Generation of (synthetic) influent data for performing wastewater treatment modelling studies

Xavier Flores-Alsina[1], Christoph Ort[2], Cristina Martin[3,4], Lorenzo Benedetti[5], Evangelina Belia[6], Laura Snip[1], Ramesh Saagi[7], Mansour Talebizadeh[4], Peter A. Vanrolleghem[4], Ulf Jeppsson[7], Krist V. Gernaey[1]

[1]Center for Process Engineering and Technology (PROCESS), Department of Chemical and Biochemical Engineering, Technical University of Denmark (DTU), Building 229, DK-2800 Kgs. Lyngby, Denmark
[2] Eawag, Swiss Federal Institute of Aquatic Science and Technology, CH 8600 Dübendorf, Switzerland
[3] Deusto Institute of Technology – University of Deusto, Avda. Universidades 24, 48007 Bilbao, Spain
[4] modelEAU, Département de génie civil et de génie des eaux, Université Laval, 1065 Avenue de la Médecine, Québec G1V 0A6, QC, Canada
[5] Waterways, Via del Ferrone 88, 50023 Impruneta (Florence), Italy
[6] Primodal Inc., 145 Aberdeen, Québec QC, G1R 2C9, Canada
[7] Division of Industrial Electrical Engineering and Automation (IEA), Lund University, Box 118, SE-221 00 Lund, Sweden

Abstract: The success of many modelling studies strongly depends on the availability of sufficiently long influent time series - the main disturbance of a typical wastewater treatment plant (WWTP) - representing the inherent natural variability at the plant inlet as accurately as possible. This is an important point since most modelling projects suffer from a lack of realistic data representing the influent wastewater dynamics. The objective of this paper is to show the advantages of creating synthetic data when performing modelling studies for WWTPs. This study reviews the different principles that influent generators can be based on, in order to create realistic influent time series. In addition, the paper summarizes the variables that those models can describe: influent flow rate, temperature and traditional/emerging pollution compounds, weather conditions (dry/wet) as well as their temporal resolution (from minutes to years). The importance of calibration/validation is addressed and the authors critically analyse the pros and cons of manual versus automatic and frequentistic vs Bayesian methods. The presentation will focus on potential engineering applications of influent generators, illustrating the different model concepts with case studies. The authors have significant experience using these types of tools and have worked on interesting case studies that they will share with the audience. Discussion with experts at the WWTmod seminar shall facilitate identifying critical knowledge gaps in current WWTP influent disturbance models. Finally, the outcome of these discussions will be used to define specific tasks that should be tackled in the near future to achieve more general acceptance and use of WWTP influent generators.

Keywords: Disturbance generators, dynamics, flow, influents, pollution loads, uncertainty

INTRODUCTION

The use of activated sludge models (ASM) (**Henze *et al*., 2000**) is constantly growing and both industry and academia are increasingly applying these tools when performing wastewater treatment plant (WWTP) engineering studies. The level of detail and the specific data required for a modelling exercise strongly depend on the project objectives. In general, the more specific the results of the simulation study, the more detailed the required set of data (**Cierkens *et al*., 2012**). However, due to the high cost of measuring campaigns, many simulation studies of full-scale WWTPs suffer from a lack of sufficiently long and detailed time series for flow rates, temperature and nutrient/pollutant concentrations representing realistic wastewater influent dynamics. For this reason, model-based influent generators are an alternative that has recently gained considerable interest (**Gernaey *et al*., 2011**).

METHODS

Literature offers a wide range of tools generating influent characteristics by means of mathematical models. The paper will analyse in detail:

- The shift in methods to generate influent dynamics from the simpler (black box) to the more complex (grey/white box) models including a more detailed description of the phenomena taking place in the urban drainage system with more/less equations/model parameters (**Fig 1**). The type of approach will determine: 1) model parsimony (limiting the number of parameters); 2) model transparency (by using model parameters that have a physical meaning); and 3) model flexibility (easily extended to other applications) (**Gernaey *et al*., 2011**).

- The data availability to create/re-create the different influents. Here, different relevant questions are to be answered: "Do I have measurements and can I therefore apply statistical analysis to obtain longer time series? "Are measurements entirely missing and do I need a model that can provide realistic patterns without measurements?"

A more in depth analysis about methods will be complemented with the critical review carried out by **Martin and Vanrolleghem** (**2013**).

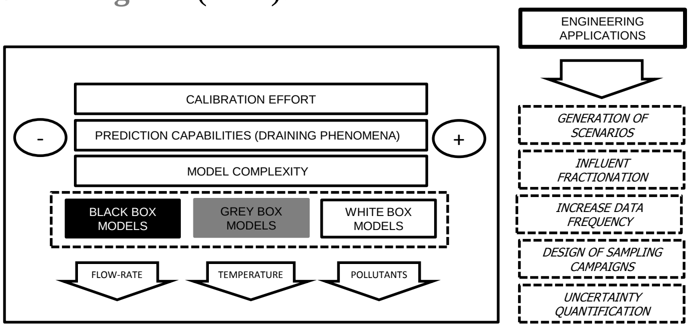

Figure 1. Methods, characteristics, modelled compounds and engineering applications of influent generators.

COMPOUNDS AND TEMPORAL RESOLUTION

Another point of discussion will focus on what kind of compounds (and their temporal resolution) can be described with the current models. For example:

- Generation/frequency of (dry weather) flow rate, temperature, traditional components (COD, TSS, TN and TP) and emerging components (pharmaceuticals, illicit drugs);
- Generation/frequency of wet weather flow, temperature and traditional/emerging components.

Most of the models used to describe traditional compounds are based on intensive measuring campaigns carried out during the 90s (**Butler *et al*., 1995**). In addition to the description of traditional pollutants some of these models can also describe emerging compounds. For example, **De Keyser *et al.*** (**2010**) developed a database summarizing different emission patterns for 26 priority pollutants (daily/weekly/seasonal /annual). **Lindblom *et al***. (**2006**) and **Snip *et al.*** (**2013**) upgraded the phenomenological influent model presented by **Gernaey *et al.*** (**2011**) including the behaviour of bisphenol A, pyrene and some pharmaceuticals (antibiotics, painkillers, mood stabilizers). **Ort *et al.*** (**2005**) developed a conceptual stochastic model to characterise short-term variations of benzotriazole concentrations (a chemical contained in dishwasher detergents), which can be easily adapted to any down-the-drain household chemical. Additional model complexity is necessary to describe the behaviour of all these elements in wet-weather conditions (**Gernaey *et al*., 2011**). Nevertheless some of the wet-weather generators are simplified and may not correctly represent the rainfall properties, the build-up/wash-off (pollution) and rainfall/run-off (water). Specifically, associated soil models currently do not include physico-chemical descriptions of moisture properties and some transport models are not capable to correctly describe the first-flush effect after a (heavy) storm event (**Martin and Vanrolleghem, 2013**). For these reasons the effect on flow rate, substances and temperature might be systematically under- or overestimated.

CALIBRATION AND VALIDATION

Another important point that will be discussed is related to several aspects that should be considered during calibration/validation of such influent generation models:

- In most black box models, parameter values are identified after processing long time series. However, these parameters are adjusted to fit the inputs and outputs and do not have any physical/biological/chemical meaning. On the other hand, grey and white box models are based on parameters that correspond to measurements or physical characteristics of the catchment.
- The traditional calibration procedure uses a trial and error process of parameter adjustments. Often, the goodness-of-fit of the calibrated model is basically a visual judgement comparing simulated and observed data. This process is subjective and can be quite long and tedious unless the process engineer has a good knowledge about the model behaviour (**Flores-Alsina *et al*., 2013**). Automatic calibration has the advantage that it can (in some cases) accelerate the process and be objective as it is based on quantitative goodness-of-fit criteria.
- Frequentist analysis has demonstrated to work quite well in identifiable systems (**Omlin and Reichert, 1999**). Nevertheless, when the models present: 1) some apparent identifiability problems (**Omlin and Reichert, 1999**); or, 2) some structural uncertainty in the model formulation (**Neumann and Gujer, 2008**), this approach is no longer valid and other approaches based on Bayesian statistics are recommended. However, the calibration effort increases substantially when using more elaborate methods (**Lindblom *et al*., 2011**; **Rieckermann *et al*., 2011**;**Talebizadeh *et al*., 2013**) (**Fig 1**).

ENGINEERING APPLICATIONS

The engineering applications of influent generators are various (**Fig 1**):

1. Increase data frequency: Sub-hour frequency of influent data is required when the model is used to test control strategies and wet-weather operation. Characterization of the influent implies a large effort and high costs when analysing samples for a series of pollutants. Recent developments in measurement technology have made sensors more reliable and cheap. Still, several standard lab analyses, such as COD, cannot be performed reliably in on-line mode in the influent of a WWTP (**Olsson *et al*., 2012**). In these situations, influent generators can certainly increase the frequency of influent data and provide additional dynamics not revealed by measurements (**Devisscher *et al*., 2006**; **Gernaey *et al*., 2011**; **Flores-Alsina *et al*., 2013**).

2. Design of sampling strategies: Grey / Black box influent generators can account for, amongst other factors, different types of dynamics, levels of occurrence and the effect of pumping strategies in the sewer when (mathematically) describing the occurrence of traditional/emerging pollutants. This feature can be extremely useful when designing sampling campaigns. **Ort *et al.*** (**2010**) demonstrated that errors of 50% or more are possible for 24-h composite samples when the compound is not sampled at a sufficiently high frequency.

3. Fractionation: Influent fractionators can easily be plugged in to the time series created by influent generators. The main idea is to correlate the model state variables used in the ASM models (**Henze *et al*., 2000**) with their analytical measurements. For example, **Grau *et al.*** (**2007**) and **Gernaey *et al.*** (**2011**) proposed two alternatives based on different principles. The first approach is based on an optimizer that finds suitable fractionation parameters according to the available data. The second approach uses (fixed) parameter values in order to convert for example COD_{sol} into non-biodegradable (S_I) and biodegradable (S_S) soluble substrates using the ASM1 (**Henze *et al*., 2000**).

4. Uncertainty/Sensitivity analysis of influent profiles: The use of probability distribution functions in some of the influent generator models combined with Monte Carlo simulations might help to quantify the range and/or uncertainty of simulated data (wastewater properties). These simulation outputs can be used to better design WWTPs using probabilistic concepts rather than safety factors (**Rousseau *et al*., 2001**; **Belia *et al*., 2009**; **Flores-Alsina *et al*., 2012; Talebizadeh *et al*., 2013**) or to test the robustness of control strategies (**Benedetti *et al*., 2006**; **Flores-Alsina *et al*., 2008**;).

5. Generation of scenarios: Dynamics and complexity of factors influencing wastewater systems make reliable predictions very difficult, i.e. the characteristics of the catchment area can change substantially over the years. For this reason, it is necessary to improve the planning and design of wastewater treatment infrastructures through methodologies that systematically account for uncertain futures (**Dominguez and Gujer, 2006**). The use of the presented tools can be very beneficial to answer "what-if" questions (**Gernaey *et al*., 2011**; **Flores-Alsina *et al*., 2013; Martin and Vanrolleghem, 2013**).

PURPOSE OF THE PAPER

The main objective of this presentation/paper is to demonstrate the advantages of influent generators (reduce the cost of measuring campaigns, fill data gaps, create additional scenarios) with several illustrative case studies. The second purpose is to identify critical knowledge gaps related to model development, calibration procedures and increasing the number of (wastewater) engineering applications. Comments received at the conference will be included in subsequent influent generator model upgrades (the authors are actively working on model development), thus addressing modellers' needs. This will finally achieve a more general acceptance and – equally important – common standards on model building and calibration of influent generators.

ACKNOWLEDGEMENTS

Dr. Flores-Alsina , Mr. Saagi and Miss Snip acknowledge the Marie Curie Program of the European Union's 7th Framework Programme FP7 2007-2013 under the REA agreements 329349 (PROTEUS) and 289193 (SANITAS). Peter Vanrolleghem holds the Canada Research Chair in Water Quality Modelling

REFERENCES

Belia E., Amerlinck Y., Benedetti L., Johnson B., Sin G., Vanrolleghem P.A., Gernaey K.V., Gillot S., Neumann M.B., Rieger L., Shaw A., Villez K. (2009). Wastewater treatment modelling: Dealing with uncertainties. Wat. Sci. Tech., 60(8), 1929-1941.

Benedetti L., Bixio D., Vanrolleghem P.A. (2006). Assessment of WWTP design and upgrade options: Balancing costs and risks of standards' exceedance. Wat. Sci. Tech., 54(6-7), 371-378.

Butler D., Friedler E., Gatt K. (1995). Characterising the quantity and quality of domestic wastewater inflows. Wat. Sci. Tech., 31(7), 13-24.

Cierkens K., Plano S., Benedetti L., Weijers S., de Jonge J., Nopens, I. (2012). Impact of influent data frequency and model structure on the quality of WWTP model calibration and uncertainty. Wat. Sci. Tech., 65(2), 233-242.

Devisscher M., Ciacci G., Fé L., Benedetti L., Bixio, D., Thoeye C., De Gueldre G., Marsili-Libelli S., Vanrolleghem, P.A. (2006). Estimating costs and benefits of advanced control for wastewater treatment plants – the MAgIC methodology. Wat. Sci. Tech., 53(4-5): 215-223.

De Keyser W., 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.

Dominguez D., Gujer W. (2006). Evolution of a wastewater treatment plant challenges traditional design concepts. Water Res., 40(7), 1389-1396.

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

Flores-Alsina X., Corominas L., Neumann M.B., Vanrolleghem P.A (2012). Assessing the use of activated sludge processs design guidelines in wastewater treatment plant projects. A methodology based on global sensitivity analysis. Environ. Modell. Softw., 38(12), 50-58.
Flores-Alsina X., Saagi R., Lindblom E., Thirsing C., Thornberg., Gernaey K.V., Jeppsson U. (2013). Calibration and validation of a phenomenological influent pollutant disturbance scenario generator using full-scale data. Water Res. Submitted.
Gernaey K.V., Flores-Alsina X., Rosen C., Benedetti L., Jeppsson U. (2011). Dynamic influent pollutant disturbance scenario generation using a phenomenological modelling approach. Environ. Modell. Softw., 26(11), 1255-1267.
Grau P., Beltrán S., de Gracia M., Ayesa E. (2007). New mathematical procedure for the automatic estimation of influent characteristics in WWTPs. Wat. Sci. Tech., 56(8), 95-106.
Henze M., Gujer W., Mino T., van Loosdrecht M.C.M. (2000). Activated Sludge Models ASM1, ASM2, ASM2d and ASM3. IWA Scientific and Technical Report No. 9, IWA Publishing, London, UK.
Lindblom E., Gernaey K.V., Henze M., Mikkelsen P.S. (2006). Integrated modelling of two xenobiotic organic compounds. Wat. Sci. Tech., 54(6-7), 213-221.
Lindblom E., Ahlman S., Mikkelsen P.S. (2011). Uncertainty-based calibration and prediction with a stormwater surface accumulation-washoff model based on coverage of sampled Zn, Cu, Pb and Cd field data. Water Res., 45(13), 3823-3835.
Martin C., Vanrolleghem P.A. (2013). Analyzing, completing and generating influent data for WWTP modelling: a critical review. Environ. Modell. Softw. Submitted
Neumann M., Gujer W. (2008). Underestimation of uncertainty in statistical regression of environmental models: Influence of model structure uncertainty. Env. Sci. Tech., 42(11), 4037-4043.
Olsson G. (2012). ICA and me: A subjective review. Water Res., 46(6), 1586-1624.
Omlin M., Reichert P. (1999). A comparison of techniques for the estimation of model prediction uncertainty. Ecol. Model., 115(1), 45-59.
Ort C., Lawrence M.G., Reungoat J., Mueller J.F. (2010). Sampling for PPCPs in sewers: Comparison of different sampling modes and optimization strategies. Env. Sci. Tech., 44(16), 6289-6296.
Ort C., Schaffner C., Giger W., Gujer W. (2005). Modeling stochastic load variations in sewer systems. Wat. Sci Tech., 52(5), 113-22.
Rieckermann J., Anta J., Scheidegger A., Ort C. (2011). Assessing wastewater micropollutant loads with approximate Bayesian computations. Env. Sci. Tech., 45(10), 4399-4406.
Rousseau D., Verdonck F., Moerman O., Carrette R., Thoeye C., Meirlaen J., Vanrolleghem P.A. (2001). Development of a risk assessment based technique for design/retrofitting of WWTPs. Wat. Sci. Tech., 43(7), 287-294.
Snip L., Flores-Alsina X., Plosz B., Jeppsson U., Gernaey K.V. (2013). Extending the BSM platform with occurrence, transport and fate of micropollutants. Environ. Modell. Softw. Submitted
Talebizadeh M., Belia L., Vanrolleghem P.A. (2014). Influent generator for probabilistic design of nutrient removal wastewater treatment plants. Submitted to 4th IWA/WEF Wastewater Treatment Modelling Seminar (WWTmod2014), Spa, Belgium, March 30 – April 2, 2014.