Influent generator for probabilistic modeling of nutrient removal wastewater treatment plants

Mansour Talebizadeh[1], Evangelia Belia[2], and Peter A. Vanrolleghem[1]

[1]model*EAU*, Département de génie civil et de génie des eaux, Université Laval, 1065 av. de la Médecine, Québec (QC) G1V 0A6, Canada (Email : *mansour.talebizadehsardari.1@ulaval.ca; Peter.Vanrolleghem@gci.ulaval.ca*)
[2] Primodal Inc., 145 Aberdeen, Québec, QC G1R 2C9, Canada (Email: *belia@primodal.com*)

Abstract
The availability of influent wastewater time series is crucial when using models to assess the performance of a wastewater treatment plant (WWTP) under dynamic flow and loading conditions. Given the difficulty of collecting sufficient data, synthetic generation could be the only option. In this paper a hybrid of statistical and conceptual modeling techniques is proposed for synthetic generation of influent time series. The time series of rainfall and influent in DWF conditions were generated using two types of statistical models (a periodic-multivariate time series model for influent in DWF conditions and a two-state Markov chain-gamma model for rainfall). These two time series serve as inputs to a conceptual sewer model for generation of influent time series during WWF conditions. The effect of total model uncertainty on the generated outputs is taken into account through a Bayesian calibration and is communicated to the user by constructing uncertainty bands with a desired level of confidence. The proposed influent generator is a powerful tool for realistic generation of the influent time series and is well-suited for risk-based design of WWTPs as it considers both the effect of input variability (i.e. variability in rainfall and influent during DWF) and total model uncertainty in the generation of the influent. Considering the fact that the proposed influent generator only requires readily-available or easy-to-obtain information and data on climate and the general characteristics of sewershed, it is an attractive tool for practical applications.

Keywords
Bayesian estimation; uncertainty analysis; urban hydrology; wastewater composition; probabilistic design

INTRODUCTION

One of the major sources of uncertainty/variability that both plant designers and operators must deal with is the dynamics of the influent. The recent advances in mathematical modeling and improved computational power have enabled researchers to better understand the performance of different WWTP design alternatives (Hao *et al*., 2001; Salem *et al.*, 2002) and/or evaluate control strategies under dynamic flow and loading conditions. However, the application of mathematical models used for simulating the performance of a WWTP could be misleading unless, among others, models are provided with representative influent time series. One of the problems that arise in this regard is the scarcity or even lack of long-term influent data. To remedy this problem, some researchers have proposed models for synthetic dynamic influent time series scenarios (Bechmann *et al.*, 1999; Gernaey *et al.,* 2011).

One of the simplest approaches in synthetic generation of influent time series is the application of empirical stochastic models (Capodaglio *et al.,* 1990; Martin *et al.*, 2007). However, these models may have a poor performance especially during wet weather flow conditions as different complex processes affect the dynamics of the influent. Indeed, such statistical models do not consider the underlying elements and processes that govern the generation and the dynamics of the influent. To

consider the underlying phenomena that are involved, some researchers have advocated the use of detailed and physically-based models (Hernebring *et al.,* 2002; Temprano *et al.,* 2007). The application of these complex models might be useful for certain purposes, (e.g. evaluating the performance of different operating strategies in a sewer system). However, in cases in which the overall behavior of the influent time series is of interest, they might not be very useful as they require detailed information on the sewage system and running them for a large number of times could be computationally expensive.
Some researchers have proposed parsimonious conceptual models as an alternative to the complex mathematical models that require detailed information and data (Achleitner *et al.,* 2007; Gernaey *et al.,* 2011). In these models a conceptual view of the main phenomena and interactive processes contributing to the influent are formulated in terms of mathematical equations. Despite successful application of these models (at least in giving an overall view of the system), the performance of these models to a great extent depends on the proper choice of model parameters. Since some of the parameters may not have a clear physical meaning they are usually estimated through model calibration. In cases in which there is no measured data available for model calibration, only a rough estimate or a range of values could be inferred from the values reported in literature. Besides, it is almost impossible to have a complete similarity between the model output(s) and the observed values owing to the inextricable uncertainties (e.g. input data uncertainty and/or model structure uncertainty) in any modeling exercise (Belia *et al.*, 2009; Freni and Mannina, 2010).
Given the importance of the issue of uncertainty, several studies have been conducted to consider its effect on both water quality and quantity in urban drainage modeling (Freni *et al.,* 2009; Dotto *et al.,* 2012). However, in these studies, only the effect of model uncertainty under a set of historical rain events (WWF conditions) has been considered (i.e. the time series of rainfall and also the contribution of wastewater in DWF conditions were known a *priori*). In this study on the contrary not only are we interested in considering the effect of model uncertainty, but also in the variability of rainfall and influent time series in DWF conditions which significantly affect both the amount and the dynamics of the influent.

PROPOSED INFLUENT GENERATOR

In this paper, a hybrid of statistical and conceptual modeling tools is proposed for synthetic generation of influent time series considering both the effect of model uncertainty and input variability. Given the importance of rainfall time series in the generation of the influent, a two-state Markov chain-gamma model (Richardson, 1981) in conjunction with two time series disaggregation methods were used for stochastic generation of rainfall time series with a high temporal resolution (i.e. 15-minute). To generate the influent time series in DWF conditions taking into account the daily periodic variation, auto-correlation, and cross-correlation in time, a multivariate time series models was developed and its parameters were estimated using the methodology proposed by Neumaier and Schneider (2001). The proposed stochastic model is superior to previous attempts in the generation of influent, as in previous studies the diurnal variation of the influent in DWF conditions was modeled using univariate time series models (Martin *et al.,* 2007), or by multiplying the daily average influent values to a set of coefficients representing the ratio of influent at different times of a day to its average value with or without addition of a noise term to the generated time series (Achleitner *et al.,* 2007; Langergraber *et al.,* 2008; Gernaey *et al.,* 2011). The outputs of the two statistical models used for the generation of rainfall and influent time series in DWF conditions are then input to a conceptual model for modeling the influent time series in WWF conditions. In this study the CITYDRAIN model (Achleitner *et al.,* 2007) was selected as the conceptual model owing to its flexibility and parsimony. The CITYDRAIN model of the sewershed is calibrated using the measured influent data through a Bayesian calibration procedure to account for the *total model uncertainty*. Finally, different realizations of the influent time series can be generated by running the calibrated CITYDRAIN model using an instance of a generated time series of rainfall and an instance of influent under DWF conditions (i.e. the two stochastic input time series). Figure shows the schematic of the proposed influent generator.

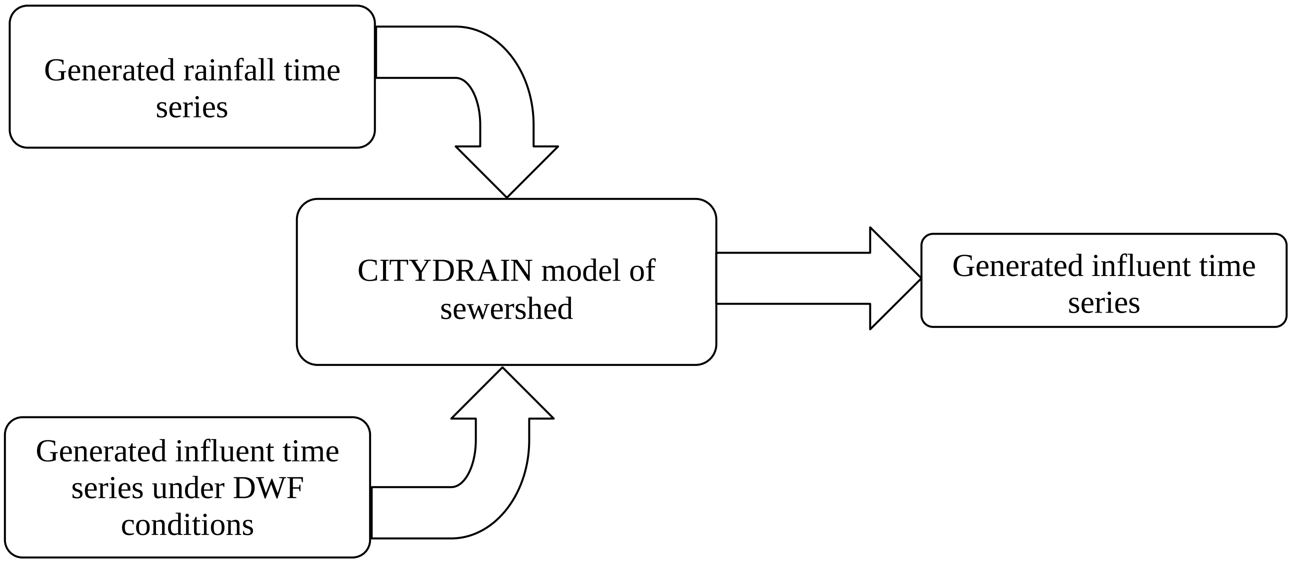

Figure 1 Schematic of the proposed influent generator

The main objective of the proposed influent generator is to produce a dynamic influent time series of flow and traditional wastewater component concentrations (TSS, COD, TN, TP, NH_4) with 15-min temporal resolution in order to capture the sub-daily time variations of the influent which could affect the operating parameters and the performance of WWTPs. One of the constraints was that the generator should only be using limited information on climate and the general characteristics of combined sewer systems. Depending on the biological model that would be used for modelling the biological processes inside a WWTP system, an influent fractionation block must be added to convert the generated traditional wastewater composition into state variables of the adopted biological models, e.g. the ASM models. The generated influent time series using the proposed tools can be used among others for the design of WWTPs under uncertainty (Martin *et al*., 2012).

Data and case study

The Eindhoven WWTP with a design capacity of 750000 population equivalent (PE) is the third largest WWTP in the Netherlands. The sewershed served by the Eindhoven WWTP has a total area of approximately 600km^2 and comprises of three main sub-sewersheds called Nuenen/Son, Eindhoven Stad, and Riool-Zuid. The influent data used in this study are related to sensor data of flow, ammonia (measured using an ion-selective sensor) soluble COD, total COD, and TSS (the latter 3 measured using an UV/VIS-based sensor) in the period of September 2011 to September 2012 at the outlet of the Nuenen/Son, Eindhoven Stad, and Riool-Zuid sub-sewersheds. It should be noted that the raw sensor data were cleaned up using visual inspection and a wavelet-based denoising strategy (details are not included in this paper).

The long-term daily rainfall data and also rainfall data with finer temporal resolution provided by KNMI (Royal Netherlands Meteorological Institute) and Waterschap De Dommel were used for estimating the parameters of the weather generator proposed in this paper.

Weather generator

Realistic generation of rainfall time series is crucial as it is one of the most important factors that affect the dynamics of the influent during WWF conditions. In this study a stochastic model proposed by Richardson (1981) was used for the synthetic generation of daily rainfall and air temperature time series. According to this method the sequence of dry and wet days is generated using a two-state Markov chain model with parameters $P(W|W)$ and $P(W|D)$ which represent the probability of having a wet day at day t given a wet day at day $t-1$ and the probability of having a wet day at time t given a dry day at time $t-1$ respectively (Figure).

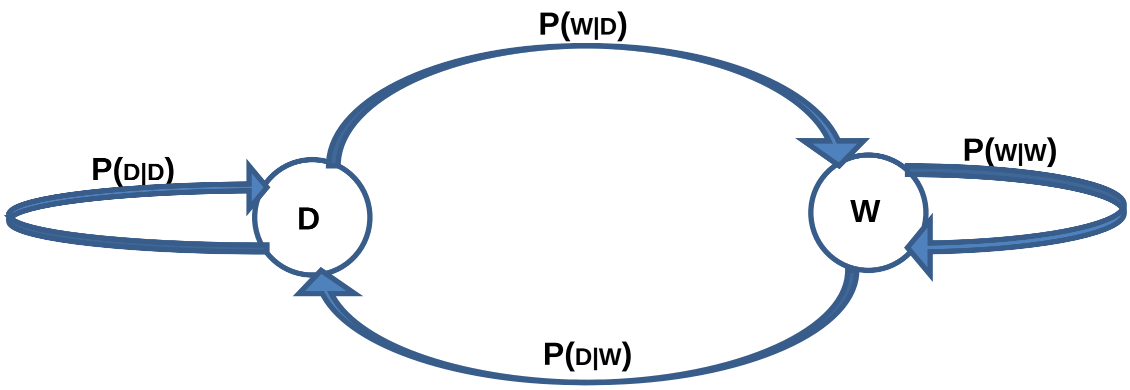

Figure 2 Schematic of a two-state Markov chain, i.e. wet (W) or dry (D)

The other two parameters of the transition matrix needed for generation of dry and wet days (i.e. $P(D|D)$ the probability of having a dry day at day t given a dry day at day $t-1$ and $P(D|W)$ the probability of having a dry day at day t given a wet day at day $t-1$) can be calculated using Equation 1 and Equation 2.

$$P(D|D) = 1 - P(W|D) \qquad \text{Equation 1}$$

$$P(D|W) = 1 - P(W|W) \qquad \text{Equation 2}$$

Once the sequence of wet and dry days is generated, the amount of rainfall in a wet day is generated by sampling from a gamma probability distribution (Equation 3)

$$f(x) = \frac{(x/\beta)^{\alpha-1}\exp(-x/\beta)}{\beta\Gamma(\alpha)} \qquad \text{Equation 3}$$

where x is the depth of daily rainfall, α and β are the two parameters of the distribution, and $\Gamma(\alpha)$ represents the gamma function evaluated at α. The time series of minimum and maximum air temperature are generated conditioned on the state of the day (i.e. wet or dry) using a multivariate linear first-order time series model (Matalas, 1967). The above weather generator is suited for random generation of daily rainfall and temperature. However, in this study we need to generate rainfall time series with a finer temporal resolution than daily resolution (15-min temporal resolution, comparable to the temporal resolution of rainfall in the BSM influent model (Gernaey *et al.*, 2011)). Some methodologies have been proposed for random generation of hourly rainfall time

series based on historical hourly rainfall data. However, long-term hourly rainfall data may not be available in every region and using a limited hourly rainfall record for random generation of long-term hourly rainfall time series may result in misrepresentation of the inter-annual variability in rainfall.
That being said, in this study it is proposed to combine the Richardson-based weather generator (i.e. which is used for daily rainfall generation) with two time series disaggregation techniques. In other words, daily rainfall time series is first generated using the Richardson (1981) method and then two time series disaggregation models, including a daily-to-hourly model (Koutsoyiannis and Onof, 2001) and an hourly-to-15-minutes model (Ormsbee, 1989) are applied for generation of long-term rainfall time series with 15-minute temporal resolution. Moreover, the original Richardson-based weather generator is also suited for the generation of daily air temperature. However, in this study not the air temperature but the wastewater temperature is of interest as it affects the rate of many biological processes taking place in the bioreactors. To estimate the wastewater temperature a simple linear regression model was fitted between the daily air temperatures and the corresponding wastewater temperature measured during the period of September 2011 to September 2012. The fitted regression model was used to calculate the daily wastewater temperature as a function of daily air temperature generated using the Richardson-based weather generator.

Influent generation in DWF conditions

The influent time series in DWF conditions usually shows specific periodic patterns which can be mainly attributed to the socio-economic fabric of society and also to the physical characteristics of the wastewater collection system. To mimic these variations in time, it is common practice to estimate representative values (e.g. multiplying flow per person to the total population for estimating flow) for flow and loads and then multiplying them to a set of normalized coefficients reflecting diurnal, weekly and seasonal time variation of the influent time series (Jeppsson *et al.*, 2007; Gernaey *et al*., 2011; Flores-Alsina *et al.*, 2014). Moreover, Gernaey *et al.* (2011) proposed to add a noise term to the deterministic influent profile in order to avoid generating the same influent time series in subsequent days. In this study the application of a multivariate auto-regressive model (Neumaier and Schneider, 2001) with periodic components is proposed.
To estimate the parameters of the proposed time series model, the influent time series during DWF conditions were extracted and analyzed for estimating the parameters of the multivariate auto-regressive model. First, the seasonal (e.g. associated to groundwater infiltration) and diurnal periodic components of flow and other wastewater constituents were estimated using different Fourier series approximations and removed from the original influent time series to calculate the residual time series. The zero-mean residual time series of influent flow and composition were furthered standardized to have an influent time series with a zero mean and unit standard deviation. The parameters of the multivariate autoregressive model in Equation 4 (i.e. p, A_l, C) were then estimated through a stepwise least square algorithm proposed by Neumaier and Schneider (2001).

$$v_t = \sum_{l=1}^{p} A_l \times v_{t-l} + \varepsilon_t \qquad \text{Equation 4}$$

In Equation 4, v_t is an m-dimensional vector (i.e. for our application m=5 which corresponds to the flow and the four wastewater compositions) containing the generated influent component at time t, p is the order of the auto-regressive model, $A_1,...,A_p$ are the coefficient matrices of the auto-regressive model, and ε_t is a noise term generated from an uncorrelated zero-mean multivariate normal distribution with the covariance matrix C (i.e. $\varepsilon_t \sim N(0,C)$). Different realizations of the residual influent time series can be generated using this time series model and converted to the original scale depending on the mean and standard deviation of the original influent time series.

Influent generation in WWF conditions

Synthetic generation of the influent time series during WWF conditions is relatively more complicated than the generation of the influent time series during DWF conditions. Difficulties arise as various phenomena are occurring during WWF conditions and as the availability of measured data is usually scarce for these periods. Hence, using a purely statistical model may result in significant discrepancies between simulated and observed time series. Therefore, we used a combination of statistical modeling techniques and a conceptual model to generate the time series of the influent during WWF conditions. The CITYDRAIN model (Achleitner *et al.*, 2007) was selected as the conceptual model as it takes into account the basic phenomena that govern the amount and dynamics of the influent and also requires only a small number of parameters whose values or ranges of values can be inferred from the basic information of a sewershed.

Flow

CITYDRAIN calculates the amount of effective rainfall by adopting the concept of virtual basins in which effective rainfall is calculated by subtracting the initial loss from rainfall and then multiplying it with the runoff coefficient. The height of the effective rainfall is then multiplied by the fraction of sewershed area which contributes to the generation of runoff to calculate flow. A simplified routing method based on the well-known Muskingum method is then used for routing flow and pollutants inside the sewer system.

Composition

For the generation of pollutant time series in WWF conditions, CITYDRAIN uses a rather simplistic approach in which a fixed pollutant concentration is imposed to the system:

$$\begin{cases} C(t) = C \text{ if } h_e > 0 \\ C(t) = 0 \text{ if } h_e = 0 \end{cases} \qquad \text{Equation 5}$$

where, $C(t)$ is the generated pollutant concentration in time, C is a model parameter representing the concentration in WWF conditions, and h_e is the effective rainfall. Given the importance of the influent time series in WWF conditions, a more appropriate conceptual model was used for simulating the accumulation-wash off processes corresponding to the particulate concentrations. To this aim, a new block was developed and implemented in CITYDRAIN to generate the pollutant concentration time series in WWF conditions. Equation 6 shows the mathematical formulation of the selected accumulation-wash off model (Kanso *et al.*, 2005).

$$\begin{cases} \text{Accumulation model: } \dfrac{dM_{(t)}}{dt} = K_a\left(m_{\lim} \times S_{imp} - M_{(t)}\right) \\ \text{Wash off model: } \dfrac{dM_{(t)}}{dt} = -W_e \times I_{(t)}{}^{w} \times M_{(t)} \end{cases} \qquad \text{Equation 6}$$

where, $M_{(t)}$ is the vailable pollutant mass on the sewershed at time t (kg), K_a is the accumulation coefficinet (1/day), $m_{\lim}$ is the maximum accumulated mass (kg/ha), S_{imp} is the impervious area (ha), $I_{(t)}$ is the rainfall intensity (mm/hr), W_e, and w are calibration parameters.

Bayesian model calibration and long-term influent generation

As explained in the previous section, the dynamics of the influent time series in WWF conditions is modeled using the CITYDRAIN model. However, one should be aware of the fact that modeling the influent time series in WWF conditions using a conceptual model may not lead to reliable results unless the model is calibrated and the effect of different sources of uncertainties on the outputs (e.g. flow and other pollutants) are taken into account. To this aim, a Bayesian framework was used to update the ranges of values that were initially assigned to the parameters of the CITYDRAIN model (i.e. estimating the *posterior distribution* of parameters using their *prior distribution* and the measured data on flow and pollutant concentrations). In general, the *posterior distribution* of parameters using Bayes' theorem can be formulated by Equation 7.

$$h(\theta \mid Data) = \frac{f(Data \mid \theta)\, p(\theta)}{f(Data)} \qquad \text{Equation 7}$$

where $h(\theta \mid Data)$ is the *posterior distribution,* $p(\theta)$ is the *prior distribution,* $f(Data)$ is merely a proportionality constant so that $\int h(\theta \mid Data) = 1$, and $f(Data \mid \theta)$ constitutes the *likelihood function* which measures the likelihood that the data correspond to the model outputs with parameter set $\boldsymbol{\theta}$. Assuming homoscedastic uncorrelated Gaussian error terms the *likelihood function* function can be formulated according Equation 8 (Bates and Campbell, 2001; Marshall *et al.,* 2004).

$$f(Data \mid \boldsymbol{\theta}) = \left(2\pi\sigma^2\right)^{-n/2} \prod_t^n exp\left\{-\frac{\left[Data_t - R(x_t;\boldsymbol{\theta})\right]^2}{2\sigma^2}\right\} \qquad \text{Equation 8}$$

where n is the number of observations, σ^2 is the variance of the residual error (i.e. the difference between model predictions and observed values), $Data_t$ is the observed variable at time t, x_t is the set of inputs at time t, $\boldsymbol{\theta}$ is the set of model parameters and $R(x_t;\boldsymbol{\theta})$ represents the model output as a function of x_t and $\boldsymbol{\theta}$.

A specific form of Markov chain Monte Carlo (MCMC) sampler known as differential evolution adaptive Metropolis or DREAM (Vrugt *et al.*, 2008) was used to efficiently estimate the *posterior distribution* of the CITYDRAIN model parameters given the time series of flow and influent composition of the Eindhoven WWTP. It should be noted that the proposed Bayesian approach is not only capable of capturing the effect of model parameter uncertainty, but also of capturing the effect of other sources of uncertainties that could result in some discrepancies between the simulated influent time series and the observed series.

Once the uncertainty ranges of the CITYDRAIN model parameters are updated, synthetic influent time series for a desired number of years considering the variability in the inputs of the CITYDRAIN model (i.e. rainfall and influent time series in DWF conditions) and also the total uncertainty can be obtained as follows:

1. Synthetic generation of the 15-minute time series of rainfall for one year
2. Synthetic generation of the 15-minute time series of the influent in DWF conditions for one year
3. Sampling a point from the *posterior distribution* of the CITYDRAIN model parameters

4. Inputting the generated time series 1) and 2) and the parameters sampled in 3) and running the CITYDRAIN model for one year
5. Repeating 1) to 4) for a desired number of years

RESULTS AND DISCUSSION

This section presents the outputs and some discussion on the results of different components of the proposed influent generator. The performance of the weather generator and the influent generator under DWF conditions are evaluated by comparing the statistical properties of the generated time series with those of the historical time series. The results corresponding to the Bayesian calibration of CITYDRAIN model are explained and at the end a 7-day snapshot of generated one year influent time series is presented and discussed.

Synthetic generation of rainfall

The parameters of the statistical Markov-gamma model were estimated using the recorded rainfall data in the studied Eindhoven catchment. The results indicate that not only are the basic yearly statistics (i.e. average and variance) of the generated rainfall time series consistent with the recorded rainfall time series, but also the seasonal variations in rainfall intensity and frequency of wet days are respected (Figure and Table 1).

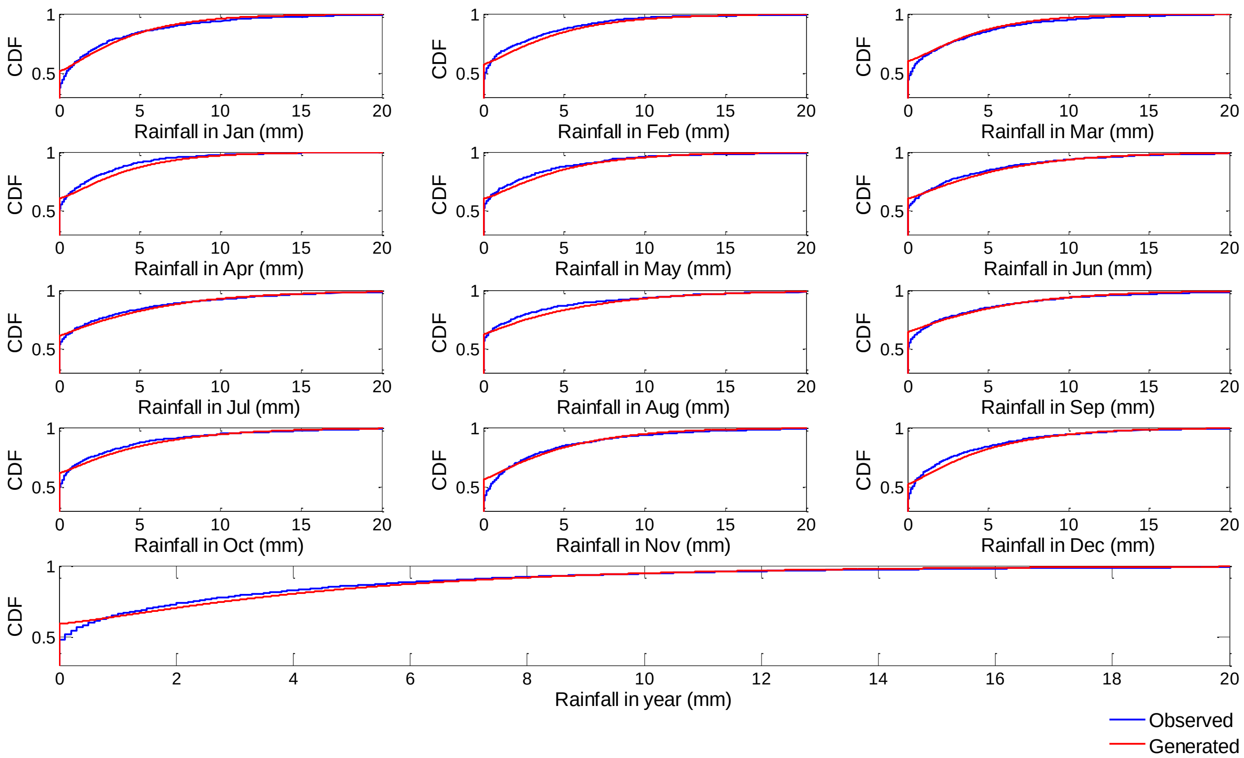

Figure 3 Cumulative distribution function of daily rainfall in the studied Eindhoven catchment

Moreover, Table shows that the hourly time series of rainfall which was generated using the time disaggregation method (i.e. disaggregation of daily to hourly time series) has the same statistical characteristics as the observed one. Overall, the synthetic generation of rainfall in which the statistical properties of the time series is respected across different time scales is a significant improvement compared to the rainfall generation in for instance the BSM influent generator in which there is no clear way for extracting and incorporating the statistical properties of available recorded rainfall data into synthetic rainfall time series generation. Besides, the flexibility of the proposed rainfall generator allows users to define different scenarios reflecting future changes in precipitation regime (e.g. due to climate change (Chen *et al*., 2010)) and its effect on the influent time series (e.g. what would happen if the amount of precipitation increases by 20%).

Table 1 Average rainfall amount and number of wet days for Eindhoven catchment

Month	Amount of Rainfall (mm)		Expected number of Wet Days	
	Observed	Generated	Observed	Generated
Jan	72.3	67.0	16	14
Feb	52.0	57.0	12	11
Mar	63.4	54.4	13	12
April	44.1	51.9	12	11
May	58.3	60.9	12	12
Jun	68.0	68.4	12	11
Jul	74.7	73.5	12	11
Aug	64.6	71.0	11	11
Sep	67.9	62.1	12	10
Oct	62.0	65.0	12	11
Nov	71.1	66.4	15	12
Dec	70.0	74.0	14	14
Annual	768	772	152	141

Table 2 Basic statistics of hourly rainfall data for Eindhoven catchment

Statistics	Unit	Observed Value	Simulated Value
Mean	mm	0.08	0.08
Standard deviation	mm	0.60	0.60
Lag 1 auto-correlation	---	0.33	0.36
Proportion of dry hours	---	0.92	0.94

Synthetic generation of influent temperature

As mentioned in the methodology section, the daily temperature of wastewater is estimated through a linear regression model which relates the daily average wastewater temperature to the daily average air temperature. Figure illustrates a random generation of air and wastewater temperature time series for one year. The linear model in Figure shows that the average wastewater temperature can be estimated reasonably ($R^2 = 0.70$) as a linear function of air temperature. To further disaggregate the daily average wastewater temperature into a time series with 15-minute temporal resolution, the average diurnal variation of wastewater temperature which was extracted and smoothed using a first order Fourier series estimate (Figure c) was multiplied to the daily average wastewater temperature. Despite the fact that the diurnal variation pattern in Figure c clearly shows a periodic behavior in time (which corresponds to the diurnal variation of wastewater temperature), there is no significant difference between the highest and lowest temperature throughout a day (i.e. the highest temperature is only around 1.001 times the daily average wastewater temperature and the lowest temperature is around 0.9985 times the daily average wastewater temperature). Therefore, in practical applications (at least for the case study in this research), the diurnal temperature variation can be ignored.

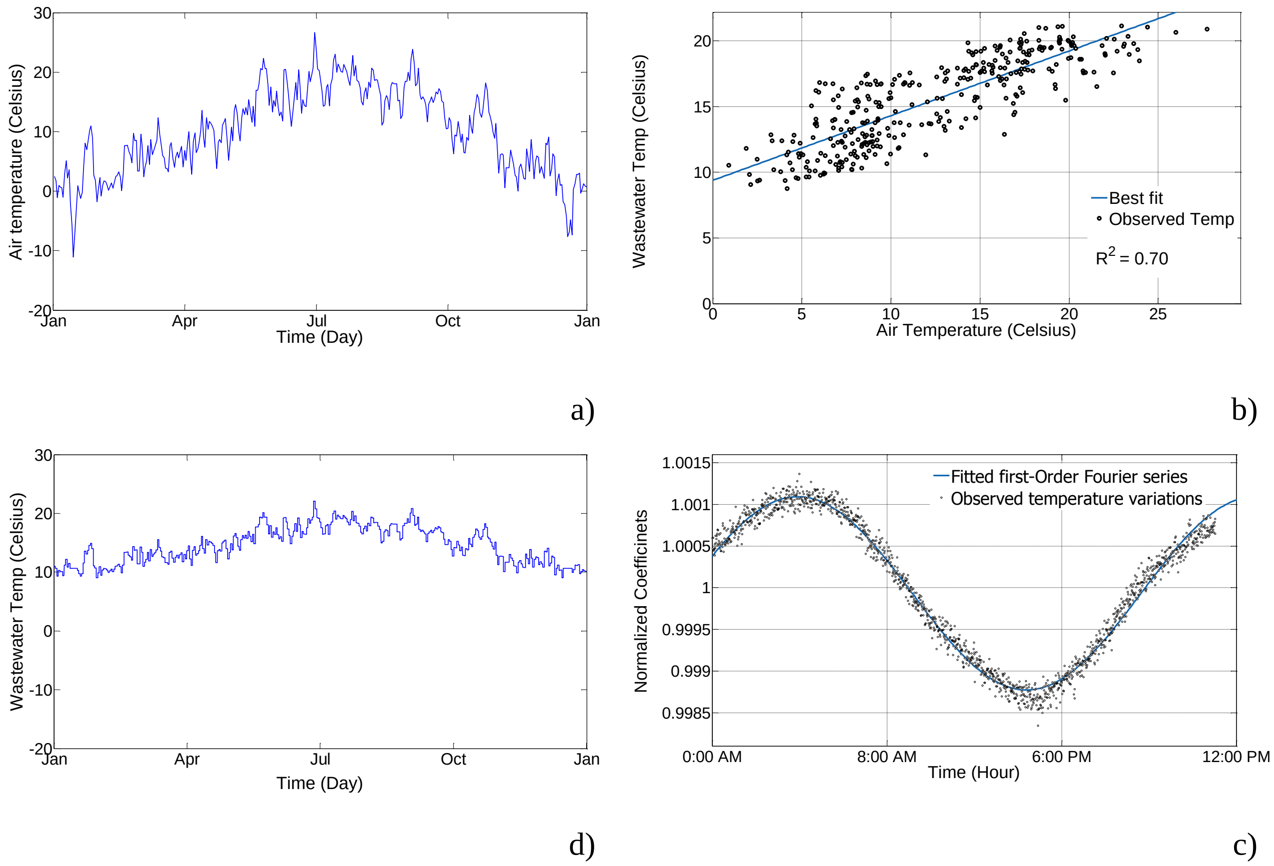

Figure 4 Random generation of air and wastewater temperature for one year for the Eindhoven WWTP: a) Randomly generated average daily air temperature for a year, b) linear regression model for calculating the average daily wastewater temperature as a function of average daily air temperature, c) the average and fitted normalized coefficients (the normalized coefficients for each day were calculated by dividing the influent temperature at different moments of a day by the daily average influent temperature in the same day) for calculating the diurnal wastewater temperature variations, and d) randomly generated wastewater temperature time series with 15-minute temporal resolution.

Multivariate auto-regressive model for DWF generation

As explained, the parameters of the multivariate auto-regressive model were estimated using a specific least square algorithm (Neumaier and Schneider, 2001). Figure shows a continuous 3-day DWF influent time series with the results corresponding to the fitted multivariate auto-regressive model. The uncertainty band was generated through random generation of the noise term (i.e. p, A_l in Equation 4 were fixed and the noise term was generated from $\varepsilon_t \sim N(0, C)$).

One of the main advantages of the proposed multivariate time series model over univariate time series models (Martin *et al.*, 2007) or the DWF generator in the BSM influent generator (Gernaey *et al.*, 2005) is that not only are the auto-correlation structures in time respected but also the cross-correlation structures. Table shows the correlation matrix for the randomly generated and observed influent time series in DWF conditions.

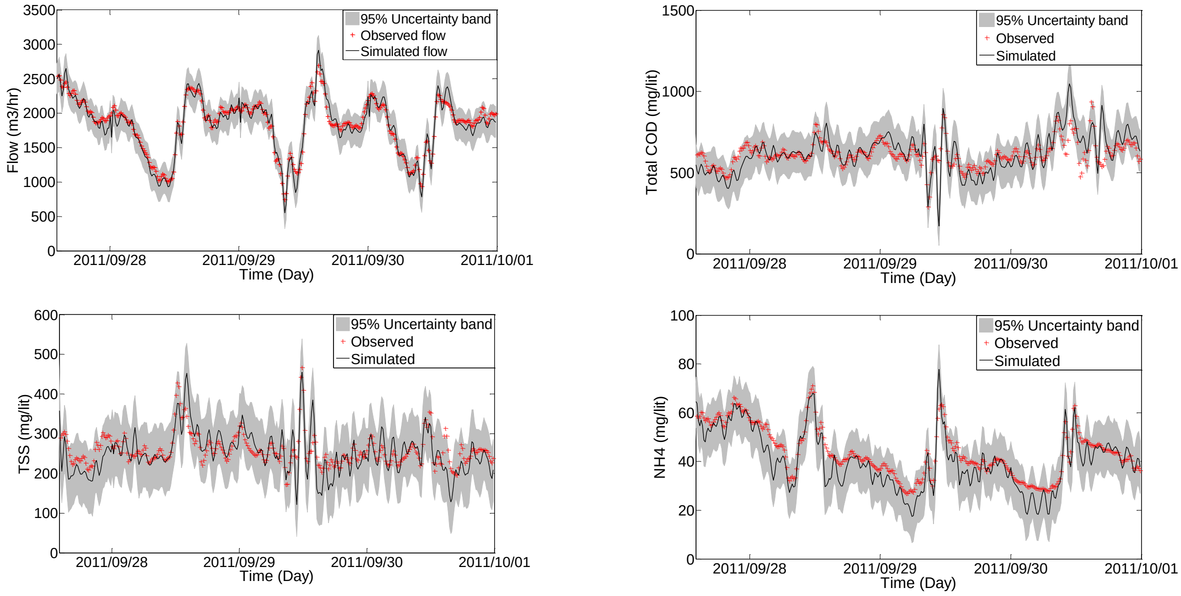

Figure 5 Observed and simulated influent time series under DWF conditions

Table 3 Correlation matrix for the generated and observed influent time series in DWF

Generated influent time series						Observed influent time series					
	Flow	Soluble COD	Total COD	TSS	NH4		Flow	Soluble COD	Total COD	TSS	NH4
Flow	1.00					Flow	1.00				
Soluble COD	-0.11	1.00				Soluble COD	-0.12	1.00			
Total COD	-0.04	0.77	1.00			Total COD	-0.06	0.77	1.00		
TSS	0.06	0.32	0.80	1.00		TSS	0.05	0.33	0.81	1.00	
NH4	-0.43	-0.04	-0.06	-0.04	1.00	NH4	-0.46	0.00	-0.02	-0.03	1.00

CITYDRAIN model calibration and synthetic influent generation

As explained in the methodology section, the CITYDRAIN model was used for modeling the dynamics of the influent time series during WWF conditions. Uniform distributions representing the initial knowledge on parameters were selected as *prior distributions* and their corresponding *posterior distributions* were estimated by sampling from Equation 7 using the DREAM sampler. Figure and Figure show the *posterior distributions* of the CITYDRAIN model after calibrating the model for flow and TSS time series in WWF conditions (three days of simulations were used as the warm-up period to set the initial conditions of the system).

As indicated in Figure and Figure , there exists some correlation among the parameters of the CITYDRAIN model. For example in Figure , the parameters that affect the generation of effective rainfall (i.e. runoff coefficient, initial loss, and permanent loss) are correlated meaning that different combinations of these parameters could result in the same amount of effective rainfall given the same inputs and values for other parameters. However, given the narrow ranges associated to the parameters that affect the amount of rainfall, the uncertainty band for flow relating to the *total model uncertainty* is mainly affected by the standard deviation of the residual error (i.e. Sigma in Figure) and not by the uncertainty of the CITYDRAIN model parameters.

The parameters that affect the accumulation of pollutant (i.e. m_lim, and Ka) and those that affect the wash-off of pollutants are also correlated. Given the different correlation structures that exist among some parameters it is very important to sample from the joint distribution of parameters to propagate the effect of parameter uncertainties to the outputs.

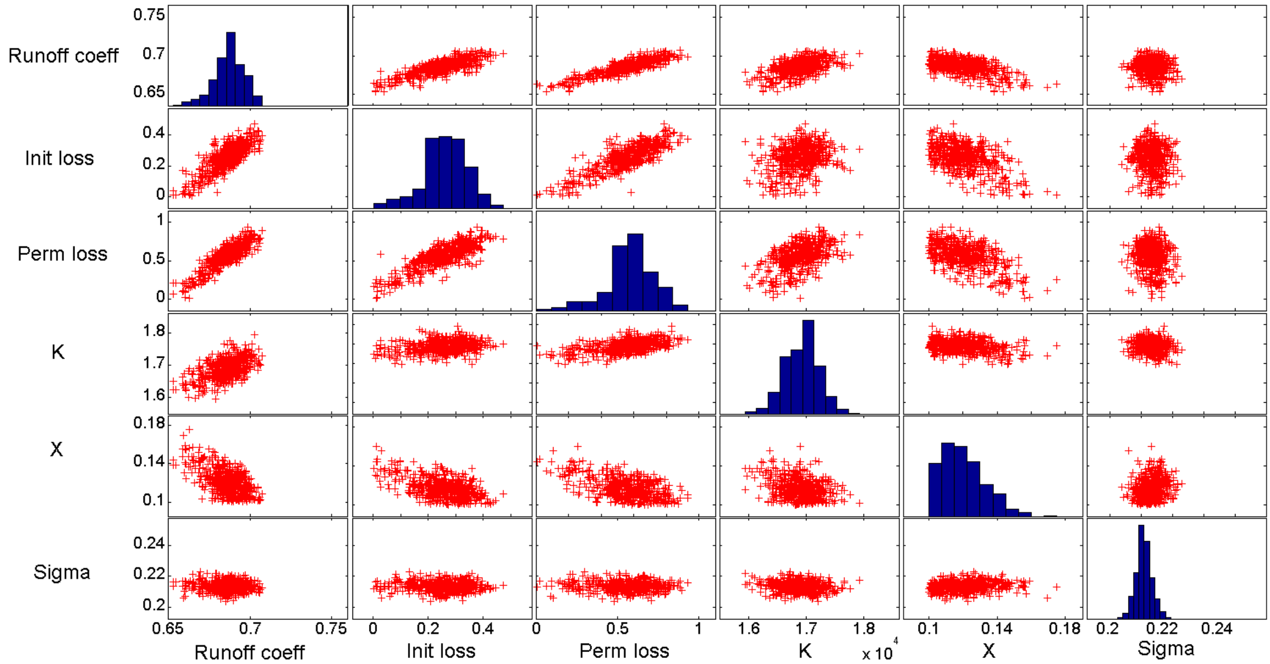

Figure 6 Posterior distribution of parameters for flow calibration where, runoff coeff, init loss, and perm loss are respectively the runoff coefficient, initial loss (mm), permanent loss (mm/day) parmeters in the virtual basins model that is used in the CITYDRAIN model, K (sec) and X are the routing parameters used in the Muskingum method, and Sigma is the standard deviation of the residual error.

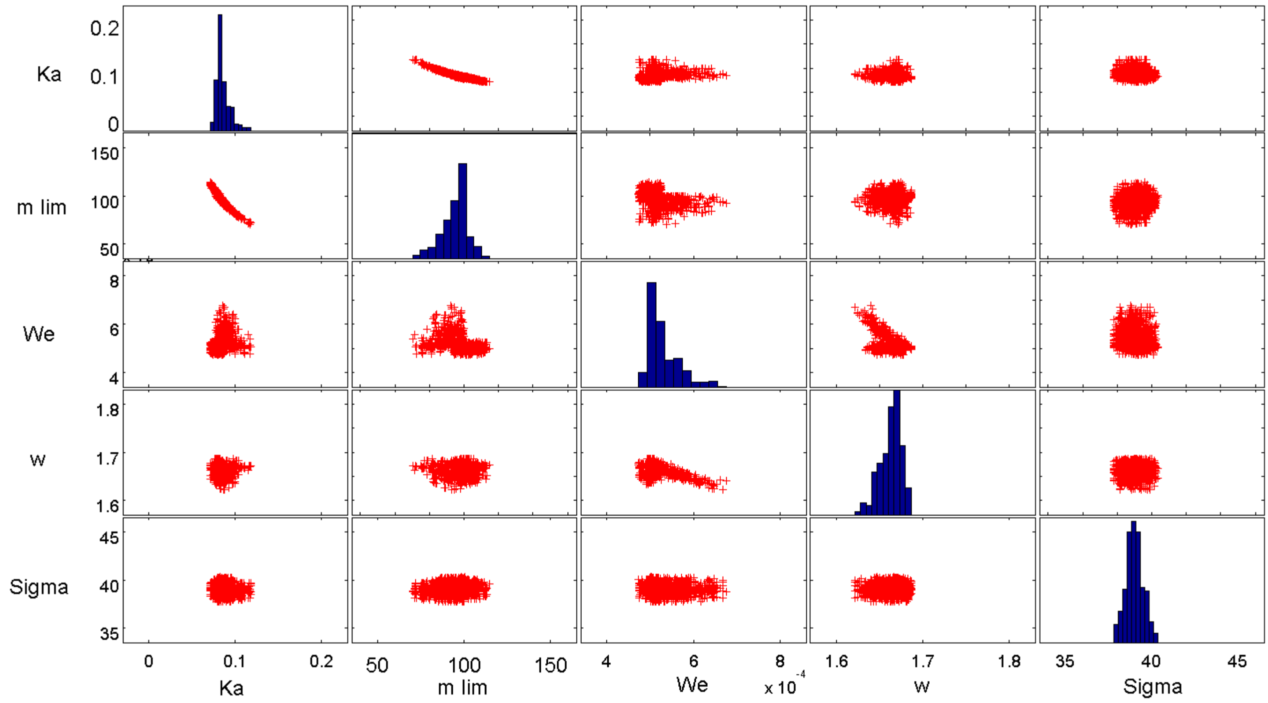

Figure 7 Posterior distribution of parameters for TSS calibration where Ka is the accumulation coefficient (1/day), m_lim is the maximum accumulated mass (kg/ha), We, and w are the calibration parameters (Equation 6).

To consider the effect of *total model uncertainty* on the outputs of CITYDRAIN model, a Monte Carlo simulation was performed by sampling from the joint *posterior distribution* of parameters and running the model for 1000 times for a particular rainfall time series. Figure illustrates the 95% uncertainty band for flow and TSS which was constructed by selecting the 2.5 and 97.5 percentiles of the cumulative distribution of flow and TSS as the lower and upper limits of uncertainty of simulation with the rainfall time series shown in the figure. The figure also presents the observed and the best simulated time series. The latter corresponds to the set of parameters that has the highest *likelihood function* value.

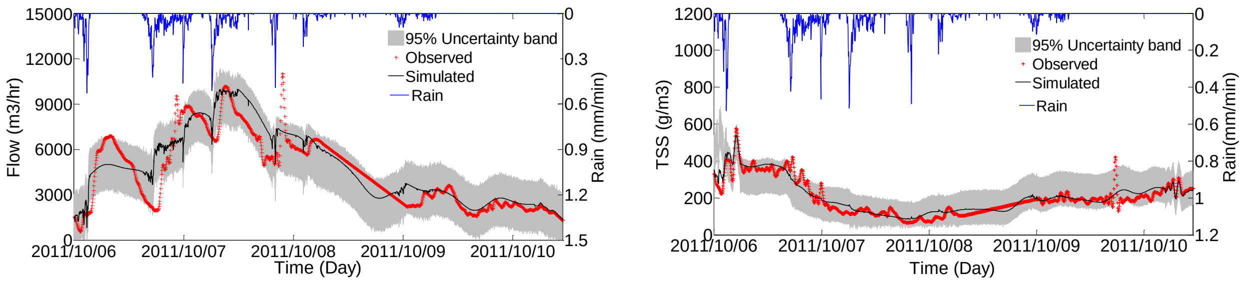

Figure 8 Uncertainty bands for flow (left) and TSS concentration (right) in a 4-day wet weather period

To further analyze the statistical properties of the simulated influent time series during both the DWF and WWF conditions, the cumulative distribution function (CDF) of the simulated and observed influent flow and pollutant load were compared in Figure and Figure 4. The simulated and observed influent time series with 15-minute temporal resolution were aggregated to construct the corresponding daily and hourly influent series. Figure and Figure 4 show that the influent generator has excellent performance when it comes to predicting the daily and hourly influent flow and pollutant load values.

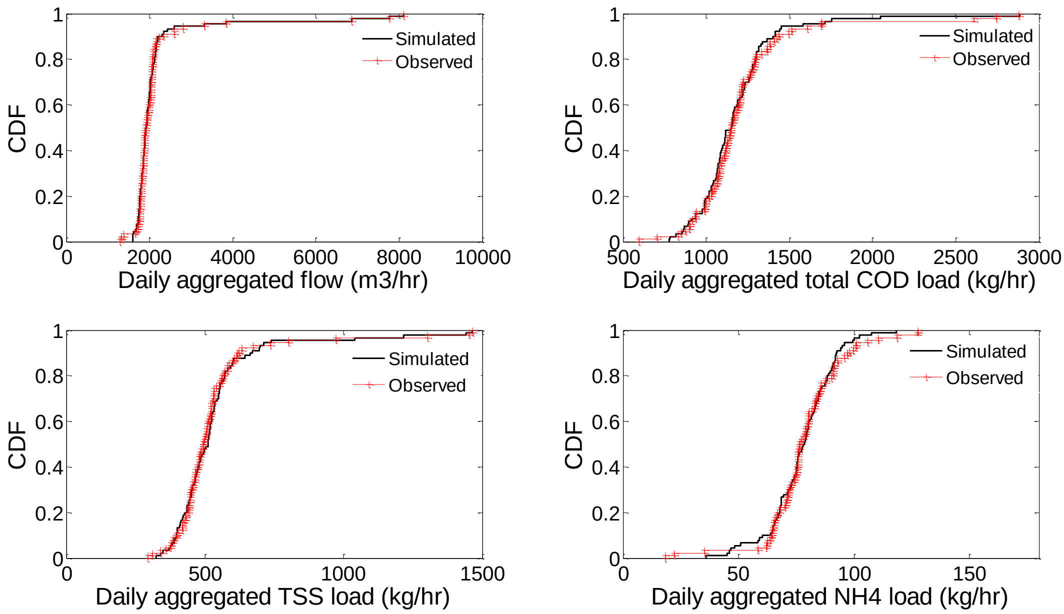

Figure 9 CDFs of daily-aggregated influent flow and load of influent pollutants

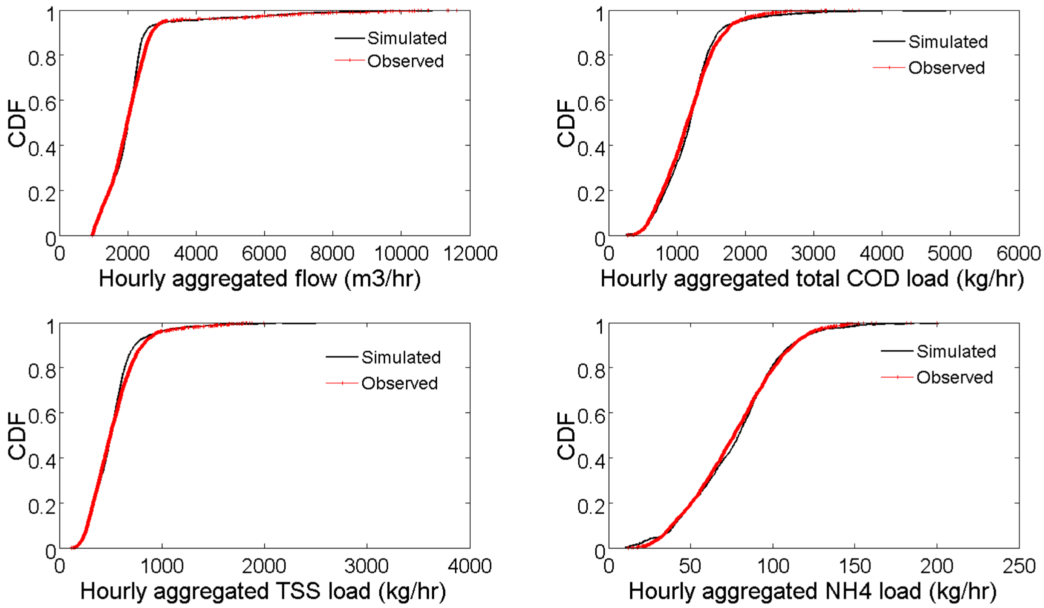

Figure 4 CDFs of hourly-aggregated influent flow and load of influent pollutants

It can be concluded from Figure and Figure 4 that the statistical properties of the simulated time series are similar to the properties of the observed series once the model is fed with the observed rainfall time series. As explained in the methodology section, synthetic generation of a one year influent time series with 15-minute temporal resolution is thus possible by sampling from the posterior distribution of the CITYDRAIN model parameters and inputting the model with synthetically-generated rainfall and influent time series for DWF conditions (both with 15-minute temporal resolution). The latter two series are to be generated using the proposed rainfall and DWF generators respectively.

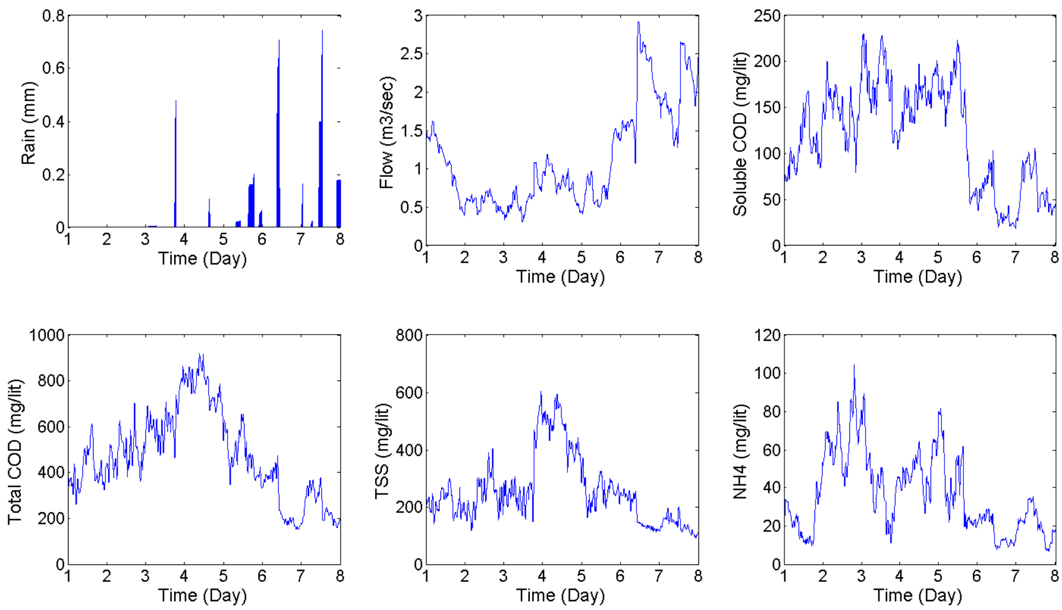

Figure 5 A 7-day realization of rainfall and influent time series

Figure 5 shows a 7-day snapshot of a generated one year influent time series. During the hours of the first day the time series of flow has a descending trend as the runoff produced by rainfall event just before the first day (not depicted in Figure 5) exits the sewer system and the flow time series reaches its DWF conditions with a typical periodic pattern (the second day in Figure 5). During the last hours of the third day another rainfall event occurs and the flow time series increases while the time series of soluble COD and ammonia drop due to dilution of wastewater by runoff. However, during the same period of time there is a sudden increase in the total COD and TSS concentrations due to the wash-off of particulate material. After the wash-off of the particulates during the last hours of the fourth day, the dilution effect starts to dominate again and the time series of total COD and TSS drop due to the dilution of the wastewater by runoff.

CONCLUSION

In this paper a combination of statistical and conceptual modeling tools was proposed for synthetic generation of dynamic influent time series of flows and pollutant concentrations with 15-miniute temporal resolution. The rainfall generator is capable of considering the annual and inter-annual rainfall regimes and keeping the consistency of the generated rainfall time series across different temporal resolutions. Comparison between observed and simulated influent time series for the Eindhoven case study proved the capability of the proposed multivariate auto-regressive model in generating realistic influent time series in DWF conditions. Moreover, long-term generation of influent time series under dry and wet weather conditions could be achieved by running the CITYDRAIN model of the sewershed using the generated stochastic inputs (i.e. rainfall and influent time series in DWF condition). Uncertainty could be captured by sampling different vectors of the model parameters from the posterior distribution obtained after Bayesian parameter estimation on the basis of the case study data.

Overall, the proposed influent generator provides a clear and coherent method to incorporate the general and easy-to-obtain information on the physical characteristics of the sewershed as well as climate conditions of the region into the synthetic generation of the influent of a treatment plant. The flexibility of the presented influent generator allows the users to define different scenarios reflecting the projected change in climate and the characteristics of the sewershed (e.g. population growth, change in pervious area) and evaluate their effect on the generated influent time series.

ACKNOWLEDGEMENT

The authors would like to thank Stefan Weijers, and Petra van Daal-Rombouts from Waterschap De Dommel (Eindhoven, The Netherlands) and Youri Amerlinck from BIOMATH-Ghent University (Belgium) for providing the influent and rainfall data for this study. This research work was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) and Primodal Inc. as well as the financial support of Hampton Roads Sanitation District, Virginia. Peter Vanrolleghem holds the Canada Research Chair in Water Quality Modeling.

REFERENCES

Achleitner S., Möderl M. and Rauch W. (2007). CITY DRAIN © – An open source approach for simulation of integrated urban drainage systems. *Environmental Modelling & Software* **22**(8), 1184-95.

Bates B. C. and Campbell E. P. (2001). A Markov Chain Monte Carlo Scheme for parameter estimation and inference in conceptual rainfall-runoff modeling. *Water Resources Research* **37**(4), 937-947.

Bechmann H., Nielsen M. K., Madsen H. and Kjølstad Poulsen N. (1999). Grey-box modelling of pollutant loads from a sewer system. *Urban Water* **1**(1), 71-8.

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. and Villez K. (2009). Wastewater treatment modelling: Dealing with uncertainties. *Water Science and Technology* **60**(8), 1929-1941.

Capodaglio A., Zheng S., Novotny V. and Feng X. (1990). Stochastic system identification of sewer-flow models. *Journal of Environmental Engineering* **116**(2), 284-98.
Chen J., Brissette F. P. and Leconte R. (2010). A daily stochastic weather generator for preserving low-frequency of climate variability. *Journal of Hydrology* **388**(3–4), 480-90.
Dotto C. B. S., Mannina G., Kleidorfer M., Vezzaro L., Henrichs M., McCarthy D. T., Freni G., Rauch W. and Deletic A. (2012). Comparison of different uncertainty techniques in urban stormwater quantity and quality modelling. *Water Research* **46**(8), 2545-58.
Flores-Alsina X., Saagi R., Lindblom E., Thirsing C., Thornberg D., Gernaey K. V. and Jeppsson U. (2014). Calibration and validation of a phenomenological influent pollutant disturbance scenario generator using full-scale data. *Water Research* **51**(0), 172-85.
Freni G. and Mannina G. (2010). Bayesian approach for uncertainty quantification in water quality modelling: The influence of prior distribution. *Journal of Hydrology* **392**(1), 31-9.
Freni G., Mannina G. and Viviani G. (2009). Uncertainty assessment of an integrated urban drainage model. *Journal of Hydrology* **373**(3), 392-404.
Gernaey, K., Rosen, C. and Jeppsson, U. (2005). BSM2: A model for dynamic influent data generation. Technical report. Department of Industrial Electrical Engineering and Automation. Lund University, Lund, Sweden.
Gernaey K. V., Flores-Alsina X., Rosen C., Benedetti L. and Jeppsson U. (2011). Dynamic influent pollutant disturbance scenario generation using a phenomenological modelling approach. *Environmental Modelling & Software* **26**(11), 1255-67.
Hao X., Loosdrecht M. C. M. V., Meijer S. C. F. and Qian Y. (2001). Model-based evaluation of two BNR processes—UCT and A2N. *Water Research* **35**(12), 2851-60.
Hernebring C., Jönsson L., Thorn U. and Møller A. (2002). Dynamic online sewer modelling in Helsingborg. *Water Science and Technology* **45**(4-5), 429-36.
Jeppsson U., Pons M., Nopens I., Alex J., Copp J., Gernaey K., Rosen C., Steyer J. and Vanrolleghem P. A. (2007). Benchmark Simulation Model no 2: General protocol and exploratory case studies. *Water Science and technology* **56**(8), 67-78.
Kanso A., Tassin B. and Chebbo G. (2005). A benchmark methodology for managing uncertainties in urban runoff quality models. *Water Science and Technology* **51**(02), 163-70.
Koutsoyiannis D. and Onof C. (2001). Rainfall disaggregation using adjusting procedures on a Poisson cluster model. *Journal of Hydrology* **246**(1–4), 109-22.
Langergraber G., Alex J., Weissenbacher N., Woerner D., Ahnert M., Frehmann T., Halft N., Hobus I., Plattes M. and Spering V. (2008). Generation of diurnal variation for influent data for dynamic simulation. *Water Science and Technology* **57**(9), 1483.
Marshall L., Nott D. and Sharma A. (2004). A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling. *Water Resources Research* **40**(2), W02501.
Martin C., Eguinoa I., McIntyre N. R., García-Sanz M. and Ayesa E. (2007). ARMA models for uncertainty assessment of time series data: application to Galindo-Bilbao WWTP. In: *Proceedings of the Seventh International IWA Symposium on Systems Analysis and Integrated Assessment in Water Management (WATERMATEX 2007).* Washington DC, USA, 7th–9th May 2007.
Martin C., Neumann M. B., Altimir J. and Vanrolleghem P. A. (2012) A tool for optimum design of WWTPs under uncertainty: Estimating the probability of compliance In: *Proceedings International Congress on Environmental Modelling and Software (iEMSs2012)*. Leipzig, Germany, 1th-5th July 2012.
Matalas N. C. (1967). Mathematical assessment of synthetic hydrology. *Water Resources Research* **3**(4), 937-45.
Neumaier A. and Schneider T. (2001). Estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Transactions on Mathematical Software (TOMS)* 27(1), 27-57.
Ormsbee L. (1989). Rainfall disaggregation model for continuous hydrologic modeling. *Journal of Hydraulic Engineering* **115**(4), 507-25.
Richardson C. W. (1981). Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research* **17**(1), 182-90.
Salem S., Berends D., Heijnen J. and Van Loosdrecht M. (2002). Model-based evaluation of a new upgrading concept for N-removal. *Water Science and Technology* **45**(6), 169-76.
Temprano J., Arango Ó., Cagiao J., Suárez J. and Tejero I. (2007). Stormwater quality calibration by SWMM: A case study in Northern Spain. *Water SA* **32**(1), 55-63.
Vrugt J. A., ter Braak C. J. F., Clark M. P., Hyman J. M. and Robinson B. A. (2008). Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resources Research* **44**(12), W00B9.