FORMALISED MODEL REPRESENTATION FOR WASTEWATER SYSTEMS

J. Alex, M. Ogurek and U. Jumar

ifak Magdeburg, Steinfeldstr. 3, D-39179 Barleben, Germany, Tel.: +49 39203 8100, Fax: +49 39203 81100, Email: jens.alex@ifak-md.de

Abstract: A fundamental pre-condition for the application of simulation is the availability of reliable models. Particular problems arise in the application of simulation for integrated process design of complex wastewater systems. The models necessary for this task have to include different sub-processes such as biological wastewater treatment, anaerobic digestion processes and also models of the control system. To deal with this problem this paper describes an approach on how (1) to follow strict mass conservation principles during model development and (2) how to use formalised model representation methods for model creation, exchange and analysis. The adaptation of the XML-based Systems Biology Mark-Up Language (SBML) to the area of wastewater modelling is proposed. The implementation of this approach in a simulation environment for practical control engineering tasks is demonstrated.

Keywords: Modelling, Formal languages, Simulation, Environmental engineering, Integrated plant control

1 INTRODUCTION

Modelling and simulation are standard methods applied in research and engineering. For research applications, it can be stated that models are used to gain new knowledge (modelling for development of hypotheses, verification using independent data) as well as to represent the available knowledge about natural and technical processes. For the translation of process knowledge from research to the practice of planning and design, models frequently play a major role. This paper demonstrates that models and simulation of wastewater treatment have matured to application in practice. However, it has to be ensured that reliable process models, comfortable simulation systems, useful guidelines and an established user community are available. Simulation in engineering practice becomes useful, in particular in cases where the interactions of the operational units involved are too complicated in order to be considered and acted upon manually. However, in these cases simulation has to deal with challenging tasks. For the description of the unit processes it is necessary to use compatible and consistent models. For water flows from one sub-model to another, appropriate interface models are required.

2 FORMALISED MODEL REPRESENTATION

2.1 Motivation and advantage of application

For the considered wastewater systems many models have been published as technical reports by the International Water Association (IWA). Examples are the Activated Sludge Models ASM1, ASM2d, ASM3 (Henze, et al., 2000) to describe bio-chemical processes of wastewater treatment. Other examples include the recently published models on processes in rivers (RWQM1, Reichert, et al., 2001) and the model for anaerobic treatment processes (ADM1, Batstone, et al., 2002). As a general trend, it can be observed that the complexity of the proposed models is increasing permanently.

For almost any application, it is necessary to implement the model in a simulation environment. The simulation environments used range from general simulation systems over open, specialised simulators to closed, specialised simulators. In any case, someone has to read the published paper describing the model and to transform it into the specific model representation language of the simulation environment. There is no big risk in stating that it is simply impossible to implement the model based solely on this information without errors. Some reasons for this situation are:

- Simplification during the translation from the original implementation
- Typing errors (e.g. copy & paste errors)
- Incomplete description (intended or by fault)
- Scattered description (different sub-parts of the model in different parts of the paper)
- misinterpretation of the written description
- errors during the transformation of the written formulas into the final modelling language (most likely).

In fact, a verified implementation can be achieved only if the following auxiliary measures are applied:

- Use of (correct) reference simulation results (paper, data, reference implementation)
- Verification of the stoichiometry by check of valid balances
- Ring tests.

But even if additional information in one of the above mentioned forms is available, implementation and verification of a new model into a simulation system still constitute a time-consuming task.

Most of the problems listed above can be solved by use of a formalised model description language. In order to find a possible candidate, several existing "standard" modelling languages have been analysed. The most developed formal modelling languages can be found in the field of general purpose simulation systems for continuous systems. As some water-related simulation systems are based on general purpose simulation systems, e.g. GPSX (http://www.hydromantis.com) on ACSL or SIMBA (http://simba.ifak-md.de) on Matlab/Simulink, one could consider to use one of these proprietary modelling languages as a standard notation. But even open modelling languages such as Modelica (http://www.modelica.org/) provide only a level of abstraction which is not significantly higher than the specification of a set of algebraic and differential equations for the description of dynamic systems. Thus, this kind of modelling language is not well suited to describe the additional model information which is given in a Petersen Matrix (Petersen, 1965), which is used in most of the model publications, such as the specification of component concentrations, the specification of processes, the separation of stoichiometry and kinetics.

Based on this situation, a formalised model notation using XML was developed (FOX format in SIMBA, ifak, 2003). Later investigations showed that this model notation was more or less identical to an initiative in the area of systems biology called SBML (Hucka, 2003, www.sbml.org). As this ongoing initiative already reached a wide acceptance, it was decided to stop the work on the own formalised format and to adopt the model description format SBML also in the field of water and wastewater. The proposed format provides all properties necessary to serve as a standard for model representation and publication of bio-chemical conversion processes in the field of urban drainage, wastewater treatment and receiving water quality modelling. The vision of this proposal is to promote the use of SBML to

- set a standard for publishing bio-chemical conversion models in the field of water
- to initiate the etiquette that a model is considered as published only if a formalised description is provided
- to establish a public model library on bio-chemical conversion models in the field of water
- to allow the export and import of conversion models using the proposed format for all relevant simulation systems (if permitted).

Similar to the application of models for research purposes, also engineering practice will benefit from such a standard. By establishing a standardised formalised model notation, it becomes much easier to implement new "standard" models on the used systems and to verify their correct implementation.

2.2 Adoption of SBML to Wastewater Systems

The Systems Biology Markup Language SBML is used to exchange models between different simulation platforms in the area of systems biology. SBML defines a computer-readable format for representing models of biochemical reaction networks. An important feature of SBML is, that it is based on the Extensible Markup Language (XML, http://www.w3.org/XML/). A major part of the required functionality of the model notation language is already available in XML. For the presentation of stoichiometric and kinetic expressions, the notation of mathematical expressions is required. A standard for this purpose already exists. This standard is called Mathematical Markup Language (MathML, http://www.w3.org/Math). The description of mathematical expressions using MathML covers the description of the contents of the expression as well

as the presentation and layout of this expression in documents.

In order to specify the model representation language as clearly as possible, a formal specification is used. This specification is based on XML Schema. Using this methodology, it is possible to describe the vocabulary completely and the internal rules to a high extent. Thus formal methods to check the validity of published models can be applied. The schema for SBML can be found at www.sbml.org.

In order to illustrate how bio-chemical conversion models are represented using a Petersen Matrix, a simple example is given here. This model describes the growth of a species of heterotrophic organisms on a single substrate. The growth of the biomass is limited by a maximum growth rate, a Monod-type substrate limitation and a limitation on the availability of dissolved oxygen. The second process considers the decay of biomass, which is assumed to be proportional to the biomass concentration. The Petersen Matrix of this model is presented in Figure 1.

comp. / process		1 S	2 X	process rate [g COD m^{-3} d^{-1}]
1	Growth	$\frac{1}{Y}$	1	$\mu_{max} \frac{S}{S+K_S} \frac{SO}{SO+K_O} X$
2	Decay	0	-1	$b\ X$
		Substrate [g COD/ m^3]	Biomass [g COD/ m^3]	Y – Yield, μ_{max}- maximum growth rate, b - decay rate, K_S, K_O - half saturation const., SO- dissolved oxygen

Fig. 1 Petersen Matrix of a simple biological conversion model

The matrix in Figure 1 contains two columns which are related to the two components of the biological conversion model (S and X) considered here. In these two columns, the name, a description text, the unit and a number of properties of this component are defined. The two rows in the upper part define the two processes considered. For each process the name and the process rate, defined by a kinetic expression, are given. The stoichiometry is defined by a set of stoichiometric factors (one for each component). The same simple model presented as a reaction network is given in Figure 2.

Fig. 2 Simple Reaction Network

The same model can be represented as a SBML file. In Figure 3 this file was opened using a WEB-Browser.

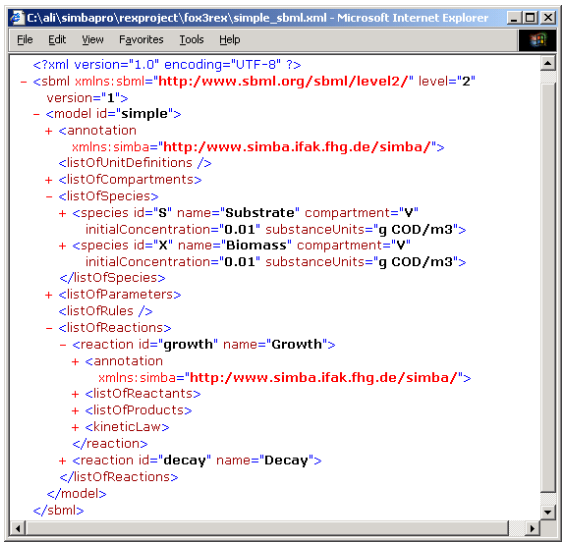


Fig. 3 SBML document for simple model

The XML file, which is a plain text file, is presented in the browser with some syntax highlighting and the option to hide some details of the text-file. The lines starting with a '+' sign contain more details which are hidden. This example shows in a very compact way the typical structure of a SBML file. Use of XML as a model description language provides some additional benefits resulting from a number of accompanying technologies such as XSLT to perform translations from XML documents into any other text based documents like HTML. This technology can be used to fully automatically generate suitable HTML documentations for a model implementation. Figure 4 shows an example of a document describing the ASM1. The support of the newest browsers for MathML allows even the presentation of complex formulas in a very well readable way. This approach guarantees, that a model description, e.g. as a Petersen Matrix, in a printed document fits to the model implementation and by this way, hopefully, to the modelling intention of the original author.

3 BALANCES AS THE FOUNDATION OF MODELS

3.1 Bio-chemical conversion models

The previous section described a methodology for ensuring a transparent presentation and exchange of models, by using a formalised model description language. But first of all, a reliable model has to be developed. With respect to modelling of wastewater treatment processes, one has to deal with very complex biological processes. Obviously it is impossible to set up models where all possible components, involved micro-organisms, different metabolic pathways etc. are included. Besides many other reasons, this is caused by only very limited options being available to measure the relevant con-

centrations and process rates. The only way to develop models, which are in principle applicable to practical applications (e.g. process engineering), is to simplify the processes and substances (or substance groups) considered significantly. But in particular in this case, where significant simplifications are necessary, it is important to have a reliable, verifiable backbone of the model. This can be achieved by the strict usage of mass (and energy) conservation principles.

With respect to bio-chemical conversion models, which can be presented as a Petersen Matrix, this conservation principle can be applied to each process defined in this model (a row in the matrix). The full Petersen Matrix of the most frequently used model for the description of biological wastewater treatment processes – the Activated Sludge Model No.1 (ASM1) - is presented in Figure 4. This model defines 9 processes, where for instance process No. 4 (p4) describes the decay of heterotrophic biomass (XBH). This process could be described using conventional chemometric conventions as

$$XBH \rightarrow fP\ XP + (1\text{-}fP)\ XS + (iXB - fP\ iXP)\ XND \quad (1)$$

where one unit of heterotrophic biomass (g COD) is converted into fP units particulate inert products (XP) plus (1-fp) units particulate slowly biodegradable substrate and (iXB – fP iXP) units particulate organic nitrogen XND. This description is a typical example of simplified modelling of biochemical processes. The units are not based on mole, the stoichiometry cannot be described by integers, and for the components involved no chemical formula (elementary composition) is available.

In order to be able to apply conservation principles also in this case, it is necessary to introduce additional properties to the components. In Figure 4 above the Petersen Matrix, a table is shown, which introduces 4 properties of the model components, the nitrogen content N [g N / unit], the chemical oxygen demand COD [g COD/unit], the total suspended solids ratio TSS [g / unit] and the electrical charge. N, COD and Charge can be considered as properties for which conservation principles can be applied. The property TSS, on the other hand, is used only for descriptive purposes – here, to calculate the TSS of a sludge sample characterised by the model fractions.

As the intention of the model is to preserve the COD and Nitrogen balances, one has to ensure that each process fulfils these balances.

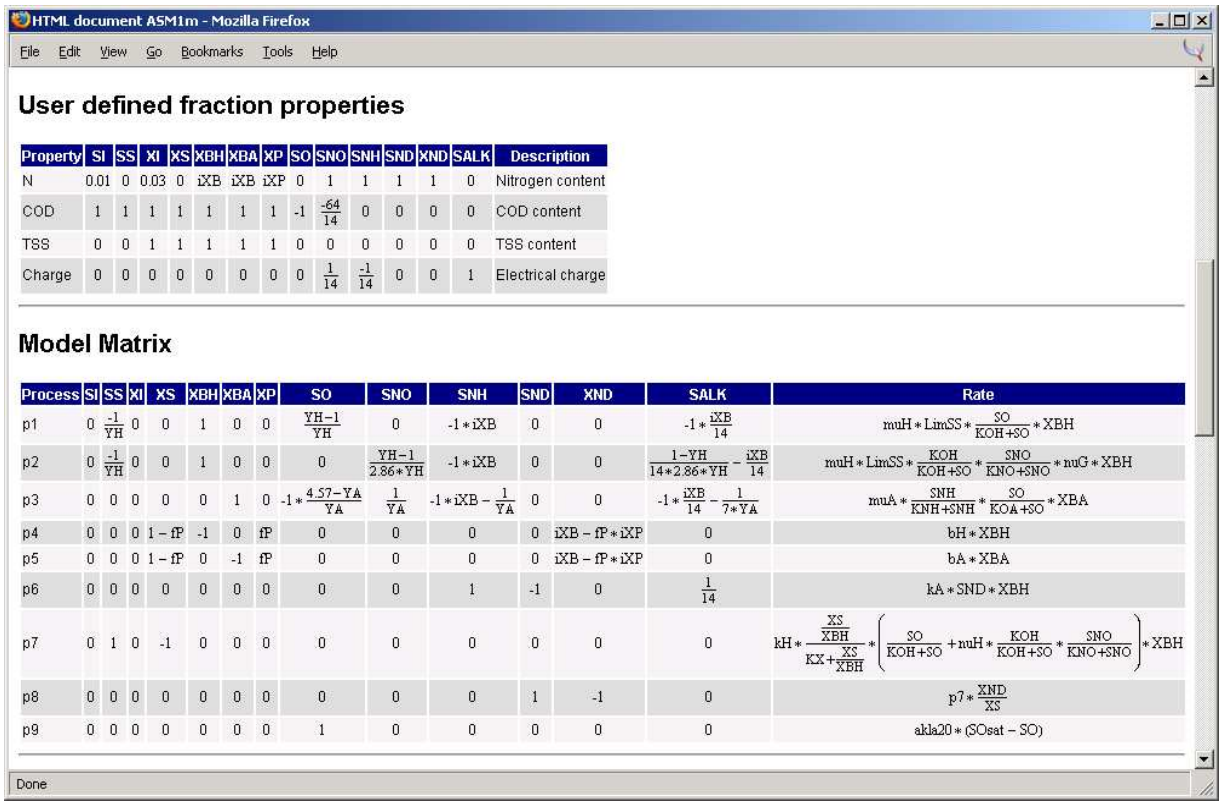
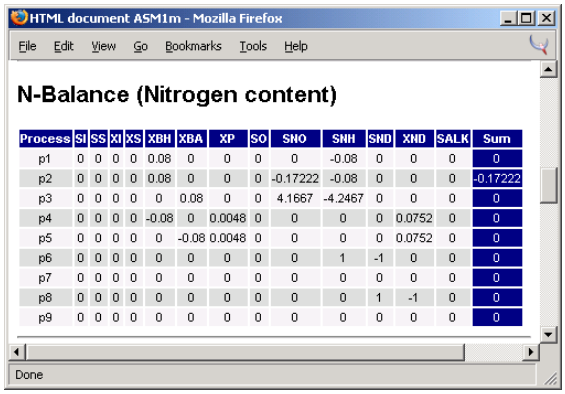


Fig. 4 Petersen Matrix of ASM1 automatically generated from SBML-document describing the ASM1

The mass balance for each conservation parameter j can be generally formulated as follows:

$$\sum_{i}^{\text{reactants}} \left(\text{-1* (stoich. factor for fraction i) * (content of parameter j for fraction i)}\right) + \sum_{k}^{\text{products}} \left(\text{(stoich.factor for fraction k) * (content of parameter j for fraction k)}\right) = 0 \qquad (2)$$

This simple principle, which is trivial for chemical reactions, needs particular attention in the modelling of biological processes and simplified modelling for engineering applications where commonly empirical methods are used. Based on the experience of the authors it turns out that it is of utmost importance to incorporate this principle systematically into a specific modelling and simulation environment. The simulation environment SIMBA (ifak 2003) based on Matlab/Simulink gives an example on how this can be achieved. SIMBA provides a model editor for bio-chemical conversion models, which is based on a symbolic representation using a Petersen Matrix format. The user is encouraged to define additional properties for the components considered in the model. Using this additional information, the modelling tool offers the opportunity to perform a check of the stoichiometric factors based on the application of conservation principles to all user defined (conservation) properties. This check is performed numerically. This means, that the stoichiometric factors are calculated based on a specific parameter set of the model. For each process and conservation parameter, these stoichiometric factors are multiplied by the fraction properties, and the mass balances for the processes following eq. (2) are calculated. The result is presented in a HTML document (Figure 5).



Process	SI	SS	XI	XS	XBH	XBA	XP	SO	SNO	SNH	SND	XND	SALK	Sum
p1	0	0	0	0	0.08	0	0	0	0	-0.08	0	0	0	0
p2	0	0	0	0	0.08	0	0	0	-0.17222	-0.08	0	0	0	-0.17222
p3	0	0	0	0	0	0.08	0	0	4.1667	-4.2467	0	0	0	0
p4	0	0	0	0	-0.08	0	0.0048	0	0	0	0	0.0752	0	0
p5	0	0	0	0	0	-0.08	0.0048	0	0	0	0	0.0752	0	0
p6	0	0	0	0	0	0	0	0	0	1	-1	0	0	0
p7	0	0	0	0	0	0	0	0	0	0	0	0	0	0
p8	0	0	0	0	0	0	0	0	0	0	1	-1	0	0
p9	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Fig. 5 Documentation of mass balances

This figure displays only a fraction of the document, where the N balance is presented. The sum in the last column should be (numerically) zero for processes fulfilling the conservation principles of the specific property. Obviously, for the nitrogen all processes besides process p2 (anoxic growth of heterotrophic biomass) fulfil the nitrogen balance. For process p2, responsible for final removal of nitrogen from the activated sludge (denitrification), this balance is not closed. Nitrogen, removed from the system, is leaving the system in a gaseous phase. This gap could be closed by introducing an additional fraction N2, describing the nitrogen gas produced. The numerical calculation of the balances can be performed in SIMBA using a random pertubation of the parameters.

3.2 Interface models

Simulation in the engineering practice becomes useful in particular in cases where the interactions of the operational units involved become too complicated to be considered and acted upon manually. However, in these cases also simulation has to deal with challenging tasks. For the description of different unit processes, it is necessary to use compatible, and consistent models. For water flows from one sub-model to another, appropriate interface models are required. With respect to simulation of wastewater systems, models are necessary at the following interfaces

- influent description → activated sludge,
- activated sludge → anaerobic digestion,
- anaerobic digestion → activated sludge or
- activated sludge → river water quality.

The set-up of these interface models is not a trivial task. During the modelling of the different units (activated sludge system, anaerobic digestion etc.) the focus of the researcher usually is on the specific process and typically not on a systematic approach which would allow the integration with models of other unit processes. Thus the considered fractions are closely related to the unit process. Even if a similar fraction (using also the same name, e.g. XI) in the other models exists, one cannot assume in a model interface that this fraction can straightforwardly be directed to the subsequent model. In order to discuss this issue, the interface from the Activated Sludge Model ASM1 to the Anaerobic Digestion Model ADM1 is considered here.

An interface model between these two models has to calculate the fraction of one model based on the information delivered by the other model. This calculation has to deal with a lack of information, over-determined information and the different context of the components defined. Again, the only possible approach is to rely on the application of conservation principles. For the interface between ASM1 and ADM1, it is possible to define the conservation rules for COD, N, electrical charge,

alkalinity and, with some minor restrictions, also for TSS and inert COD. Based on these conservation principles, the interface models can be developed. As a result of the different context of the two models, it becomes necessary to introduce conversion processes also in the interfaces, which in reality will occur in the following process step, but are not explicitly modelled by specific process model.

This short discussion on some of the issues related to the set-up of interface models underlines the necessity to support the modeller by convenient tools for this complex task. A specific model block has been introduced in SIMBA especially to implement interface models. The model is defined by specifying equations (symbolically) for the fractions of the destination model based on the concentrations of the origin model. To allow an error-free and comfortable documentation of the model, the editor provides a feature for the generation of a HTML-documentation. With respect to conservation principles, the tool includes the option to define an arbitrary number of conservation parameters (here N, COD, TSS, charge) and to calculate the values of these conservation variables numerically for a defined influent sample (Figure 6). If the values are equal for both models, it can be assumed that the aim of mass conservation is reached by the interface model developed.

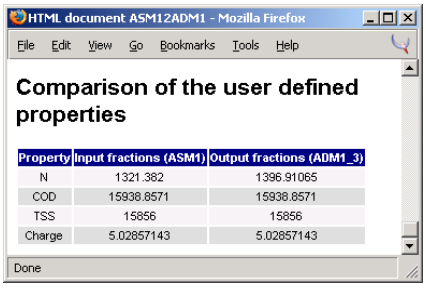


Fig. 6 Balance check for interface model

4 CONCLUSION

It can be stated that modelling and simulation have reached a mature state for practical application in the field of wastewater system design and optimisation. One reason for this status is the availability of reliable and de-facto standardised models. With regard to the further development in this area, it can be observed that recent models are becoming more complex and that simulation is applied more and more for tackling integrated problems. This paper discussed two approaches for dealing with the increased level of complexity. The first suggested approach is to apply systematically formalised model representation methods for model definition, exchange and analysis. The adaptation of the XML-based Systems Biology Mark-Up Language (SBML) to the area of wastewater modelling was proposed. The implementation of this approach into an advanced simulation system demonstrates significant advantages of the methods.

The second topic of this paper focussed on the application of strict mass conservation principles during model development to provide a reliable backbone for modelling and integrated simulation of coupled unit processes. A specific problem of integrated simulation, the development of interface models, was discussed, based on an example for the interfaces from an activated sludge model to an anaerobic digestion model. These interface models are based on strict mass conservation principles, again. If this principle is followed in a stringent way, the application of simulation can deliver reliable results, even if several aspects of the processes considered have to be accepted as being uncertain. Based on this consideration, it can be concluded that integrated simulation can be used for practical applications, even if a number of specific questions are yet open.

5 REFERENCE

Batstone D.J., J. Keller, I. Angelidaki, S.V. Kalyuzhnyi, S.G. Pavlostathis, A. Rozzi, W.T.M. Sanders, H. Siegrist and V.A. Vavilin (2002). Anaerobic Digestion Model No. 1. IWA Scientific and Technical Report No. 13, IWA Publishing, London, UK.

Gernaey, K.V., C. Rosen, D.J. Batstone and J. Alex (2004). Efficient modelling necessitates standards for model documentation and exchange. accepted for 6th WATERMATEX, November 3rd to 5th, 2004, Beijing.

Henze M., W. Gujer, T. Mino and M. van Loosdrecht (2000). Activated sludge models ASM1, ASM2, ASM2D and ASM3. IWA Scientific and Technical Report No.9. IWA Publishing, London, UK.

Hucka (2003). The systems biology markup language (SBML): A medium for representation of biochemical network models. Bioinformatics, 19(4), 524-531.

ifak (2003) SIMBA 4.2. Manual. Institut f. Automation und Kommunikation ifak Magdeburg e. V., Barleben.

Petersen E.E. (1965). Chemical Reaction Analysis. Prentice-Hall, Englewood Cliffs, NJ.

Reichert P., D. Borchardt, M. Henze, W. Rauch, P. Shanahan, L. Somlyody and P.A. Vanrolleghem (2001). River Water Quality Model No.1. IWA Scientific and Technical Report No. 12, IWA Publishing, London, UK.