Environmental Modelling & Software 77 (2016) 32–49

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

# Influent generator for probabilistic modeling of nutrient removal wastewater treatment plants

Mansour Talebizadeh [a, *], Evangelia Belia [b], Peter A. Vanrolleghem [a]

[a] modelEAU, Département de génie civil et de génie des eaux, Université Laval, 1065 av. de La Médecine, Québec, QC G1V 0A6, Canada
[b] Primodal Inc., 145 Aberdeen, Québec, QC G1R 2C9, Canada

ARTICLE INFO

*Article history:*
Received 22 November 2014
Received in revised form
27 August 2015
Accepted 16 November 2015
Available online 12 December 2015

*Keywords:*
Bayesian estimation
Probabilistic design
Uncertainty analysis
Urban hydrology
Wastewater composition
Weather generator

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 (a Markov chain-gamma model for stochastic generation of rainfall and two different multivariate autoregressive models for stochastic generation of air temperature and influent time series in dry conditions) and conceptual modeling techniques is proposed for synthetic generation of influent time series. The time series of rainfall and influent in dry weather conditions are generated using two types of statistical models. These two time series serve as inputs to a conceptual sewer model for generation of influent time series. The application of the proposed influent generator to the Eindhoven WWTP shows that it is a powerful tool for realistic generation of influent time series and is well-suited for probabilistic design of WWTPs as it considers both the effect of input variability and total model uncertainty.



**Software availability**

Name of the software: WWTP influent advisor
Developer: Mansour Talebizadeh, Evangelia Belia, Peter A. Vanrolleghem
Programing language: Matlab 2012
Availability: The software can be obtained upon request by contacting Evangelia Belia, Primodal Inc., 145 Aberdeen, Québec, QC G1R 2C9, Canada. Email:belia@primodal.com

## 1. Introduction

One of the major sources of uncertainty/variability that both plant designers and operators must deal with is the dynamics of the influent (Belia et al., 2009). 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; Hyland et al., 2012) 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 other reasons, 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). The development of a tool capable of generating dynamic influent time series that is representative of the climate and characteristics of the sewershed could have several applications. Synthetically-generated influent time series can serve as input to a dynamic model of a plant for checking the performance of different configurations, sizings, as well as devising an optimum control strategy regarding the treatment of wastewater (Benedetti et al., 2006; Devisscher et al., 2006; Guerrero et al., 2011; Ciggin et al., 2012). In addition, realistic generation of different realizations of influent time series is one of the most important component of studies that take into account the issue of uncertainty in design and operation of WWTPs (Rousseau et al., 2001; Bixio et al., 2002; Martin et al., 2012; Talebizadeh et al., 2014).

Various approaches have been adopted by different researchers

* Corresponding author.
*E-mail addresses:* mansour.talebizadehsardari.1@ulaval.ca (M. Talebizadeh), belia@primodal.com (E. Belia), Peter.Vanrolleghem@gci.ulaval.ca (P.A. Vanrolleghem).

http://dx.doi.org/10.1016/j.envsoft.2015.11.005
1364-8152/

for influent generation (for a review see Martin and Vanrolleghem (2014)). One of the simplest approaches in synthetic generation of influent time series is the application of stochastic or regression models with or without periodic components (Capodaglio et al., 1990; Martin et al., 2007; Rodríguez et al., 2013). 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 processes that govern the generation and the dynamics of the influent. Langeveld et al. (2014) proposed an empirical model for predicting the influent pollutant time series as a function of influent flow for both dry and wet weather flow conditions (simulation of pollutant time series as a function of flow was also adopted by Rousseau et al. (2001) and Bixio et al. (2002)). Although the proposed model could be used for prediction of pollutant influent time series, it requires a stochastic input (i.e. influent flow) generator if it is to be used for generating different realizations of influent time series.

To consider the underlying phenomena, some researchers have advocated the use of detailed conceptual and/or 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 (i.e. the overall variation of influent time series, not all the different phenomena that have resulted in that time series), they might not be very useful as they require very detailed information on the sewer system and running them for a large number of times could be computationally expensive. Besides, even if a detailed sewershed model proves to have a good performance regarding the simulation of the influent time series under a given set of inputs, it cannot be called an influent generator unless a procedure is available for the generation of different realizations of stochastic inputs (e.g. rainfall time series, wastewater time series in dry weather flow (DWF) conditions).

Some researchers have proposed parsimonious conceptual models as an alternative to the complex mathematical models that require detailed information and data (Gernaey et al., 2011). In these models a conceptual view of the main phenomena and interactive processes 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. In addition, 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 that consider its effect on both water quality and quantity prediction 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 (wet weather flow, WWF, conditions) has been considered (i.e. the time series of rainfall and also the contribution of wastewater in DWF conditions were assumed known a priori). In this study not only are we interested in the effects 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 loads.

Considering the shortcomings of the previous studies, this study aims to develop an influent generator which is capable of producing dynamic influent time series of flow and traditional wastewater component concentrations (TSS, COD, TN, TP, $NH_4$) with 15-min temporal resolution (15-min temporal resolution was assumed to be enough for capturing sub-daily time variations of the influent which could affect the operating parameters and the performance of WWTPs). The proposed methodology will enable users to generate dynamic influent time series that have the same statistical properties as the observed ones using a set of statistical and conceptual modeling tools that only require basic information on climate and characteristics of the sewershed under study. It should be noted that the proposed influent generator is capable of considering the effect of uncertainty in model parameters on the generated influent time series whether the uncertainty can be reduced using observed data (e.g. for the current study) or not (uncertainty in model parameters is characterized by a range of values, determined through expert elicitation or the data from similar sewersheds). In the current study, the variability in inputs (captured by generating different realizations of rainfall and influent time series in DWF conditions, explained in Section 2.1 and Section 2.2, respectively) as well as the uncertainty in model parameters (explained in Section 2.4) on the generated dynamic influent time series are other important issues that will be covered.

## 2. Methodology

In this paper, a hybrid of statistical and conceptual modeling tools is proposed for synthetic generation of influent time series considering both model parameter uncertainty and input variability. A two-state Markov chain-gamma model (Richardson, 1981) in conjunction with two time series disaggregation methods were used for the stochastic generation of rainfall time series with a high temporal resolution (i.e. 15-min). To generate the influent time series during 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 expected to be superior compared to previous attempts in the generation of influent, as in previous studies the diurnal variation of the influent under DWF conditions was modeled using only univariate time series models (Martin et al., 2007), or by multiplying the daily average influent values to a set of coefficients representing the normalized dynamics of the influent at different times of a day 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 problem resulting from the application of univariate time series models is that the cross-correlation structure that exists among different wastewater constituents may not be respected, as the different constituents are generated independently from the others.

In DWF conditions, the influent time series is generated using multivariate time series models. Conversely in WWF conditions, the outputs of the two statistical models used for the generation of the rainfall and influent time series in DWF conditions are input to a conceptual model for modeling the influent time series in WWF conditions (Fig. 1). 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 under study is calibrated using measured influent data through a Bayesian calibration procedure to account for the total model uncertainty (uncertainty stemming from both model parameters and the distribution of error, i.e. the difference between the observed and simulated time series).