

Global sensitivity analysis of the BSM2 dynamic influent disturbance scenario generator

Xavier Flores-Alsina, Krist V. Gernaey and Ulf Jeppsson

ABSTRACT

This paper presents the results of a global sensitivity analysis (GSA) of a phenomenological model that generates dynamic wastewater treatment plant (WWTP) influent disturbance scenarios. This influent model is part of the Benchmark Simulation Model (BSM) family and creates realistic dry/wet weather files describing diurnal, weekend and seasonal variations through the combination of different generic model blocks, i.e. households, industry, rainfall and infiltration. The GSA is carried out by combining Monte Carlo simulations and standardized regression coefficients (SRC). Cluster analysis is then applied, classifying the influence of the model parameters into strong, medium and weak. The results show that the method is able to decompose the variance of the model predictions ($R^2 > 0.9$) satisfactorily, thus identifying the model parameters with strongest impact on several flow rate descriptors calculated at different time resolutions. Catchment size (PE) and the production of wastewater per person equivalent (Q_{perPE}) are two parameters that strongly influence the yearly average dry weather flow rate and its variability. Wet weather conditions are mainly affected by three parameters: (1) the probability of occurrence of a rain event (L_{rain}); (2) the catchment size, incorporated in the model as a parameter representing the conversion from mm rain \cdot day $^{-1}$ to m $^3 \cdot$ day $^{-1}$ (Q_{permm}); and, (3) the quantity of rain falling on permeable areas (aH). The case study also shows that in both dry and wet weather conditions the SRC ranking changes when the time scale of the analysis is modified, thus demonstrating the potential to identify the effect of the model parameters on the fast/medium/slow dynamics of the flow rate. The paper ends with a discussion on the interpretation of GSA results and of the advantages of using synthetic dynamic flow rate data for WWTP influent scenario generation. This section also includes general suggestions on how to use the proposed methodology to any influent generator to adapt the created time series to a modeller's demands.

Key words | activated sludge modelling, benchmarking, BSM, influent modelling, Monte Carlo simulations, standardized regression coefficients (SRC)

Xavier Flores-Alsina (corresponding author)
Ulf Jeppsson
 Division of Industrial Electrical Engineering and Automation (IEA),
 Department of Measurement Technology and Industrial Electrical Engineering (MIE),
 Lund University,
 Box 118, SE-221 00 Lund,
 Sweden
 E-mail: xavier.flores@iea.lth.se

Krist V. Gernaey
 Center for Process Engineering and Technology (PROCESS),
 Department of Chemical and Biochemical Engineering,
 Technical University of Denmark,
 Building 229, DK-2800 Kgs. Lyngby,
 Denmark

NOMENCLATURE

| | | | |
|-------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|------------------------------------------------------------------------------------------------------|
| A_k | Vector of model parameters used in the SRC analysis. The number of model parameters is k | b_k | Slope in the SRC model. The number of slopes corresponds to the number of studied parameters (k) |
| AADF | Average annual daily flow rate [m $^3 \cdot$ d $^{-1}$] | BSM1 | Benchmark Simulation Model no. 1 |
| aH | A parameter determining the direct contribution of rainfall falling on impermeable surfaces in the catchment area to the flow rate in the sewer ('Rain generator' model block) [%] | BSM2 | Benchmark Simulation Model no. 2 |
| ASMs | Activated Sludge Models | CK | Coefficient of kurtosis [–] |
| b_0 | Off-set in the SRC model | CS | Coefficient of skewness [–] |
| | | GSA | Global Sensitivity Analysis |
| | | DISG | Dynamic influent scenario generator |

| | |
|----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| HH | Households model block |
| H_{inv} | Height of the invert level ('Soil' model block) [m] |
| IndS | Industries model block |
| <i>InfAmp</i> | Amplitude of the sine wave for generating seasonal effects due to infiltration ('Seasonal correction factor' model block) [$m^3 \cdot d^{-1}$] |
| <i>InfBias</i> | Mean value of the sine wave signal for generating seasonal effects due to infiltration ('Seasonal correction factor' model block) [$m^3 \cdot d^{-1}$] |
| IWA | International Water Association |
| K | Soil permeability constant ('Soil' model block) [$m^3 \cdot m^{-2} \cdot d^{-1}$] |
| K_{down} | Gain for adjusting the flow rate to downstream aquifers ('Soil' model block) [$m^2 \cdot d^{-1}$] |
| K_{inf} | Infiltration gain ('Soil' model block) [$m^{2.5} \cdot d^{-1}$] |
| <i>Length</i> | A parameter that forms a measure of the size of the catchment area. It will determine the number of variable volume tanks in series that will be used for describing the sewer system ('Sewer' model block) [-] |
| <i>Llrain</i> | A constant converting the random number generator output to a value representing rainfall intensities ('Rain generator' model block) [$mm \text{ rain} \cdot d^{-1}$] |
| MaxD | Average maximum flow sustained for a period of one day [$m^3 \cdot d^{-1}$] |
| MaxH | Average maximum flow sustained for a period of one hour [$m^3 \cdot d^{-1}$] |
| MaxM | Average maximum flow sustained for a period of one month [$m^3 \cdot d^{-1}$] |
| MC | Monte Carlo |
| MinD | Average minimum flow sustained for a period of one day [$m^3 \cdot d^{-1}$] |
| MinH | Average minimum flow sustained for a period of one hour [$m^3 \cdot d^{-1}$] |
| MinM | Average minimum flow sustained for a period of one month [$m^3 \cdot d^{-1}$] |
| <i>PE</i> | Person equivalent ('Households' model block) [-] |
| <i>Qpermm</i> | Flow rate per mm rain ('Rain generator' model block) [$m^3 \cdot mm^{-1}$] |
| <i>QperPE</i> | Wastewater flow rate per person equivalent ('Households' model block) [$m^3 \cdot d^{-1}$] |
| <i>QInd</i> | Average wastewater flow rate from industry on normal week days (Monday to Thursday) ('Industry' model block) [$m^3 \cdot d^{-1}$] |

| | |
|-----------------|---------------------------------------------------------------------|
| RAIN | Rain model block |
| RangeD | Difference between MaxD and MinD [$m^3 \cdot d^{-1}$] |
| RangeH | Difference between MaxH and MinH [$m^3 \cdot d^{-1}$] |
| RangeM | Difference between MaxM and MinM [$m^3 \cdot d^{-1}$] |
| SD | Standard deviation of daily flow rate [$m^3 \cdot d^{-1}$] |
| SEWER | Sewer model block |
| SOIL | Soil model block |
| SCF | Seasonal correction factor |
| SRC | Standardized regression coefficient |
| <i>subareas</i> | Catchment size [-] |
| \hat{X}_j | Regression model prediction. The number of model predictions is j |
| WWTP | Wastewater treatment plant |

SOFTWARE AVAILABILITY

The source code for generation of WWTP influent files can be obtained for free. Contact Krist V. Gernaey, Center for Process Engineering and Technology (PROCESS), Department of Chemical and Biochemical Engineering, Technical University of Denmark, Building 229, DK-2800 Kgs. Lyngby, Denmark. E-mail: kvg@kt.dtu.dk

RESEARCH HIGHLIGHTS

- Monte Carlo simulations and standardized regression coefficients can satisfactorily decompose the variance of the influent flow rate model predictions.
- Cluster analysis allows identifying the set of parameters explaining the main contributions to the total variance.
- The relative importance of the different model parameters changes depending on: (1) dry/wet weather conditions; and (2) the evaluated time scale (hour, day or month).
- GSA is useful to gain deeper knowledge of the model behaviour. This insight can be exploited to: (1) adapt the model to the modeller's needs; and (2) to generate additional WWTP influent scenarios.

INTRODUCTION

Dynamic influent disturbance scenario generators (DIDSG) have recently gained interest in the field of wastewater treatment plant (WWTP) modelling. In essence, synthetic data

can overcome one of the main limitations when performing simulation studies: a sufficiently long set of influent data representing the inherent natural variability of the flow rate and pollutant concentrations at the plant inlet is often not available. If inadequate dynamic influent disturbances are applied to the WWTP in a simulation study, the system will not be sufficiently excited and thus the simulations will result in a too optimistic picture of the plant performance (Ráduly *et al.* 2007).

During recent years, several DIDSG have been developed with multiple applications (see for example Muschalla *et al.* 2006; Langergraber *et al.* 2008; Alex *et al.* 2009). de Keyser *et al.* (2010) developed a model that creates time series of traditional and micro-pollutants from their emission sources in the urban catchment. Similarly, Ort *et al.* (2005) developed a stochastic model describing short-term variations of benzotriazole concentrations (a chemical in dishwasher detergents). Additionally, Rosen *et al.* (2008) used a Markov chain approach to describe the occasional occurrence of either toxic or inhibitory influent shock loads. One successful application of an influent wastewater generator was developed by Gernaey *et al.* (2011), and was used to generate influent data for the different Benchmark Simulation Models (BSM) (Spanjers *et al.* 1998; Copp 2002; Rosen *et al.* 2004; Jeppsson *et al.* 2007; Nopens *et al.* 2010; Gernaey *et al.* 2012), widely used in the wastewater modelling community. The latter influent generator will be investigated in more detail in this paper.

The BSM2 DIDSG is comprised of a set of generic model blocks and takes into account the contributions of households, industries, infiltration and run-off from impermeable surfaces. The model also includes the 'smoothing' effect of the sewer network. Although the BSM2 DIDSG is applied to create the disturbance influent file used to evaluate different control strategies in the BSM2 platform, the tool is general and has a wide range of applications. Benedetti *et al.* (2008) combined a preliminary version of this approach with different rainfall time series to come up with a wide range of influent conditions including variation in weather scenarios (Alpine, Oceanic, Continental, Mediterranean), loading (ratio between households and industry) and holiday activities (tourism) to evaluate the robustness of different plant designs. Another example can be found in Flores-Alsina *et al.* (2012) where the influent model was combined with the ASM-X approach (Plósz *et al.* 2010) to describe occurrence and transport of micro-pollutants, i.e. pharmaceuticals, taking into account: (1) oral administration patterns; (2) body residence time; and (3) discharge patterns. Other applications of the influent

model can be found elsewhere (Gernaey *et al.* 2006; Lindblom *et al.* 2006; Béraud *et al.* 2007).

The software for generating influent data is intended to be flexible. Nevertheless, the full potential of the influent model, the possibility to export the model to describe different systems, i.e. catchments, or to explore additional scenarios cannot be exploited to the full extent possible unless a comprehensive global sensitivity analysis (GSA) is made. This analysis will help to elucidate which inputs are the major causes of variability in the outputs. Also, the study can give an idea about the effort that the modeller has to invest in finding realistic values for a certain parameter if he/she does not have any measurements/expert knowledge available for the design study under investigation.

The purpose of this paper is to present the global sensitivity analysis (GSA) of the model parameters used to create the BSM2 influent flow rate produced by the DIDSG presented in Gernaey *et al.* (2011). As such, this manuscript should also be seen as a first step in adopting the influent model for applications that fall outside the influent model benchmarking applications. The analysis is carried out by combining Monte Carlo (MC) simulations with Standardized Regression Coefficients (SRC), and decomposes the variance of the flow rate predictions under different weather conditions. Finally, for different flow rate descriptors, calculated at different time resolutions, the influence of the model parameters on the generated flow rate data is classified into strong, medium and weak. The manuscript is organized as follows: First, the problem framing for the DIDSG for flow rate generation, the parameter ranges and the different techniques for GSA are described. Next the results for both dry and wet weather conditions are presented and interpreted. The study is complemented by a critical discussion of the results, focusing on the practical implications of the GSA results. Finally, general guidelines about how to use the presented methodology to other influent generators are given.

METHODS

Dynamic WWTP influent disturbance scenario generator

The DIDSG is based on the work presented in Gernaey *et al.* (2011). The proposed phenomenological approach produces dynamic influent flow rate, pollutant concentrations and temperature profiles using different model blocks and it was used during the development of the BSM2 (Nopens

et al. 2010). The generation of the influent flow rate is achieved by combining the contributions of households (HH), industry (IndS), rainfall (Rain) and infiltration (SCF) (see Figure 1). Rainfall (RAIN) contributes to the total flow rate in two ways: the largest fraction ($aH/100$) originates from the run-off of impermeable surfaces, and is thus transported directly to the sewer. Rainfall on permeable surfaces, a fraction $(100-aH)/100$, will influence the groundwater level, and thus the contribution of infiltration to the influent flow rate. An additional snow melting module could be included as well, and can have important implications in certain geographical regions, but for simplicity it is not considered in this case study. Assuming a cold and a warm season, the seasonal correction factor modifies the amount of infiltration, which is attributed to changes in the groundwater level over the year, i.e. different evapotranspiration regimes. The seasonal correction factor is combined with the rainfall falling on permeable surfaces, and the sum of both flows is passed through the soil model. The soil model (SOIL) is described as a variable volume tank model and it is used to represent the assumed storage of water in the soil. Afterwards, the net contribution of infiltration is combined with the overall flow rate resulting from households and industry and the flow contribution from rainfall on impermeable surfaces. The conceptual principle of the sewer model block (SEWER) is based on a number of variable volume tank models in series, which modifies the dynamics of the generated time series. Due to the model simplicity neither CSOs nor backflow effects are included during liquid transport. The list of model parameters ruling each model block (HH, IndS, RAIN, SCF, SOIL and SEWER) is specified in Table 1.

Figure 2 (left) shows a dynamic profile of the dry weather flow rate generated with the influent model using the default set of parameters, i.e. the ones used to create the BSM2 influent file (Gernaey *et al.* 2011). The flow rate time series presents daily, weekly and seasonal variations (e.g. holiday period and closing of industries). The starting

date of the flow rate time series is July 1st. In addition, a slight increase of the flow rate during winter due to the effect of infiltration is visible. During the cold season, it is indeed assumed that the groundwater level is high resulting in increased infiltration to the sewer system. Figure 2 (right) shows the dynamic profile of the wet weather flow rate generated with the influent model using the default set of parameters. Besides the above mentioned daily, weekly, yearly and seasonal phenomena there are sudden increases of the flow rate due to rain events.

Problem statement

The problem framing for the uncertainty and sensitivity analysis follows the approach presented by Sin *et al.* (2011). Given the catchment characteristics, such as ratio households/industry, the quantity of infiltration, the seasonal variation, the rainfall periodicity, the length of the sewer, etc., the main question to be answered is: what are the most significant model inputs that contribute to the variability of the different flow rate descriptors? Default parameters and their assumed variability ranges are summarized in Table 1. These ranges of variability are defined based on a combination of expert knowledge and process understanding. All the probability distributions are assumed to be uniform, mainly due to lack of information on the real distributions.

MC simulation and SRCs

The MC simulation methodology is based on four steps: (1) specification of the input ranges, i.e. model parameters (Table 1); (2) sampling from the input ranges using the Latin Hypercube Sampling method (Iman *et al.* 1981); (3) propagation of the sampled values through the DIDSG (Gernaey *et al.* 2011); and finally (4) analysis of the results to obtain distributions of the values for the outputs, i.e. flow rate descriptors (Tchobanoglous *et al.* 2003).

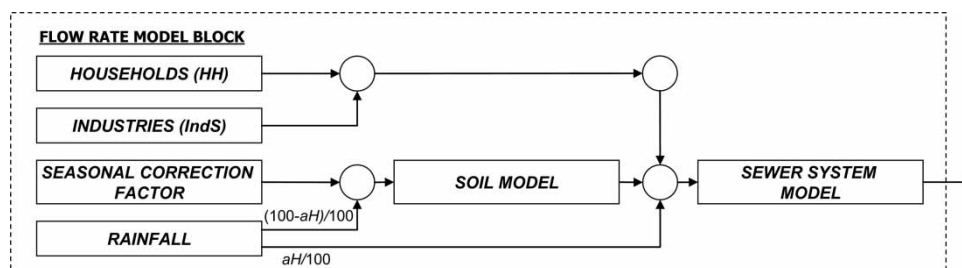
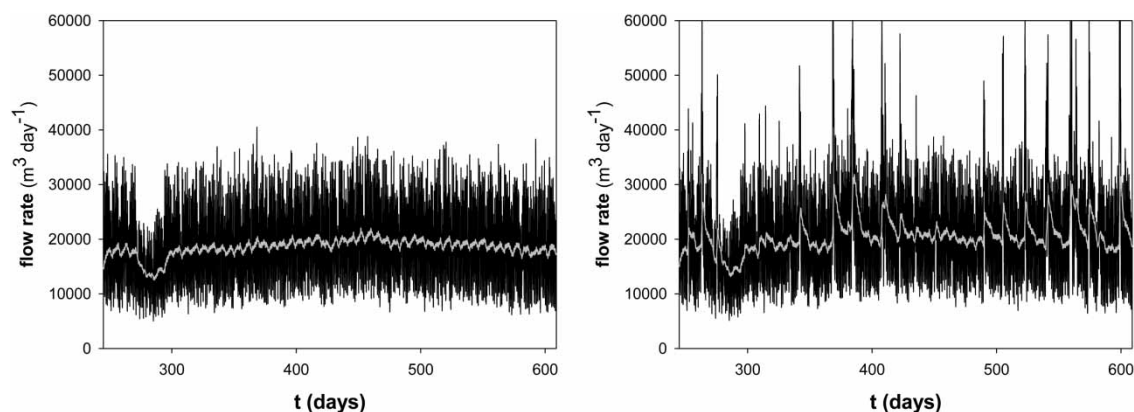


Figure 1 | Schematic representation of the WWTP influent flow rate generator modelling approach (Gernaey *et al.* 2011).

Table 1 | Model (default) parameters and input ranges

| Parameter | Model compartment (bold) / parameter description (normal font) | Default value | Range (%) |
|-------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------|---------------|-----------|
| 'Households (HH)' model block | | | |
| <i>QperPE</i> | Wastewater flow rate per person equivalent [$\text{m}^3 \cdot \text{d}^{-1}$] | 150 | ± 10 |
| <i>PE</i> | Person equivalent [-] | 80,000 | ± 10 |
| 'Industry (IndS)' model block | | | |
| <i>QInd</i> | Average wastewater flow rate from industry on normal week days [$\text{m}^3 \cdot \text{d}^{-1}$] | 2,500 | ± 10 |
| 'Seasonal correction factor (SCF)' model block | | | |
| <i>InfAmp</i> | Amplitude of the sine wave for generating seasonal effects due to infiltration [$\text{m}^3 \cdot \text{d}^{-1}$] | 1,200 | ± 10 |
| <i>InfBias</i> | Mean value of the sine wave signal for generating seasonal effects due to infiltration [$\text{m}^3 \cdot \text{d}^{-1}$] | 7,100 | ± 10 |
| 'Rain generator (RAIN)' model block | | | |
| <i>LLrain</i> | A constant converting the random number generator output to a value representing rainfall intensities [$\text{mm rain} \cdot \text{d}^{-1}$] | 3.5 | ± 25 |
| <i>Qpermm</i> | Flow volume per mm rain [$\text{m}^3 \cdot \text{mm}^{-1}$] | 1,500 | ± 25 |
| <i>aH</i> | Direct contribution of rainfall falling on impermeable surfaces in the catchment area to the flow rate in the sewer [%] | 75 | ± 25 |
| 'Soil (SOIL)' model block | | | |
| <i>H_{inv}</i> | Height of the invert level [m] | 2 | ± 50 |
| <i>Subareas</i> | Number of subareas (discrete), measure of the size of the catchment area [-] | 4 | ± 50 |
| <i>K</i> | Soil permeability constant [$\text{m}^3 \cdot \text{m}^{-2} \cdot \text{d}^{-1}$] | 1 | ± 50 |
| <i>K_{down}</i> | Gain to adjust the flow rate to downstream aquifers [$\text{m}^2 \cdot \text{d}^{-1}$] | 1,000 | ± 50 |
| 'Sewer system (SEWER)' model block | | | |
| <i>Length</i> | Number of sewer sections (discrete), measure of the length of the sewer system [-] | 4 | ± 50 |

**Figure 2** | Dry (left) and wet (right) weather profiles generated with the influent model based on default parameters. An exponential smoothing filter (time constant = 1 day) is used to improve visualization of the data (in white). The starting date of both flow rate time series is July 1st.

The SRC method involves performing a linear regression on the output of the MC simulation revealing the relationship between the model parameters and the flow rate characteristics (Saltelli *et al.* 2004) as follows:

$$\hat{X}_j = b_0 + \sum_{k=1}^n b_k A_k \quad (1)$$

where \hat{X}_j is the regression model prediction for a flow rate calculation j , b_0 is the offset, b_k are the slopes and n is the total number of model parameters A_k . The standardized regression coefficients (SRCs) are then obtained by scaling the slopes b_k using the standard deviations of the parameters (σ_{A_k}) and the outputs (σ_{X_j}) as

Table 2 | Summary statistics of the MC propagation

| Item conditions | Mean | | Coefficient of variation | | 5% percentile | | 95% percentile | |
|----------------------------------|----------|----------|--------------------------|------|---------------|----------|----------------|----------|
| | DRY | WET | DRY | WET | DRY | WET | DRY | WET |
| Average annual daily flow (AADF) | 1,8569.9 | 2,0607.5 | 7.3 | 8.2 | 1,6316.4 | 18,009.5 | 20,879.1 | 23,529.4 |
| Standard deviation (SD) | 6,277.8 | 8,300.1 | 10.6 | 15.8 | 5,233.1 | 6,421.4 | 7,443.5 | 1,0768.9 |
| Coefficient of skewness (CS) | 0.4 | 1.2 | 32.5 | 31.7 | 0.2 | 0.6 | 0.5 | 1.9 |
| Coefficient of kurtosis (CK) | 2.5 | 5.9 | 6.2 | 27.0 | 2.3 | 3.6 | 2.8 | 8.8 |
| Maximum hour (MaxH) | 29,498.8 | 35,441.1 | 8.1 | 12.9 | 25,759.5 | 28,918.4 | 33,648.8 | 44,336.3 |
| Maximum day (MaxD) | 21,276.8 | 30,999.5 | 6.7 | 17.5 | 18,909.5 | 23,454.2 | 23,719.9 | 40,884.6 |
| Maximum month (MaxM) | 20,017.8 | 23,211.8 | 6.9 | 9.7 | 17,783.8 | 19,976.2 | 22,383.8 | 27,218.4 |
| Minimum hour (MinH) | 9,410.7 | 9,983.7 | 10.2 | 10.1 | 7,804.8 | 8,307.8 | 10,937.2 | 11,615.3 |
| Minimum day (MinD) | 15,018.2 | 15,560.0 | 8.5 | 8.3 | 12,892.8 | 13,439.4 | 17,152.3 | 17,718.8 |
| Minimum month (MinM) | 17,647.5 | 18,184.3 | 7.5 | 7.5 | 15,411.7 | 15,943.4 | 19,903.2 | 20,462.2 |
| MaxH-MinH (RangeH) | 20,088.1 | 25,457.4 | 10.4 | 16.9 | 16,750.6 | 19,501.7 | 23,610.4 | 33,885.3 |
| MaxD-MinD (RangeD) | 6,258.6 | 15,439.4 | 4.3 | 33.0 | 5,841.4 | 8,633.0 | 6,703.4 | 24,857.4 |
| MaxM-MinM (RangeM) | 2,370.3 | 5,027.5 | 4.8 | 28.8 | 2,189.0 | 3,209.2 | 2,552.9 | 7,790.4 |

stated in Equation (2).

$$SRC = b_k \frac{\sigma_{A_k}}{\sigma_{X_j}} \quad (2)$$

According to Saltelli *et al.* (2004) the SRC is a valid measure of sensitivity if the coefficient of determination $R^2 > 0.7$. When the model is linear, the $\sum_k (SRC)^2 = 1$, but the normal non-linear case is $\sum_k (SRC)^2 \leq 1$ and equal to R^2 (Saltelli *et al.* 2008). The higher the absolute values of the SRC, the stronger the influence of the corresponding input A_k on determining the output \hat{X}_j .

The results of the absolute values of the SRCs are then classified in three groups using *k*-means clustering (Hair *et al.* 1998), and as a result parameters are grouped into three categories which correspond to a strong, medium and weak influence on the output, respectively. The cluster with the highest and lowest (absolute) SRC values are labelled as strong and weak, respectively, while the intermediate values are categorized as medium. Direct or indirect correlations are specified using positive (+) and negative (−) signs of the regression coefficients. Coefficients close to zero mean that the output is not sensitive to a certain input.

RESULTS

Monte Carlo simulation results

In Table 2, the respective distributions for all flow rate descriptors are summarized by their mean, coefficient of variation, 5% and 95% percentile values. The mean values of the average annual daily flow (AADF), the standard deviation (SD), the coefficient of skewness (CS), the coefficient of kurtosis (CK) and the hourly, daily and monthly maxima, minima and ranges are higher in wet weather conditions. However, comparatively, the relative differences between (dry/wet) maximum average values are more extreme compared with the (dry/wet) minimum values for the different statistics summarized in Table 2. For example, the dry/wet weather differences between the maximum flow rate values are 20% (MaxH), 46% (MaxD) and 3% (MaxM) when considering a time scale of hours, days and months, respectively. On the other hand, differences between the minimum values are 6% (MinH), 4% (MinD) and 3% (MinM) for a time scale of hours, days and months, respectively. This is mainly due to: (1) the buffering effect of the soil model in the influent generator; and (2) the possibility to divert rain water directly (via run-off) into the sewer system (see Figure 1). The 5% and 95% percentiles can be interpreted in a probabilistic way, e.g. the 95% percentile

of MaxH for the wet weather scenario means there is a probability of 95% that the average (hourly) flow rate is below $44336.3 \text{ m}^3 \cdot \text{d}^{-1}$. Finally, the differences between Max and Min (ranges) decrease when the length of the time scale increases (for the different statistics). For example, the (mean) range of dry and wet weather flow rate values is decreased when the scale is changed from hours (RangeH) to months (RangeM). For example, in dry conditions RangeH is $20088 \text{ m}^3 \cdot \text{d}^{-1}$ and RangeM is only $2370 \text{ m}^3 \cdot \text{d}^{-1}$.

GSA of the WWTP influent generator during dry weather conditions

The parameters with the strongest influence on the dry weather flow rate are summarized in Table 3. These parameters are the result of clustering the absolute SRC values for the different flow rate outputs. It is important as well to highlight that R^2 values were above 0.9 in all the cases reported. Average annual daily (AADF), maximum hourly (MaxH), daily (MaxD) and monthly (MaxM) flow rate values are strongly (positively) influenced by the HH model block parameters. These parameters represent the flow rate generated per person equivalent (*QperPE*) and the number of person equivalents in the catchment area (*PE*). The length of the sewer system (*Length*) has a considerable effect on the standard deviation (SD), skewness (CS), kurtosis (CK) and MaxH/MinH/Range hourly values. When the value of the parameter *Length* is higher, the sewer system is assumed to increase in size and there is consequently a larger smoothing effect on the flow rate

values. In the Appendix section (Figure A1, available online at <http://www.iwaponline.com/wst/065/089.pdf>) additional information can be found regarding the influence that: (1) *PE* & *QperPE*; and (2) the parameter *Length* may have on one year of simulated dry weather flow rate data. In addition, the SRC and R^2 values for the AADF and the SD can also be found there (Table A1, available online at <http://www.iwaponline.com/wst/065/089.pdf>).

The Seasonal Correction Factor (SCF) and the soil (SOIL) model parameters mainly influence the quantities of water: (1) originating from upstream aquifers; (2) undergoing evapo-transpiration; (3) accumulated in soil; (4) passing to the sewer via infiltration; and (5) diverted to downstream aquifers. The amplitude of the sine wave (*InfAmp*) basically modifies process (2) and has a strong effect on RangeM because it increases/decreases the differences between winter and summer periods, corresponding to different evapo-transpiration regimes for the cold and the warm season. The height of the invert level in soil (H_{inv}) and the gain for adjusting the flow rate to the downstream aquifers (K_{down}) influence processes (3), (4) and (5) and have a strong impact on the quantity of water leaving the soil model block (see Figure 1). As a result, there is a dramatic reduction of the infiltration flow to the sewer system and as a consequence a decrease of minimum hourly (MinH), daily (MinD) and monthly (MinM) flow rate values. The rest of the SCF/SOIL model parameters show poor sensitivity. The limited amounts of industrial wastewater compared with the HH contribution means that the effect of the IndS related model parameters on the flow rate are almost unnoticeable.

Table 3 | GSA results (strong parameters, group 1 in *k*-means clustering) for dry and wet weather conditions. Negative (–, indirect correlation) and positive (+, direct correlation) signs represent the correlation of the model parameters with the different evaluation criteria. In all cases $R^2 > 0.9$

| Item | Dry weather | Wet weather |
|----------------------------------|-------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------|
| Average annual daily flow (AADF) | <i>PE</i> (+), <i>QperPE</i> (+), K_{down} (–), H_{inv} (–) | <i>Llrain</i> (–), <i>PE</i> (+), <i>QperPE</i> (+), K_{down} (–) |
| Standard deviation (SD) | <i>Length</i> (–), <i>PE</i> (+), <i>QperPE</i> (+) | <i>Llrain</i> (–), <i>Qpermm</i> (+), <i>aH</i> (+) |
| Coefficient of skewness (CS) | <i>Length</i> (–) | <i>Llrain</i> (–), <i>Qpermm</i> (+), <i>aH</i> (+) |
| Coefficient of kurtosis (CK) | <i>Length</i> (–) | <i>Llrain</i> (–), <i>Qpermm</i> (+), <i>aH</i> (+) |
| Maximum hour (MaxH) | <i>PE</i> (+), <i>QperPE</i> (+), <i>Length</i> (–) | <i>Llrain</i> (–) |
| Maximum day (MaxD) | <i>PE</i> (+), <i>QperPE</i> (+), K_{down} (–) | <i>Llrain</i> (–) |
| Maximum month (MaxM) | <i>PE</i> (+), <i>QperPE</i> (+), K_{down} (–) | <i>Llrain</i> (–), <i>PE</i> (+), <i>QperPE</i> (+), K_{down} (–), H_{inv} (–) |
| Minimum hour (MinH) | K_{down} (–), H_{inv} (–), <i>Length</i> (+) | H_{inv} (–), K_{down} (–), <i>Length</i> (+) |
| Minimum day (MinD) | K_{down} (–), H_{inv} (–), <i>PE</i> (+), <i>QperPE</i> (+) | H_{inv} (–), K_{down} (–), <i>PE</i> (+), <i>QperPE</i> (+) |
| Minimum month (MinM) | K_{down} (–), H_{inv} (–), <i>PE</i> (+), <i>QperPE</i> (+) | H_{inv} (–), K_{down} (–), <i>PE</i> (+), <i>QperPE</i> (+) |
| MaxH-MaxH (RangeH) | <i>Length</i> (–), <i>PE</i> (+), <i>QperPE</i> (+) | <i>Llrain</i> (–), <i>Length</i> (–) |
| MaxD-MaxD (RangeD) | <i>PE</i> (+), <i>QperPE</i> (+) | <i>Llrain</i> (–) |
| MaxM-MaxM (RangeM) | <i>InfAmp</i> (+) | <i>Llrain</i> (–) |

GSA of the WWTP influent generator during wet weather conditions

RAIN related model parameters (Q_{permm} , L_{rain} and aH) have a strong impact on the total flow rate quantity and variability (see Table 3). On the one hand, the parameter ruling the occurrence of rain events (L_{rain}) directly influences the probability of having wet weather episodes (low L_{rain} values results in increased occurrence of rain events). On the other hand, the parameter converting mm rain to flow rate (Q_{permm}) and the % impermeable area (aH) strongly influence: (1) the intensity of these rain events; (2) the quantity of water entering the soil; and (3) the quantity of water that is diverted directly to the sewer (see Figure 1).

As can be expected when the number of rain events is increased (parameter L_{rain} decreases), AADF, SD and hourly (RangeH), daily (RangeD) and monthly (RangeM) range values are higher. Q_{permm} and aH are catchment dependent and have a major influence on the shape of the resulting flow rate distribution, i.e. increasing its asymmetry (CS) and peak height (CK). This situation is mainly due to the fact that most of the flow rate values are moved to the right-hand side of the distribution when the values of those parameters are increased. It is important to highlight that RAIN related parameters have a strong influence on flow rate descriptors that are sustained for a short period of time, for example MaxH and MaxD (peak values). For longer time horizons, i.e. MaxM and MinM, the parameters with the strongest impact are the same for both wet and dry weather conditions, i.e. HH and SOIL. Finally, minimum values, i.e. MinH, MinD and MinM, are strongly influenced by the soil model parameters, similar to dry weather conditions. Again, it is possible to see the buffer effect of the soil model. Unless there is either a dramatic decrease in the quantity of water accumulated by the soil (parameter H_{inv}) or an increase in the quantity of water going to downstream aquifers, minimum values are more or less constant for both dry and wet weather conditions (see Table 3). In the Appendix section (Figure A2, available online at <http://www.iwaponline.com/wst/065/089.pdf>) additional information can be found about the influence that (1) L_{rain} and (2) Q_{permm} and aH may have on one year of simulated wet weather flow rate data. In addition, the SRC and R^2 values for the AADF and the SD are provided (Table A1).

DISCUSSION

The presented results necessitate a thorough discussion. First of all, it is important to highlight that the results

presented herein will to a large extent depend on the framing of the problem (selection of model parameters, definition of input ranges, sampling methodologies) (Sin et al. 2011). The results of the GSA presented in this paper are specific for this case study and they should be interpreted within that context, i.e. the analysis should be repeated if the problem framing is set up in a different way (Helton & Davis 2003).

Secondly, the GSA provides a better picture about how the DIDSG behaves, by determining the strength of the relation between the input ranges (model parameters) and the different outputs (flow rate descriptors in this case). For example, Figure 3 shows the effect (for dry weather flow rate) of some model parameters on the flow rate descriptors. SCF parameters can increase the monthly differences between summer and winter times (see Figure 3 top, see Table 2 RangeM). The parameters PE and Q_{perPE} increase AADF, MaxH, MaxD and MaxM (see Figure 3 middle, see Table 2). Finally, a stronger smoothing effect can be obtained if the length of the sewer network is increased (Figure 3 bottom, Table 2 MaxH). In wet weather conditions, the periodicity of rain events is mainly determined by the parameter L_{rain} (see Figure 4). Figure 4 shows the generated time series in dry (rain generator is turned off) and wet (default) weather conditions. In addition, when the parameter L_{rain} is decreased, the number of rain events is increased and so are the peak flow rate values. In the case of using pluviometric data as input – an option that is available for the user of the influent model – the parameter L_{rain} is no longer used and the adjustment should be carried out by means of modifying the parameters aH and Q_{permm} . As in the dry weather case, hourly peaks in the wet weather scenario can be influenced by adjusting the parameter $Length$.

The result of the GSA will allow users of the influent model to adapt the generated time series to each modeller's demands. We are indeed convinced that the influent generator developed for the BSM2 can be useful outside the original benchmarking applications as well. An interesting example is shown in Figure 3, where a stepwise procedure to create the different dry weather time series is demonstrated. PE , Q_{perPE} , K_{down} , H_{inv} can be used to adjust the AADF to the user desired conditions. Additional features, such as influent variability (SD, CS and CK) or seasonal – i.e. summer/winter – variation, could be adjusted by modifying the parameters $Length$ and $InfBias$, respectively. Finally, for wet weather influent profile generation, the occurrence of rain events (AADF) and their intensity (MaxH, MaxD, MaxM) can be adjusted with L_{rain} . In addition, alternative scenarios can be generated following the same catchment

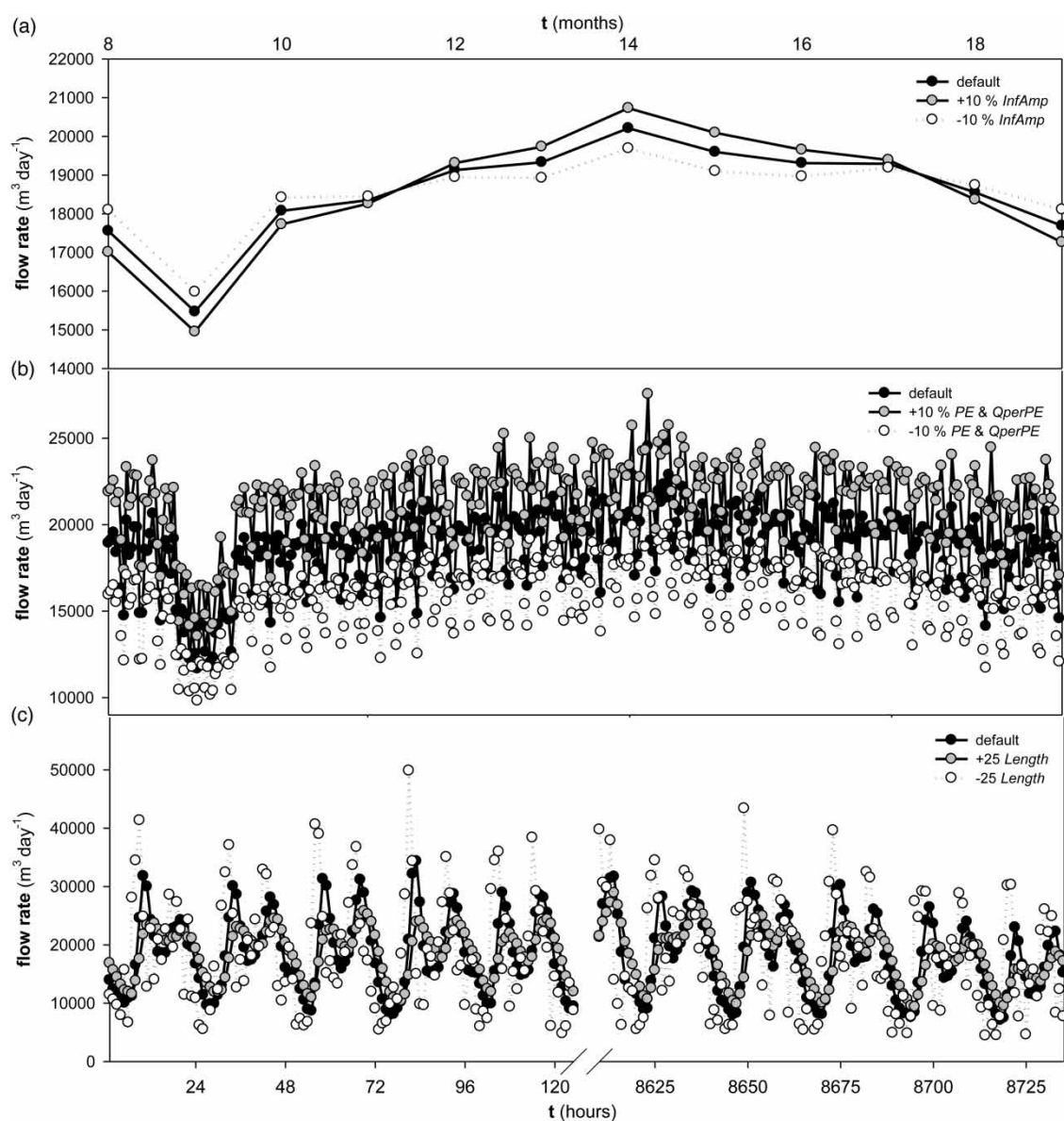


Figure 3 | Effects of the parameter SCF (a), the HH model block (b), and the SEWER model block (c) on the dynamic dry flow rate at different temporal resolutions (frequency of calculations through the time series): monthly (a), daily (b) and hourly (c). Top x-axis (months) refers to plots (a) and (b) while bottom x-axis (hours) refers to plot (c). The starting date of the time series is July 1st.

principles: What would be the effect of changing the rain regime or infiltration dynamics (due to for example climate change) in the wastewater influent profile? What is the effect of a population increase (changes in the number of population equivalents) or higher industrial activity on the influent wastewater flow rate? All these scenarios can in principle be investigated with the influent model, on condition that the model user has a proper understanding of the meaning and the importance of each parameter in the influent model. This paper provides a first clue towards obtaining this understanding.

Finally, it should be emphasized that even though the authors run the analysis using the influent generator developed by BSM Task Group (Gernaey *et al.* 2011) the methodology presented herein is general. Other influent generators could be analysed as well (e.g. Muschalla *et al.* 2006; Langergraber *et al.* 2008; Alex *et al.* 2009; de Keyser *et al.* 2010), obtaining valuable information about the parameters with the strongest influence related to short-term and long-term dynamics. In the case of using a different influent generator, the same flow rate descriptors could be calculated, e.g. AADF, MaxM, MinM ..., identifying the

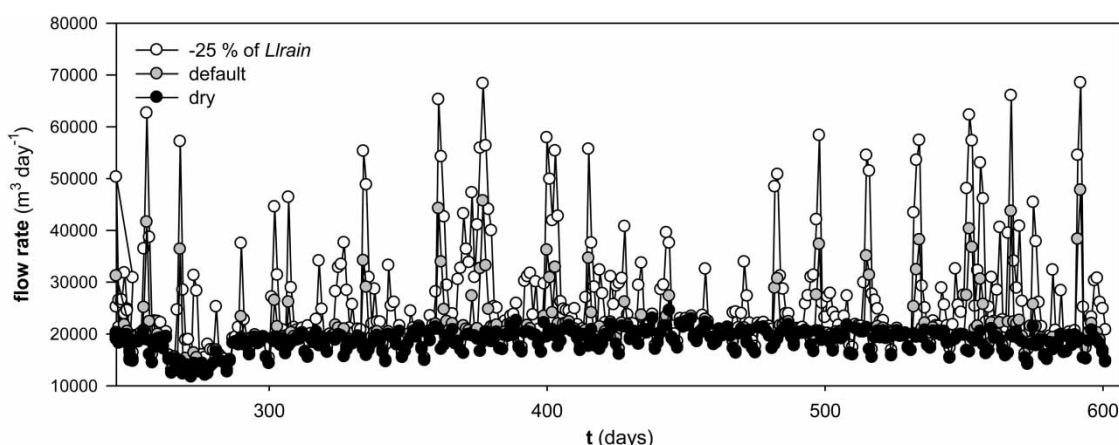


Figure 4 | Effect of the parameter L_{rain} on the dry and wet weather influent flow rate profiles. The starting date of the time series is July 1st.

responsible parameters for the time series dynamics by means of combining MC simulations and SRC.

CONCLUSIONS

This paper presented the results of performing a Global Sensitivity Analysis on the output of a DIDSG. The high R^2 values showed that variance decomposition for a range of different flow rates was possible. The key findings can be summarized as follows:

In dry weather conditions:

- (1) HH parameters (Q_{perPE} , PE) have the strongest influence on dry weather influent flow rate quantity and variability. The sewer parameter $Length$ has a direct influence on the peaks;
- (2) SCF ($InfBias$) and SOIL related (H_{inv} , K_{down}) parameters influence the quantity of a) infiltration and b) water accumulation in soil, decreasing (to a certain extent) average, ranges and minimum flow rate values.

In wet weather conditions:

- (1) RAIN related parameters (L_{rain} , Q_{permm} and aH) have a strong influence on wet weather influent flow rate quantity and variability;
- (2) RAIN related parameters (L_{rain}) clearly affect short-term evaluation criteria (hour, day), but in the long run (month) the HH related parameters have a stronger influence;
- (3) SOIL related parameters (H_{inv} , K_{down}) influence (as for dry weather conditions) minimum values, no matter the time scale.

The GSA has been extended with: (i) a critical discussion about how to interpret the results of the GSA;

- (ii) some hints on how to use the influent model to adapt the generated time series to each modeller's demands;
- (iii) a section focused on the advantages of using synthetic flow rate data for WWTP influent scenario generation; and
- (iv) a perspective on how the methodology can be used with other influent generators.

REFERENCES

- Alex, J., Hetschel, M. & Ogurek, M. 2009 [Simulation study with minimized additional data requirements to analyze control and operation of WWTP Dorsten, Germany](#). *Wat. Sci. Tech.* **60** (6), 1371–1377.
- 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 systems that considers uncertainty](#). *Environ. Modell. Softw.* **23** (8), 1082–1091.
- Béraud, B., Steyer, J. P., Lemoine, C., Gernaey, K. V. & Latrille, E. 2007 Model-based generation of continuous influent data from daily mean measurements available at industrial scale. In: *Proceeding of the 3rd International IWA Conference on Automation in Water Quality Monitoring (AutMoNet2007)*, 5–7 September, Gent, Belgium.
- Copp, J. B. (ed.) 2002 *The COST Simulation Benchmark – Description and Simulator Manual*. Office for Official Publications of the European Communities, Luxembourg.
- Flores-Alsina, X., Plósz, G. B., Lindblom, E., Gernaey, K. V. & Jeppsson, U. 2012 Extending the BSM platform with occurrence, transport and fate of micro-pollutants using the ASM-X framework. *IWA World Water Congress 2012*, Busan, Korea (submitted).
- Gernaey, K. V., Rosen, C. & Jeppsson, U. 2006 [WWTP dynamic disturbance modelling – an essential module for long-term benchmarking development](#). *Wat. Sci. Tech.* **53** (4–5), 225–234.

- Gernaey, K. V., Flores-Alsina, X., Rosen, C., Benedetti, L. & Jeppsson, U. 2011 [Phenomenological modelling approach to generate WWTP disturbance scenarios](#). *Environ. Modell. Softw.* **26** (11), 1255–1267.
- Gernaey, K. V., Jeppsson, U., Vanrolleghem, P. A., Copp, J. B. & Steyer, J.-P. (eds) 2012 *Benchmarking of Control Strategies for Wastewater Treatment Plants*. IWA Scientific and Technical Report, IWA Publishing, London, UK (to appear).
- Hair, J. F., Anderson, R. E., Tatham, R. L. & Black, W. C. 1998 *Multivariate Data Analysis*, 5th edition. Prentice Hall, London, UK.
- Helton, J. C. & Davis, F. J. 2003 [Latin hypercube sampling method and the propagation of uncertainty in analyses of complex systems](#). *Reliab. Engng. Syst. Saf.* **81**, 23–60.
- Iman, R. L., Helton, J. C. & Campbell, J. E. 1981 An approach to sensitivity analysis of computer models, Part 1. Introduction, input variable selection and preliminary variable assessment. *J. Qual. Technol.* **13** (3), 174–185.
- Jeppsson, U., Pons, M.-N., Nopens, I., Alex, J., Copp, J., Gernaey, K. V., Rosen, C., Steyer, J.-P. & Vanrolleghem, P. A. 2007 [Benchmark Simulation Model No 2 – General protocol and exploratory case studies](#). *Wat. Sci. Tech.* **56** (8), 67–78.
- Keyser, W. de, Gevaert, V., Verdonck, F., De Baets, B. & Benedetti, L. 2010 [An emission time series generator for pollutant release modelling in urban areas](#). *Environ. Modell. Softw.* **25** (4), 554–561.
- Langergraber, G., Alex, J., Weissenbacher, N., Woener, D., Ahnert, M., Frehman, T., Halft, N., Hobus, I., Plattes, M., Spering, V. & Winkler, S. 2008 [Generation of diurnal variation for influent data for dynamic simulation](#). *Wat. Sci. Tech.* **50** (9), 1483–1486.
- Lindblom, E., Gernaey, K. V., Henze, M. & Mikkelsen, P. S. 2006 [Integrated modeling of two xenobiotic organic compounds](#). *Wat. Sci. Tech.* **54** (6–7), 213–221.
- Muschalla, D., Ostrowski, M. W., Pereira, C. & Schneider, S. 2006 *Dokumentation des Schmutzfrachtmodells SMUSI Version 5.0*. Darmstadt, Institut für Wasserbau und Wasserwirtschaft, Technische Universität Darmstadt, Germany.
- Nopens, I., Benedetti, L., Jeppsson, U., Pons, M.-N., Alex, J., Copp, J. B., Gernaey, K. V., Rosen, C., Steyer, J.-P. & Vanrolleghem, P. A. 2010 [Benchmark simulation model No 2: Finalization of plant layout and default control strategy](#). *Wat. Sci. Tech.* **62** (9), 1967–1974.
- Ort, C., Schaffner, C., Giger, W. & Gujer, W. 2005 Modeling stochastic load variations in sewer systems. *Wat. Sci. Tech.* **52** (5), 113–122.
- Plósz, B. G., Leknes, H. & Thomas, K. V. 2010 [Impacts of competitive inhibition, parent compound formation and partitioning behaviour on antibiotic micro-pollutants removal in activated sludge: Measurements and modelling](#). *Environ. Sci. Technol.* **44** (2), 734–742.
- Ráduly, B., Gernaey, K. V., Capodaglio, A. G., Mikkelsen, P. S. & Henze, M. 2007 [Artificial neural networks for rapid WWTP performance evaluation: Methodology and case study](#). *Environ. Modell. Softw.* **22** (8), 1208–1216.
- Rosen, C., Jeppsson, U. & Vanrolleghem, P. A. 2004 Towards a common benchmark for long-term process control and monitoring performance evaluation. *Wat. Sci. Tech.* **50** (11), 41–49.
- Rosen, C., Jeppsson, U., Rieger, L. & Vanrolleghem, P. A. 2008 [Adding realism to simulated sensors and actuators](#). *Wat. Sci. Tech.* **57** (3), 337–344.
- Saltelli, A., Tarantola, S., Campolongo, F. & Ratto, M. 2004. *Sensitivity Analysis in Practice. A Guide to Assessing Scientific Models*. John Wiley & Sons, Chichester, UK.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. & Tarantola, S. 2008 *Global Sensitivity Analysis: The Primer*. John Wiley & Sons, West Sussex, UK.
- Sin, G., Gernaey, K. V., Neumann, M. B., van Loosdrecht, M. C. M. & Gujer, W. 2011 [Global sensitivity analysis in wastewater treatment plant model applications: prioritizing sources of uncertainty](#). *Water Res.* **45** (2), 639–651.
- Spanjers, H., Vanrolleghem, P. A., Nguyen, K., Vanhooren, H. & Patry, G. G. 1998 [Towards a simulation-benchmark for evaluating respirometry-based control strategies](#). *Wat. Sci. Tech.* **37** (12), 219–226.
- Tchobanoglous, G., Burton, F. L. & Stensel, H. D. 2003 *Wastewater Engineering: Treatment, Disposal and Reuse*. McGraw-Hill, New York, USA.