  Water Science & Technology | 63.10 | 2011

Assessing the convergence of LHS Monte Carlo simulations of wastewater treatment models

Lorenzo Benedetti, Filip Claeys, Ingmar Nopens and Peter A. Vanrolleghem

ABSTRACT

Monte Carlo (MC) simulation appears to be the only currently adopted tool to estimate global sensitivities and uncertainties in wastewater treatment modelling. Such models are highly complex, dynamic and non-linear, requiring long computation times, especially in the scope of MC simulation, due to the large number of simulations usually required. However, no stopping rule to decide on the number of simulations required to achieve a given confidence in the MC simulation results has been adopted so far in the field. In this work, a pragmatic method is proposed to minimize the computation time by using a combination of several criteria. It makes no use of prior knowledge about the model, is very simple, intuitive and can be automated: all convenient features in engineering applications. A case study is used to show an application of the method, and the results indicate that the required number of simulations strongly depends on the model output(s) selected, and on the type and desired accuracy of the analysis conducted. Hence, no prior indication is available regarding the necessary number of MC simulations, but the proposed method is capable of dealing with these variations and stopping the calculations after convergence is reached.

Key words | computation time, global methods, sensitivity and uncertainty analysis

Lorenzo Benedetti (corresponding author)
Ingmar Nopens
BIOMATH, Ghent University,
Coupure links 653, B-9000,
Ghent, Belgium
E-mail: *lbene@biomath.ugent.be*

Lorenzo Benedetti (corresponding author)
Filip Claeys
MOSTforWATER N.V.,
Sint-Sebastiaanslaan 3a,
B-8500 Kortrijk,
Belgium

Peter A. Vanrolleghem
model*EAU*, Département de génie civil et génie des eaux,
Université Laval, Pavillon Pouliot,
1065 av. de la Médecine,
Québec QC,
G1V 0A6,
Canada

INTRODUCTION

The recent mounting demand for uncertainty analysis (UA) and global sensitivity analysis (GSA) in wastewater treatment modelling (Belia *et al.* 2009) entails the use of appropriate (methodological) tools to perform such studies (Benedetti *et al.* 2008). The former, UA, concerns the propagation of parameter uncertainty to output uncertainty. The latter, GSA, aims at quantifying the influence of parameter variation, within their whole (global) domain, to model output changes.

The basis of all the (rare) UA and GSA studies so far conducted in wastewater treatment is Monte Carlo (MC) simulation (e.g. Benedetti *et al.* 2010; Bixio *et al.* 2002; Flores-Alsina *et al.* 2008; McCormick *et al.* 2007; Martin 2009; Sin *et al.* 2009).

MC simulation requires a large number of runs of wastewater treatment models, which are complex, non-linear and dynamic with high computational demands. It is, therefore, very important to define and apply a stopping rule to have the minimum number of runs sufficient to satisfactorily obtain the result pursued by the modeller. This is actually a quite unexplored aspect of MC simulation, also in domains other than wastewater treatment plant (WWTP) modelling (Ballio & Guadagnini 2004; Ata 2007). Usually a large number of runs is executed and subsequently it is checked whether the quantity of interest has converged or not (e.g. Donigian & Love 2007; Rousseau *et al.* 2001).

In this work, Latin hypercube sampling (LHS) was used, a pseudo-random sampling technique that allows to evenly explore the parameter space and hence reduce the number of MC runs compared to pure random sampling (McKay *et al.* 1979). A first indication for the required number of LHS simulations was given by Iman & Helton (1985), who suggest running a minimum number of runs of at least 4/3 times the number of uncertain parameters. The same rule is suggested by Manache & Melching (2007), although also noting that it may not be sufficient for models with highly non-linear properties (e.g. WWTP models). Some indication about the necessary number of runs, to be calculated *a priori* as a function of the desired percentile and confidence interval, can be found in Morgan & Henrion (1990), but,

doi: 10.2166/wst.2011.453

according to the authors' experience, this is also a function of the complexity of the model and of the number of its parameters.

In this article, a very simple and intuitive empirical method to define stopping rules of LHS-MC simulation is introduced and illustrated by means of a case study of WWTP modelling.

THE PROPOSED CONCEPT

MC simulation should stop when the UA/GSA results "converge". When defining a stopping rule of MC simulation, it is really up to the modeller to decide when the MC simulation results have reached a satisfactory convergence. Convergence can be defined as small variability of MC simulation results (e.g. the average effluent concentration) obtained from N to $N+n$ runs, where n is the increment of runs for which the MC simulation results are evaluated. This small variability must prove to remain small after a number of runs.

The main user inputs of the method would be the selection of:

1. the model outputs;
2. the number of runs n per batch of MC simulations;
3. the maximum number of batches b;
4. one or more criteria (linked to the model outputs) to evaluate the MC convergence; in this article the adopted criteria to be quantified are:
   - the width of the band within which the variability is deemed acceptable;
   - the length of the band (number of runs) necessary to consider the small variability as stable.

The concept proposed in this article is similar to the more complicated method of Ata (2007), who also suggests using width and length of a band, but requires a preliminary MC experiment to estimate quantities.

MATERIALS AND METHODS

The case study used to illustrate the method is the model of the Eindhoven WWTP (Benedetti *et al.* 2009), designed for nutrient removal. It is a rather complex model, with 16 activated sludge units modelled with a modified ASM2d model (Gernaey & Jørgensen 2004), five settlers, two buffer tanks and five complex controllers (two for aeration systems and three for flow regulation systems). The model is fed with an 8-day input file with data every 15 min and very dynamic conditions of dry and wet weather, generating output data with the same frequency. The model is implemented in the WEST® simulator (MOSTforWATER, Belgium) and each run takes on average 6.5 s.

The model outputs used for the GSA and evaluated for each single run were:

- average NH_4 concentration in the effluent;
- maximum NH_4 concentration in the effluent;
- average total suspended solids (TSS) concentration in the effluent;
- maximum TSS concentration in the effluent.

The MC simulation was carried out by assigning uniform distributions to 15 operational parameters of the model and sampling from those distributions with LHS. Batches of 100 runs ($n=100$) of 8 days were performed for 30 times ($b=30$) for a total of 3,000 model runs; this value for n was deemed a reasonable trade-off, as having n too large would prevent achieving significant savings in the case that convergence is quickly reached (running too many runs), and choosing n too small would prevent properly exploring the parameter space, therefore losing the advantages of LHS (reducing the stratification of sampling), unless incremental LHS sampling techniques are used (Stein 1987). The notation in the following will be: $N_i = N_{i-1}+n$ with $i = 1, \ldots, b$.

For each batch of 100 runs the seed of the LHS algorithm was changed to have completely independently sampled parameter values.

A GSA with linear regression and calculation of the standardized regression coefficient (SRC) values (for more details on the method see Saltelli *et al.* 2004) was executed after each batch on the cumulative number of output files (100, 200, … , 3,000) for a total of 30 times to check the convergence of the GSA results.

The two convergence criteria evaluated after each batch of 100 runs for the cumulative number of runs were the following:

- *model output variability*: expressed as the percentage of change of model output from N_{i-1} to N_i runs, calculated for the average, 5th, 50th and 95th percentiles, for each of the four model outputs; a possible width of variability band to define small differences may be ±1%; the variability is stable with N_i runs if the variability stays within the band with $N_i, N_{i-1}, \ldots, N_{i-j}$ runs, where j is a number of batches considered sufficient to prove stability (length of the variability band).

- *stability in parameter selection*: the ranking and selection of significant parameters according to their SRCs; the selection is stable with N_i runs if the selection of parameters is the same with N_i, $N_{i-1} \ldots N_{i-j}$ runs.

To summarize, the user inputs in this example are the following:

1. outputs: average and maximum NH_4 and TSS effluent concentrations;
2. $n = 100$;
3. $b = 30$;
4. criteria: *model output variability* and *stability in parameter selection*:
   - band width for variability: ±1%;
   - band length for variability: a suggestion will be provided in the results.

RESULTS AND DISCUSSION

The results of the *model output variability* criterion for the four model outputs are shown in Figures 1 and 2. A few observations:

- in all cases, the calculated quantities seem to converge after 3,000 runs;
- in general, in the first few batches the 5th and 95th percentiles tend to show more variability;
- output averages converge faster than 50th percentiles;
- TSS shows more convergence problems than NH_4;
- for NH_4 average and maximum, the number of runs after which the percentage relative difference (output variability) is always below 1% is 800 and 600 respectively;
- for the TSS average and maximum, the output variability is always below 1% after 1,100 and 1,300 runs respectively;
- the maximum number of batches between the first time that all quantiles are within the ±1% band and the next time at least one quantile is outside that band is five, which happens for NH_4 average between batch 3 and batch 7; this means that, in this case, the suggested band length for *model output variability* is five batches of 100 runs each.

In the following paragraphs, the *stability in parameter selection* is analysed for the four model outputs. Table 1 shows the parameter ranking and the cumulative sensitivity fraction (CSF) for the four selected model outputs, both for 30 batches and for the number of batches after which the *model output variability* remains within the ±1% band (see Figures 1 and 2). The higher the absolute value of SRC, the more sensitive the output is towards variations of that parameter. The CSF for a given parameter is (see Equation (1)) the sum of the absolute values of the

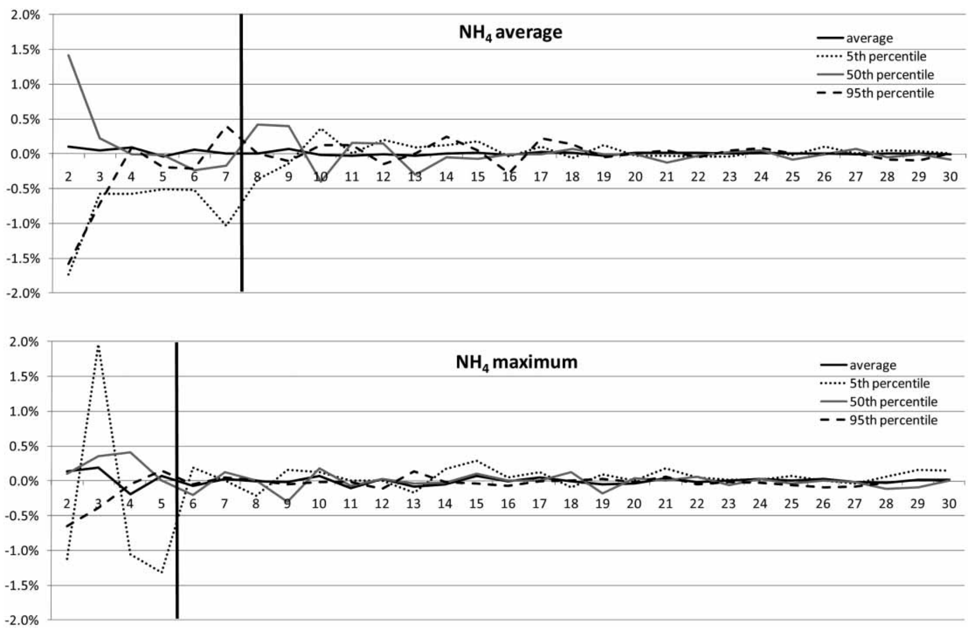

Figure 1 | *Model output variability* for NH_4 concentration in the WWTP effluent; on the *y* axis the percentage relative difference of model output between batch $i-1$ and i; on the *x* axis the batch number i (the difference between i and $i-1$ is calculated only from the second batch on); after the vertical line the variability stays within the ±1% band.

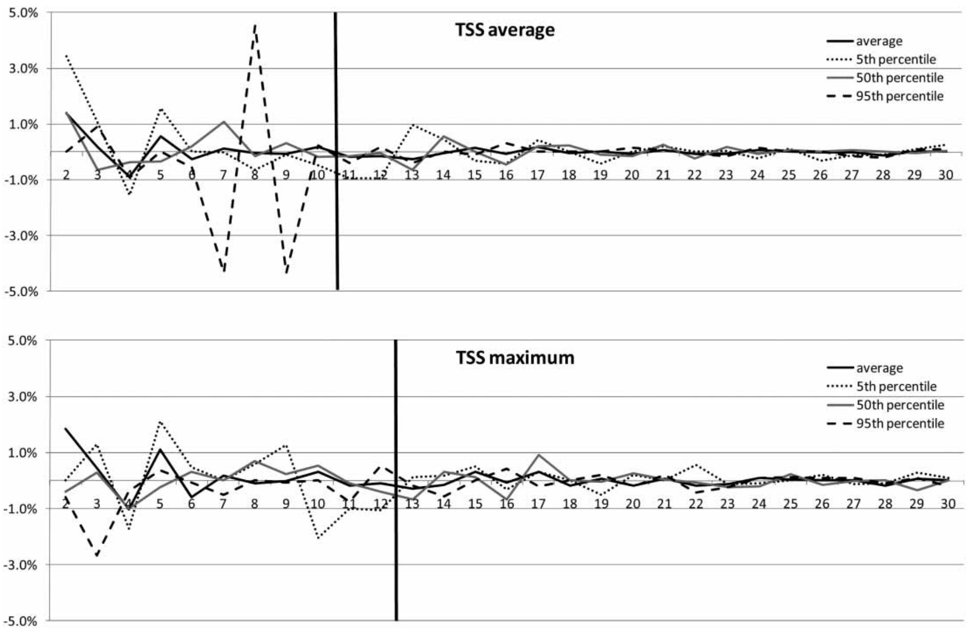

Figure 2 | *Model output variability* for TSS concentration in the WWTP effluent; on the *y* axis the percentage relative difference of model output between batch $i-1$ and i; on the *x* axis the batch number i (the difference between i and $i-1$ is calculated only from the second batch on); after the vertical line the variability stays within the ±1% band.

Table 1 | *Stability in parameter sensitivity ranking*: parameter rank and cumulative sensitivity fraction (CSF) for the four selected model outputs after 30 batches and for the number of batches after which the *model output variability* is always below the ±1% band (see Figures 1 and 2); in grey cells the parameters which have a total CSF up to at least 0.9 (explaining 90% of the output variability); "par." stands for "parameter name".

NH_4 average				NH_4 maximum				TSS average				TSS maximum			
30		8		30		6		30		11		30		13	
par.	CSF	par.	CSF	par.	CSF	par.	CSF	par.	CSF	par.	CSF	par.	CSF	par.	CSF
J	0.21	N	0.21	I	0.28	I	0.27	H	0.46	H	0.45	H	0.54	H	0.52
N	0.41	J	0.41	H	0.51	H	0.51	G	0.73	G	0.71	I	0.82	I	0.79
K	0.58	K	0.58	N	0.63	N	0.63	I	0.92	I	0.89	G	0.91	G	0.88
L	0.65	O	0.65	J	0.73	J	0.72	A	0.94	K	0.91	A	0.93	E	0.90
O	0.72	L	0.73	K	0.83	K	0.81	K	0.95	E	0.93	L	0.94	J	0.92
M	0.78	M	0.79	F	0.87	O	0.86	D	0.96	C	0.94	E	0.96	A	0.93
F	0.83	C	0.85	O	0.90	F	0.91	F	0.96	A	0.95	D	0.96	O	0.94
C	0.88	F	0.89	M	0.92	L	0.93	C	0.97	O	0.96	M	0.97	C	0.95
E	0.92	E	0.93	L	0.94	M	0.95	E	0.98	N	0.97	C	0.98	L	0.96
H	0.95	H	0.95	E	0.96	C	0.96	L	0.99	F	0.98	K	0.98	K	0.97
I	0.97	I	0.97	D	0.98	D	0.97	N	0.99	B	0.98	J	0.99	F	0.98
D	0.98	D	0.98	A	0.99	G	0.98	B	1.00	J	0.99	N	0.99	D	0.99
B	0.99	A	0.99	B	1.00	B	0.99	M	1.00	M	0.99	F	1.00	B	0.99
A	1.00	B	1.00	C	1.00	A	1.00	J	1.00	D	1.00	O	1.00	N	1.00
G	1.00	G	1.00	G	1.00	E	1.00	O	1.00	L	1.00	B	1.00	M	1.00

sensitivity coefficients (in this case the SRCs) from the highest absolute value down to the absolute value of the SRC of that given parameter, divided by the sum of the absolute values of the SRCs of all parameters. A criterion to select significant parameters could be to consider as significant the set of parameters that have a CSF value up to 0.9, i.e., the parameters that all together describe 90% of the output variability.

$$CSF_m = \frac{\sum_{k=1}^{m} |SRC_k|}{\sum_{k=1}^{p} |SRC_k|} \quad (1)$$

where the SRCs are ranked in decreasing order of absolute value (e.g. SRC_1 has the largest absolute value), m is the rank of the parameter for which the CSF is calculated and p is the total number of parameters (15 in this case).

From Table 1 it can be observed that for NH_4 average and NH_4 maximum *stability in parameter selection* is already achieved when *model output variability* within the ±1% band is achieved. The number of batches necessary to reach *stability in parameter selection* for TSS average and TSS maximum is larger (12 and 23 respectively, not shown) than the one to have *model output variability* within the ±1% band (11 and 33 respectively), indicating that the stability of the TSS-influential parameter set is less than the stability of the statistical properties of TSS. Stopping after a lower number of batches would have led anyway to a conservative decision for the selection of parameters significant for TSS, as in both cases one more parameter is included as significant with fewer batches. This property of stability cannot be proven, but it has always been noticed by the authors, with several models of different complexity.

It is clear that the achievement of convergence is strongly dependent on the criterion adopted to evaluate the convergence, and that for the given criteria the number of runs necessary to reach convergence varies between 40 and 150 times the number of uncertain parameters, in contrast to what was found in previous works (only by checking the convergence, without the incremental approach of this work). In particular, Rousseau *et al.* (2001) accepted 15 times, Donigian & Love (2007) 20 times, Manache & Melching (2007) 4/3 to three times and Benedetti *et al.* (2010) 50 times.

CONCLUSIONS

A method to find the minimum required number of runs in MC simulations of WWTP models has been presented and illustrated by means of a case study. Its pragmatism and easy automation make it suitable for engineering applications.

The main conclusion from the case study is that different criteria to assess the MC convergence lead to very different numbers of runs. This is actually in contrast to what is reported in previous works.

In particular, with the complex non-linear model used and with 15 uncertain parameters, the number of runs necessary to reach convergence for the different combinations of model outputs and convergence criteria varies between 40 and 150 times the number of uncertain parameters.

ACKNOWLEDGEMENTS

Lorenzo Benedetti is a post-doctoral researcher of the Special Research Fund (BOF) of Ghent University. Peter A. Vanrolleghem holds the Canada Research Chair in Water Quality Modelling. They are respectively Secretary and Vice-Chair of the IWA Task Group on Design and Operations Uncertainty.

REFERENCES

Ata, M. Y. 2007 A convergence criterion for the Monte Carlo estimates. *Simulation Modelling Practice and Theory* **15**, 237–246.

Ballio, F. & Guadagnini, A. 2004 Convergence assessment of numerical Monte Carlo simulations in groundwater hydrology. *Water Resources Research* **40**, W04603.

Belia, E., Amerlinck, Y., Benedetti, L., Johnson, B., Sin, G., Vanrolleghem, P. A., Gernaey, K. V., Gillot, S., Neumann, M. B., Rieger, L., Shaw, A. & Villez, K. 2009 Wastewater treatment modelling: dealing with uncertainties. *Water Science and Technology* **60** (8), 1929–1941.

Benedetti, L., Bixio, D., Claeys, F. & Vanrolleghem, P. A. 2008 Tools to support a model-based methodology for emission/immission and benefit/cost/risk analysis of wastewater treatment systems which considers uncertainties. *Environmental Modelling & Software* **23** (8), 1082–1091.

Benedetti, L., De Baets, B., Nopens, I. & Vanrolleghem, P. A. 2010 Multi-criteria analysis of wastewater treatment plant design and control scenarios with the Benchmark simulation model No. 2 under uncertainty. *Environmental Modelling & Software* **25** (5), 616–621.

Benedetti, L., Jonge, J. D., Amerlink, Y., Plano, S., Nopens, I. & Weijers, S. 2009 Wet-weather treatment upgrade scenarios with sensitivity and uncertainty analysis at the Eindhoven WWTP. In *Proceedings of the 1st IWA BeNeLux Regional Young Water Professionals Conference, 30 September–2 October 2009, Eindhoven, The Netherlands*.

Bixio, D., Parmentier, G., Rousseau, D., Verdonck, F., Meirlaen, J., Vanrolleghem, P. A. & Thoeye, C. 2002 A quantitative risk analysis tool for design/simulation of wastewater treatment plants. *Water Science and Technology* **46** (4–5), 301–307.

Donigian Jr, A. S. & Love, J. T. 2007 The Housatonic River watershed model: model application and sensitivity/ uncertainty analysis. In *Proceedings of the 7th International Symposium on Systems Analysis and Integrated Assessment in Water Management (WATERMATEX2007), 7–9 May 2007, Washington, District of Columbia, USA*.

Flores-Alsina, X., Rodríguez-Roda, I., Sin, G. & Gernaey, K. V. 2008 Multi-criteria evaluation of wastewater treatment plant control strategies under uncertainty. *Water Research* **42** (17), 4485–4497.

Gernaey, K. & Jørgensen, S. B. 2004 Benchmarking combined biological phosphorous and nitrogen removal wastewater treatment processes. *Control Engineering Practice* **12**, 357–373.

Iman, R. L. & Helton, J. C. 1985 *A Comparison of Uncertainty and Sensitivity Analysis Techniques for Computer Models*. Report NUREGICR-3904, SAND 84–1461, Sandia National Laboratories, Albuquerque, New Mexico.

Manache, G. & Melching, C. S. 2007 Sensitivity of Latin hypercube sampling to sample size and distributional assumptions, In *Proceedings of the 32nd Congress of the International Association of Hydraulic Engineering and Research, 1–6 July 2007, Venice, Italy*.

Martin, C. 2009 *Integrated Monte Carlo Methodology for Parameter Estimation in Biochemical Models*. PhD Thesis, University of Navarra, Spain.

McCormick, J. F., Johnson, B. & Turner, A. 2007 Analyzing risk in wastewater process design: using Monte Carlo simulation to move beyond conventional design methods. In *Proceedings of the 80th Annual Water Environment Federation Technical Exhibition and Conference, 13–17 October 2007, San Diego, California, USA*.

McKay, M. D., Beckman, R. J. & Conover, W. J. 1979 A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21** (2), 239–245.

Morgan, M. G. & Henrion, M. 1990 *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press, Cambridge, UK.

Rousseau, D., Verdonck, F., Moerman, O., Carrette, R., Thoeye, C., Meirlaen, J. & Vanrolleghem, P. A. 2001 Development of a risk assessment based technique for design/retrofitting of WWTPs. *Water Science and Technology* **43** (7), 287–294.

Saltelli, A., Tarantola, S., Campolongo, F. & Ratto, M. 2004 *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models.* John Wiley and Sons.

Sin, G., Gernaey, K. V., Neumann, M. B., van Loosdrecht, M. C. M. & Gujer, W. 2009 Uncertainty analysis in WWTP model applications: a critical discussion using an example from design. *Water Research* **43** (11), 2894–2906.

Stein, M. L. 1987 Large sample properties of simulations using Latin hypercube sampling. *Technometrics* **29**, 143–151.