Quo Vadis Benchmarking Simulation Models?


¹ Div. of Industrial Electrical Engineering and Automation (IEA), Lund University, Box 118, SE-221 00 Lund, Sweden (Corresponding author, E-mail: {ulf.jeppsson,xavier.flores}@iea.lth.se)
² ifak e.V. Magdeburg, Werner-Heisenberg-Str. 1, D-39106 Magdeburg, Germany (E-mail: jens.alex@ifak.eu)
³ AWMC, Advanced Wastewater Management Centre, University of Queensland, Brisbane, QLD 4072 Australia (E-mail: damienb@awmc.uq.edu.au)
⁴ Waterways srl, Via del Ferrone 88, 50023 Impruneta (FI), Italy (E-mail: lorenzobenedetti@waterways.it)
⁵ Laboratory of Chemical and Environmental Engineering, University of Girona, Campus Montilivi s/n, 17071 Girona, Catalonia, Spain (E-mail: joaquim.comas@udg.edu)
⁶ Primodal, Inc., 122 Leland Street, Hamilton, Ontario L8S 3A4, Canada (E-mail: copp@primodal.com)
⁷ ICRA (Catalan Institute for Water Research), Scientific and Technological Park of the University of Girona, H2O Building, Emili Grahit 101, 17003 Girona, Spain (E-mail: {irodriguezroda,lcorominas}@icra.cat)
⁸ Dept. of Chemical and Biochemical Engineering, Technical University of Denmark, Building 229, DK-2800 Kgs. Lyngby, Denmark (E-mail: kvg@kt.dtu.dk)
⁹ BIOMATH, Ghent University, Coupure Links 653, B-9000 Ghent, Belgium (E-mail: ingmar.nopens@ugent.be)
¹⁰ LRGP-CNRS, Nancy University, INPL, 1 rue Grandville, BP 451, F-54001 Nancy cedex, France (E-mail: marie-noelle.pons@ensic.inpl-nancy.fr)
¹¹ Veolia Water - Solutions & Technology, AnoxKaldnes AB, Klosterängsvägen 11A, SE-226 47 Lund, Sweden (E-mail: christian.rosen@veoliawater.com)
¹² INRA, UR050, Laboratoire de Biotechnologie de l’Environnement, Avenue des Etangs, Narbonne, F-11100, France (E-mail: steyer@ensam.inra.fr)
¹³ modelEAU, Dépt de génie civil et de génie des eaux, Université Laval, Québec, G1V 0A6, QC, Canada (E-mail: peter.vanrolleghem@uci.ulaval.ca)
¹⁴ Dept. of Biosystems Engineering, Ghent University, Coupure Links 653, B-9000 Ghent, Belgium (E-mail: eveine.volcke@ugent.be)
¹⁵ Dept. of Systems and Control, Josef Stefan Institute, Jamova 39, SI-1000 Ljubljana, Slovenia (E-mail: darko.vrecko@ijs.si)

Abstract
As the work of the IWA Task Group on Benchmarking of Control Strategies for WWTPs is coming towards an end, it is essential to disseminate the knowledge gained. For this reason, all authors of the IWA Scientific and Technical Report on benchmarking have come together to provide their insights, highlighting areas where knowledge may still be deficient and where new opportunities are emerging, and to propose potential avenues for future development and application of the general benchmarking framework and its associated tools. The paper focuses on the topics of temporal and spatial extension, process modifications within the WWTP, the realism of models, control strategy extensions and the potential for new evaluation tools within the existing benchmark system.

Keywords: Benchmarking; BSM; control; evaluation; modelling; process monitoring; simulation; wastewater treatment

INTRODUCTION
Over the past decade, considerable investments have been made in acquiring knowledge of how to best perform objective benchmarking of control and monitoring strategies for wastewater treatment plants (WWTPs) and how to evaluate the results using a detailed simulation protocol. The success of the COST/IWA benchmark simulation models BSM1, BSM1_LT and BSM2 (e.g. Spanjers et al., 1998; Copp, 2002; Rosen et al., 2004; Jeppsson et al., 2007; Nopens et al., 2010; Corominas et
al., 2011; Gernaey et al., 2011a; http://www.benchmarkwwpt.org) for control strategy and monitoring system development and evaluation clearly illustrates the usefulness of such tools for the wastewater research community. More than 300 papers, conference presentations and theses on work related to the benchmark systems have been published to date. The freely available simulation models are used by numerous research groups around the world for various purposes and are available as predefined software tools in several commercial WWTP simulator packages (e.g. GPS-X™, SIMBA®®, WEST®) – as well as in a stand-alone FORTRAN implementation and for the general MATLAB®/SIMULINK® platform. Implementations with varying success have also been achieved in STOAT™, BioWin™, AQUASIM, JASS, SciLab and EFOR™.

Efforts have focussed on providing tools for analysing and solving real problems for real WWTPs and establishing a general platform and simulation protocol that can be further extended in the future. As the IWA Task Group on Benchmarking of Control Strategies prepares to publish the official Scientific and Technical Report (STR) in 2011 (Gernaey et al., 2011a), it is important to take advantage of the experience gained by the researchers that have been involved in the BSM development over the years. This paper has been written to define potential avenues for future work, as well as to suggest potential uses for the BSM platform and its associated tools. For this purpose, all authors of the BSM STR have come together in this paper to highlight areas where knowledge may still be deficient and where new opportunities are emerging for future BSM development and application.

Although valuable tools, the current BSM systems do not include all aspects of importance for benchmarking WWTP control and monitoring strategies. A number of potential pathways for extensions have been identified and are discussed in this paper. These include: (1) Temporal extension; (2) Spatial extension; (3) Process extensions within the WWTP; (4) Realism of the models used in the BSM; (5) Control strategies extension; and (6) Extended evaluation tools.

TEMPORAL EXTENSION

In BSM1, only 14-day influent data series for dry, rain and storm conditions were necessary. These data series were generated from a real data set combined with some hypotheses on correlated influent characteristics (Spanjers et al., 1998). This was sufficient for BSM1, but insufficient for long-term simulations as desired for BSM1_LT and BSM2 (1-2 years). To deal with longer term simulations, a phenomenological influent wastewater generator model was developed (Gernaey et al., 2011b) to provide realistic influent data to the BSMs. Should the objective of future BSMs become more ambitious, it might require even longer influent data files to perform complete scenario analyses. For example, one might want to include the impact that climate change will have on the precipitation regimes (e.g. extreme rainfall events or drought periods), air/water temperature or snow melting periods (Semadeni-Davies, 2004; Semadeni-Davies et al., 2008; Plósz et al., 2009). These factors will strongly impact the quantity and quality of the influent wastewater as well as the way future WWTPs might be operated. Another possible example requiring temporal extension could be changes in the urban catchments. This might include changing from combined to separated systems or even to source treatment, separate storm water treatment, use of rainwater for non-potable water use, longer sewer networks or the appearance/increase/decrease of new pollutants (Ashley et al., 2007). All such changes take place over a long time scale and the evaluation period may therefore need to be extended to ten or even fifty years to include the capability for this sort of evaluation. Despite ever-increasing computational power, it does not seem reasonable to extend the evaluation period to ten or fifty years to include all the detailed effects when performing simulation-based scenario analysis, unless parts of the simulation model can be speeded up (e.g. Ráduly et al., 2007). Other long-term phenomena that should be considered, but that are not fitting into the input
file, are the fouling of the aeration system with consequent efficiency decrease, and events which may occur not every year but with high impact like maintenance line closures and equipment failures (Rosso and Stenstrom, 2006). Such modifications would require the consideration of some practical aspects including simulation speed and model accuracy. Development in this direction can be found in Benedetti (2006).

**SPATIAL EXTENSION**

The family of benchmark systems are defined as ‘within-the-fence’ systems, *i.e.* the model descriptions and simulations do not extend outside the borders of the WWTP. The importance of the sewer system and processes in the receiving waters was recognised by the Task Group but including these complicating factors in the original BSMs was deemed to be beyond the Task Group’s scope.

- *Sewer network*

  From a control and monitoring perspective the inclusion of the sewer network into the benchmark system would open up a range of new possibilities for interactions and manipulation of the combined sewer/WWTP system (*e.g.* back-flow effects, storm tanks and pumping stations, combined sewer overflows, pollution contributions from run-off). For example, the KOSIM sewer model (ITWH, 2000) which was designed to calculate pollutant loads to the WWTP and the receiving waters in the context of planning and dimensioning of sewer systems and storage tanks has already been used in an integrated context (Solvi *et al*., 2006). The KOSIM model does not include biochemical reactions and transformations, but several ASM like models are available to describe the chemical and microbial transformations of organic matter, nitrogen and sulfur within the sewer (*e.g.* Hvitved-Jacobsen *et al*., 1998; Sharma *et al*., 2008, Jiang *et al*., 2010).

- *Receiving waters*

  As for receiving waters, existing models such as the River Water Quality Model No. 1 (Reichert *et al*., 2001) or simplifications thereof (Schütze *et al*., 2011), can be added or linked to the BSMs without too much difficulty, given the proper interfaces (Benedetti *et al*., 2007). This kind of approach would be particularly beneficial for more detailed evaluation of the environmental impact of wastewater pollutants. As well, this combination would promote the use of the benchmark system as a decision support tool in agreement with current river basin management approaches, as pursued by the EU Water Framework Directive (*i.e.* immission-based rather than emission-based). Integrated evaluation experience can already be found in Benedetti *et al*., (2010) and Brehmer *et al*., (2009). As with all other benchmarking tools developed so far, consensus will have to be reached on objective evaluation criteria that assess the urban water quality impacts in receiving waters, but ideas for this are not lacking (*e.g.* Bauwens *et al*., 1996; Benedetti *et al*., 2010).

**PROCESS EXTENSIONS WITHIN THE WWTP**

The original purpose of the benchmark system was to allow for the objective comparison of control and monitoring strategies of a treatment plant removing organic carbon and nitrogen, and therefore a fixed plant layout was defined and used. In many cases, however, users have experienced a need to modify the layout (plant configuration) or have added additional treatment process models, thereby creating a WWTP more suited for their specific application, *e.g.* to benchmark potential plant upgrades (see for example the EU-CD4WC project, Benedetti *et al*., 2010) while maintaining the original set of benchmark performance evaluation criteria. Specifically for nitrogen removal, process models of an oxidation ditch plant configuration (Abusam *et al*., 2002), the combined SHARON-Anammox process (Dapena-Mora *et al*., 2004), membrane bioreactors (Maere *et al*., 2011) and many more have been added to the BSM platform. Extensions of the plant configuration towards bio-P removal have been reported as well (Gernaey and Jørgensen, 2004).
In many cases these add-ons have been defined and implemented without any insight from the Task Group and have remained the property of the individual research groups. This contrasts the Task Group philosophy, which has always aimed to freely distribute verified implementations of the benchmark plants. The global research community would certainly benefit if those additional models could be collected, standardised, verified and then made generally available as an extended BSM model library. One option for a formalised model library, at least for ASM type models, was suggested in Alex et al. (2005) using an XML format description. Indeed, for many potential benchmark users, the amount of work involved in developing their own process extensions is an important factor when considering whether or not to use one of the existing benchmark plants to a specific situation or plant.

It is clear from the above that process extensions within the WWTP are related to the appearance of models for new unit processes such as the SHARON and the ANaerobic AMMonium Oxidation (Anammox) processes (Volcke et al., 2006a). We expect that this evolution will continue in the future. At this moment, additional process extensions related to integrated fixed-film processes (Vanhooren et al., 2002) and SBR configurations are needed to address current requirements. One issue that is often forgotten is that process extensions coincide with the need for suitable model interfaces when the state variables in the model of a new unit process are different from the state variables in the original benchmark models. The function of these interfaces is to ensure that material mass balances and continuity principles are met, and ensures the proper mapping of the output variables of one model to the most appropriate input variables of another model (Alex et al., 2005; Vanrolleghem et al., 2005b; Volcke et al., 2006b; Nopens et al., 2009).

REALISM OF THE MODELS USED IN THE BSMs

The process extensions outlined above rely on the availability of models for new unit processes. However, it is also recognised by the authors that the models currently used in the BSMs might undergo changes in the future. Indeed, the mathematical models used in the BSMs today were chosen because they were internationally accepted and well-established, such as ASM1 (Henze et al., 1987), a 10-layer one-dimensional settler model (Takács et al., 1991) and ADM1 (Batstone et al., 2002). There is, however, an almost unlimited possibility to extend and upgrade the models within the existing BSM plant configurations, including the models describing sensors and actuators. Obviously the aim of any changes would be to enhance realism of the systems rather than to simply increase the level of detail and complexity. In some cases, improved models have become available since development of BSM1 and BSM2 was initiated. Although the Task Group decided not to change models during development, most of these updated models are well described, and thus can be (easily) interchanged in the current BSM framework. Furthermore, the BSM framework can be used to test models under development. An overview of possible model extensions and future inclusions are listed below.

- **ASM2d (phosphorus removal)**
  The advantage of including ASM2d is that phosphorus (P) removal (both biochemically and chemically) is added to the BSM framework. This would allow for the inclusion of P-limits to the Effluent Quality Index and chemical dosing for P-precipitation to the Operational Cost Index (see later section for discussion of EQI and OCI). These inclusions would then force P-related ‘costs’ to be accounted for when developing general control strategies or allow for the benefits to be quantified if specific strategies for P-removal were being investigated. Given the fact that low effluent P-limits are common, this extension would be timely. It should, however, be noted that the addition of P would require also an update to ADM1 (see below) for P-related components and processes (Harding et al., 2011). Furthermore, ASM2d does not come with realistic default parameters (Henze et al., 2000; Hauduc et al., 2011). Last but not least, the fact that decay rates are
not electron acceptor dependent in ASM2d might lead to an overestimation of the decay rate in the system, especially when the anaerobic volume is considerable. ASM2d can be modified to obtain electron acceptor dependent decay rates (Gernaey and Jørgensen, 2004; Benedetti et al., 2010; Flores-Alsina et al., 2011b), an ASM2d modification that has been successfully used on full-scale plants as well (Ingildsen et al., 2006).

- **ASM3**

ASM3 introduces the concept of storage and applies the endogenous respiration concept to describe the decrease of biomass and storage products over time. Some argue that this is a better approach in certain instances than the approach used in ASM1. However again, a drawback is the high number of parameters for which no real default set is defined (Henze et al., 2000; Hauduc et al., 2011).

- **Multi-step nitrification/denitrification and N₂O production**

The current BSM relies on single-step nitrification and single-step denitrification, which is based on the assumption that nitrite does not accumulate in typical WWTPs. However, nitrite is known to accumulate during unstable operation, at high temperatures, within side-stream processes and in industrial WWTPs (Sin et al., 2008). Low oxygen levels applied to WWTPs in view of energy savings also increase the chance of nitrite accumulation. The need to include nitrite in the future will result from the need to better estimate the exact effluent nitrogen load. Also novel nitrogen removal principles based on nitritation and nitritation-denitritation processes will require a substantially increased level of detail for modelling the transformation of nitrogen components (Gustafsson, 2011). In addition, N₂O production in WWTPs currently receives considerable attention because of its greenhouse gas potential (e.g. Flores-Alsina et al., 2011a; Porro et al., 2011). Kampschreur et al. (2009) found that both the nitrification and denitrification stages contribute to the production of N₂O under certain specific conditions, such as low oxygen concentrations, increased nitrite concentrations and low COD/N ratios in the denitrification stage. If the BSM performance criteria are extended to greenhouse gasses then N₂O will have to be modelled (Flores-Alsina et al., 2011a). Inclusion of multi-step nitrification/denitrification models has been implemented in a benchmark framework (Porro et al., 2011). Studying this in more detail is one of the tasks defined in an IWA Task Group focusing on “The use of water quality and process models for minimizing wastewater utility greenhouse gas footprints” (see http://www.iwataskgroupghg.com/).

- **Sulfur reducing/oxidizing reactions**

Sulfate is a key electron sink in anaerobic systems. Sulfate will reduce to sulfide under anaerobic conditions and will progressively re-oxidise to poly-sulfide, sulfur and sulfate under aerobic conditions. Its direct impacts are numerous and include a reduction in methane flow (due to the loss of electrons), inhibition of anaerobic microbes and sulfate contamination of the gas phase (Fedorovich et al., 2003). While levels in domestic sewage are very low, generally resulting in gas phase sulphide below 1000 ppm, this is enough to decrease the value of the gas produced and will be included in the models eventually. Sulfate also has a subtle impact on the phosphorus system. In the presence of sulfide, iron phosphate resolubilises in anaerobic digesters (unpublished results). Making the task of incorporating sulfur into the BSM structure easier is the fact that there are various published models available to describe sulfate reduction (e.g. Fedorovich et al., 2003, Poinapen and Ekama, 2010).

- **Thickener and dewatering models**

To date, ideal models have been used for these units in the BSMs. The implemented models are based on steady-state mass balances for specific thickening/dewatering efficiencies and given amounts of suspended solids in the sludge streams (Jeppsson et al., 2007). However, because these processes are dynamic in reality and can impact the sludge balance of the system, they can impact the biodegradation rates in the activated sludge reactors, the gas production and composition from the anaerobic digester as well as the amount of sludge stored in the secondary clarifier. Such BSM
outputs are embedded in the evaluation criteria and can potentially impact development of control strategies. More complete models could be developed and included.

- **Settler model**
  It has been shown consistently that the 10-layer Takács model does not perform well during high flow rate hydraulic events (Jeppsson and Diehl, 1996). Alternative models exist and could easily be integrated (Plósz et al., 2011; Bürger et al., 2011).

- **Reactive settler processes**
  The current BSMs have biologically inactive primary and secondary settler models included. In the case of a long residence time in the primary settler, a portion of the influent hydrolysis will not be accounted for, which could impact the behaviour of the activated sludge reactors and, hence, all controllers developed based on that. Reactive primary and secondary settler models have been proposed before and linked to ASM-models (Gernaey et al., 2001, 2006; Flores-Alsina et al., 2011b). An important reaction in the secondary settler when residence times increase is the occurrence of denitrification. This phenomenon has important implications in P-removal systems since the quantity of returning nitrates to the anaerobic section via external recirculation might be overestimated (Flores-Alsina et al., 2011b). Secondary settler denitrification may also hamper the settling process. This was partially anticipated by defining a risk index for settling in the BSMs (Comas et al., 2008). However, a general problem with ASM1-based reactive settler models is the overestimation of the importance of decay, since decay simply continues in the settler no matter the electron acceptor present (Gernaey et al., 2006; Flores-Alsina et al., 2011b).

- **Time-varying parameters regarding biodegradation and settling**
  To date, BSM models have used default parameter sets, which were taken from the original references. However, in practice, system behaviour can be different and these parameter values might be varying as a function of time. This is quite commonly known for settling but models describing the direct link between sludge settleability and settling behaviour are not yet available. Work on a fuzzy rule-based system to infer the risk of settling problems is ongoing (Comas et al., 2008). Simulation of poor or good settling characteristics can be established by modifying the model parameters on-line based on the estimated risk of settling problems (Flores-Alsina et al., 2009). With respect to biodegradation, current ASM models contain temperature and pH corrections for process kinetics. One could easily imagine other impacts to include, e.g. inhibition. This can be accomplished by modifying the typical Monod expressions by Haldane expressions (Rosen et al., 2008a).

- **Degradation processes for micro-pollutants**
  New regulations stress the importance of estimating the loads and the fate of micro-pollutants (MPs) in the water cycle. The models available to simulate the transport and removal of ‘traditional’ pollutants, such as organic matter, nutrients and suspended solids in WWTPs can be extended with processes describing the fate of MPs (i.e. physical, chemical and biological processes) (e.g. Lindblom et al., 2006; Schönertke et al., 2009; Barret et al., 2010; Plósz et al., 2010). The fate models may be used to determine the distribution of the regulated MPs between solid, liquid and gas phases, so that monitoring for the contaminants can be done more efficiently (De Keyser et al., 2010). Should regulatory limits be imposed with maximum or “never to exceed” effluent concentrations or loadings, dynamic modelling is required to determine under what set of operating conditions compliance with the limits would be maintained (or alternatively under which conditions effluent limits might be exceeded, and for how long).

- **Physicochemical processes**
  Physicochemical processes occurring in wastewater treatment indirectly affect the biological conversions taking place. The descriptions in popular models, such as the ASM1 and ADM1, are limited to essential elements only, so pH for example is not described using the existing models. The ASM1 uses a global alkalinity state, but alkalinity does not properly consider the continuum that exists when there are both weak and strong acids present. The ADM1, however, includes pH
calculation. The fact that pH is described in some but not all BSM sub-models requires particular
attention during model coupling (Volcke et al., 2006b; Nopens et al., 2009) and provides a strong
argument to develop a common physicochemical model across the whole system (Grau et al.,
2007). The effect of non-ideal behaviour (i.e. activity, ion pairing etc.) needs to be included for
concentrated wastewater streams. Precipitation is also highly important and currently not included
in the BSMs. ASM2d considers empirical relationships for precipitation or re-dissolution of metal
phosphate complexes. ADM1 does not consider metal precipitation, although a potential approach
is provided by Batstone et al. (2002). Jones et al. (2007) have presented a physicochemical model
that is valid for the whole WWTP. Recently, an IWA Task Group working on physico-chemical
aspects in biological (waste)water treatment modelling has been established. This group (see
www.iwahq.org/Home/networks/task_groups/task_group_on_physicochemical_framework) is
addressing various physicochemical aspects besides the abovementioned acid-base and precipitation
reactions (Batstone et al., 2010) and working towards a generalized physicochemical model widely
applicable to water and wastewater. Once established, such a generalised physicochemical model
can then also be applied to the BSM platform.

Work has been initiated for about half of the above topics during the last few years, using the BSM
platform. The added realism to simulation results through such model extensions will also promote
the use of the benchmarking framework for more practical applications, on condition that the new
models are properly verified, calibrated and validated. Availability of a considerable number of new
or extended models will potentially mean that the distribution of the BSMs will have to be
reconsidered, for example by making a general library of ring-tested unit process models available
instead of models of pre-defined full-plant configurations.

**CONTROL STRATEGIES EXTENSION**

The advances in instrumentation and automation allow us to have access to information regarding
the urban wastewater system (UWWS) in real-time and at high levels of accuracy. This information
can be acquired not only from the WWTP but also on-line sensors can be installed in the sewer
systems and monitoring stations are being developed for monitoring river water quality. In this
sense, large quantities of data are now becoming available and this data can be used for fault-
tolerant, uncertainty-aware and system-wide control design.

- **Fault-tolerant control**
  The use of on-line sensors in control and automation for optimised operation of WWTPs is common
  practice. However, it is necessary to use methods to check the quality of the signals provided by
  these sensors as they are subject to failures (drift, shift, calibration etc.) (Rieger et al., 2003; Rosen
  et al., 2008b). Poor signal quality can lead to undesired control actions causing severe effluent limit
  violations or increase of operating costs. In view of control implementation in full scale it is
  necessary to develop new tools and new strategies to increase the reliability. Therefore, fault
detection methods and fault diagnosis (to identify the root cause of the fault) should be coupled to
  control methods to assure fault tolerant control (Alcaraz-González et al., 2005; Mhaskar, 2006;
  Zumofen and Basualdo, 2008; Corominas et al., 2011). The benchmark platform can be used to
demonstrate the validity of such control systems and allow for the evaluation of the effects of faults
in active controllers and the resulting overall plant performance.

- **Uncertainty-aware control**
  Proper control of wastewater treatment systems strongly depends on reliable input information,
  which is usually obtained from fast and simple measurements (e.g. DO, nitrates, ammonium, biogas
  flow rate, biogas composition, pH) or estimated from mathematical structures called observers.
  Several identification techniques can be used for the design of state observers. Kalman Filtering
  represents a rigorous and powerful methodology that has been applied for on-line state estimation in
WWTPs (e.g. Beck, 1981). Nevertheless, there are other observer and estimation techniques that can be beneficial for wastewater treatment (e.g. Alcarez-Gonzalez et al., 2002; Lardon et al., 2004). Hence, in a similar way as is done for monitoring methods, the BSM platform could be used to evaluate and compare different types of estimation methods. 

- *System-wide control*

System-wide control can be used to manage the UWWS as one integrated unit. The conventional approach is to design and operate each system component separately, e.g. sewer system, storm tanks, WWTP, receiving water bodies. However, the optimal performance of the UWWS cannot be realised using such an approach. This concept was described already in Beck (1976) and can now be applied because of the improvement of monitoring systems, mathematical models and increased computational power. Butler and Schütze (2005) demonstrated that the application of conventional criteria (e.g. overflow volumes, discharged pollutant loads) can result in misleading conclusions when assessing the performance of the UWWS under various scenarios and therefore, immission-based approaches are required (Benedetti, 2006). However, system-wide control is a difficult task because of the interactions between the different elements of the system (Rauch et al., 2002) and requires the development of new tools that provide information about scenario analyses and operational procedures, which improve the performance of the overall UWWS. Efforts towards this objective have already been initiated by the sewer system research community: realistic virtual sewer systems are created (now even automatically, Sitzenfrei et al., 2010) and control strategies are objectively compared within a benchmarking context (e.g. Borsanyi et al., 2008). From an integrated perspective, control actions taken within a combined sewer system can reduce combined sewer overflows (CSOs) and increase the load to the WWTP, at the expense perhaps, of the WWTP’s performance (Bauwens et al., 1996; Rauch and Harremoës, 1999; Schütze et al., 1999; Vanrolleghem et al., 2005a; Benedetti et al., 2009). This is still not a fully accepted approach and lacks widespread application in practice. Its future use may increase if the approach was incorporated into the existing BSM initiative and more widely promoted in the associated research community. The available BSM methodology may be adapted for combined sewer system control and system-wide control approaches could be investigated by integrating the sewer system, WWTP and the receiving waters into one extended benchmark system.

**EXTENDED EVALUATION TOOLS**

The basic premise on which benchmarking is based are the metrics used in the evaluation phase. The availability and reliability of the evaluation tools to effectively ‘score’ the process under study is essential for the success of any benchmark system. It matters not whether the benchmarking involves delivery of government services or activated sludge systems, the evaluation criteria (the metrics) must efficiently simplify a complex comparison into a few meaningful index values that capture the relative strengths and weaknesses of the items being compared. The approach adopted during the BSM development was to develop criteria that were independent of location so that the BSM application was not constrained by jurisdiction. It can of course always be questioned if we succeeded in the latter. Use of the current BSM criteria and recent advances in research knowledge have highlighted some deficiencies in the current evaluation criteria. Nevertheless, the vision for extending the evaluation criteria follows this basic approach; namely, that the criteria must be as much as possible independent of jurisdiction.

The current BSM platform is based on three main types of evaluation criteria (effluent quality, operational cost issues and risk). Effluent quality is considered through an Effluent Quality Index (\(EQI\)), which is defined to quantify into a single term the effluent pollution load to a receiving water body. This combined with an effluent violation metric gives a reasonable overview of the ability of the benchmarked system to meet a particular effluent requirement whatever that might be. Energy ‘costs’ are considered through pumping, mixing and aeration energy calculations. Sludge ‘costs’ are
considered through sludge production and disposal calculations and costs related to chemical additions are also included (external carbon source). Together these ‘costs’ form an Operational Cost Index (OCI). Finally, process risk is considered through a fuzzy logic calculation of microbiology-related operational problems to create a Risk Index. Each of these will remain a part of the protocol, although further validation and extension of the cost and risk indices need to be performed.

- **Energy consumption models**
  In particular, energy consumption and production (and the associated costs) should be modelled in more detail. Energy efficiency (and even self-sufficiency) is an important decision driver in a modern day WWTP (Siegrist *et al.*, 2008). Strategies that show improved process efficiency will be considered at full-scale, but only if the model used to calculate those efficiencies is sufficiently accurate. For example, new digestion processes which show promise in solids destruction and energy production are coming on-line at full-scale. As the BSM framework will be expanded to new plant layouts, such as SBRs, oxidation ditches or membrane bioreactors, the evaluation criteria will have to be adapted to these new configurations. To keep pace, the BSM evaluation criteria will have to be expanded and validated to ensure that these criteria accurately predict the relative change in energy consumption and production.

- **Microbiology-related TSS separation problems**
  Process risk is defined within the BSM context as the risk of failure due to settling problems of microbiological origin, such as the proliferation of foaming and bulking organisms. The risk index is calculated based on a fuzzy logic approach using model-based process variables (Comas *et al.*, 2008). This index is a promising development with the potential to predict conditions that could lead to unfavourable settling, but because the calculation is a significant simplification of a complex biological process further validation of the model is required and this could lead to refinement of the approach. Full-scale validation of the predictions will add credibility to the approach and this validation could potentially also provide valuable information about the role of things like channel design, weir design and solids loading on filament proliferation. This risk index approach has the potential to answer questions such as ‘Why does one plant bulk and why does another not bulk?’ or at the very least identify the issues that lead to settling problems. The most urgent thing to do now is that the approach needs to be validated and extended to include risks related to other potential problems and processes within the WWTP (Dalmau Solé, 2009).

- **Capital and maintenance costs**
  To date, the issue of capital costs has been neglected in the BSM platform. Work in this area is needed because capital expenditures play such a large role in the decision making process (Gillot *et al.*, 1999). Simply because a process is more efficient or less susceptible to process upsets does not mean it is the proper course of action if the capital costs to implement the change are prohibitive. A similar argument can be made with respect to operational control strategies if the capital cost to implement the control is more heavily weighted than the benefit being achieved by the control. As currently defined, the BSMs would indicate the benefits from the control, but not consider the capital cost of implementation nor the maintenance cost or its effect on personal cost. A capital and maintenance cost index is required, but care will have to be taken to implement it in a jurisdictionally neutral way.

- **Uncertainty-based evaluation**
  In addition to the above traditional metrics, there has been a significant recent research effort in the field of model uncertainty. Uncertainty is a central concept when dealing with biological systems like activated sludge because they are inherently subject to large natural variations (Belia *et al.*, 2009). Traditionally, WWTP process simulators assume constant rather than variable model parameters, and are thus not capable of accounting for the inherent randomness. Even though some of the processes taking place in the UWWS are well known, most of the model parameters are
uncertain. Examples of uncertain parameters include the parameters describing the influent COD fractionation, or the parameters describing the effect of temperature or toxic compounds on the kinetics, and all of these have a significant influence on the model predictions. The assessment and presentation of uncertainty is recognized as an important part of the analysis of control strategies for wastewater systems (Beck et al., 1987). The variability and uncertainty in the model results might be captured in a ‘robustness’ index (Vanrolleghem and Gillot, 2002). That is, because all mathematical models in use represent simplifications of the treatment processes, it is often of interest to know how reliable or robust the predictions are. ‘Will small changes in the inputs, parameters and even model structure (ASM1-ASM3, different settler models) result in significantly different results?’ (Belia et al., 2009). If control is involved, will the operational strategy be able to deal with events appropriately if the models on which it is based are incorrect? In essence, how robust are the model predictions given the unknowns in the model. Consideration of uncertainty during the evaluation of control strategies makes it possible to answer questions such as ‘What would happen if there is a change in the influent composition?’ and ‘What are the expected effects of either temperature changes or toxic spills and will the controller handle them appropriately?’. Considering uncertainty when evaluating control strategy performance comes with the advantage that it gives an indication on the robustness of a proposed control strategy, i.e. it will become clear whether a proposed alternative is valid for a narrow range of conditions only or performs well for a broad range of situations (Flores-Alsina et al., 2008).

- **Sustainability**

Sustainability is a more complicated problem as any calculations depend on where the system borders are drawn. Nevertheless, carbon footprinting and greenhouse gas emission modelling is coming and should, therefore, be incorporated as an additional dimension during the evaluation procedure (e.g. Flores-Alsina et al., 2011a). Some work has been dedicated to include the time factor into life cycle assessment (Collet et al., 2011) and could be applied. Research into nitrous oxide emissions from treatment plants is also ongoing and although the emissions represent a small mass of nitrogen, these emissions have a significant global warming potential. Similarly, methane and carbon dioxide emissions are tied to the relative sustainability of a given process. Assuming that accurate models of these emissions can be developed, then the emissions should be captured by a kind of ‘sustainability’ index that might well include some of the energy cost calculations currently implemented. Exergy, which can encompass information related to energy quality and to resource availability (such as chemicals), could also form an important contribution to such an index (Belhani et al., 2008).

- **Geographically dependent regulations**

Although the evaluation criteria are meant to be geographically independent, the structure of the indices have always allowed for location specific criteria to be defined in subsequent analyses. For example, emphasis or weighting terms can be placed on specific performance items depending on location specific criteria. To formalise this analysis, the inclusion of a legislative module should be considered. Such a tool would allow users to specify regional or national requirements, which in turn may greatly influence what plant configuration and control strategy is the most appropriate for a specific case (e.g. if effluent quality demands are based on 2-hour grab samples or yearly averages, electricity tariffs and costs, sources of energy production available). Clearly such a module would make it impossible to perform objective comparisons of results on a global scale and should therefore not be used if the purpose of the research is general benchmarking. However, its availability would most certainly enhance the use of the benchmark platform to investigate options, solve problems and potentially enhance the overall performance of real wastewater treatment plants by allowing for stakeholders in charge of the services to include their local requirements and demands.
CONCLUSIONS
The BSM systems serve as a very useful and freely available software platform and simulation protocol for research groups all over the world. Whether used for their initially intended purpose of objective benchmarking of control strategies and monitoring methods or as a starting point for other types of investigations is of minor importance. As the IWA Task Group is coming to an end, it is the group’s obligation and responsibility to promote potential avenues for future development. A significant number of possible extensions and improvements have been defined in this paper. It is the sincere hope of the Task Group that this will inspire other research groups to continue the development of the BSM systems, thereby allowing it to flourish and remain a state-of-the-art tool for research, development and practical application within the fascinating field of wastewater treatment.

ACKNOWLEDGEMENTS
A large number of excellent researchers and close friends have contributed to the development of the benchmark systems during the last 15 years and the authors wish to express their sincere gratitude to all of them. Lluís Corominas received the ‘Juan de la Cierva’ scholarship from the Ministry of Spain. Peter Vanrolleghem holds the Canada Research Chair on Water Quality Modelling. Eveline Volcke is a post-doctoral research fellow of the Research Foundation - Flanders (Belgium) (FWO). The support of the International Water Association (IWA) is gratefully acknowledged.

REFERENCES
**Directive context.** PhD thesis, Ghent University, Belgium.


