QualViz: a tool for visual representation of water quality models

G. T. Parker

ABSTRACT

As water quality models and their implementation have become increasingly diverse, complex and proprietary, a need for more thorough understanding of the differences between each alternative arises. The work presented here proposes a novel visualization paradigm for water quality applications which can be used to understand difference between implementations of identical and different conceptual models. A proof-of-concept visualization tool was developed and tested again three scenarios for four different conceptual models of biochemical kinetics. Results show representative figures illustrating how the approach can communicate differences in model complexity and dynamic behaviour. The proposed tool should help ensure more suitable application of water quality models in varied contexts. A discussion of quantifying model complexity in a single metric is also presented, and recommendations are made on the selection of various representational forms for communicating and exploring specific model characteristics.

Key words | model, activated Sludge, river, BSM, schematic, water quality, visualization

G. T. Parker (corresponding author)
University of Cambridge,
Department of Engineering,
Trumpington Street,
Cambridge CB2 1PZ
E-mail: *geoff.parker@eng.cam.ac.uk*

INTRODUCTION

Water quality models are commonly used tools to assess, design and manage wastewater treatment plants (WWTPs) as well as the aquatic systems to which they discharge. A wide variety of simulation software exists and is in use worldwide. As an example, the benchmark simulation model 1 (BSM1, see Copp 2002), which describes a relatively simple activated sludge process with clarifier, has been implemented on at least 9 simulation platforms, namely BioWin, EFOR, FORTRAN, GPS-X, MATLAB/Simulink, SIMBA, STOAT, WEST, and JAVA (Samuelsson *et al.* 2001; Copp 2002). Additional models can be used to consider new and revised conceptual understanding of key processes, which has led to numerous activated sludge model (ASM) variants (e.g. ASM2D, ASM3, ASM3-P, UCTPHO+ and ASDM) that variously describe organic matter and nitrogen removal, phosphorus removal, and even anaerobic digestion and chemical equilibrium kinetics. Models for receiving waters, including for instance river quality models, describe many of the same processes and, in an analogous fashion, encompass equally diverse conceptual elements.

Managing this range of models is a non-trivial task, and yet a clear understanding of model difference is essential for simulation use, interpretation, and result duplication. Much progress has been made, starting chiefly with a rigourous format for model description in matrix form as proposed by Petersen (1965) and implemented in Henze *et al.* (1987). This has become the de facto standard in describing water quality both for river (Reichert *et al.* 2001) and treatment process (Comeau & Takács 2008) modeling applications, and it is assumed the reader is familiar with this representation. Several issues remain, including the use of varied symbols between researchers for the same compounds and the logistics of publishing very large matrices associated with numerous processes and state variables. To address these issues Comeau & Takács (2008) presented a proposed standardized list of state variables as well as a schematic representation of ASM processes. Other issues relevant to the discussion include the differences between implementation of the same conceptual model on many platforms (see for example Copp 2002) and the still widespread use of models with either unpublished or only partially published Petersen matrices (as discussed in e.g. Claeys 2008).

In this paper we present a framework, and a visualization tool implementing it, that can be used to better communicate

doi: 10.2166/wst.2011.028

the structural and behavioural distinctions between various water quality models. The proposed framework implements the Comeau and Takacs symbol notation standard, but also leverages modern visualization technology and revised paradigms (e.g. colour, projection along cardinal axes) to simplify presentation. In addition, a standardized 'model benchmarking' framework is proposed to allow for communication of model dynamics under several conditions.

METHODS

Approach

The QualViz tool was developed on the 'Processing' platform. 'Processing' is an open-source, cross-platform programming language (and compiler), built on JAVA, which is particularly well suited, and thus widely used, for visualization tasks (see for instance Reas & Maeda 2007). A hypothetical scenario was examined for visualization of activated sludge and river water quality models of increasing complexity.

Computation

During initialization, the QualViz code identifies state variables tracked by the conceptual model and allocates positions along the cardinal axes, in the horizontal for soluble substrate and along the vertical for particulate matter and consumer organisms. Kinetics are then simulated for each time-steps, a setting of 2000 time-steps per day is used, so each increment is approximately equal to 43 seconds. The selected RWQM1 model variant, including the assumptions on the composition of organic matter and kinetic constants as specified in Vanrolleghem *et al.* (2001), is implemented at each time-step to simulate kinetics at that point. This is purposefully analogous to the BSM1 implementations described in Claeys 2008 and draws on readily available sources. For the purposes of a river problem, this can be conceived as one-dimensional steady flow model, connecting a sequence of well-mixed cross-sectional elements.

Visualization

Figure 1 shows a detail of a representative output schematic from a visualization run.

Constituents in each schematic are designated by circles, alongside standardized symbols as suggested by Comeau & Takács (2008), with substrate components grouped along a centered horizontal axis and particulate/biomass components grouped vertically. Links between components are

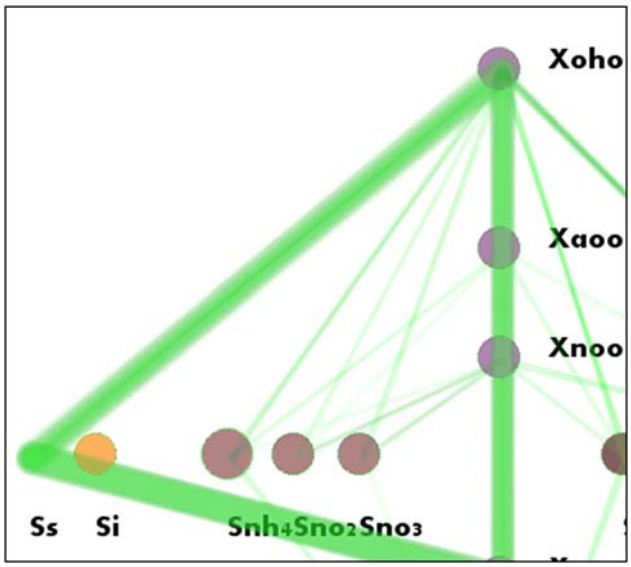

Figure 1 | Detail of QualViz output for 'Simplified RWQM no. 1' conceptual model at Simulation Time of 1 hr for 'Dry' Condition. Subscribers to the online version of *Water Science and Technology* can access the colour version of this figure from http://www.iwaponline.com/wst.

dynamically drawn, with greater width at higher process rates. Model 'snapshots' in this paper are presented at 1-hour (hr), 12-hr and 24-hr time-steps, with a colour shift (to blue, from green) after hour 12 used to help distinguish phases.

Circles representing state variables are scaled at each time-step to note their relative change in concentration. Flow lines between state variables are redrawn at each step and are also scaled to show significance. Lines are drawn progressively in very light ('transparent') layers so thick dark lines are processes that are denote significant mass flows for a substantial duration. Each state variable is also tagged with a symbol which is consistent with the notation recommended in Comeau & Takács (2008).

Case study

Benchmark Conditions

The case study scenario is based on scenarios developed as part of the BSM1 (Copp 2002) for activated sludge systems. In the hypothetical scenario, the discharge from the BSM1 model enters a river system, which is also modelled. Table 1 presents average benchmark values for WWTP effluents from BSM1 Scenarios.

The 3 scenarios examined are termed 'Dry'. 'Rain' and 'Storm' in the original reference and the standard. This nomenclature will also be used here.

Table 1 | Benchmark Values (FORTRAN) for WWTP effluents from BSM1 Scenarios *

			Simulated Condition		
Symbol	**Units**	**Variable**	**Dry**	**Rain**	**Storm**
S_S	(mg COD/L)	Readily biodegradable organic matter	0.9730	1.1300	1.1100
S_I	(mg COD/L)	Soluble unbiodegradable organic matter	30.0000	22.8000	26.3000
S_{NH4}	(mg/L)	Total ammonia	4.8000	5.0100	5.3900
S_{NO}	(mg/L)	Total nitrite + nitrate	8.8000	6.9300	7.4500
S_{O2}	(mg/L)	Dissolved oxygen	0.7978	0.8860	0.7990
S_{ALK}	(mg/L)	Alkalinity	4.4600	5.1500	4.8800
X_{OHO}	(mg COD/L)	Ordinary heterotrophic organisms	10.2000	12.9000	11.9000
X_{NIT}	(mg COD/L)	Nitryfying Organisms	0.5420	0.6430	0.5890
X_S	(mg COD/L)	Slowly biodegradable Organic Matter	0.2290	0.3460	0.3240
X_I	(mg COD/L)	Particulate unbiodegradable endogeneous products	4.5878	5.6500	5.6400
X_P	(mg HPO_4/L)	Stored polyphosphates	1.7600	2.0700	1.9200

*(Compiled from Copp 2002).

Conceptual Model Selection

The riverine water quality models used to demonstrate the visualization tool were derived from the River Water Quality Model no. 1 (RWQM1) model presented in Reichert *et al.* (2001), and are dubbed 'Minimal Streeter-Phelps', 'Extended Streeter-Phelps', 'Similar to Qual2e' and 'Simplified RWQM1', in increasing order of complexity. Model processes and parameter kinetics were as specified in Vanrolleghem *et al.* (2001), with addition of a Churchill *et al.* (1962) reaeration term.

Table 2 presents a description of the model examined in this paper, alongside two reference models. While it can be a simple, even self-evident, task to rank two or more models in order of increasing complexity (e.g. Lindenschmidt 2005), describing model complexity in a quantified manner is not as trivial. Such a measure is useful however, as it can provide some clues as to where in a posited complexity spectrum a model may lie and is in some senses a useful a-priori visualization tool itself, especially in conjunction with the dynamic visualization discussed in the next section.

Table 2 | Matrix Size and Model Complexity of Riverine Water Quality Models

				Complexity Measures		
Model Designation	**Complexity Rank**	**Number of State Variables**	**Number of Processes**	**Matrix Dimensions**	I_c	I_c'
Minimal Streeter-Phelps	1	2	1	6	562	34
Extended Streeter-Phelps	2	6	2	24	1683	79
Mike11	N/A*	6	11	66	78	78
QUAL2	N/A*	9	15	135	153	151
Similar to Qual2e	3	7	5	77	3948	210
Simplified RWQM1	4	13	15	195	7882	471

(Larger values imply more complexity).*Provided for comparison. See also Parker (2009).

The simplest way to calculate a complexity measure for models with a published Petersen matrix is, probably, to describe the size of this matrix. Generally speaking, more complex models will consider more state variables (i.e. columns) and kinetic processes (i.e. rows), so multiplying these two dimensions will provide both the total matrix 'area' and some indication as to the model complexity. One disadvantage to this measure is that the matrix formulation is, to some degree, a subjective task, and therefore several potential matrices can be formulated for identical models. This might occur for instance when processes are consolidated into a single row or alternatively by incorporating additional parameters in the matrix cells themselves. In a paper discussing the effect of model complexity on model uncertainty, Snowling & Kramer (2001) quantified model complexity in a metric encompassing the total number of parameters and operations contained within a matrix:

$$I_c = \sum_{j=1}^{N} \sum_{i=1}^{n_l} p_i r_i \quad (1)$$

where I_c is Snowling and Kramer's metric, N is the number of columns in the Petersen matrix, n_j is the number of rows (processes) active for that column, p_i is the number of parameters and r_i is the number of mathematical operation in the formulation of each process i. While this method has the advantage of being less sensitive to the matrix formulation, it is instead sensitive to process kinetic formulation and an increasing number of operations does not necessarily translate into a substantial increase in model complexity. This limitation becomes obvious when comparing very similar models with, however, significant differences in formulation (e.g. QUAL2, 'Similar to QUAL2E' and 'Extended SP', 'Mike11' in Table 2). Also presented here is a new metric, 'I_c', that includes only the parameter component, pi, and is therefore defined as:

$$I_c = \sum_{j=1}^{N} \sum_{i=1}^{n_j} p_i \quad (2)$$

As shown in Table 2, this metric seems to be more consistent with qualitative model similarity across the examined model formulations. The rationale is that thoughtfully constructed models will tend to add (and remove) complexity in their formulation primarily through parametrization, and further, that models that are useful (and therefore most commonly used) will tend to be those that do not add additional parameters simply for their own sake. The metric is also more straightforward to compute for an operation such as x^y, which could be considered as either a single operation or as a series of y multiplications (in which case the Snowling and Kramer metric will change depending on the value of parameter y).

However useful a measure of the complexity, such as that presented above, a Petersen matrix, or a schematic showing the elements in each specific component process (as presented in Comeau & Takács 2008) may be, these tools serve only to describe the conceptual model prior to implementation. More complexity should not be considered a merit in and of itself (Lindenschmidt 2005) and poses no advantage if the added complexity makes little kinetic difference given a particular set of conditions to be examined (e.g. parameters and data constraints). This becomes particularly relevant in problems where model response under uncertainty must be considered (as shown in Parker *et al.* 2009). In these cases, it is important to understand not only how the conceptual model can behave, depending on parameter and data, but the dynamics of how it actually is behaving in the parameter and data regions of interest. In this way, a visualization can provide insight into the data component of the aggregate model consisting of a conceptual framework, parameter decisions and calibration, and input data.

RESULTS AND DISCUSSION

Results

Results from 3 model types over 3 time-frames are presented in Figure 2. All models in Figure 2 were derived from the 'Dry' conditions as detailed in Table 1. Other conditions examined were the 'Rain' and 'Storm' conditions, though they are not presented in this figure.

Figure 2 is an illustrative example of the visualization tool applied to each conceptual model. While there are significant similarities between the models, the differences in the model response become apparent as time progresses. The Minimal SP presented in the first row is very simple. and behaves as such. The only two state variables that are affected by model kinetics are the readily biodegradable organic matter, which degrades over time, and the dissolved oxygen variables on the horizontal axis. in the second row, state variables which were tracked but not active in the first visualization sequence now begin to play a small role. This model considers the growth of algae, which has implications for both the nitrate and oxygen state variables, and this is shown. In the third row, additional complexity has been

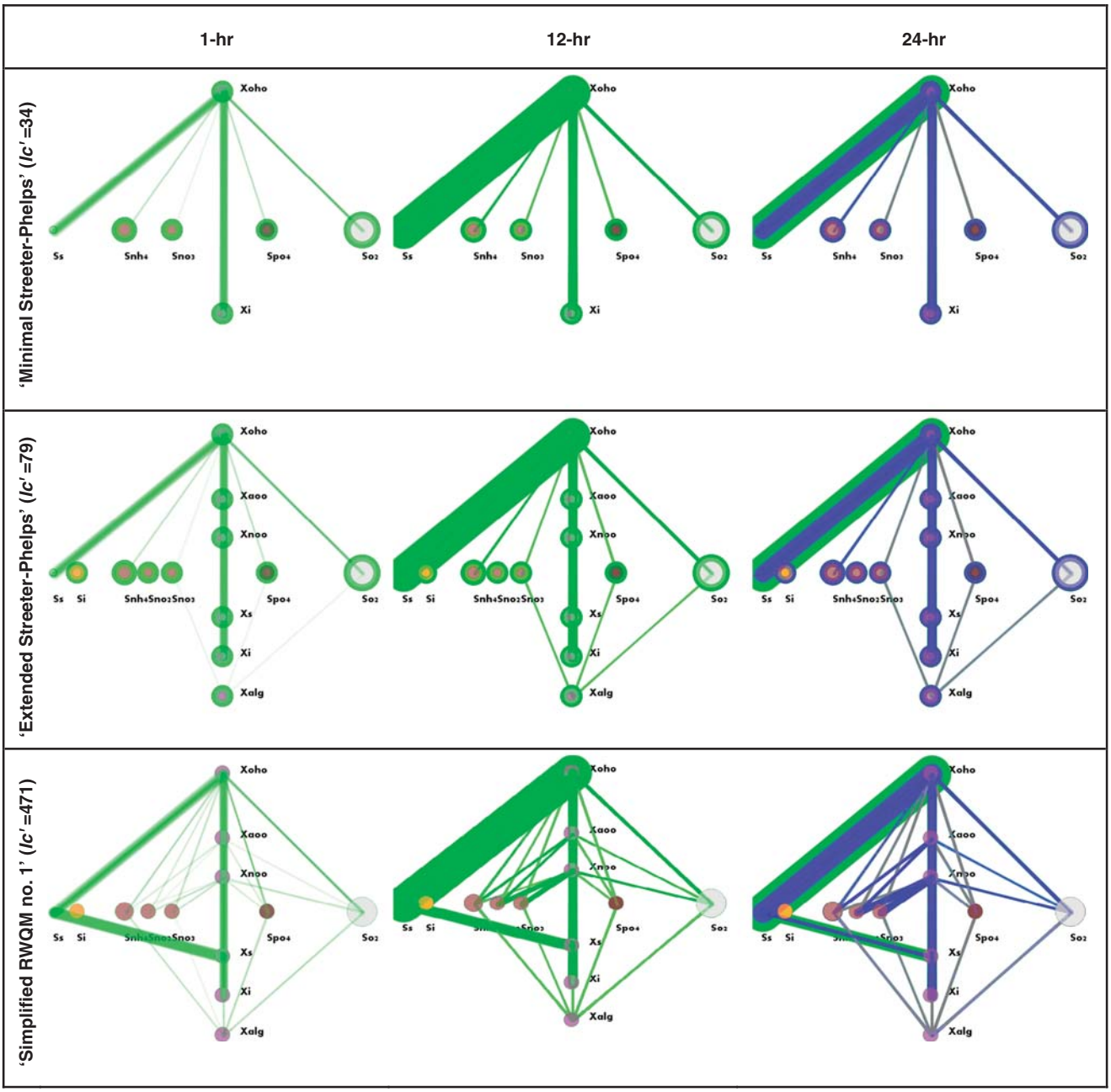

Figure 2 | Schematic Visualization of River Water Quality Models for 'Dry' Conditions at Simulation Times of 1, 12 and 24 hrs Subscribers to the online version of Water Science and Technology can access the colour version of this figure from http://www.iwaponline.com/wst.

introduced and the vertical axis is now tracking six active particulate fractions. Several of these sta variables are playing a significant role when compared with the previous with steady state assumption, within the vertical axis species themselves as well as with the tracked soluble fractions (evidenced by the thickness of the transfer lines). Hydrolysis is now playing a significant role in controlling the readily biodegradable organic matter. The kinetics of nitrifier and heterotroph growth and respiration on the nitrogen species is also significantly more complex here.

Figure 3 shows the Similar to QUAL2E' conceptual model, examined after a simulated 24 hr period for the 3 different data sets presented in Table 1.

Though there is some variation in this range, it is apparent that the model is behaving rather consistently in the 3 different cases examined. Differences are apparent in the

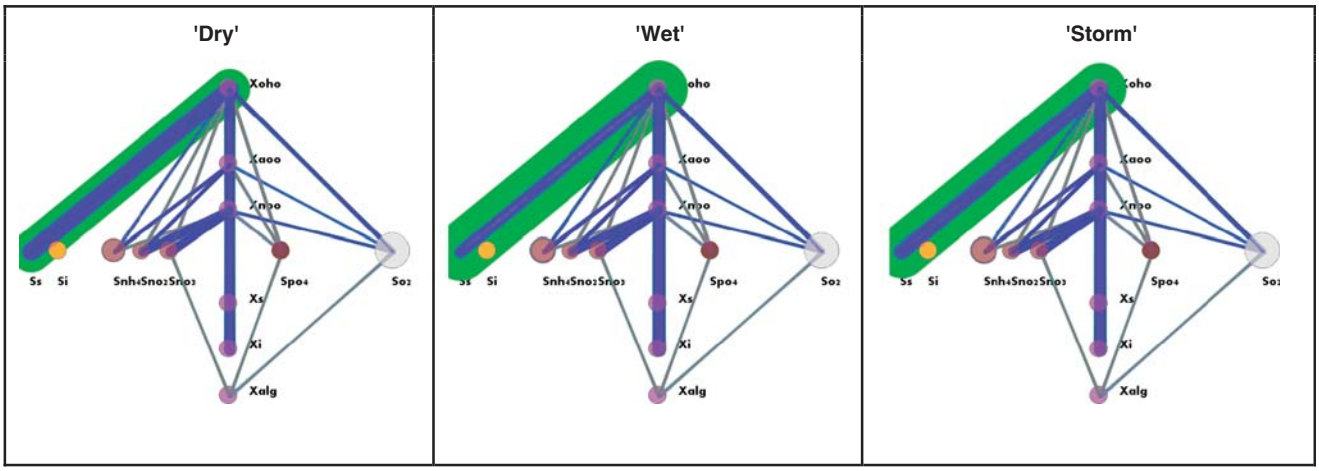

Figure 3 | Schematic Visualization of 'Similar to QUAL2E' (I_c' = 210) River Water Quality Model for "Dry','Wet' and 'Storm' Conditions at Simulation Time of 24 hrs. Subscribers to the online version of *Water Science and Technology* can access the colour version of this figure from http://www.iwaponline.com/wst.

degradation of soluble organics, especially when conditions at the 12 hr and 24 hr points are compared. A visualization similar to Figure 2 would more clearly illustrate that, as in those cases, the role of 'additional' processes in the overall model state becomes more significant as time elapses.

DISCUSSION

Taken over several time-frames and conditions, the proposed approach permits quick and simple communication and differentiation of active state variables and the importance of exchanges between them in a water quality model. Previous schemes, as presented for instance in Comeau & Takács (2008), focus more on issues of constituent grouping and pathway description. The focus in this paper has been to demonstrate a much more simple visualization framework which can supplement rather than replace the model matrix description. In cases where no complete model matrix is published, a QualViz or similar simple visualization affords an opportunity for model users to identify strengths and weaknesses as well as relevance of the underlying conceptual models. The same can be said of identical model matrices implemented variously on several platforms, as in Copp (2002).

The use of a well-known benchmark to generate simulation conditions for the discussion in this paper was, it should be noted, not incidental. What is presented here is, effectively, also potentiall a means to benchmark a water quality simulation model itself. While in Copp (2002) the model and upstream data are held constant with the aim of evaluating performance of varied control strategies, in this case we have held data, and the conditions under which the model operates, constant, The result, in some sense, is a benchmark of the conceptual model performance, relative to alternative formulation, under these conditions.

There are of course limitations to the suitability of the proposed tool. It is true, for instance, that these plots may not lend themselves well to monochromatic reproduction (e.g. publication), however, given the continued rise in model complexity, this is increasingly true for other representations as well. Nevertheless, in this article, four different models under three conditions at three timesteps have been presented. In articles where demonstrating the representation itself is not the main thrust, the limitations of scale and space would perhaps be less significant. Additionally, there are many forums, including the classroom, online and support documentation and of course the modeling tool interfaces themselves, that do not (or should not) limit themselves to the constraints imposed by 16th century technology, as the saying goes. With this in mind, Table 3 presents an outline of where, it is thought, this representation can be of some use when compared with existing alternatives mentioned in this paper.

CONCLUSIONS

This paper has drawn on previous efforts to simplify visual representation of differences between increasingly complex water quality models. The plots presented here are in no way meant to supplant a formal model description in matrix form, which remains very necessary for model users. However, in cases where the matrix is unavailable, or is so unwieldy as to

Table 3 | Guidance for Representation Selection for Exploring Specific Model Characteristics

For Communicating	The representation(s)* that are best suited include:	less well suited include:	not well suited include:
General Model			
Exact Kinetic Formulation	Petersen		Comeau and Takacs, QualViz
Major Processes Modelled	Comeau and Takács	Petersen	QualViz
Important Model Parameters	Petersen		Comeau and Takacs, QualViz
A priori			
Important State Variables	Comeau and Takacs,	Petersen, QualViz	
Differences in Model Complexity	Petersen, Comeau and Takacs, QualViz		
A posterior			
Dynamic Behaviour and Complexity	QualViz		Petersen, Comeau and Takacs,
Important State Variables	QualViz		Petersen, Comeau and Takacs,

*'Petersen' refers to the matrix formulation from Petersen (1965) as used in Henze *et al*. (1987)
'Comeau and Takacs' refers to the schematics as presented in Comeau & Takács (2008).
'QualViz' refers to a dynamic visualization as presented in this paper.

raise questions on dynamic model issues, a visualization framework such as that proposed here can be readily employed by model users (and developers) to contrast and compare alternative models and expected behaviour. This provides an opportunity for benchmarking decisions in conceptual model structure at the implementation level (which even a matrix does not).

REFERENCES

Churchill, M. A., Elmore, H. L. & Buchingham, R. A. 1962 The prediction of stream reaeration rates. *Journal of the Sanitary Engineering Division, American Society of Civil Engineers* **88**(SA4): 1–46.

Claeys, F. 2008. A Generic Software Framework for Modelling and Virtual Experimentation with Complex Environmental Systems. PhD. Thesis. Faculty of Bioscience Engineering. Ghent University. pp. 303.

Comeau, Y. & Takács, I. 2008. Schematic Representation of Activated Sludge Models. Proceedings of the Water Environment Federation 81st Annual Technical Exhibition & Conference, Chicago, IL, USA, October 18–22.

Copp, J. B., 2002. *The COST simulation benchmark–description and simulator manual*. Luxembourg: Office for Official Publications of the European Communities, 92-894-1658-0.

Henze, M., Grady, C. P. L., Gujer, W., Marais, G. V. R. & Matsuo, T. 1987. Activated sludge model no. 1. IAWPRC Scientific and Technical Report No. 1, IAWPRC, London.

Lindenschmidt, K. E. 2005. The effect of complexity on parameter sensitivity and model uncertainty in river water quality modelling, *Ecol. Modell.* **190**(1–2), 72–86.

Parker, G. T., Droste, R. L. & Rennie, C. D. 2009 An objective test of stochastic behaviour in riverine water quality models. *Wat. Sci. Tech.* **59**(1), 159–165.

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

Reas, B. & Maeda, J. 2007. *Processing: A Programming Handbook for Visual Designers and Artists* (1st ed.), The MIT Press, pp. 736, 0262182629.

Reichert, P., Borchardt, D., Henze, M., Rauch, W., Shanahan, P., Somlyódy, L. & Vanrolleghem, P. A. 2001 River Water Quality Model No. 1 (RWQM1): II. Biochemical Process Equations, *Wat. Sci. Tech.* **43**(5), 11–30.

Samuelsson, P., Ekman, M. & Carlsson, B. 2001. A JAVA based simulator of activated sludge processes. *Mathematics and Computers in Simulation* **56**(4–5), 333–346.

Snowling, S. D. & Kramer, J. R. 2001. Evaluating modelling uncertainty for model selection, *Ecol. Modell.* **138** (1), 17–30.

Vanrolleghem P., Borchardt D., Henze M., Rauch W., Reichert P., Shanahan P., & Somlyódy L. 2001 River Water Quality Model No. 1 (RWQM1): II. Biochemical Submodel Selection, *Wat. Sci. Tech.* **43**(5), 31–40.