Ecological Engineering 42 (2012) 283–294

Contents lists available at SciVerse ScienceDirect

Ecological Engineering

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

Does increased model complexity improve description of phosphorus dynamics in a large treatment wetland?

Rajendra Paudel[a,b,*], James W. Jawitz[a]

[a] Soil and Water Science Department, University of Florida, Gainesville, FL 32611, United States
[b] Fort Lauderdale Research and Education Center, University of Florida, Fort Lauderdale, FL 33314, United States

ARTICLE INFO

Article history:
Received 21 August 2011
Received in revised form 27 January 2012
Accepted 8 February 2012

Keywords:
Model complexity
Effectiveness
Mechanistic
Biogeochemical
Wetland
Everglades
Phosphorus

ABSTRACT

As the structure of ecological models grows more complex, it becomes increasingly important to identify the appropriate level of complexity for reliable process description and prediction. Here, a suite of mechanistic biogeochemical models with different levels of complexity for representing phosphorus cycling processes was developed and tested against observations from a large treatment wetland. The study site was the 147-ha Cell 4 of Stormwater Treatment Area 1 West, which was designed to help protect the greater Everglades, FL, USA, from nutrient over-enrichment. Six biogeochemical models of differing complexity were coupled with a pre-calibrated two-dimensional hydrodynamic model of Cell 4 and tested against field data. We provide guidance for evaluating a set of models with varying level of complexity using key model attributes that influence the suitability of a model or a set of models. Considerations of model accuracy, complexity, and explanatory depth are combined into a single indicator of model effectiveness. Results revealed that the most complex model structure may not necessarily be the most effective in simulating the dynamics of total phosphorus (TP) concentrations in the wetland. The rate of improvement in the model performance decreased as model complexity increased. Although the most complex model reproduced the field observations best, the marginal improvement in model performance compared to simpler models was outweighed by the higher costs of increased complexity. Highly detailed representations of system structures may not be useful to simulate TP dynamics in treatment wetlands if comprehensive data sets are not available to constrain each pathway. It is crucial for model developers and users to evaluate model structures of differing complexity to identify the appropriate level of complexity for given data and questions of interest.

1. Introduction

Mechanistic biogeochemical models are becoming critical tools for predicting nutrient behavior in constructed treatment wetlands to address a wide range of management and research questions (Christensen et al., 1994; Wang and Mitsch, 2000; Walker and Kadlec, 2011; Paudel et al., 2010). However, there is a continuing challenge of constructing or selecting an appropriate model structure that adequately represents the nutrient cycling mechanisms and establishes a rigorous link to the measured field data to provide reliable predictions. Nutrient cycling processes in wetland systems are mainly governed by complex, heterogeneous, microscale, physical, chemical and biological processes, which are often hard to discern and characterize. Given the inherent complexity of wetlands, models developed with simplistic assumptions do not adequately represent the reality of a dynamic system, losing physical meaning to make numerically reliable predictions (Kadlec, 2000; Fulton et al., 2003; Haws et al., 2006). Therefore, overly simplistic models will likely fail under different field conditions that are beyond the bounds of the calibration data (Martin and McCutcheon, 1999). These failures are attributable to the simplifications in the model formulation that do not sufficiently account for complex interactions among various wetland components, such as water, soil, biota, and dissolved/particulate constituents. Conversely, if the model has too many parameters that surpass the type of data needed, calibration to measured data can be non-unique such that no particular combination of parameters represents the solution (Beven, 1996; Martin and McCutcheon, 1999), and we may not be able to discern the underlying causes of system behaviors and make useful predictions.

When formulating a model structure of an aquatic ecosystem, such as a treatment wetland, a common misconception is that if one model structure fails to reasonably predict the data or

* Corresponding author at: Soil and Water Science Department, University of Florida, Gainesville, FL 32611, United States. Tel.: +1 352 392 1951x218.
E-mail addresses: rpaudel@ufl.edu, rajen.water@gmail.com (R. Paudel).

0925-8574/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecoleng.2012.02.014

experimental scenario, increasing the level of complexity (i.e., greater process descriptions) will improve the model performance. However, a review of 153 aquatic biogeochemical models published from 1990 to 2002 failed to support this commonly held notion (Arhonditsis and Brett, 2004). These authors found that adding complexity to an aquatic biogeochemical model did not guarantee improved prediction accuracy. A similar finding was presented by Min et al. (2011) in the context of Everglades wetlands.

Complex biogeochemical models usually have fewer restricting assumptions and exhibit more flexibility (Snowling and Kramer, 2001); however, increasing the level of complexity in the model leads to an increased sensitivity of the output to the input (Snowling and Kramer, 2001; Lindenschmidt, 2006). This is primarily because large uncertainties may arise due to the increased number of interactions between state variables and unconstrained parameters (Robson et al., 2008). Incorporating comprehensive representations of biogeochemical processes and their effects into models also entails practical limitations. For example, mechanistic biogeochemical models require large amounts of data (Robson et al., 2008), which may be relatively scarce. If the model is not constrained by the available field data, the cost associated with fitting noise could lead to diminished performance (Friedrichs et al., 2006). Complex models also need huge human and computer resources (Jorgensen, 2002); therefore, increasing the level of complexity by incorporating more state variables and processes may not be cost effective, because the majority of the modeling resources may then be devoted to developing and maintaining the model, rather than its application (Fulton et al., 2003). In addition, the computational cost of adding more detail may effectively inhibit the utility of the model. As a consequence, there is a conflict between the desire to constrain the model complexity and to incorporate more processes mechanistically.

When developing a biogeochemical model for a specific modeling purpose – such as synthesizing the data/knowledge, testing hypotheses, assessing management strategies, and forecasting future outcomes – a critical question is: how complex should the model structure be to produce the most reliable numerical predictions that balances the tradeoffs between undesirable details and unjustified simplifications (Flynn, 2005). Based on the principle of parsimony (Box and Jenkins, 1970), an appropriate model for a given condition is minimally parameterized with adequate representation of the available data so that the model can be tested more comprehensively, given the intrinsic limitations of the available data (Kirchner, 2006). As wetland ecosystem structures are intrinsically complex, the task of formulating an appropriate level of complexity for simulation models remains a critical and challenging one. Generally, while formulating the model structure, judgments (which are often implicit) are made about the level of details that need to be considered (Cox et al., 2006). However, consideration of some standard qualitative as well as quantitative model evaluation criteria in an easily applicable and holistic approach is essential to assess the best model approximation for any specific purpose.

Evaluating a set of models of wetland systems requires a comparative assessment of different model attributes with respect to the model complexity. Costanza and Sklar (1985) considered model accuracy (based on the quantitative goodness-of-fit of models to historical data) in their review of several published freshwater wetland models. They measured complexity by 'articulation' which incorporated the size of the model in terms of components (number of state variables), and spatial and temporal resolutions. Model performance was measured by 'effectiveness' (a quantity used to represent the trade-off between complexity and model accuracy). These authors found that the maximum effectiveness was for models of moderate complexity. Fulton (2001) presented a similar finding in the context of the published literature on marine ecosystem models. However, Myung et al. (2009) emphasized the importance of considering other model complexity attributes, such as explanatory adequacy, interpretability, descriptive adequacy, and generalizability. Here, we used the following model properties to compare a candidate set of phosphorus biogeochemical models: (a) modeling cost and effort; which is basically the level of model complexity (Haraldsson and Sverdrup, 2004), (b) descriptive adequacy (whether the model describes the existing knowledge and field data of the system); (c) predictive adequacy (whether the model predicts the behavior of a system or data outside the existing boundaries); and (d) explanatory depth (whether the model describes more underlying phenomena to provide the knowledge/information about the system structure). Even though each of these criteria describes a property (i.e., element) of a model that can be evaluated on its own, they are intricately related to each other. We suggest that concomitant consideration of all four criteria is essential to fully evaluate the suitability (i.e., effectiveness) of a model.

Prior studies evaluating the relative effectiveness of models of varying complexity have relied on reviews of published literature, in which the set of models were not tested against the same field data to qualitatively test the model structure. Here, we quantitatively analyzed six phosphorus biogeochemical models with different levels of mechanistic complexity applied to a single consistent data set. The data were from a large-scale constructed treatment wetland, the 147-ha Cell 4 of Stormwater Treatment Area 1 West, that was designed to help protect the greater Everglades, FL, USA, by intercepting runoff and associated nutrients from upstream agricultural areas. The models were tested against 7 years (1995–2001) of outlet total phosphorus (TP) concentrations and 5 years (1995–1999) of spatial soil TP data within the wetland.

This manuscript begins with a review of applications of the model properties evaluated here, with an emphasis on environmental and ecological studies. This is followed by the formulation of model structures applied in Cell 4, with hierarchal level of process-complexity (from low to high level of mechanistic descriptions). The effectiveness of each model for simulating phosphorus dynamics in a stormwater treatment wetland was then assessed based on the selected model properties.

2. Evaluating model effectiveness

The process of developing a model of complex and dynamic systems, such as wetlands, requires a robust way to evaluate the model predictions relative to observations, as well as an equally robust way to compare across models of differing complexity. The purpose of this section is to systematically define and describe key model attributes (i.e., evaluation criteria) that are critical in constructing or selecting a suitable model structure of wetland environments.

2.1. Model complexity (modeling cost and effort)

There are wide ranges of conceptions about model complexity depending on the type of difficulty focused on, and the type of formulations desired for any specific goal (Edmonds, 2000). Defining the complexity level of a model is confounded by the diverse nature of the problems addressed, modeling purpose, assumptions, and limitations. Snowling and Kramer (2001) related model complexity to its structure and the level of details in the processes (i.e., number of parameters, state variables, and the sophistication of the mathematical relationships that explain the modeling processes). Other studies have measured the complexity level only in proportion to the number of optimized parameters (Gan et al., 1997; Perrin et al.,

Table 1
Summary of phosphorus cycling model structures.

	Number of components/state variables	Number of processes	Number of process-specific parameters
Model 1	1	1	1
Model 2	2	2	2
Model 3	3	4	4
Model 4	3	6	6
Model 5	4	9	9
Model 6	4	9	11

2001). That is, the greater the number of optimized parameters, the greater the complexity of the model. In this study, we categorized the complexity level of each model structure in terms of the number of process-specific parameters, p (Table 1), which reflects the number of degrees of freedom and therefore the modeling cost and effort in estimating these parameters. Here both hydrodynamic and biogeochemical parameters were considered in the complexity determination.

2.2. Descriptive adequacy

To evaluate the robustness of a model, it is obvious that the simulated results need to be compared with observations. In general, the term 'descriptive' refers to models that describe an existing knowledge and a known behavior of the system (Costanza and Sklar, 1985). The degree of adequacy with which a particular model describes existing structures and behaviors can be measured in a number of ways. Frequently, a residual (misfit) is used to characterize the model performance based on how well the model is capable of producing a particular data set that is used in the calibration exercise (Friedrichs et al., 2006). Myung et al. (2009) considered the misfit between the model and data as a critical factor in evaluating the relative descriptive adequacy of a set of models. In this study, we used mean squared error (*MSE*) to quantify the misfit between the measured variables (O_{jk}) and simulated variables (S_{jk}) during the model calibration as a sum of residual squares divided by the number of observations of each variable, k:

$$MSE_k = \frac{1}{r}\sum_{j=1}^{r}[O_{jk} - S_{jk}]^2 \quad (1)$$

where r is the number of sampling points in space and time for which simulated values correspond to the available observations of each variable k. In order to make a dimensionless index of misfit error, MSE_k was normalized by the maximum error ($\max[MSE_k]$) among all complexity levels of each variable k, for which observations were available to compare with model-predicted values. The descriptive adequacy index (*DAI*) of variable k for each candidate model was then expressed as:

$$DAI_k = 1 - \frac{MSE_k}{\max[MSE_k]} \quad (2)$$

This formulation results in a relative index that ranges from 0 to 1. The highest value of 1 indicates perfect accuracy and the lowest value of 0 indicates accuracy of the worst performing model among the set of candidate models.

2.3. Predictive adequacy

The term 'predictive adequacy' refers to the ability of a calibrated model to simulate the structural characteristics or behavior of a system outside the existing data boundaries. Descriptive adequacy only describes the performance of a model during calibration for a specific data set; however, it is essential to test the model against the data that are beyond the calibration boundaries (Friedrichs et al., 2006; Myung et al., 2009). A validation exercise is just a robust approach to evaluate the predictive adequacy of a model. The best model should be consistently good when extending the simulations across different conditions. Hence, we employed 'predictive adequacy' as one of the elements of model evaluation. This was quantified by extending model simulations to a new, independent data set that was not used in the calibration. The predictive adequacy index (*PAI*) was determined in the same manner as the *DAI* for validation data.

2.4. Model performance index

The overall performance of each candidate model was represented by a model performance index (*MPI*) that incorporates both descriptive and predictive adequacy indices:

$$MPI = \frac{1}{n_k}\sum_{i=1}^{n_k}(W_d DAI_k) + \frac{1}{n_k}\sum_{i=1}^{n_k}(W_p PAI_k) \quad (3)$$

where n_k is number of state variables for which the measured data were available for the comparison, and W_d and W_p are the weighting constants for *DAI* and *PAI*. The values of these constants are subjectively chosen by considering the importance and availability of spatio-temporal data. In this study, equal weight was given to *DAI* and *PAI* based on the assumption that descriptive and predictive adequacies are equally important. As with *DAI* and *PAI*, *MPI* is a relative index that ranges between 0 and 1.

2.5. Explanatory depth

Modeling goals for environmental systems include not just future forecasting, but also testing hypotheses, analyzing cause-effect relationships, and gaining deeper understanding of system behaviors. Generally, the explanatory ability of a model can be related to the amount of detail it provides about the system components (state variables), and governing processes. A model is an abstraction of a real system; hence it will not reflect all of the reality (Wainwright and Mulligan, 2004). Given the complexity of a system, models developed with higher levels of abstraction generally explain more features of a system structure, not only in principle but also in practice (Beven, 2001). A complex model can provide information about the system that is not available from a simple model (Hannah et al., 2010). We suggest that in evaluating overall model effectiveness, models that describe the system structure with greater detail should be weighted more heavily than simple models. Many modeling studies in environmental science have exhaustively focused on the goodness-of-fit at the expense of evaluating the effects of model complexity on the performance. A narrow focus on data fitting can ignore the capacity of a model to provide more relevant information toward understanding the system behavior.

The term 'explanatory depth' (*ED*) is a more qualitative concept than goodness-of-fit, and is thus often difficult to assess objectively (Beven, 2001). Here, to decrease the subjectivity, we developed an index to reflect the level of detail for describing the system based on the relative number of parameters of each candidate model. It was assumed that increased complexity will increase the *ED* of model m exponentially:

$$ED_m = \frac{\exp(c\Delta p)}{\sum_{m}^{i=1}[\exp(c\Delta p)]} \quad (4)$$

$$\Delta p = (p_m - p_f) \quad (5)$$

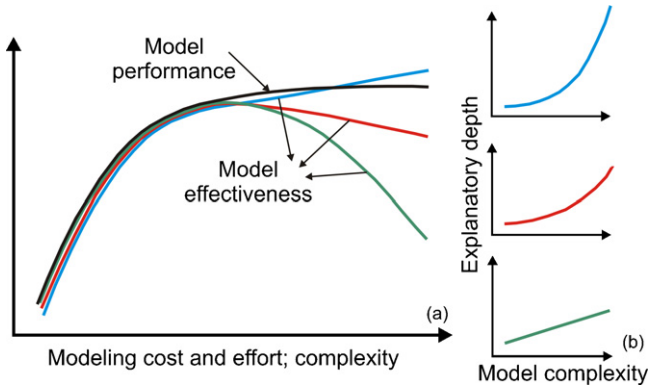

Fig. 1. (a) Conceptual diagram of effect of model performance and explanatory depth on model effectiveness as a function of model complexity. Model performance (black dashed line) is the weighted combination of descriptive and predictive adequacy, in which a larger value indicates a better performance. (b) Exponential increase in explanatory depth with model complexity for *c*-values (Eq. (4)) of 0.4, 0.15, and 0.05 (top to bottom).

where p_m and p_f are the number of process-specific parameters for, respectively, each model m and the model that consists of fewest parameters among the set of models; c is the exponential scale factor that determines the shape of the *ED* curve in relation to the model complexity. The shape of the exponential curve reflects how efficiently additional model processes increase the ability to describe the system, as will be discussed further below.

2.6. Model effectiveness

Model parsimony has become a guiding principle for multi-model selection in environmental science (Burnham and Anderson, 2002). A parsimonious model is the one with the greatest explanation and/or predictive performance in relation to the least process-complexity (Wainwright and Mulligan, 2004). As a first principle, the best model is the one that is most parsimonious in state variables and parameters as well as satisfying the modeling objective. We suggest that the 'best' (most effective) model is the one that optimizes the benefits of increased model performance, as well as providing greater *ED* in relation to the model complexity (modeling cost and effort) (Fig. 1). Effectiveness is an attempt to quantify the trade-offs between these model attributes and answer the question, 'Given the data, modeling questions and resources, how much complexity in the model is appropriate?'

Here, we developed an index to evaluate the effectiveness of each candidate model, represented as a 'coefficient of effectiveness' (*CE*) determined from the weighted resultant of *MPI* and *ED* normalized by p for model m:

$$CE_m = \frac{\sqrt{(K_p MPI_m^2 + K_e ED_m^2)/(K_p + K_e)}}{p_m} \times 100 \quad (6)$$

where K_p and K_e are weighting constants for *MPI* and *ED*, respectively. If both elements are equally valued, $K_p = K_e = 0.5$. The values of CE_m are relative, and can be used only to compare models within the same study.

2.7. Relationship between model performance, explanatory depth, and effectiveness

Our general hypothesis is that as the process-complexity of a phosphorus biogeochemical model of a treatment wetland increases, the model performance also increases. However, at a certain complexity level, the benefits of increased performance become outweighed by the cost of added complexity. Thus, there will be an optimal level of model complexity, quantifiable as the highest *CE* (Fig. 1).

As illustrated in Fig. 1, additional complexity improves model performance. Model effectiveness also increases with complexity – up to a point, after which it may or may not be positively correlated to the model performance as complexity increases. Because of the subjective nature of *ED*, there will be several possible effectiveness curves for corresponding *ED* curves. If the model provides greater useful information (sharp exponential curve; Fig. 1b), such as in the case of process-based climate models, higher values of *ED* result in a high effectiveness scores at high complexity levels. However, if the complex model provides little additional information (close to linear curve; Fig. 1b) then the effectiveness curve sharply declines after a certain complexity level. The coefficient '*c*' determines the shape of the *ED* curve in Eq. (4). In this study, we proposed $c = 0.15$ assuming that in distributed wetland biogeochemical models, additional useful information increases moderately with higher complexity levels, with a type of *ED* curve depicted in the Fig. 1b (middle plot).

2.8. Akaike's information criterion

Akaike (1973) derived an information criterion that has become a fundamental basis for selecting the best model from a set of candidate models (Burnham and Anderson, 2002). Akaike's information criterion (AIC) considers the goodness-of-fit and number of parameters in the model, and is therefore particularly useful in selecting the most parsimonious, best performing model structure, where models use the same data set. This approach has been widely applied in ecological studies (Burnham and Anderson, 2001).

Here, we viewed mechanistic biogeochemical models as a probabilistic model of a dynamic system (Ljung, 1987; McDonald and Urban, 2010), in which each model describes the probability distribution of the data. Residuals are assumed to be normally distributed with a constant variance, σ^2. Generally, when the ratio of the number of data points to the number of parameters (r/p) is small (roughly <40), a bias adjustment term is added to the AIC (Hurvich and Tsai, 1989). Data are often relatively scarce for wetland biogeochemical models such that $r/p < 40$, and this was the case in this study (Models 4–6; see Appendix A). The bias-corrected AIC can be expressed as:

$$AIC_c = -2\ln(L(\hat{\theta})) + 2p + \frac{2p(p+1)}{r-p-1} \quad (7)$$

where r is the number of data points and $L(\hat{\theta})$ is the maximum likelihood function of the parameter vector θ. Note that there is some room for interpretation about how to correctly count p. Burnham and Anderson (2002) emphasized that the parameters that are uniquely estimable from the data should be counted for p. In this study, we used process-specific parameters (Table 2) to estimate AIC_c values of each model. The log-likelihood is (Burnham and Anderson, 2002)

$$ln\ L(\hat{\theta}) = -\frac{1}{2}ln(\sigma^2) - \frac{r}{2}ln(2\pi) - \frac{r}{2} \quad (8)$$

The second and third terms in Eq. (8) are additive constants that can be omitted when using identical data for a set of candidate models (Burnham and Anderson, 2002). AIC_c is a relative value over the candidate set of models considered and the lowest value corresponds to the best model structure.

Table 2
Mean squared error (*MSE*), descriptive adequacy index (*DAI*), predictive adequacy index (*PAI*), model performance index (*MPI*), model efficiency index (R_{eff}), explanatory depth index (*ED*), and AIC_c scores (using outlet TP concentration data for the calibration period, $r = 179$). Note that p includes the phosphorus cycling parameters from Table 1 plus 10 hydrodynamic model parameters described in Paudel et al. (2010).

Model	p	Descriptive adequacy				Predictive adequacy		Outlet TP		*ED*	AIC_c
		Outlet TP		Soil TP		Outlet TP		Calibration	Validation		
		MSE	*DAI*	*MSE*	*DAI*	*MSE*	*PAI*	R_{eff}	R_{eff}		
Model 1	11	142.0	0.00	0.00	0.00	474.0	0.00	−0.22	−2.36	0.07	1395
Model 2	12	107.0	0.25	3.78	0.00	241.0	0.49	0.08	−0.70	0.09	1370
Model 3	14	99.0	0.30	2.90	0.23	190.0	0.60	0.15	−0.35	0.11	1360
Model 4	16	92.0	0.35	2.48	0.34	159.0	0.66	0.21	−0.12	0.16	1353
Model 5	19	91.0	0.36	2.11	0.44	141.0	0.70	0.22	0.00	0.24	1358
Model 6	21	90.0	0.37	1.94	0.49	140.0	0.71	0.23	0.01	0.33	1360

$R_{eff} = 1$ – variance of residuals/variance of observations. R_{eff} ranges from $-\infty$ to 1; and 1 represents the perfect match.

3. Case study: Cell 4 of Stormwater Treatment Area 1 West

3.1. Study site

Stormwater Treatment Area 1 West (STA-1W) is located in the subtropical region of South Florida (Fig. 2). Cell 4 is one of four treatment cells of STA-1W, that formerly were known as the Everglades Nutrient Removal Project (ENRP, Chimney and Goforth, 2006). Cell 4 is a 147 ha marsh that, during the study period of 01/1995 to 10/2001, was dominated by submerged aquatic vegetation (SAV). The bottom elevation in Cell 4 ranged from approximately 2.7 to 3.2 m (NGVD 29). Cell 4 was also the most effective of the four large ENRP cells for phosphorus removal based on areal TP load removal (Chimney et al., 2000). The ENRP was initiated in August 1994 to act as a buffer to reduce concentrations of nutrients such as phosphorus from Everglades Agricultural Area drainage waters before entering the adjacent natural wetlands (Chimney et al., 2000).

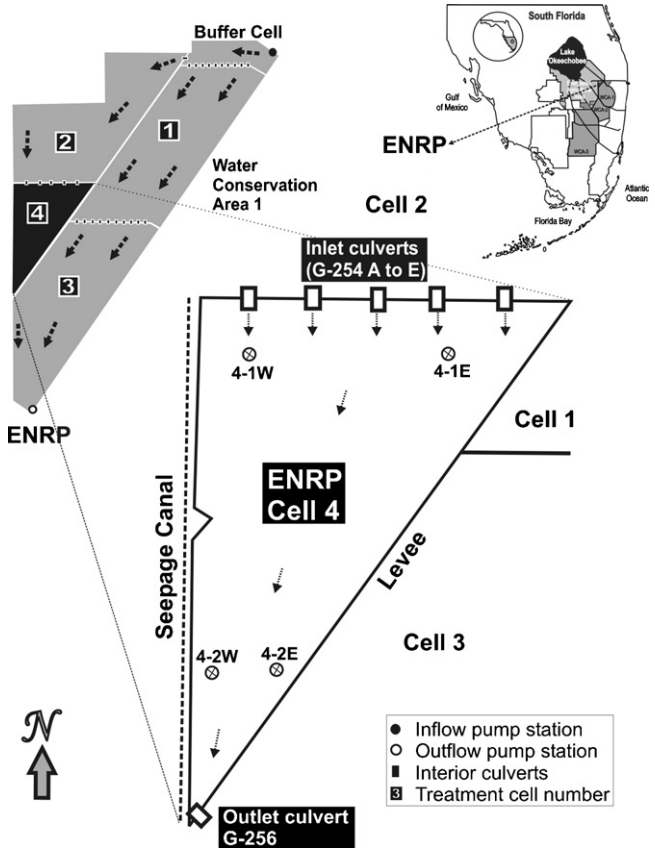

Fig. 2. Location and plan view of study area, Cell 4 of Stormwater Treatment Area 1 West.

3.2. Modeling framework

A pre-calibrated two-dimensional (2-D) hydrodynamic model of Cell 4, STA-1W (Paudel et al., 2010) was used to provide hydrodynamic data, such as depth and velocity fields, to candidate biogeochemical models. Hydrologic simulations were conducted with the physically based Regional Simulation Model (RSM, Lal et al., 2005; SFWMD, 2005), which simulates the coupled movement and distribution of overland and groundwater flow. A water quality module that was coupled with RSM (RSMWQ, James and Jawitz, 2007; Jawitz et al., 2008; James et al., 2009) was used to simulate phosphorus transport and reaction equations.

3.3. Biogeochemical models

Predicting the phosphorus behavior is often challenging and uncertain in wetland systems such as Stormwater Treatment Areas (STAs) that comprise many interconnected components that interact nonlinearly with many feedback loops. Given the inherent complexity of these systems, it is often difficult to set an appropriate level of abstraction in the model. Here, we developed a set of six mechanistic phosphorus biogeochemical models with hierarchal levels of complexity in their process representations. The model processes were selected on the basis of available data, existing knowledge about the system, and experience from previous wetland models (Kadlec, 1997; DBEL, 2002a,b; Dierberg et al., 2005; Kadlec and Wallace, 2008; Walker and Kadlec, 2011). Each model was implemented in the framework of RSMWQ to simulate phosphorus dynamics using available data from Cell 4 of STA-1W. Components, physical processes or parameters were added in a sequence of relevance to increase the level of complexity. All models were applied to the same flow and biogeochemical data; therefore, the model structures were directly compared.

Conceptually, phosphorus cycling in a wetland is the transfer (or flow) of this element in various forms between different stores, often referred to as 'state variables' (e.g., water column, macrophyte, soil, and periphyton). Process diagrams of all six model structures are depicted in Fig. 3 in terms of transfers between stores. Transfer processes, such as settling, release, plant uptake, and burial of TP from senesced plants were described either by linear (e.g., first-order) or nonlinear (typically second-order, and Monod types of transformations) differential equations (see Appendix A). The number of components/state variables, processes and parameters of each model are listed in Table 1.

In the simplest complexity level (Model 1) all phosphorus removal processes were 'lumped' together as a net settling of TP from the water column. This representation included only a single

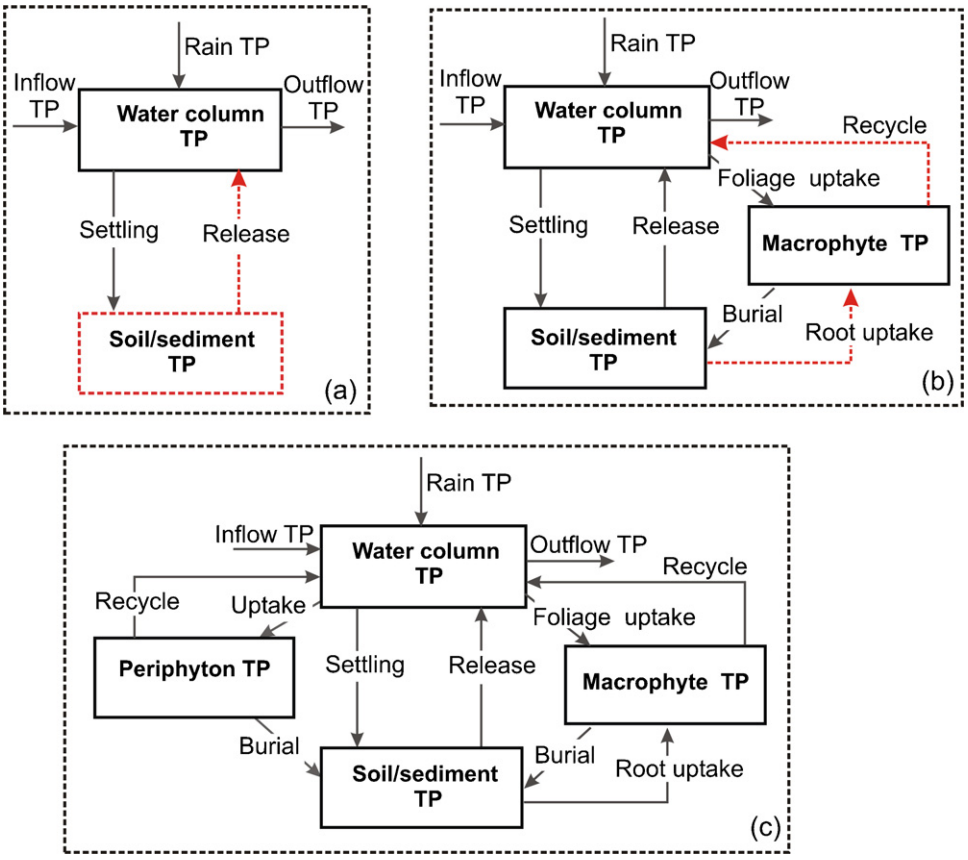

Fig. 3. Conceptual process diagram of a set of phosphorus biogeochemical models. Each rectangle represents the component (i.e., state variable), and the arrow represents the process. (a) Models 1 and 2: Model 1 considers only the water column TP store and settling process (black rectangle and arrows), but Model 2 extended Model 1 by including soil TP store and a release mechanism (indicated in red); (b) Models 3 and 4: Model 3 does not consider macrophyte TP recycle, and root uptake (shown in red dashed lines) but Model 4 considers all components and processes shown in the figure; (c) Models 5 and 6: Model 6 simulates periphyton and macrophyte TP uptake with phosphorus-limited Michaelis–Menten type equation with more parameters than Model 5. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

component (state variable) and a single equation. The apparent net settling rate approach has been frequently used to describe phosphorus removal in wetlands (Walker, 1995; Raghunathan et al., 2001; Kadlec and Wallace, 2008). Model 2 added the effect of an additional component (e.g., soil TP), in which all phosphorus reactions were lumped as soil-water uptake and release (Paudel et al., 2010). These processes were described by first-order reactions, and reflect the effects of a combination of several physical and biochemical processes between the water column and soil TP stores. Water column TP was considered as a mobile (transportable) component, whereas the soil TP was assigned to be stabile (stationary). Walker and Kadlec (2011) used a similar model structure in the STAs by simulating TP transfer between two storage compartments (water column and biomass); however, they included three transfer processes. Model 3 included the processes of Model 2 and added TP storage in macrophytes and transfer processes such as foliage uptake and burial. Model 4 built on Model 3 by adding macrophyte TP recycle back to the water column, and root uptake from the soil TP store (Min et al., 2011). Model 5 extended Model 4 by including periphyton TP as an additional component and three transfer processes. The uptake of TP by periphyton was assumed to be proportional to the water column TP concentration. It was presumed that TP losses from macrophyte and periphyton primarily occur during senescence/decay; therefore, TP loss from these components was represented by first-order reactions. A fraction of both macrophyte and periphyton TP loss was recycled back into the water column, while the remaining TP fraction was buried and assimilated with soil TP (Kadlec, 1997). In Model 6, the complexity level was further increased by considering the growth dynamics of macrophyte and periphyton. The growth dynamics of these plant communities were modeled using Michaelis–Menten type kinetics (Scinto and Reddy, 2003) and TP transfer during growth was estimated based on their corresponding TP fractions (see Appendix A). The TP transfer processes between components were modeled either by using linear first-order or non-linear equations.

3.4. Model setup

The wetland (Cell 4) was represented by a 2-D finite element mesh of 298 unstructured triangular elements (average area: 5100 m^2) and 192 nodes for all simulation models. Biogeochemical models were fit to weekly auto-sampled composite (time-averaged) TP concentrations, monitored at outlet hydraulic structure (Fig. 4). In each simulation, TP inflow concentrations monitored at inlet hydraulic structures (G-254B, G-254D; Fig. 2) were specified as a source boundary condition (Fig. 4). Initial water column TP concentration of 40 $\mu g\,L^{-1}$ was specified as a spatially constant value throughout the model domain, which was the average measured value from two sampling locations (ENR401 and ENR402; Fig. 2) on 10 January 1995, the beginning date of the simulations. A spatially constant TP wet deposition of 10 $\mu g\,L^{-1}$ was applied over the entire domain based on measurements at STA-1W (Ahn and James, 2001).

Initial values for TP concentration in the upper 10 cm of soil were based on samples collected by South Florida Water Management District (SFWMD) on 20 January 1995 at four sampling stations

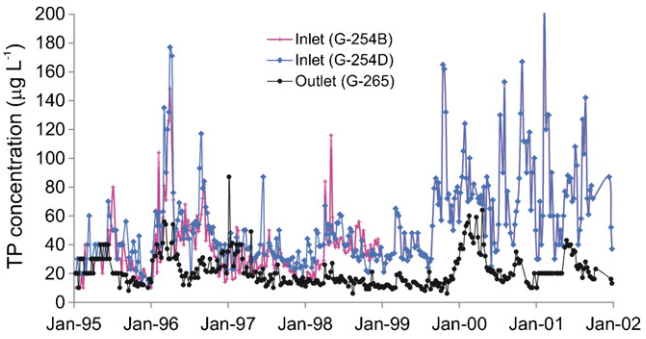

Fig. 4. Inflow and outflow TP concentrations used in the model. Inflow TP concentration data at G-254B structure were only available up to 15 December 1998.

within Cell 4 (4-2E, 4-2W, 4-1W, and 4-1E; Fig. 2). Peat accretion was monitored from mid-1995 to mid-1999 using feldspar horizon markers throughout STA-1W (Chimney et al., 2000). Soil TP measured in samples collected at 4-1E and 4-2W on 12 November 1998 and 20 October 1999 were compared to model-simulated soil TP levels at these dates.

For Models 3–6 that included macrophyte dynamics, the initial macrophyte TP concentration was based on a study at the outlet zone of Cell 4, where *Najas*- and *Ceratophyllum* vegetation species were dominant (DBEL, 2004). These species were considered as representative of the SAV biomass for the entire Cell 4 (DBEL, 2002b; Dierberg et al., 2005) and were also considered as dominant for the simulation period. The initial areal periphyton TP for Models 5 and 6 was based on data from eutrophic regions of the Everglades wetland (McCormick et al., 1998). Simulated macrophyte and periphyton TP components were maintained at close to equilibrium at any given date, as time-series field observations of these variables were not available.

In all model structures, phosphorus cycling processes were described by ordinary differential equations that were solved by fourth-order Runge–Kutta numerical integration methods. The initial values of kinetic rate constants and other parameters in most pathways were based on the previous models of STAs, available knowledge, and the treatment wetland literature (Buzzelli et al., 2000; Paudel et al., 2010; Walker and Kadlec, 2011). Later, these initial values were adjusted during the model calibration as field-measured empirical values were not available. As spatial phosphorus data were limited, biogeochemical model parameters were specified as spatially constant values throughout the model domain.

3.5. Calibration

The unconstrained model parameters were adjusted through calibration against observations to produce an adequate fit between model results and measured data (see Appendix B). In practice, mechanistic biogeochemical model parameters are generally calibrated by trial and error (Robson et al., 2008). The level of expertise (understanding about the system dynamics and model structures) of a modeler and data availability often dictate the ability to produce best fit with near optimal parameters. Here, we adopted a trial and error approach to calibrate the model by adjusting individual parameters within literature ranges because spatially explicit models, as in our case, are computationally intensive and prohibited the use of formal comprehensive calibration techniques.

The long-term water column TP concentration profiles measured at weekly intervals from 1995 to 1998 at the wetland outlet hydraulic structure (G-256) and soil TP concentrations measured

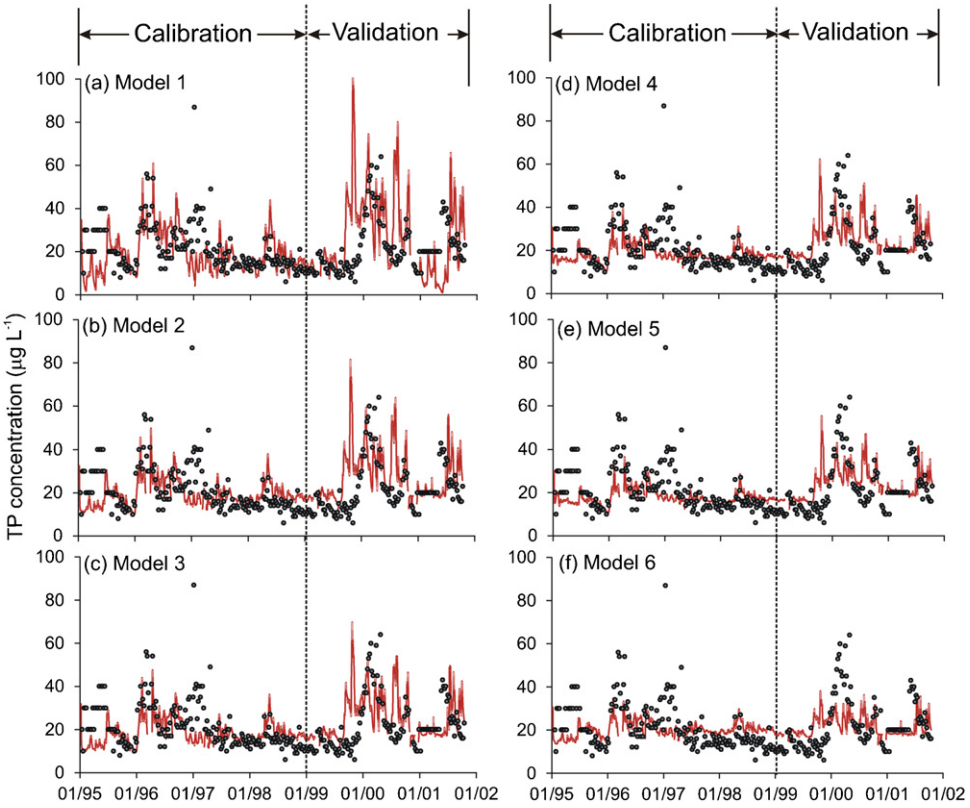

Fig. 5. Observed (gray symbols) and modeled (red lines) for the outlet TP concentrations during calibration, and validation periods. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

over the 5-year period 1995–1999 were simultaneously used to fit the models. Soil TP data included three sampling events (20 January 1995, 12 November 1998, and 20 October 1999) at two locations within Cell 4 (4-1E and 4-2W). Model parameters were adjusted until optimal or near optimal fits were obtained based on the lowest *MSE* of each state variable for which field observations were available. Final values of model parameters are listed in Appendix B.

3.6. Validation

For the validation, all simulations were performed against independent outlet TP concentration data for about 3 years (1 January 1999–31 October 2001) to evaluate the predictive adequacy of calibrated models. Model validation was limited to water column TP data because of the limited availability of soil TP data. We re-initialized the model inputs and extended simulations without modifications in the calibrated parameters.

4. Results

4.1. Model performance

The simulated and measured outlet TP profiles during the model calibration period are presented in Fig. 5. In general, descriptive error decreased with increasing level of model complexity. Model 6 (most complex model) resulted in the lowest misfit error (*MSE* = 90) and Model 1 (simplest model) resulted in the highest misfit error (*MSE* = 142). The introduction of an additional state variable (soil TP) in Model 2 significantly improved the descriptive performance (Table 2). The gain in *DAI* by adding the periphyton TP state variable in Model 5, however, is clearly much less pronounced comparing to the gain in *DAI* by adding macrophyte TP in Model 3. A comparison between measured and simulated soil TP shows that all levels of complexity successfully reproduced the spatio-temporal variation in soil TP over a 5-year period (Fig. 6), with higher concentrations near the cell inlet and lower concentrations near the outlet, both increasing over time. As the complexity level increased, there was a slight improvement in the model fit to the data from station 4-1E; however, the improvement was insignificant for the 4-2W station (Fig. 6).

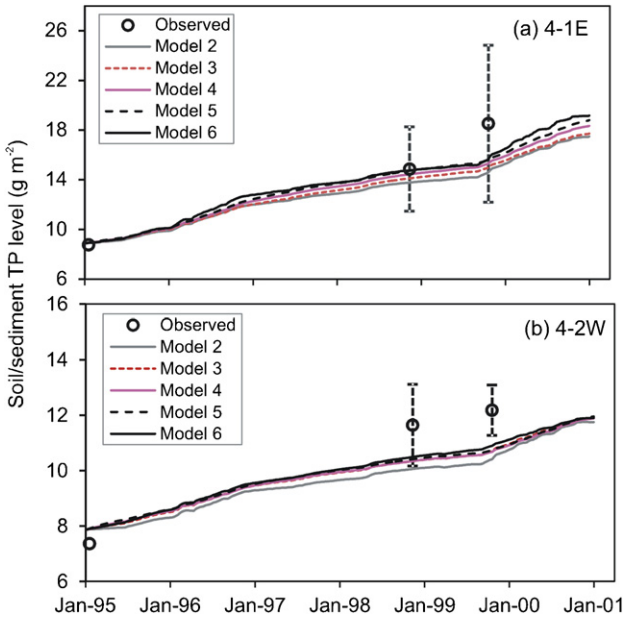

Fig. 6. Simulated and observed soil TP content at two locations within Cell 4 for the upper 10 cm of the soil profile: (a) upstream area, 4-1E and (b) downstream area, 4-2W. Observed values are shown as Mean ± 1SD.

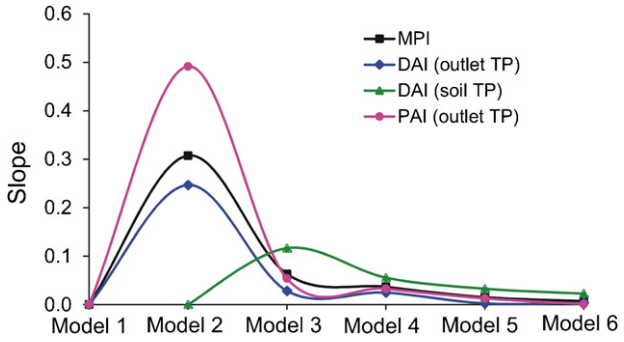

Fig. 7. Slope of each model attributes over model complexity for different variables. For example, slope of model performance (*MPI*) is the $\Delta MPI/\Delta p$ between consecutive models.

A similar trend of decreased error with increased complexity level was also observed in the prediction errors during the validation period (Fig. 5). The misfit error was the highest for the simplest model (Model 1; *MSE* = 474) and lowest for the most complex model (Model 6; *MSE* = 140) (Table 2). The corresponding *PAI* values ranged from 0.0 (Model 1) to 0.71 (Model 6), where zero is the worst performing model and 1 indicates perfect accuracy. As the complexity level further increased, *PAI* increased rapidly at the beginning (low complexity level); whereas the *PAI* increased monotonically at the end (high complexity level), with a diminished rate of improvement in the performance (Table 2). These results indicate that the influence on predictive performance of increasing the level of complexity is significant at the lower end of the complexity spectrum. A variable slope (Fig. 7) illustrates that every free parameter did not contribute equally to the ability of the model to fit the observations.

The overall model performance was evaluated in terms of *MPI*, which combines both calibration and validation adequacies for each state variable for which observations were available to compare with simulated values. The performance increased substantially from Models 1 to 2, as shown by the steep *MPI* curve in Fig. 7. The curve flattened notably from Models 5 to 6, indicating that increasing model complexity was not supported by the available field data at upper end of the complexity spectrum.

4.2. Model selection

The coefficient of effectiveness (*CE*) is shown as a function of model complexity index with or without considering the *ED* in

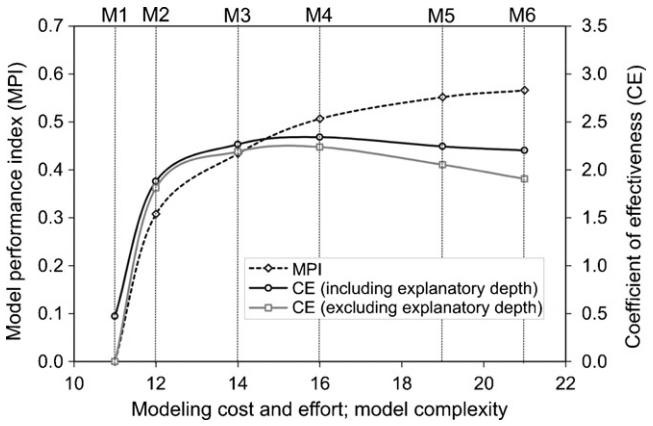

Fig. 8. Model performance index, *MPI* (left *y*-axis) and coefficient of effectiveness, *CE* (right *y*-axis) as a function of modeling cost and effort (i.e., model complexity). M1–M6 corresponds to Models 1–6.

Fig. 8. In this study, the most effective models (i.e., Models 3 and 4, Fig. 8) were found to be those of intermediate complexity. Without considering the *ED*, the effectiveness curve shifted down particularly at higher complexity zone.

AIC_c decreased from Models 1 to 4, but increased again from Models 4 to 6 (Table 2). The inflection point of AIC_c was identified at intermediate complexity, as indicated by the minimum value with Model 4 ($AIC_c = 1353$). Due to the large number of parameters and uncertainties in the structure of the complex model, the miss-fit error due to variance outweighed the miss-fit error due to bias; hence, the AIC_c-value increased again. Based on this criterion, Model 4 was found to be the best model.

5. Discussion

5.1. Influence of complexity level on model performance

Our analysis illustrates that sequentially increasing the level of complexity in a phosphorus cycling model of a wetland system also increased the model performance. Adding each biogeochemical process in the model contributed to the performance but not in an equal proportion. The largest increase in the performance was observed when the number of model compartments increased from one to two (i.e., soil TP was added as a state variable from Models 1 to 2). This is primarily because the simplest model did not include adequate mechanisms to describe the key features of the TP concentration dynamics. Thus, consideration of two TP transfer processes (e.g., settling and release) between two compartments is critical for representing wetland TP biogeochemical cycling. Model 2 included the release mechanism that represented the aggregated effects of TP release processes to the water column. For example, soil TP can be released by mechanisms such as diffusion and resuspension of sediments to the water column, and may include the exchange of both labile as well as non-labile phosphorus. As settling and release are represented here as aggregated processes between soil and the water column, similar performance may have been achieved if the two exchange processes had been between the water column and vegetation stores, as in the DMSTA model described by Walker and Kadlec (2011).

Including vegetation TP as a state variable with net uptake and burial mechanisms as in Model 3 significantly improved the model performance (Table 2). Thus, with the given data set, expanding the model size from Models 1 to 3 is justified. Further expansion to Models 4–6 may need to be carefully considered because the rate of improvement in the performance was gradually reduced (Fig. 7). This was mainly due to the uncertainty associated with additional variables and parameters (Snowling and Kramer, 2001; Friedrichs et al., 2006) and unavailability of field data to independently constrain each process parameter.

The relative performance of the models evaluated here increased substantially with model complexity. However the absolute performance of even the most complex model was only comparable to predictions using the mean of the observed data (i.e., $R_{eff} \sim 0$, Table 2). Another way to evaluate consistent behavior between observed data and modeled trends is to check if the inputs and outputs are correlated. Outflow phosphorus profiles in the STAs have been found to be only weakly correlated with inflow phosphorus loading rate (Juston and DeBusk, 2006). Because simple settling rate models, such as Model 1, exhibit strong correlation between inlet and outlet concentrations, such models are unlikely to adequately predict the outflow concentration variability. These authors subsequently found (Juston and DeBusk, 2011) that adding treatment wetland TP cycling process complexity at levels generally consistent with our Model 2 (two compartments and approximately 3 process parameters), did not improve model predictive performance when concentrations were at very low levels, concluding that a suite of other unidentified processes controlled an apparent 'background concentration'. Here, we evaluated several additional levels of complexity in Models 3–6, and found that the performance increased substantially, but there still remain factors or mechanisms that induced the observed variability in the data that are not correctly reflected in the models.

5.2. Model effectiveness

For the set of models considered here, the model effectiveness showed a 'humped' curvilinear relationship with respect to the model complexity. This indicates that there may exist an optimal level of model complexity corresponding to the maximum effectiveness (Fig. 8). As effectiveness is a function of both model performance and explanatory depth, the shape of the curve is partly determined by how much information the model can provide. For example, three dimensional global climate models (e.g., GCMs) that couple components of atmosphere, ocean, cryosphere, and biosphere may include increasingly realistic representations of numerous processes of the climate system (Harvey, 2004). Such models tend to be of complex structure; however, they provide a number of useful insights concerning the dynamics of large number of output of interests and interactions among components. In such a case, *ED* would follow a sharp exponential curve (top plot of Fig. 1b, with a relatively high *c*-value in Eq. (4), e.g., 0.4) and likely result in complex models being highly effective. In contrast, simple wetland treatment performance models are thought to be capable of consistently describing the system performance (Kadlec, 2000; Walker and Kadlec, 2011). A complex model structure may not provide much more useful information than a simple one. In such a case, *ED* would follow a near-linear curve (bottom plot of Fig. 1b) with relatively low *c* (e.g., 0.05), and the resulting effectiveness curve would sharply decline after a point on the complexity spectrum. Based on calculated *ED* values for each model complexity (Table 2), a maximum point on the concave-downward effectiveness curve was found, which corresponds to the optimal level of complexity (Fig. 8). Even though, complex models (e.g., Models 5 and 6) included better process descriptions than Model 4, the diminished improvement in the performance resulted in lower effectiveness scores in the expense of higher complexity. Although the results and interpretation of this analysis was dependent on the models under consideration, available data, and the definition of model effectiveness used here, the pattern in the results may be valid across other type of mechanistic biogeochemical modeling approaches.

Based on the AIC_c analysis, Model 4 was found to be the most parsimonious model for simulation of phosphorus dynamics in the STAs. Model 4 compromised between accuracy and complexity and was best supported by the calibration data with the fewest parameters. In a statistical sense, Model 4 balanced the bias (goodness-of-fit) and variance (uncertainties in estimated parameters) of the misfit error. Models 5 and 6 were penalized for the inclusion of additional parameters in relation to the diminished improvement in model fit. The finding of this approach is consistent with the model effectiveness scores (Fig. 8). Neither approach suggested the most complex model as a best approximating model. Although both approaches identified the same model (Model 4) as one of the best models, AIC_c technique may not be adequate if the modeling goal is to provide understanding about the system dynamics because it does not account for *ED* which may limit the scope in wetland biogeochemical modeling.

While we rigorously evaluated the important properties of model complexity evaluation, there are still other attributes not included in this study that can potentially influence the evaluation

process, such as quality and availability of data, uncertainties in model parameters, understandability, modeling goal, available resources, and the expertise of a modeler. As noted by Pitt and Myung (2002) "objectively comparing competing models is no easier or less subjective than choosing between competing theories." Hence, the evaluation approaches used here should be viewed as a tool to deepen our understanding about the assessment of competing wetland models.

5.3. Future directions for treatment wetland phosphorus biogeochemical modeling

A primary challenge in formulating a phosphorus biogeochemical model of a treatment wetland is how to provide a reasonable representation of a phosphorus cycling structure to address the modeling questions at hand. This challenge is often characterized as a choice between adding complexity to explicitly represent more of the components of the system and using the fewest number of components and parameters to make the model simple. Often, a modeling goal or the issue under investigation defines the complexity level in the model (Haraldsson and Sverdrup, 2004). For example, is the model intended for use as a management or a scientific research tool? A complex model with greater *ED*, such as Model 5 or 6, may be an appropriate choice to use as a scientific research tool for testing research hypotheses or gaining understanding about the system components and interactions. More complex models include more components and processes, and generally represent complex systems more comprehensively than simple models. Beven (2001) emphasized that processes that are perceived to have an effect in the real system should be included in the model. Even if data are not available to comprehensively test the model, complex models may be useful, particularly for testing a wide range of interactions/mechanisms, and predicting more state variables.

If the models are intended for management purposes, simple model structures such as in Models 1 and 2 may also satisfy the modeling needs because such models consist of few parameters and are relatively easy to analyze. Simple models can serve as hands-on tools for managers to evaluate wetland performance as well as performing multiple diagnostic tests of management alternatives. For example, DMSTA (Walker and Kadlec, 2011), which has only two storage compartments (water column and biomass) and three processes (settling, recycle and burial) in a hydraulic reactor modeling framework (i.e., implicit with no spatial representation) is widely used for this purpose. Because of fewer spatial details and fast run times associated with numerical computations, the model is tractable, and involves less effort for calibration and application. However, if the management questions demand complex, heterogeneous systems to be simulated, the case for spatially distributed model is strong. In such circumstances, Models 1–6 may provide greater utility because these models can represent spatial characteristics and predict local variability within the wetland. The choice of a specific phosphorus model structure may depend on the requirements for a given research or management objective. It should also be noted that an important limitation of distributed models is that their computational intensity constrains how comprehensively uncertainty analysis of parameters can be performed.

A fundamental approach for identifying the appropriate level of complexity is to start from the simplest model representations and test whether field observations are reproduced reasonably, and introduce added complexity only when the model performance consistently improves and adequate information is obtained for given questions. Such approaches help identify the potential convergence to an optimal level of model complexity, where the field data are fully exploited with adequate information while the model complexity is minimized (Schoups and Hopmans, 2006). From theoretical perspectives, it may be interesting to incorporate more complete representations of phosphorus cycling structures into wetland models; however, adding more of the wrong processes will not improve the performance. Therefore, a mechanistic understating of key processes that mediate phosphorus cycling in wetlands is critical to identify whether additional processes in the model are justified. For these reasons, model developers are recommended to evaluate models across varying level of complexity.

Acknowledgments

This research was supported by the United States Geological Survey and the South Florida Water Management District. Dr. Andrew I. James was the primary developer of the transport and reaction simulation code. Also, Drs. Zaki Moustafa and Naiming Wang of SFWMD provided spatial data, and valuable comments and suggestions to improve this work.

Appendix A. Phosphorus biogeochemical models of increasing complexity

Model 1

$$\frac{dC_{tp}}{dt} = -k_{st1}C_{tp}$$

Model 2

$$\frac{dC_{tp}}{dt} = -k_{st2}C_{tp} + \frac{k_{rs2}}{z_d\theta_{wc}}S_{tp}$$

$$\frac{dS_{tp}}{dt} = k_{st2}z_d\theta_{wc}C_{tp} - k_{rs2}S_{tp}$$

Model 3

$$\frac{dC_{tp}}{dt} = -k_{st3}C_{tp} + \frac{k_{rs3}}{z_d\theta_{wc}}S_{tp} - k_{up\_mf3}C_{tp}$$

$$\frac{dM_{tp}}{dt} = k_{up\_mf3}C_{tp}z_d\theta_{wc} - k_{b3}M_{tp}$$

$$\frac{dS_{tp}}{dt} = k_{st3}z_d\theta_{wc}C_{tp} - k_{rs3}S_{tp} + k_{b3}M_{tp}$$

Model 4

$$\frac{dC_{tp}}{dt} = -k_{st4}C_{tp} + \frac{k_{rs4}}{z_d\theta_{wc}}S_{tp} + \frac{k_{rc4}}{z_d\theta_{wc}}M_{tp} - k_{up\_mf4}C_{tp}$$

$$\frac{dM_{tp}}{dt} = k_{up\_mf4}C_{tp}z_d\theta_{wc} - k_{rc4}M_{tp} - k_{b4}M_{tp} + k_{up\_mr4}S_{tp}$$

$$\frac{dS_{tp}}{dt} = k_{st4}z_d\theta_{wc}C_{tp} - k_{rs4}S_{tp} + k_{b4}M_{tp} - k_{up\_mr4}S_{tp}$$

Model 5

$$\frac{dC_{tp}}{dt} = -k_{st5}C_{tp} + \frac{k_{rs5}}{z_d\theta_{wc}}S_{tp} + \frac{k_{d\_m5}\alpha M_{tp}}{z_d\theta_{wc}} + \frac{k_{d\_p5}\beta P_{tp}}{z_d\theta_{wc}} - k_{up\_mf5}C_{tp} - k_{up\_p5}C_{tp}$$

$$\frac{dM_{tp}}{dt} = k_{up\_mf5}C_{tp}z_d\theta_{wc} - k_{d\_m5}M_{tp} + k_{up\_mr5}S_{tp}$$

$$\frac{dP_{tp}}{dt} = k_{up\_p5}C_{tp}z_d\theta_{wc} - k_{d\_p5}P_{tp}$$

$$\frac{dS_{tp}}{dt} = k_{st5}z_d\theta_{wc}C_{tp} - k_{rs5}S_{tp} + k_{d\_m5}M_{tp}(1-\alpha) + k_{d\_p5}P_{tp}(1-\beta) - k_{up\_mr5}S_{tp}$$

Model 6

$$\frac{dC_{tp}}{dt} = -k_{st6}C_{tp}^2 + \frac{k_{rs6}}{z_d\theta_{wc}}S_{tp} + \frac{k_{rec\_m6}M_{tp}^2}{z_d\theta_{wc}} + \frac{k_{rec\_p6}P_{tp}^2}{z_d\theta_{wc}} - r_m\left(\frac{C_{tp}}{C_{tp}+k_{c\_m}}\right)\frac{M_{tp}}{z_d\theta_{wc}} - r_p\left(\frac{C_{tp}}{C_{tp}+k_{c\_p}}\right)\frac{P_{tp}}{z_d\theta_{wc}}$$

$$\frac{dM_{tp}}{dt} = r_m\left(\frac{C_{tp}}{C_{tp}+k_{c\_m}}\right)\frac{M_{tp}}{z_d\theta_{wc}} - k_{b\_m6}M_{tp} - k_{rec\_m6}M_{tp}^2 + k_{up\_mr6}S_{tp}$$

$$\frac{dP_{tp}}{dt} = r_p\left(\frac{C_{tp}}{C_{tp}+k_{c\_p}}\right)\frac{P_{tp}}{z_d\theta_{wc}} - k_{b\_p6}P_{tp} - k_{rec\_p6}P_{tp}^2$$

$$\frac{dS_{tp}}{dt} = k_{st6}z_d\theta_{wc}C_{tp} - k_{rs6}S_{tp} + k_{b\_m6}M_{tp} + k_{b\_p6}P_{tp} - k_{up\_mr6}S_{tp}$$

where S_{tp} = soil TP storage ($g\,m^{-2}$); C_{tp} = water column TP storage ($g\,m^{-3}$); z_d = water column depth (m); θ_{WC} = water column porosity (unitless); M_{tp} = macrophyte TP storage ($g\,m^{-2}$); and P_{tp} = periphyton TP storage ($g\,m^{-2}$).

Note: (1) The subscript at the end of each parameter corresponds to the level of complexity models (Models 1–6).

(2) Parameter descriptions are given in Appendix B.

Appendix B. Parameters determined through model calibration for the set of phosphorus cycling models

Parameter	Definition	Units	Value
k_{st1}	TP settling rate	d^{-1}	0.1425
k_{st2}	TP settling rate	d^{-1}	0.2678
k_{rs2}	TP release rate as a function of the soil TP	d^{-1}	1.97×10^{-4}
k_{st3}	TP settling rate	d^{-1}	0.19
k_{rs3}	TP release rate as a function of the soil TP	d^{-1}	1.97×10^{-4}
k_{up_mf3}	Macrophyte foliage TP uptake rate	d^{-1}	0.1
k_{b3}	Macrophyte TP burial rate	d^{-1}	2.70×10^{-3}
k_{st4}	TP settling rate	d^{-1}	0.22
k_{rs4}	TP release rate as a function of the soil TP	d^{-1}	1.97×10^{-4}
k_{up_mf4}	Macrophyte foliage TP uptake rate as a function of the water column TP	d^{-1}	0.21
k_{b4}	Macrophyte TP burial rate	d^{-1}	2.70×10^{-3}
k_{up_mr4}	Macrophyte root TP uptake rate as a function of the soil TP	d^{-1}	2.20×10^{-5}
k_{rc4}	Macrophyte TP recycle rate	d^{-1}	3.50×10^{-3}
k_{st5}	TP settling rate	d^{-1}	0.18
k_{rs5}	TP release rate as a function of the soil TP	d^{-1}	1.97×10^{-4}
k_{up_mf5}	Macrophyte foliage TP uptake rate as a function of the water column TP	d^{-1}	0.3
k_{up_p5}	Periphyton TP uptake rate	d^{-1}	0.03
k_{up_mr5}	Macrophyte root TP uptake rate	d^{-1}	2.20×10^{-5}
k_{d_m5}	Decay rate of macrophyte TP	d^{-1}	9.00×10^{-3}
k_{d_p5}	Decay rate of periphyton TP	d^{-1}	0.01
α	Recycled fraction of macrophyte TP	–	0.56
β	Recycled fraction of periphyton TP	–	0.50
k_{st6}	TP settling rate	$m^3\,mg^{-1}\,d^{-1}$	0.01036
k_{rs6}	TP release rate	d^{-1}	1.97×10^{-4}
r_m	Macrophyte intrinsic growth rate	d^{-1}	4.00×10^{-3}
r_p	Periphyton intrinsic growth rate	d^{-1}	7.00×10^{-3}
k_{c_m}	Half saturation constant for P as limiting nutrient in macrophyte growth	$g\,m^{-3}$	0.01
k_{c_p}	Half saturation constant for P as limiting nutrient in periphyton growth	$g\,m^{-3}$	0.015
k_{up_mr6}	Macrophyte root TP uptake	d^{-1}	2.20×10^{-5}
k_{b_m6}	Macrophyte TP burial rate	d^{-1}	2.70×10^{-3}
k_{b_p6}	Periphyton TP burial rate	d^{-1}	4.80×10^{-3}
k_{rec_m6}	Macrophyte TP recycle rate	$m^2\,g^{-1}\,d^{-1}$	3.11×10^{-3}
k_{rec_p6}	Periphyton TP recycle rate	$m^2\,g^{-1}\,d^{-1}$	3.45×10^{-3}

References

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Proceedings of the 2nd International Symposium on Information Theory Tsahkadsor, Armenian SSR, pp. 267–281.

Ahn, H., James, R.T., 2001. Variability, uncertainty, and sensitivity of phosphorus deposition load estimates in South Florida. Water Air Soil Pollut. 126, 37–51.

Arhonditsis, G.B., Brett, M.T., 2004. Evaluation of the current state of mechanistic aquatic biogeochemical modeling. Mar. Ecol. Prog. Ser. 271, 13–26.

Beven, K., 2001. On explanatory depth and predictive power. Hydrol. Process. 15, 3069–3072.

Beven, K., 1996. A discussion of distributed hydrological modeling. In: Abbott, M.B., Refsgaard, J.C. (Eds.), Distributed Hydrological Modeling. Kluwer Academic, pp. 255–278.

Box, G.E.P., Jenkins, G.M., 1970. Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco, USA.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach. Springer-Verlag, NY, USA, p. 488.

Burnham, K.P., Anderson, D.R., 2001. Kullback–Leibler information as a basis for strong inference in ecological studies. Wildl. Res. 28, 111–119.

Buzzelli, C.P., Childers, D.L., Dong, Q., Jones, R.D., 2000. Simulation of periphyton phosphorus dynamics in Everglades National Park. Ecol. Model. 134, 103–115.

Chimney, M.J., Goforth, G., 2006. History and description of the Everglades Nutrient Removal Project, a subtropical constructed wetland in south Florida (USA). Ecol. Eng. 27, 268–278.

Chimney, M.J., Nungesser, M., Newman, J., Pietro, K., Germain, G., Lynch, T., Goforth, G., Moustafa, M.Z., 2000. Chapter 6: Stormwater Treatment Areas—status of research and monitoring to optimize effectiveness of nutrient removal and annual report on operational compliance. In: SFWMD, 2000 Everglades Consolidated Report, West Palm Beach, FL. SFWMD, pp. 6-1–6-127.

Christensen, N., Mitsch, W.J., Jorgensen, S.E., 1994. A first generation ecosystem model of the Des-Plaines river experimental wetlands. Ecol. Eng. 3, 495–521.

Costanza, R., Sklar, F.H., 1985. Articulation, accuracy and effectiveness of mathematical-models—a review of fresh-water wetland applications. Ecol. Model. 27, 45–68.

Cox, G.M., Gibbons, J.M., Wood, A.T.A., Craigon, J., Ramsden, S.J., Crout, N.M.J., 2006. Towards the systematic simplification of mechanistic models. Ecol. Model. 198, 240–246.

DB Environmental Laboratories, Inc. (DBEL), 2002. Submerged aquatic vegetation/limerock treatment system technology: follow-on study: STSOC Report. Report Prepared for SFWMD and Florida Department of Environmental Protection, West Palm Beach, FL.

DB Environmental Laboratories, Inc. (DBEL), 2002. Submerged aquatic vegetation/limerock treatment system technology for removing phosphorus from Everglades Agricultural Area waters follow on study—final report. Prepared for the SFWMD, West Palm Beach, FL and the Florida Department of Environmental Protection, Tallahassee, FL.

DB Environmental Laboratories, Inc. (DBEL), 2004. Appendix 4B-4: vegetation biomass and nutrient analysis for Stormwater Treatment Area 1 West (STA-1W). In: SFWMD, 2004 Everglades Consolidated Report. SFWMD, West Palm Beach, FL.

Dierberg, F.E., Juston, J.J., DeBusk, T.A., Pietro, K., Gu, B.H., 2005. Relationship between hydraulic efficiency and phosphorus removal in a submerged aquatic vegetation-dominated treatment wetland. Ecol. Eng. 25, 9–23.

Edmonds, B., 2000. Complexity and scientific modelling. Found. Sci. 5, 379–390.

Flynn, K.J., 2005. Castles built on sand: dysfunctionality in plankton models and the inadequacy of dialogue between biologists and modellers. J. Plankton Res. 27, 1205–1210.

Friedrichs, M.A.M., Hood, R.R., Wiggert, J.D., 2006. Ecosystem model complexity versus physical forcing: quantification of their relative impact with assimilated Arabian Sea data. Deep Sea Res. II: Top. Stud. Oceanogr. 53, 576–600.

Fulton, E.A., Smith, A.D.M., Johnson, C.R., 2003. Effect of complexity on marine ecosystem models. Mar. Ecol. Prog. Ser. 253, 1–16.

Fulton, E.A., 2001. The effects of model structure and complexity on the behaviour and performance of marine ecosystem models. PhD thesis, School of Zoology, University of Tasmania, Hobart, Australia.

Gan, T.Y., Dlamini, E.M., Biftu, G.F., 1997. Effects of model complexity and structure, data quality, and objective functions on hydrologic modeling. J. Hydrol. 192, 81–103.

Hannah, C., Vezina, A., St John, M., 2010. The case for marine ecosystem models of intermediate complexity. Prog. Oceanogr. 84, 121–128.
Haraldsson, H.V., Sverdrup, H.U., 2004. Finding simplicity in complexity in biogeochemical modelling. In: Wainwright, J., Mulligan, M. (Eds.), Environmental Modelling—Finding Simplicity in Complexity. John Wiley & Sons, pp. 211–223.
Harvey, L.D.D., 2004. Climate and climate-system modeling. In: Wainwright, J., Mulligan, M. (Eds.), Environmental Modelling—Finding Simplicity in Complexity. John Wiley & Sons, pp. 77–91.
Haws, N.W., Bouwer, E.J., Ball, W.P., 2006. The influence of biogeochemical conditions and level of model complexity when simulating cometabolic biodegradation in sorbent-water systems. Adv. Water Resour. 29, 571–589.
Hurvich, C.M., Tsai, C.-L., 1989. Regression and time series model selection in small samples. Biometrika 76, 297–307.
James, A.I., Jawitz, J.W., 2007. Modeling two-dimensional reactive transport using a Godunov-mixed finite element method. J. Hydrol. 338, 28–41.
James, A.I., Jawitz, J.W., Muñoz-Carpena, R., 2009. Development and implementation of a transport method for the transport and reaction simulation engine (TaRSE) based on the Godunov-mixed finite element method. U.S. Geological Survey Scientific Investigations Report 2009-5034, p. 40.
Jawitz, J.W., Muñoz-Carpena, R., Muller, S., Grace, K.A., James, A.I., 2008. Development, testing, and sensitivity and uncertainty analyses of a transport and reaction simulation engine (TaRSE) for spatially distributed modeling of phosphorus in South Florida peat marsh wetlands. USGS Scientific Investigations Report 2008-5029, p. 109.
Jorgensen, S.E., 2002. Integration of Ecosystem Theories: A Pattern. Kluwer Academic Publishers, Dordrecht, The Netherlands, p. 432.
Juston, J.M., DeBusk, T.A., 2006. Phosphorus mass load and outflow concentration relationships in stormwater treatment areas for Everglades restoration. Ecol. Eng. 26 (3), 206–223.
Juston, J.M., DeBusk, T.A., 2011. Evidence and implications of the background phosphorus concentration of submerged aquatic vegetation wetlands in Stormwater Treatment Areas for Everglades restoration. Water Resour. Res. 47, W01511.
Kadlec, R.H., 1997. An autobiotic wetland phosphorus model. Ecol. Eng. 8, 145–172.
Kadlec, R.H., 2000. The inadequacy of first-order treatment wetland models. Ecol. Eng. 15, 105–119.
Kadlec, R.H., Wallace, S., 2008. Treatment Wetlands, second ed. CRC Press, Boca Raton, FL, p. 952.
Kirchner, J.W., 2006. Getting the right answers for the right reasons: linking measurements, analyses, and models to advance the science of hydrology. Water Resour. Res. 42, W03S04.
Lal, A.M.W., Van Zee, R., Belnap, M., 2005. Case study: model to simulate regional flow in South Florida. J. Hydraul. Eng. ASCE 131, 247–258.
Lindenschmidt, K.E., 2006. The effect of complexity on parameter sensitivity and model uncertainty in river water quality modelling. Ecol. Model. 190, 72–86.
Ljung, L., 1987. System Identification: Theory for the User. Prentice-Hall, Englewood Cliffs, NJ, USA, p. 519.
Martin, J.L., McCutcheon, S.C., 1999. Hydrodynamics and Transport for Water Quality Modeling. CRC Press, Boca Raton, FL, USA, p. 794.
McCormick, P.V., Shuford, R.B.E., Backus, J.G., Kennedy, W.C., 1998. Spatial and seasonal patterns of periphyton biomass and productivity in the northern Everglades, Florida, USA. Hydrobiology 362, 185–208.
McDonald, C.P., Urban, N.R., 2010. Using a model selection criterion to identify appropriate complexity in aquatic biogeochemical models. Ecol. Model. 221, 428–432.
Min, J., Paudel, R., Jawitz, J.W., 2011. Mechanistic biogeochemical model applications for Everglades restoration: a review of case studies and suggestions for future modeling needs. Crit. Rev. Environ. Sci. Technol. 41, 489–516.
Myung, J.I., Tang, Y., Pitt, M.A., 2009. Evaluation and comparison of computational models. Methods Enzymol. Comput. Methods 454, 287–304.
Paudel, R., Min, J., Jawitz, J.W., 2010. Management scenario evaluation for a large treatment wetland using a spatio-temporal phosphorus transport and cycling model. Ecol. Eng. 36, 1627–1638.
Perrin, C., Michel, C., Andreassian, V., 2001. Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. J. Hydrol. 242, 275–301.
Pitt, M.A., Myung, I.J., 2002. When a good fit can be bad. Trends Cogn. Sci. (Regul. Ed.) 6, 421–425.
Raghunathan, R., Slawecki, T., Fontaine, T.D., Chen, Z.Q., Dilks, D.W., Bierman, V.J., Wade, S., 2001. Exploring the dynamics and fate of total phosphorus in the Florida Everglades using a calibrated mass balance model. Ecol. Model. 142, 247–259.
Robson, B.J., Hamilton, D.P., Webster, I.T., Chan, T., 2008. Ten steps applied to development and evaluation of process-based biogeochemical models of estuaries. Environ. Model. Softw. 23, 369–384.
Schoups, G., Hopmans, J.W., 2006. Evaluation of model complexity and input uncertainty of field-scale water flow and salt transport. Vadose Zone J. 5, 951–962.
Scinto, L.J., Reddy, K.R., 2003. Biotic and abiotic uptake of phosphorus by periphyton in a subtropical freshwater wetland. Aquat. Bot. 77, 203–222.
SFWMD, 2005. RSM theory manual—HSE v1.0. South Florida Water Management District, West Palm Beach, FL, p. 308.
Snowling, S.D., Kramer, J.R., 2001. Evaluating modelling uncertainty for model selection. Ecol. Model. 138, 17–30.
Wainwright, J., Mulligan, M., 2004. Introduction. In: Wainwright, J., Mulligan, M. (Eds.), Environmental Modelling: Finding Simplicity in Complexity. John Wiley & Sons, pp. 1–4.
Walker, W.W., 1995. Design basis for Everglades Stormwater Treatment Areas. Water Resour. Bull. 31, 671–685.
Walker, W.W., Kadlec, R.H., 2011. Modeling phosphorus dynamics in Everglades wetlands and Stormwater Treatment Areas. Crit. Rev. Environ. Sci. Technol. 41 (S1), 430–446.
Wang, N.M., Mitsch, W.J., 2000. A detailed ecosystem model of phosphorus dynamics in created riparian wetlands. Ecol. Model. 126, 101–130.