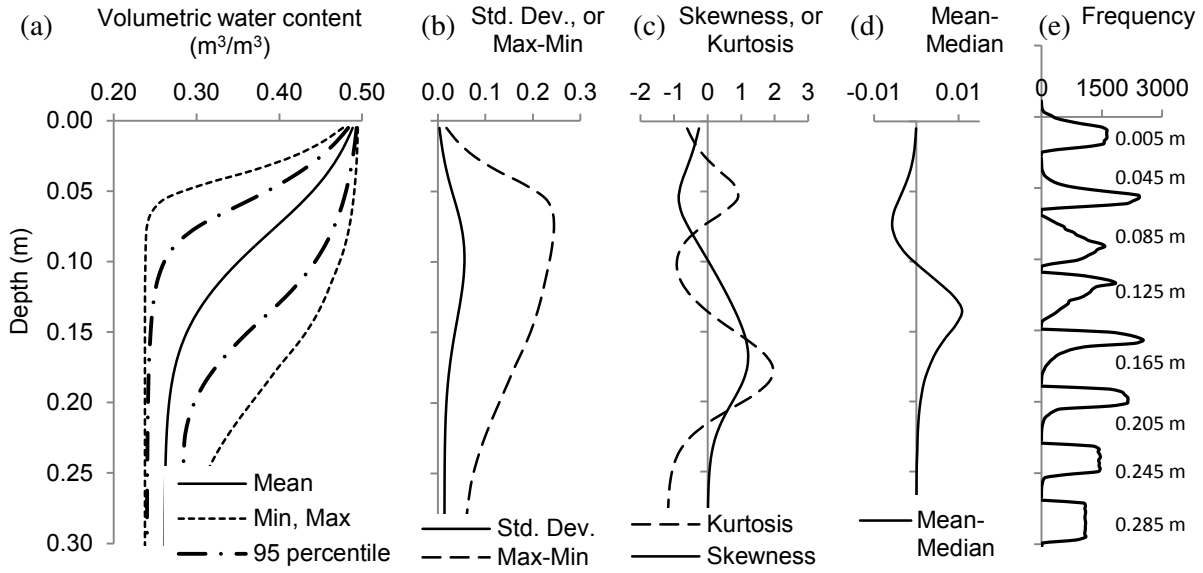

Figure-6. Go approximately here.

A high standard deviation and Max-Min values of Case C than Case B were not unusual given that 20 % increment of input parameter values of the former case, than the latter case of only 10 % increment. The corresponding largest values of Max-Min were found to be 0.245 and 0.202 m^3/m^3 that were equivalent to 51.6 and 45.7 % of error from the base case water content. The significant digit approximation method in Case A resulted in 0.06 m^3/m^3 that can be translated into 15.4 % of error from base case water content.

3.2 Simulation outcomes of simple genetic algorithm (SGA)

In SGA, there are three basic input data required, which were population size, mutation number, and generation number. In the first stage of the studies, the generation number was set at 100 while mutation number was varied between 0 and 90, and population size was studied between 20 and 300. The input parameters of Case C were used at this stage because it has the largest uncertainty values of input parameters compared to the other two cases. The simulation results were compared to Philip's semi-analytical solution from Haverkamp *et al.* (1977), and from now onis known as Philip (H). The objective function of the model was to minimize the sum of the absolute difference between simulated and Philip (H), from now onis known as OF. Each chromosome (or individual) in the population would be assigneda particular OF value based on its simulation output. The primary objective of the SGA is to minimize the OF value. The results showed that the OF value reduced to a minimum value after a number of generations (Figures 7(a)-(b)), i.e. approximately 10 generations. The OF value reached the minimum value at later generation as the population size increases from 50 to 300, as shown in Figures 7(b)-(d). In addition, it was found that neither extremely high nor low mutation number was required to obtain the lowest OF value at each population size. In the case of 20 population size, mutation number was found to be an optimum at 18. The mutation sites needed to obtain small OF values reduced as the population size increased from 50 to 300. At population size of 50, a mutation number of 18 (Figure-7(b)) was found sufficient to result in the smallest OF value, whereas at 200 and 300 population sizes the the lowest OF values were achieved respectively at 9 (Figures 7(c)) and 4 (Figure-7(d)) mutation numbers.

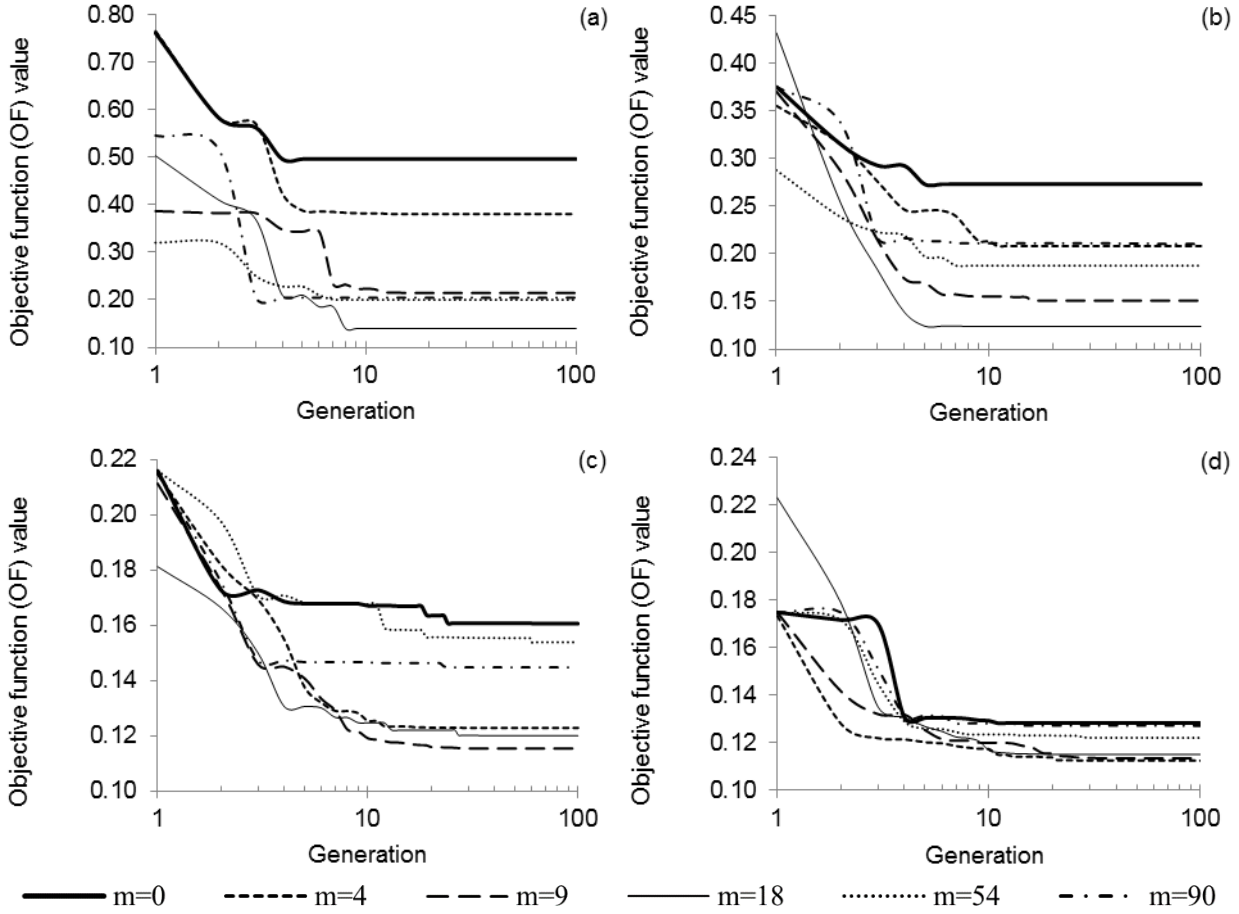

Figure-7. Goes approximately here.

Moreover, the global search for lower OF values indicates that at higher population size the minimum OF value would reduce from 0.1395 to 0.1123 at each consecutive population size from 20 to 300, respectively. The water infiltration front data points from Philip(H) was plotted in Figure-8. They were inverse searched with input parameter uncertainty range of Case C, which was stated earlier to have the most wide uncertainty range that its success in inverse search should presumably be applicable to Cases A and B. The Philip(H) was inverse searched with different population sizes from 20 to 300. The results of the inverse search by population sizes of 50 and 300 were showed in Figure-8. A similar trend of water content was observed from the two different population sizes from the surface of infiltration model until the depth of 0.25 m. However, a significant discrepancy was observed between the predictions of both population sizes at water infiltration depth after 0.25 m. Although both population sizes of 50 and 300 have showed approximately equal prediction of water content at the near surface, their values were also significantly deviated from the Philip (H), after 0.25 m infiltration depth.

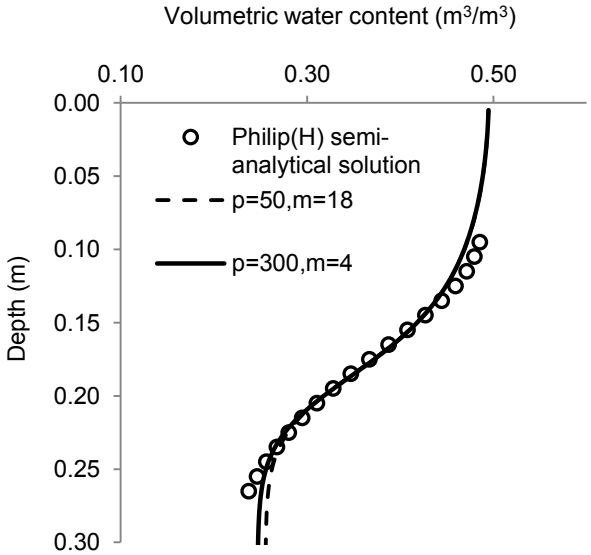

Figure-8. Go approximately here.

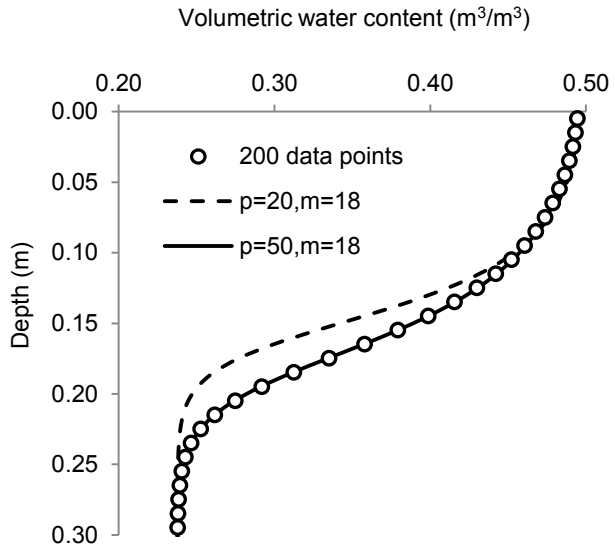

Figure-9. Go approximately here.

Apparently, the result up to the current stage, as in Figure-8 indicates that SGA was not able to reproduce a good result at two end tails of water infiltration front. A hypothesis was established that the insufficient data at both ends were the result of unabling the SGA to search properly and subsequently result in failure to represent the characteristic of water infiltration adequately. Hence, the hypothesis was tested by increasing the data points from 18 to 200 and re-run the SGA. The result in Figure-9 shows that the population size of 50 was already sufficient to obtain a good agreement between numerical data and simulation. Similarly, at higher population sizes, e.g. 100-500, were able to reproduce a good fitting result. It suggests that sufficient water content data at the upper and lower plains of water infiltration front were necessary, and they dictate the accuracy of SGA prediction. Without such information, the SGA would not be able to predict a leveling off of the saturation water content and the residual water content.

In addition, SGA was subjected to artificial noise to test its ability to operate on non-smooth water content data. Three different Gaussian noises were imposedin each case of A, B, and C, which is equivalent to a total of nine independent investigations. An example of Gaussian noise application on data has been demonstrated in Balasubramanian and Schwartz (2002). SGA was found able to generate smooth curve for each type of noise for each case. Figure-10 shows the selected few cases of the water infiltration front where a reasonably good fitting result was achieved for each case. It was noticeable that noise data do not impede the function of SGA as an inverse method, but the noise data were rather influence the curvature shape of the water infiltration front.

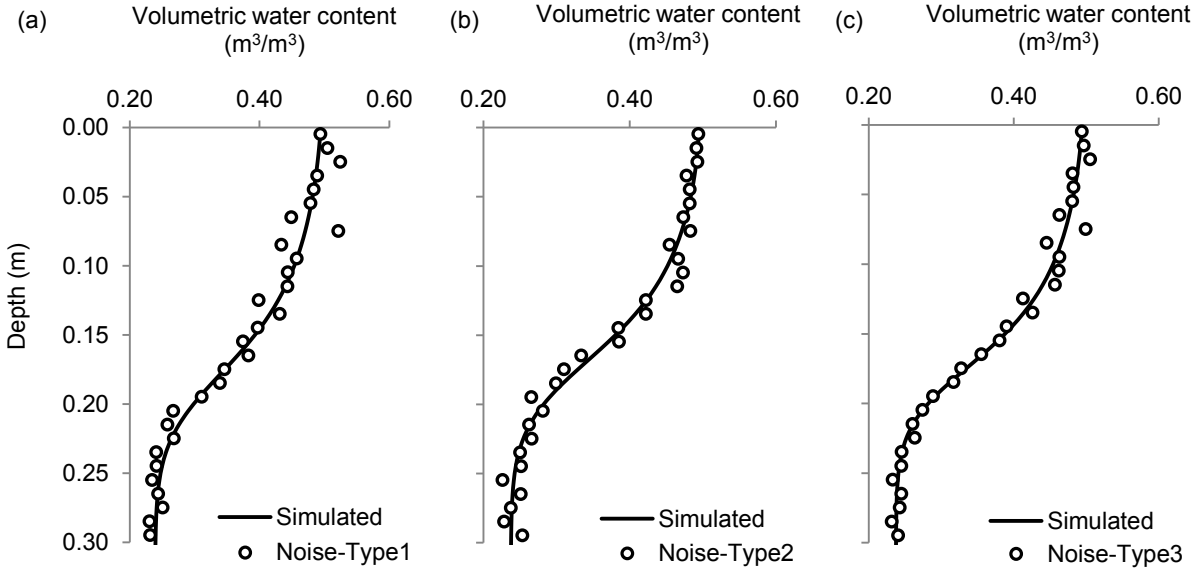

Figure-10. Goes approximately here.

Additional tests on the robustness of the SGA were also carried out. Apart from the input parameter range in Cases A, B, and C, random variation of input parameter was tested. Again, the population size of 50 was found sufficient to reproduce the water infiltration front. At extreme case when all the lower boundaries of the input parameters uncertainty values were to reduce to zero, except the saturated water content parameter, the SGA was found able to reproduce a reasonably good result at the expense of higher population size, i.e. 300. When the upper limit of each input parameter was increased to few hundreds of percentage, on top of the previous case condition, a population size of 500 was needed to reproduce a reasonable good result.

4. CONCLUSIONS

Saltelli and Annoni (2010) suggested that all models should be tested by global sensitivity analysis in order to avoid perfunctory sensitivity analysis, i.e. local sensitivity analysis, due to an incomplete investigation of

input spaces. Identification of relevant parameters from those of least importance is necessary. The uncertainty values of input parameters could be subjected to the SGA to find a set of globally optimized input parameters values that was best described the water infiltration model. Those models with large number of input parameters may prefer to first screening out uninfluential parameters using elementary effect method (EEM) before processed by Sobol' variance-based method and SGA. Hence, the computational time would tremendously reduce under reduced number of input parameters.

In Sobol' variance-based method, the equal percentage variation on input parameters was showed to lead to sensitivity analysis (Cases B and C), while different percentages variation of each input parameter would result in uncertainty analysis (Case A). It is simply because those in different case studies with differentpercentage variations on each input parameter would result in different first order and total effect indices values. In addition, the simulated output was significantly influenced by saturated volumetric water content at near surface depth; and at deeper depth after the water infiltration front, the output was solely influenced by initial water content used in the simulation. It thus suggested that the accuracy of water content prediction in these regions would depend on a narrower range of uncertainty values of the related parameters. However, in a situation when a particular parameter, e.g. β in Case A, has a greater values of uncertainty than other parameters, the previous observation does not apply. A simple case study of Case A has demonstrated that significant digit approximation that was developed based on existing significant digit of published data was found responsible for 15.4 % of error. Such error could be reduced by providing a greater number of significant digits on the input parameters values.

The SGA was found capable of searching for globally optimized input parameter values. A population size of 50 was found successful of searching and reproduced water infiltration front in all cases tested, provided that sufficient data at the upper and lower plains of the water infiltration front were provided. SGA was able to handle a random range of input parameter values and also noise data. At extreme cases when extremely low and high values were used as search space for input parameters, the population size needed to perform the inverse method of SGA would be increased.

REFERENCES

Avagliano S., Parrella L. 2009. Managing Uncertainty in Risk-Based Corrective Action Design: Global Sensitivity Analysis of Contaminant Fate and Exposure Models Used in the Dose Assessment. Environ. Model Assess. 14(1): 47-57. doi:10.1007/s10666-008-9163-5.

Balasubramanian M., Schwartz E.L. 2002. The Isomap Algorithm and Topological Stability. Science 295(5552), 7. doi:10.1126/science.295.5552.7a.

Campolongo F., Cariboni J., Saltelli A. 2007. An effective screening design for sensitivity analysis of large models. Environmental Modelling & Softw. 22(10): 1509-1518. doi:http://dx.doi.org/10.1016/j.envsoft.2006.10.004.

Chu-Agor M.L., Muñoz-Carpena R., Kiker G., Emanuelsson A., Linkov I. 2011. Exploring vulnerability of coastal habitats to sea level rise through global sensitivity and uncertainty analyses. Environmental Modelling & Softw. 26(5): 593-604. doi:http://dx.doi.org/10.1016/j.envsoft.2010.12.003.

Drouet J.L., Capian N., Fiorelli J.L., Blanfort V., Capitaine M., Duretz S., Gabrielle B., Martin R., Lardy R., Cellier P., Soussana J.F. 2011. Sensitivity analysis for models of greenhouse gas emissions at farm level. Case study of N_2O emissions simulated by the CERES-EGC model. Environmental Pollut. 159(11): 3156-3161. doi:http://dx.doi.org/10.1016/j.envpol.2011.01.019.

EPA: 2009. Guidance on the Development, Evaluation, and Application of Environmental Models, Technical report, Office of the Science Advisor, Council for Regulatory Environmental Modeling, March 2009, EPA/100/K-09/003, http://www.epa.gov/crem/library/cred_guidance_0309.pdf, accessed 2 June 2014 In.

Fayer M.J. 2000. UNSAT-H Version 3.0: Unsaturated Soil Water and Heat Flow Model - Theory, User Manual, and Examples. In., vol. PNNL-13249. Pacific Northwest Laboratory, Richland, Washington.

Fayer M.J., Jones T.L. 1990. UNSAT-H Version 2.0: Unsaturated soil water and heat flow model. In., vol. PNL-6779. Pacific Northwest Laboratory, Richland, Washington.

Goh E.G., Noborio K. 2013. Sensitivity Analysis on the Infiltration of Water into Unsaturated Soil. In: Proceedings of Soil Moisture Workshop 2013, Hiroshima University Tokyo Office in Campus Innovation Center. pp. 66-68.

Haverkamp R., Vauclin M., Touma J., Wierenga P.J., Vachaud G. 1977. A Comparison of Numerical Simulation Models for One-Dimensional Infiltration1. Soil Sci. Soc. Am. J. 41(2): 285-294. doi:10.2136/sssaj1977.03615995004100020024x.

Holland J.H. 1975. Adaptation in Natural and Artificial Systems. University of Michigan Press.

Ines A.V.M., Droogers P. 2002. Inverse modelling in estimating soil hydraulic functions: a Genetic Algorithm approach. Hydrol. Earth Syst. Sci. 6(1): 49-66. doi:10.5194/hess-6-49-2002.

Kelleners T.J., Soppe R.W.O., AyarsJ.E., Šimůnek J., Skaggs T.H. 2005. Inverse Analysis of Upward Water

Flow in a Groundwater Table Lysimeter. Vadose Zone J. 4(3): 558-572. doi:10.2136/vzj2004.0118.

Mollerup M., Hansen S., Petersen C., Kjaersgaard J.H. 2008. A MATLAB program for estimation of unsaturated hydraulic soil parameters using an infiltrometer technique. Computers & Geosciences. 34(8): 861-875. doi:http://dx.doi.org/10.1016/j.cageo.2007.12.002.

Morris M.D. 1991. Factorial sampling plans for preliminary computational experiments. Technometrics. 33(2): 161-174. doi:10.2307/1269043.

Noborio K., McInnes K.J., Heilman J.L. 1996. Two-Dimensional Model for Water, Heat, and Solute Transport in Furrow-Irrigated Soil: II. Field Evaluation. Soil Sci. Soc. Am. J. 60(4): 1010-1021. doi:10.2136/sssaj1996.03615995006000040008x.

Nossent J., Elsen P., Bauwens W. 2011. Sobol' sensitivity analysis of a complex environmental model. Environmental Modelling & Softw. 26(12): 1515-1525. doi:http://dx.doi.org/10.1016/j.envsoft.2011.08.010.

Pachepsky Y., Timlin D., Rawls W. 2003. Generalized Richards' equation to simulate water transport in unsaturated soils. J. of Hydrology 272(1-4): 3-13. doi:http://dx.doi.org/10.1016/S0022-1694(02)00251-2.

Richards L.A. 1931. Capillary Conduction of Liquids through Porous Mediums. J. of Appl. Phys. 1(5): 318-333. doi:http://dx.doi.org/10.1063/1.1745010.

Saltelli A., Annoni P. 2010. How to avoid a perfunctory sensitivity analysis. Environmental Modelling & Softw. 25(12): 1508-1517. doi:http://dx.doi.org/10.1016/j.envsoft.2010.04.012.

Saltelli A., Annoni P., Azzini I., Campolongo F., Ratto M., Tarantola S. 2010. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Computer Phys. Communications. 181(2): 259-270. doi:http://dx.doi.org/10.1016/j.cpc.2009.09.018.

Saltelli A., Ratto M., Andres T., Campolongo F., Cariboni J., Gatelli D., Saisana M., Tarantola S. 2008. Variance-Based Methods. In: Global Sensitivity Analysis. The Primer. pp. 155-182. John Wiley & Sons, Ltd.

Saltelli A., Tarantola S., Campolongo F., Ratto M. 2004. The Screening Exercise. In: Sensitivity Analysis in Practice. pp. 91-108. John Wiley & Sons, Ltd.

Schneider S., Jacques D., Mallants D. 2013. Inverse modelling with a genetic algorithm to derive hydraulic properties of a multi-layered forest soil. Soil Res. 51(5): 372-389. doi:http://dx.doi.org/10.1071/SR13144.

Shin Y., Mohanty B.P., Ines A.V.M. 2012. Soil hydraulic properties in one-dimensional layered soil profile using layer-specific soil moisture assimilation scheme. Water Resources Res. 48(6): W06529. doi:10.1029/2010WR009581.

Šimůnek J., Bradford S.A. 2008. Vadose zone modeling: Introduction and importance. Vadose Zone J. 7(2): 581-586.

Sobol I.M.: 1990. On sensitivity estimates for nonlinear mathematical models. Matematicheskoe Modelirovanie. 112-118.

Vrugt J.A., Stauffer P.H., Wöhling T., Robinson B.A., Vesselinov V.V. 2008. Inverse Modeling of Subsurface Flow and Transport Properties: A Review with New Developments. Vadose Zone J. 7(2): 843-864. doi:10.2136/vzj2007.0078.

Yang J. 2011 Convergence and uncertainty analyses in Monte-Carlo based sensitivity analysis. Environmental Modelling & Softw. 26(4): 444-457. doi:http://dx.doi.org/10.1016/j.envsoft.2010.10.007.

Younes A., Mara T.A., Fajraoui N., Lehmann F., Belfort B., Beydoun H. 2013 Use of Global Sensitivity Analysis to Help Assess Unsaturated Soil Hydraulic Parameters. Vadose Zone J. 12(1): doi:10.2136/vzj2011.0150.