A Numerical Study of Accounting for Structural Error and Uncertainty in Environmental Models

Lin, Zhulu
North Dakota State University, Department of Agricultural and Biosystems Engineering
Dept 7620, PO Box 6050
Fargo, ND 58108-6050, USA
E-mail: zhulu.lin@ndsu.edu

Beck, M. Bruce
University of Georgia, Daniel B. Warnell School of Forestry and Natural Resources
180 E Green Street
Athens, GA 30602-2152, USA
E-mail: mbbeck@uga.edu

INTRODUCTION

Analysis of the uncertainty in environmental models and their forecasts has now a long tradition (O'Neill, 1973; Beck and van Straten, 1983; Jakeman *et al.*, 2006; Warmink *et al.*, 2010). In short, we know well how to assess the consequences of uncertainty in the estimates of a model's parameter ($\boldsymbol{\theta}$) values, as they appear in the set of relationships between the observed inputs ($\mathbf{u}$), the state variables ($\mathbf{x}$), and the observed outputs ($\mathbf{y}$) of the system (Beck and Halfon, 1991). It is not difficult to extend such assessments to include uncertainty in the estimates of the initial values of the states ($\mathbf{x}(t_0)$) at the start of the forecasting period (time t_0). Subject to parameterizing the variations in the inputs ($\mathbf{u}(t)$) over time — for example, using a time-series model with parameters ($\boldsymbol{\psi}$) identified from historical observations — even the uncertainty in these streams of (future) disturbances of the system's behavior can be accounted for (Osidele and Beck, 2004; see also Kavetski *et al.*, 2006a,b). Significant problems have always arisen from the chronic paucity of the requisite field data, on [$\mathbf{u}$, $\mathbf{y}$]. From a pragmatic perspective, however, the almost inevitable over-parameterization of models and, therefore, their lack of identifiability (relative to the available data), should not be a problem in making an ensemble of forecasts upon which decisions are to be based. After all, what matters is not so much the ability to explain past behavior unambiguously, without uncertainty, but to assess whether the preferred policy action is robust against forecast uncertainty and — by implication — the lack of identifiability of the model.

Yet the real challenge in all of this, it has long been argued (van Straten and Keesman, 1991; Beven, 2005; Refsgaard *et al.*, 2006; Gaganis and Smith, 2008), is that of accounting for the errors and uncertainty in the structure of the model, or the conceptual errors in the model, which arise because the choices of [$\mathbf{x}$, $\boldsymbol{\theta}$] are inappropriate in some sense and, more familiarly, the relationships amongst them and with [$\mathbf{u}$, $\mathbf{y}$] have been incorrectly specified in the model. How is this source of uncertainty to be accounted for in employing the model to make forecasts (Butts *et al.*, 2004)? How, more challengingly, is the nature of this structural error to be identified (Bulygina and Gupta, 2009)? And how is it to be enumerated in the first place, in particular, from empirical field data (Osidele *et al.*, 2006)? It is on these questions that we focus herein.

Generating responses to these questions matters greatly (Beck, 2007; Beck *et al.*, 2009). They are not merely esoteric or theoretical issues, to be kept within the confines of technical discussions amongst the professional modeling fraternity. They matter to policy and to our collective futures (Petersen, 2006). The environment is just too complex for us to reason through the needs of policy without models. Yet the more complex the models themselves become, paradoxically the less they may be trusted by the public, and the greater the surprise (to some) when models fail to embrace — as they surely will — what comes to pass in

actuality. Assembling models to illuminate how to proceed under the prospect of climate change are no exception. Climate models inevitably have incomplete structures and the various alternative models tend to have *similarly* erroneous model structures. Consensus can seem stronger and more brightly illuminated than it ought, while significant unknowns and possibilities at the periphery of our understanding are left without any light being shed upon them (Oppenheimer *et al.*, 2007).

In this paper, we propose an algorithmic means of exploring (and quantifying) the sources and elements of the model structural error and uncertainty (MSEU). The key to appreciation of the problem is to view the parameters ($\boldsymbol{\theta}$) of the model's structure *not* as random variables, i.e., as constants (albeit not known with certainty), but as stochastic processes, i.e., quantities that vary with time, in part in a systematic manner and in part in an essentially random manner. Central to the means of solution is the algorithmic framework of recursive estimation. But this is not the only algorithmic framework in which to conceive of solutions; we also comment briefly on some of the alternatives, specifically, the familiar use of a "model fitting error", the inclusion of some facets of MSEU as model parametric uncertainty, and a Bayesian treatment of a plurality of candidate model structures. With reference to a hypothetical case study of a nonlinear biological system, the paper provides a quantitative inter-comparison of the performance of these three alternatives, together with that of the approach proposed herein based on recursive estimation of stochastic processes.

INNOVATIONS MODEL PARAMETERS AS STOCHASTIC PROCESSES

Although rarely practiced, many models of the behavior of environmental systems can be defined according to the following continuous-discrete innovations format (Ljung, 1987), i.e.,

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{f}(\mathbf{x}, \mathbf{u}, \boldsymbol{\theta}; t) + \mathbf{K}\boldsymbol{\nu}(t) \tag{1}$$

$$\mathbf{y}(t_k) = \mathbf{h}(\mathbf{x}, \mathbf{u}, \boldsymbol{\theta}; t_k) + \boldsymbol{\nu}(t_k) \tag{2}$$

where t and t_k are continuous and discrete time, respectively. $\mathbf{u}$ and $\mathbf{y}$ are the input and output vectors with dimensions *r* and *m*, respectively. $\mathbf{x}$ is a *n*-dimensional state vector; and $\boldsymbol{\theta}$ is a *p*-dimensional vector of model parameters. $\boldsymbol{\nu}(t_k)$ normally represents the (output) observation noise and is estimated using the innovation $\boldsymbol{\varepsilon}(t_k)$ – the mismatch between the predicted and observed values of the output at the next sampling instant in discrete time, i.e. $\boldsymbol{\varepsilon}(t_k) = \mathbf{y}(t_k) - \hat{\mathbf{y}}(t_k)$. This innovation will reflect all the undifferentiated sources of error/uncertainty that contribute to the divergence between the observed and modeled quantification of the system's outputs ($\mathbf{y}$) – with the system's inputs ($\mathbf{u}$) always having been accepted as being *exactly* as observed. $\mathbf{K}\boldsymbol{\nu}(t)$ is customarily considered to represent the system noise, comprised primarily of unobserved, incoming, input disturbances of the system and, secondarily, of the errors of observation associated with $\mathbf{u}$. The gain matrix $\mathbf{K}$ is a weighting matrix and can be thought of as a device for distributing the impacts of the innovations among the constituent representations of the various state variable dynamics, i.e. the representations $f_z(\cdot)$ for each state x_z. Provided $\mathbf{K}$ can be estimated or enumerated in some way, $\mathbf{K}\boldsymbol{\nu}(t)$ in Eq. (1) is a notional representation of a *computable* quantity, $\mathbf{K}\boldsymbol{\varepsilon}(t_k)$, reflecting those attributes of the system's behavior that are either omitted or not to be included in the model in more specific mathematical forms.

$\mathbf{f}$ and $\mathbf{h}$ are vectors of nonlinear functions. The *structure* of the model is most succinctly conveyed in terms of [$\mathbf{f}$, $\mathbf{h}$], which denote the logical interconnections among $\mathbf{u}$, $\mathbf{x}$, $\mathbf{y}$, while $\boldsymbol{\theta}$ signifies parameterization of the particular mathematical expressions of all the hypothetical mechanisms believed to underpin these interactions. The gap between [$\mathbf{f}$, $\mathbf{h}$] (i.e., what is included in the model) and the "truth of the matter" (i.e., what the model approximates) is what we should label as the *structural error* or *structural uncertainty* in the model.

If we are to detect in any way the presence of such inadequacy or approximation – the structural error/uncertainty in the model – some computable account, or indication, of it *must* be manifest in the model.

For it is *only* with the device of the model that we are able to explore the nature of (and engage with) the behavior of the real thing. In fact, structural error thus defined enters into Eq. (1) and (2) through $\boldsymbol{\theta}$, $\mathbf{K}\mathbf{v}(t)$, and $\mathbf{v}(t_k)$, although these points of entry differ in their interpretation and significance. The principal distinction is between $\boldsymbol{\theta}$, embedded within the choices for [$\mathbf{x}$, $\boldsymbol{\theta}$, $\mathbf{f}$, $\mathbf{h}$], which signify that which we presume (or wish) to know of the system's behavior, to be denoted as {presumed known}, and [$\mathbf{K}\mathbf{v}(t), \mathbf{v}(t_k)$], which fall outside the scope of what we believe we know, to be denoted as {acknowledged unknown}. The former can be wrongly presumed in the event, i.e., be found to be in error (a "structural error"), while the latter is more intuitively what we would label "structural uncertainty".

Furthermore, the elements of the parameter vector $\boldsymbol{\theta}$ and the matrix $\mathbf{K}$ can respectively be used as tags for the two sub-spaces. In other words, any significant variations in the reconstructed estimates of the elements of $\boldsymbol{\theta}$ specifically reflect changes or inadequacies of the model's constituent hypotheses in the model structure. Excursions of the reconstructed estimates of the elements of the matrix $\mathbf{K}$ significantly from their prior, presumed values of 0.0 indicate a change, or incompleteness, in the model as a whole. We might better conceive of the nature of our model's parameters, therefore, *not* as random variables, i.e., as uncertain *constants*, but as stochastic processes, whose variation through time is largely systematic (and capable, in principle, of interpretation), but also random and accordingly ascribable only to the actions of pure chance. Clearly, invoking this perspective is essential to detecting apparent structural *change* in the behavior of the system; and the capacity to detect this is a prerequisite for accounting for it, diagnosing its nature and, ultimately, taking steps to approach the essentially unattainable, yet highly desirable, property of invariance in the model's structure. The notion of treating model parameters as stochastic processes is implicit in filtering theory and, more generally, the algorithms of recursive estimation (Beck and Young, 1976). Such estimation of $\boldsymbol{\theta}$ and $\mathbf{K}$ can be simultaneously carried out in the Recursive Prediction Error (RPE) algorithm proposed in Lin and Beck (2007). It is worth noting that it is possible to detect structural change over various sub-blocks of time within an observed, empirical record of behavior, without recourse to a recursive estimation algorithm (e.g., Beven, 2002; Chen and Beck, 2002; Wagener *et al.*, 2003)

If we are going to resort to the use of recursive estimation algorithms – to detect changes over time in $\boldsymbol{\theta}$ and $\mathbf{K}$, with respect to past observed behavior – there arises then the matter of how to account for such structural error/uncertainty in generating forecasts of future behavior from the model. Since we shall invoke the use of a Generalized Random Walk (GRW) model for the description of a stochastic process, it is apparent that once parameterized – through a general set of parameters $\boldsymbol{\beta}$ (as in the generic form of time-series model that might be used to account for the future inputs ($\mathbf{u}$) of the system) – the propagation of structural error/uncertainty into the model's forecasts could be accounted for through the ensemble of such GRW models, the uncertainty attaching to the estimates of $\boldsymbol{\beta}$, and the stochastic perturbations in the (white-noise) sequences that drive the GRW models.

COMPUTATIONAL RESULTS: HYPOTHETICAL CASE STUDY

A hypothetical "real" system and its model

Suppose that we have a hypothetical nonlinear system of biomass-substrate interaction, for example, in a biological wastewater treatment facility, taking place in a single (idealized) continuously stirred tank reactor (CSTR). The true biological system is defined by the following set of ordinary differential equations:

$$\frac{dx_1(t)}{dt} = \mu_{\max} \frac{x_2(t)}{K_S + x_2(t)} x_1(t) - q_0 x_1(t) \tag{3}$$

$$\frac{dx_2(t)}{dt} = -\frac{1}{Y} \mu_{\max} \frac{x_2(t)}{K_S + x_2(t)} x_1(t) + q_0 [u(t) - x_2(t)] \tag{4}$$

where $x_1(t)$ and $x_2(t)$ are the concentrations of biomass and substrate in the bioreactor at time t, $\mathrm{M \cdot L^{-3}}$;

μ_{max} is the maximum specific growth rate of the biomass, T^{-1}; K_S is the half-saturation constant of the substrate, $M \cdot L^{-3}$; Y is the yield coefficient of biomass on substrate, $M \cdot M^{-1}$; $u(t)$ is the inlet concentration of substrate, $M \cdot L^{-3}$; and q_0 is the constant dilution rate of the CSTR, T^{-1}. Given the specific parameter values presented in Table 1, the system is simulated using SIMULINK® software, generating the input and output signals of $y_1(t)$ and $y_2(t)$, which are the observed time-series of $x_1(t)$ and $x_2(t)$, respectively. The input and outputs are plotted in Fig. 1. The output signals were corrupted with white Gaussian measurement noise with variance-covariance matrix $\mathbf{\Lambda}$=diag(0.5, 0.1).

Now, suppose that our conceptual model of the system's observed behavior is of a simpler structure, being linear in its assumptions about the kinetic interactions between biomass and substrate. That is to say, we have the following approximation

$$\theta(t) = -\frac{1}{Y}\frac{\mu_{max} \cdot x_1(t)}{K_S + x_2(t)} \tag{5}$$

which, when substituted into Eq. (3) and (4), yields our conceptual model structure as

$$\frac{dx_1(t)}{dt} = -Y\theta(t)x_2(t) - q_0 x_1(t) \tag{6}$$

$$\frac{dx_2(t)}{dt} = \theta(t)x_2(t) + q_0[u(t) - x_2(t)] \tag{7}$$

where $\theta(t)$ is a *time-varying* parameter reflecting the possibility of a changing structure in the dynamics of the state variables ($x_1(t)$ and $x_2(t)$), in line with Eq. (5). Apparently, the model described by Eq. (3) and (4) represents the "truth of the matter", while the model of Eq. (6) and (7) is an approximation of the "truth of the matter". As defined above, the gap between these two entities is the model error/uncertainty, which is to be quantified in the following paragraphs using various methods.

Table 1. Parameter values in the model of the hypothetical system

Parameter	Value
μ_{max}	0.3
K_S	3.0
Y	0.6
q_0	0.1
$[\, x_1(0), x_2(0)\,]^T$	$[8.0, 4.0]^T$

Conventional and recursive model parameter estimation

The model of the real system, described in Eq. (6) and (7), was first calibrated using the uniform covering by probailistic rejection (UCPR) algorithm (Hendrix and G.-Tóth, 2010), which is a typical Monte Carlo smpaling-based approach to parameter estimation, treating model parameters as temporally constant random variables. The goal of UCPR is to search for sub-regions of the parameter domain that contain the best-fitting parameter values, and eventually converge an ensemble of randomly generated parameter sets onto the global "optimal" parameter sets. As shown in Fig. 2, the objective function to be minimized by model calibration has a valley-shaped response surface along the direction of parameter $\theta(t)$, implying that the time-varying parameter, $\theta(t)$, is less identifiable than the time-invariant parameter, Y. We can reasonably assume this is caused by the temporally varying nature of parameter $\theta(t)$ as defined by Eq. (5), which allows parameter $\theta(t)$ to take values between -2.07 to -0.53, as indicated by the two vertical lines in Fig. 2. Nevertheless, the UCPR algorithm still manages to calibrate the model, converging to a cluster of 5000 parameter pairs with $\Phi_{max}/\ \Phi_{min}$ = 1.1 (Fig. 2). The mean of $\hat{\theta}$ – the estimates of θ – is -1.28 with a

standard error (SE) of 0.12, while the mean and the SE of $\hat{Y}$ are 0.60 and 0.016. Obviously, the estimate of θ is more variable than that of Y. The histograms (not shown) also demonstrate that the sample distributions for both parameters are double-moded, indicating the ambiguity in determining *the* global optimal parameter set for θ and Y. Despite with an erroneous model structure, the calibrated models with different pairs of value for θ and Y randomly selected from the converged cluster are surprisingly successful in simulating the biomass concentrations and the lower substrate concentrations, but fail to simulate the high substrate concentrations (not shown).

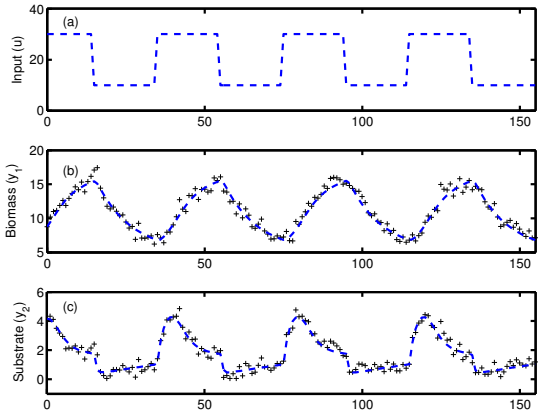

Fig. 1. Simulated data of the hypothetical system using parameter values in Table 1 (dashed, true; cross, noise-corrupted).

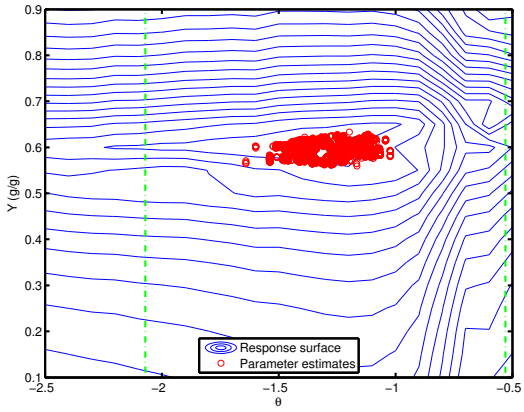

Fig. 2. Calibrated parameter estimates with contour map of the response surface of the objective function.

Given the temporally varying nature of parameter $\theta(t)$, the RPE algorithm (Lin and Beck, 2007) was then used to estimate the time-varying parameter. The significance of the RPE algorithm lies in its treatment of $\theta(t)$ as a Random Walk stochastic process, rather than a temporally constant random variable. Although estimating this specific, time-varying parameter is our primary goal, all the parameters of the innovations representation (Eq. (1) and (2)) are estimated recursively with the RPE algorithm, i.e., here, the time-invariant parameter Y in Eq. (6) and the elements in the matrix **K** matrix (which are also treated as time-invariant parameters). Fig. 3(a-b) show that the estimates of the time-invariant parameter Y converge to its true value very quickly, while the estimates of the *time-varying* parameter $\theta(t)$ track its true pattern of cyclic variation successfully. It is apparent that the performance of the algorithm is indeed successful, in the sense that comparison with the known-true situation is demonstrable. In Fig. 3 (c-f) the recursive estimates for the elements of the **K** matrix are also presented, where it is apparent that all four elements deviate from their initial values of zero. The behavior of the recursive estimates of these elements may have been influenced by the errors associated with the incorrect initial assumptions about the model's parameter estimates.

The variation in the estimates of parameter $\theta(t)$ signals the *structural error* of the linear model (Eq. (6) and (7)) in the space of the {presumed known}. It is apparent from the reconstructed trajectory for $\theta(t_k)$ that this parameter must be correlated in some way with the oscillations in the state variables, as clearly expected through Eq. (5), possibly more so with those of the substrate concentration (x_2). Furthermore, the excursions in the estimated trajectories of the elements of the matrix **K** from zero are supposed to indicate the significance of the *structural uncertainty* in the space of the {acknowledged unknown}. However, the structure of the linear model (Eq. (6) and (7)) in this instance has not so much overlooked (or omitted) something of significance, as presumed an incorrect specification of the nature of the kinetic interaction between substrate and biomass. Nevertheless, the non-zero estimates of $k_{i,j}(t_k)$ is suggestive of something of significance *not* having been included in that model's structure. Alternatively, we must acknowledge that these results may be an artifact of treating the elements of **K** as *invariant* quantities. More extensive

discussion of the role of the **K** matrix elements in the RPE algorithm and in our approach to model structure identification are given in Lin and Beck (2007). Their detailed interpretation is not central to the present discussion of accounting for MSEU.

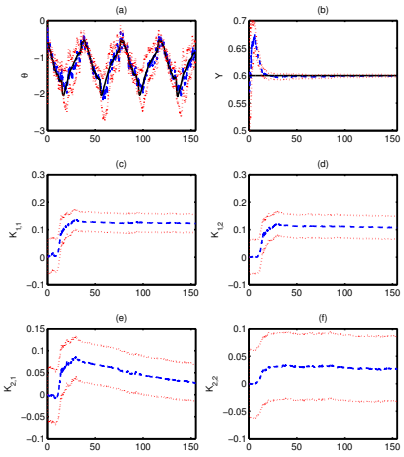

Fig. 3. Recursive estimation results: (a) and (b), parameter estimates (estimated value, blue dashed line; standard error, red dotted line; true value, black solid line); (c) – (f), **K** matrix elements (estimated value, blue dashed line; standard error, red dotted line).

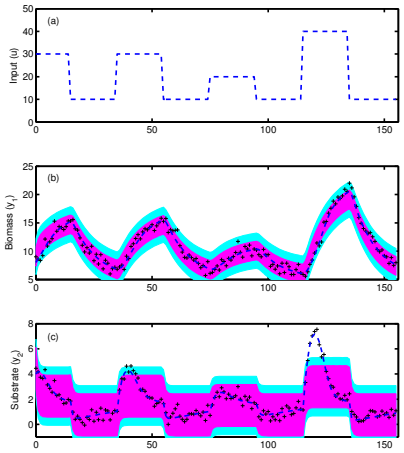

Fig. 4. (a) Input function during prediction period; (b) – (c) The model fitting error method: comparisons of true (dashed lines) and observed (crosses) system's responses with the 95% confidence intervals for the average model prediction (magenta) and the 95% prediction intervals for the model prediction of an individual observation (cyan).

Accounting for model structural error/uncertainty in model predictions

Given the inevitability of being obliged to use a simplified model to approximate the behavior of the real system, the challenge we are facing is how to exploit the available finite-order resolution models to make useful predictions for the future *in the presence of model structure error/uncertainty*. Four strategies are considered in the following: (1) adding a 'model fitting error' to model predictions, denoted for brevity as the model fitting error (MFE) method; (2) expanding parameter uncertainty, or parameter uncertainty expansion (PUE) method; (3) extrapolation with multiple model structures, i.e., the (MMS) method; and (4) the innovations representation with parameters as stochastic processes, or the time-varying parameter (TVP) method in short. An inter-comparison of the performances of these four approaches follows, with reference to the foregoing hypothetical case study.

The predictive test is defined as: prediction of the output variables under a future condition similar to that of the past, i.e., to the set of past data employed for model calibration. We readily acknowledge that this is not the most stringent of such tests, or that which will normally attract the greatest concern in attempting to formulate robust environmental policy: of making projections into a future where behavior may be radically different from that observed in the past (Beck, 2002). But it is a necessary and sufficiently challenging test with which to open up more systematic and extensive study of MSEU. For the present predictive test, therefore, the chosen input (forcing) function has the same long-term trend and pattern of seasonal variation as that used for model calibration (i.e., the square wave input function displayed in Fig. 1). To test the robustness of the model error/uncertainty accounting methods, however, the amplitude of the square-wave signal of the influent substrate concentration over the period of prediction is variable. This is notably different from the period of model calibration (comparing Fig. 1(a) and 4(a)). Since we are working with a hypothetical system, the system's responses to the future forcing function can also be obtained, such these known, true output responses will allow a suitable benchmarking of the relative

performances of the four methods of accounting for model structural error/uncertainty.

MFE – adding model fitting error to predictions

We take the means of the parameter estimates obtained from the previous conventional model calibration process and employed them in the model (Eq. (6) and (7)), together with the future input function (Fig. 4(a)), to generate an average future response of the system. The 'model fitting errors' were estimated to be the root mean squared errors (RMSE) found for the *model calibration* period, which are 0.84 and 0.85 $M{\cdot}L^{-3}$ for the biomass and substrate concentrations, respectively. These 'model fitting error' estimates are then subtracted from, and added to, the average of the system's responses to the future input function to form the confidence intervals (or regions) for the system's average future responses.

It should be noted that the *confidence interval* for an average future response differs from the *prediction interval* for an individual observation of the system's response (Rencher, 2000). It is obvious that the latter will have to be wider than the former, since the prediction interval should also take into account (future) errors of observation. Without necessarily losing the underlying mathematical rigor, we adopt the following practical approach to estimating the confidence intervals for the average model prediction in the future ($\bar{y}$) and the prediction intervals for the model prediction of an individual future observation (y). Specifically, the 95% confidence intervals for ($\bar{y}$) are estimated as $\bar{y} \pm 1.96 \cdot \mathrm{RMSE}$ and the 95% prediction intervals for (y) are estimated as

$$\bar{y} \pm 1.96 \cdot \left(\mathrm{RMSE} + \sqrt{\hat{\sigma}_y^2} \right)$$

where $\hat{\sigma}_y^2$ is the variance of the observed times series during the model calibration period, which was estimated to be 0.55 for the biomass concentrations and 0.11 for the substrate concentrations, using a method provided in Wadsworth (1998). The true variances of the observation noises are 0.5 and 0.1 for the biomass and substrate concentration time series, respectively.

In Fig. 4 we show the true and the noise-corrupted observations of the system's responses under the future forcing function together with the 95% confidence intervals and the 95% prediction intervals for the predictions of the biomass and the substrate concentrations in the CSTR. For the biomass, the true and the noise-corrupted observations for most of the prediction period are covered by the 95% confidence intervals and the 95% prediction intervals (Fig. 4(b)). For substrate, however, two of the three peaks are not encompassed by the confidence intervals, while the two highest peaks are not covered by the prediction intervals (Fig. 4(c)). This method assumes that the model uncertainty remains constant throughout the entire modeling period. As a result, the confidence and the prediction intervals for the high substrate concentrations have the same magnitude as those for the low substrate concentrations, even though we have more confidence in the model when predicting the low substrate concentrations relative to the high concentrations.

PUE – expanding parameter uncertainty

In the PUE method, we assume that the model structure error/uncertainty can be compensated for by expanding the uncertainty associated with the model parameter estimates (here, $\hat{\theta}$ and $\hat{Y}$), as customarily accounted for by the standard errors of the model's parameter estimates. We further assume that the model parameter estimates are normally distributed. The expansion of PUE is herein taken to be that of amplifying the parametric uncertainty to be accounted for from that enclosed in the standard-error bounds to that enclosed in three times the standard-error bounds. Fig. 5 displays the histograms of the distributions of the model parameter estimates $\hat{\theta}$ and $\hat{Y}$ resulting from 2000 realizations of their associated normal distributions,

$$N\,(\bar{\theta} \text{ or } \bar{Y}, \quad \hat{\sigma}_{\theta}^2 \text{ or } \hat{\sigma}_{Y}^2)$$

in which, $\bar{\theta} = -1.28$ and $\bar{Y} = 0.6$, which are the means of $\hat{\theta}$ and $\hat{Y}$ resulting from the conventional model calibration; $\hat{\sigma}_i (i = \theta \text{ or } Y)$ = SE, or under the expansion of parametric uncertainty, 3·SE of $\hat{\theta}$ or $\hat{Y}$.

The SEs of $\hat{\theta}$ and $\hat{Y}$ are 0.12 and 0.016, respectively.

Five hundred (500) out of the 2000 model parameter sets of each distribution ($\hat{\sigma}_i$ = SE or 3·SE) were then substituted into the model of Eq. (6) and (7) to generate 500 model-predicted time series of the biomass and substrate concentrations in the CSTR driven by the future input function. Subsequently, the 95% confidence intervals for the model predictions of the average biomass and substrate concentrations (shown in Fig. 6) were calculated based on these 500 model-generated time series as $\bar{y} \pm 1.96 \cdot s_{\bar{y}}$, where $\bar{y}$ and $s_{\bar{y}}$ are the time series of the mean and the standard deviation of the 500 model-generated time series of biomass and substrate concentrations. For clarity, the corresponding 95% prediction intervals are not computed or plotted in Fig. 6.

As shown in Fig. 7, the confidence intervals for the average model predictions resulting from the expanded parametric uncertainty are much wider than those from the non-expanded parametric uncertainty. As for the MFE method, the true responses of biomass concentration are largely covered by the 95% confidence intervals, while those of the high concentrations of substrate are not covered by the corresponding 95% confidence intervals. Unlike the results from the MFE method, however, the widths of the confidence intervals estimated with the PUE method are not constant over time. They are wider for the high concentrations of substrate, where we have less confidence in the model, and vice versa for the low concentrations. On this basis, we would prefer the PUE method over the MFE method in accounting for MSEU. However, PUE is still not satisfactory, since: (1) the model structure errors and uncertainties are not adequately compensated for, since two peaks of the substrate concentrations are completely missed by the 95% confidence intervals; and (2) the model parameters may assume unrealistic values, should the parametric uncertainty be expanded indefinitely (see Fig. 5(a)).

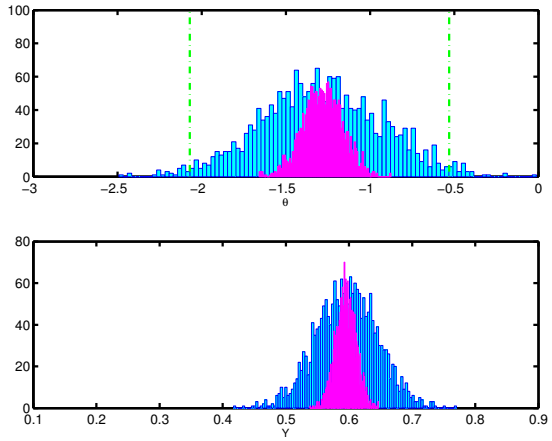

Fig. 5. Histograms of the model parameter distributions with standard deviations being the standard errors (magenta) and three times the standard errors (cyan) of the parameter estimates: (a) $\hat{\boldsymbol{\theta}}$, and (b) $\hat{\boldsymbol{Y}}$.

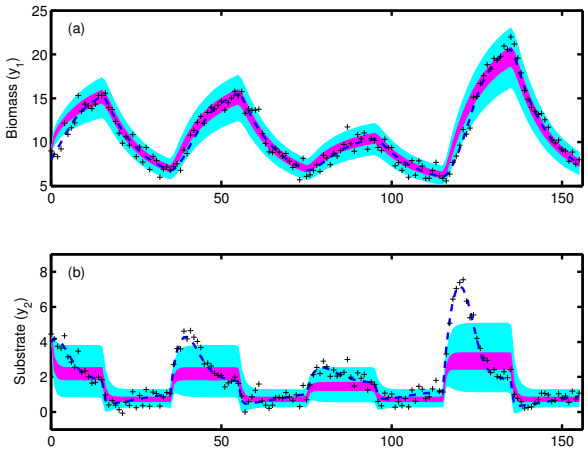

Fig. 6. The parameter uncertainty expanded method: comparisons of true (dashed lines) and observed (crosses) system's responses with 95% confidence intervals of model predictions (magenta, parametric uncertainty not expanded; cyan, parametric uncertainty expanded).

MMS – exploiting multiple model structures

To illustrate the MMS method, we need to introduce the possibility of another model for the hypothetical system, which has a simpler nonlinear structure than that described by Eq. (3) and (4), so that this new, nonlinear model has two parameters only, thus eliminating the notoriously non-identifiable, 3-parameter structure of the Monod equation. We now denote the linear model as Model-I, the two-parameter nonlinear model as Model-II, and the original three-parameter nonlinear model as Model-III. It is important to note that the first two models have erroneous structures, whereas Model-III has the true model structure.

To set up the test of the MME approach, we must first estimate the parameters in all three model structures, again using the UCPR method mentioned previously. The resulting parameter ensembles from calibration are displayed in Fig. 7. For Model-III, Y and μ were estimated using the UCPR method, while K_S was set equal to its true value of 3.0 M·L^{-3} (see Table 1). In the subsequent MMS analysis, the means of the parameter estimates were employed in the corresponding model structures to generate the model predictions of the system's responses under the future input functions.

Several model-averaging methods have been proposed in the literature to tackle the problem of MSEU, including Bayesian Model Averaging methods (Draper, 1995; Hoeting, *et al.*, 1999; Rojas *et al.*, 2008; Diks and Vrugt, 2010; Tsai, 2010). For the purposes of illustration, we adopt herein an approach proposed by Poeter and Anderson (2005), to which reference should be made for details of the method. In short, their method combines predictions of multiple competing models using Akaike weights, which are calculated mainly based on the model's performance and the number of model parameters in the model. Then, the model-averaged prediction ($\hat{\bar{y}}$) is computed as $\sum_{i=1}^{R} w_i \hat{y}_i$, where $\hat{y}_i$ are the model predictions generated by model i and w_i are the weights calculated for model i based on the Akaike Information Criterion for each model; and R is the number of candidate model structures. The variance of $\hat{\bar{y}}$ is therefore calculated as

$$\hat{\sigma}_{\hat{\bar{y}}}^2 = \left\{\sum_{i=1}^{R} w_i [\widehat{\mathrm{var}}(\hat{y}_i|\mathrm{model}_i) + (\hat{y}_i - \hat{\bar{y}})^2]^{0.5}\right\}^2 \quad (8)$$

in which the first term represents the variance, given one calibrated model, and the second term represents the among-model variance, given the set of models. Thus, the approximate 95% confidence intervals for $\hat{\bar{y}}$ can be expressed as $\hat{\bar{y}} \pm 1.96 \cdot \hat{\sigma}_{\hat{\bar{y}}}$.

In Fig. 8, two 95% confidence intervals are plotted against the true and noise-corrupted observations of the biomass and substrate concentrations in the CSTR. The narrower confidence interval was obtained using the two erroneous model structures (i.e., Model-I and Model-II), where the Akaike weights for the two models were 0.3 and 0.7 respectively. The wider confidence interval was obtained using all three model structures, i.e., including the true model structure, Model-III. The Akaike weights for the three models were 0.176, 0.412 and 0.412, respectively. For clarity, the 95% prediction intervals are not computed or plotted in Fig. 8.

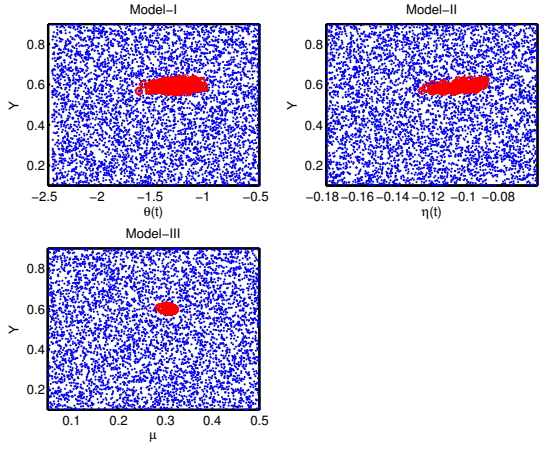

Fig. 7. Parameter ensembles for the three models (blue dots, initial parameter ensembles; red circles, estimated parameter ensembles).

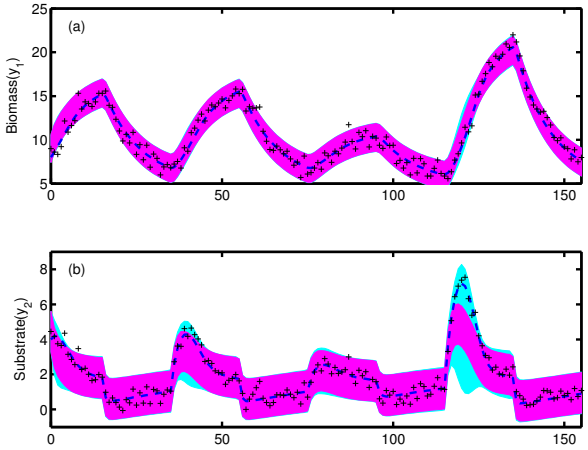

Fig. 8. The multiple model structure method: comparisons of true (dashed lines) and observed (crosses) system's responses with 95% confidence intervals of model-averaged predictions (magenta, Model-III excluded; cyan, Model-III included).

With the MMS method the confidence intervals (narrow or wide) are able to cover the second highest peak of the substrate concentration time series, which is a noticeable improvement over the performance of

the previous MFE and PUE methods. Nevertheless, the narrower confidence intervals do not then cover the highest peak, although the wide confidence intervals do. This is unsurprising, because these wider intervals draw upon inclusion of the true model structure in the set of candidates. As for PUE, the widths of the confidence intervals in the MMS method vary with time, although significantly less so. Such variation is ascribed to the second term in Eq. (8), which reveals how the MMS method is a compromise between the MFE and the PUE methods.

TVP: Innovations respresentation parameters as stochastic processes

In the TVP method, a time-series model with long-term trend and seasonal components is employed to represent the pattern of variation exhibited in $\theta(t)$, since seasonality is evident in the trajectory of the recursive estimates of this parameter (Fig. 3(a)). The time-series model consists of three terms: a long-term trend, a seasonal component, and an AR(1) (first-order autoregressive) component (see Eq. (9) below). The coefficients in the time-series model (Eq. (9) and (10)) are estimated using the time-series analysis software ITSM 2000 (Brockwell and Davis, 2002). Note that the first two terms are deterministic while the third term is stochastic.

$$\theta(t_k) = m(t_k) + S_{40}(t_k) + X(t_k) \tag{9}$$

where $m(t_k) = -0.0001t_k - 1.22$; $S_{40}(t_k)$ is the seasonal component with period 40, i.e., $S(t_k + 40) = S(t_k)$, and $\sum_{k=1}^{40} S(t_k)$=0; $X(t_k)$ is an AR(1) model described by

$$X(t_k) = \varphi_1 X(t_{k-1}) + Z(t_k) \tag{10}$$

where $\varphi_1 = 0.92 \pm 0.01$; $\{Z(t_k)\}$is a white noise sequence with a variance vector set to be proportional to the variance of $\hat{\theta}(t)$ estimated by the RPE method previously.

Eq. (9) and (10) were used to generate 500 time-series realizations of $\theta(t)$ over the period of prediction (see Fig. 9(a)). It is worth mentioning that the 500 time series of $\theta(t)$ gradually diverge from each other after an initial transient period, starting around time 15 (refer to Fig. 9(a)). This indicates that there is less certainty about the time-varying parameter's ability to represent the system's changing dynamics as time proceeds into the more distant future. For the time-invariant parameter *Y*, 500 realizations were also sampled from a normal distribution, with the mean and the standard deviation being 0.6021 and 0.0009 as a result of the prior recursive estimation (refer to Fig. 3(b) and Fig. 9(b) and note the scale of the horizontal axis of Fig. 9(b)). Each of the projected future trajectories for the time-varying parameter $\theta(t)$ were then paired with one realization of the time-invariant parameter *Y* and inserted into the linear model of Eq. (6) and (7) to generate predictions under the future input function.

Subsequently, the 95% confidence intervals for the model predictions of the average biomass and substrate concentrations ($\bar{y}$) were calculated, based on the 500 model-generated time series. As argued earlier, model structure error/uncertainty enters into a model through $\boldsymbol{\theta}$, $\mathbf{K}\boldsymbol{\nu}(t)$, and $\boldsymbol{\nu}(t_k)$ (refer to Eq. (1) and (2)). In contrast, model prediction uncertainty (in the form of confidence intervals) is here calculated as a function solely of the propagated parametric uncertainty (i.e., the uncertainty associated with $\boldsymbol{\theta}$). Uncertainty deriving from the elements of $\mathbf{K}\boldsymbol{\nu}(t)$ could also be numerically integrated over time and added to the model prediction uncertainty. However, compared with the contribution to prediction uncertainty deriving from parametric uncertainty, that from the system's noise $\mathbf{K}\boldsymbol{\nu}(t)$ is negligible in this particular case study (in fact, it is shown as the white gaps between the confidence intervals and the prediction intervals in Fig. 10) . This is due to the fact that the time-varying parameter estimate ($\hat{\theta}(t_k)$) has successfully captured the divergence between [**f**, **h**] and the "truth of the matter". For this reason, $\mathbf{K}\boldsymbol{\nu}(t)$ was not included in our calculation of the confidence and prediction intervals. To complete the analysis, therefore, a form of $\boldsymbol{\nu}(t_k)$ (i.e., $\hat{\sigma}_y$) is added to estimate the prediction intervals. The 95% prediction intervals for (y) are thus calculated as $\bar{y} \pm 1.96 \cdot \left(s_{\bar{y}} + \hat{\sigma}_y\right)$ with $\bar{y}$, $s_{\bar{y}}$, and $\hat{\sigma}_y^2$ as defined previously.

Fig. 10 displays the comparisons of the observations of the system's responses under the future forcing

function and the 95% confidence intervals and the 95% prediction intervals for the model predictions of the biomass and substrate concentrations in the CSTR. Three salient features are worth mentioning. First, the confidence intervals are almost able to cover the entire time series of the true biomass and substrate concentrations, including the highest peak of the substrate concentration. The exception is where the true substrate concentrations just fall outside the confidence intervals during period 90 through 110 (see Fig. 10(b)). Our conclusion is that the estimated time-varying parameter $\theta(t)$ of the finite-order model structure of Eq. (6) and (7) is able to reflect very well the underlying (true) system's dynamics, consequently reducing the uncertainty associated with estimating the other model parameter, *Y*.

The second salient feature of these results is associated with the fact that the confidence intervals are considerably narrower than their counterparts, as estimated in the three preceding methods. This implies that the TVP method has significantly reduced the predictive uncertainty as a result of reducing model structure error and uncertainty. Third, we note that the widths of the confidence intervals (thus the prediction intervals) for both biomass and substrate are wider at the high substrate concentrations, yet narrower at the low substrate concentrations. This faithfully reflects the uncertainty associated with the model's predictions.

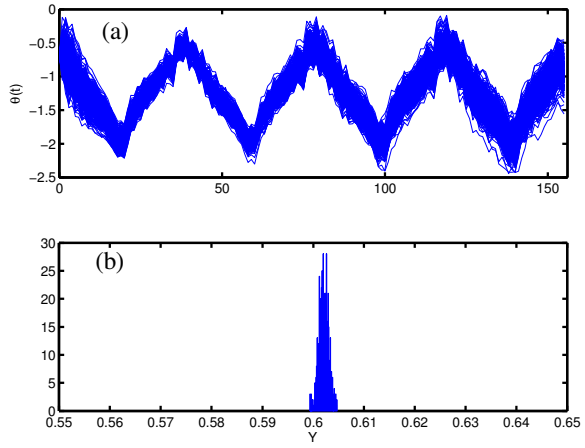

Fig. 9. 500 realizations of the time-varying parameter ($\boldsymbol{\theta(t)}$) and the time-invariant parameter (*Y*) used in model predictions.

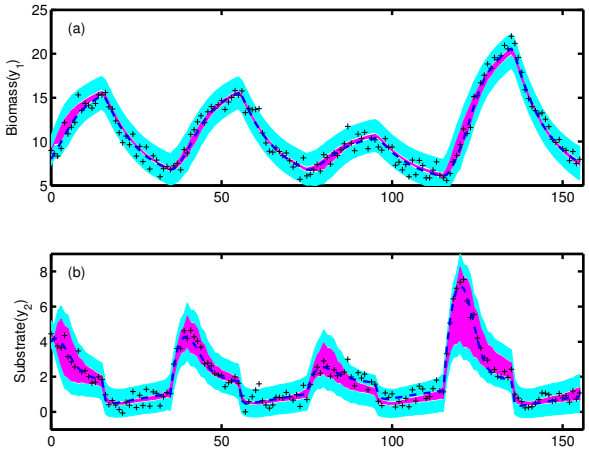

Fig. 10. The time-varying parameter method: comparisons of true (dashed lines) and observed (crosses) system's responses with the 95% confidence intervals for the average model prediction (magenta) and the 95% prediction intervals for the model prediction of an individual observation (cyan).

To summarize, our proposed approach, deriving from the innovations representation of a model's structure, can in principle account for both error in the {presumed known} and uncertainty in the {acknowledged unknown} and does so through the treatment of an expanded vector of model parameters as stochastic processes (TVP), with elements of systematic variation (trend, seasonality) and random variability (white-noise process). From the numerical results of our hypothetical case study, we judge that the TVP and MMS approaches both perform reasonably well in accounting for the propagation of MSEU in predictions, although MMS presumes a candidate model structure that is indeed the true structure. The comparative performances of the MFE and PUE approaches are both inferior to those of the MMS and TVP. Further, for this particular hypothetical case study, there happens to be very little uncertainty attaching to the {acknowledged unknown} of the model's structure.

CONCLUSIONS

An innovations representation, with its origins in mathematical filtering theory, has been used to

specify more precisely than previously the nature and sources of error and uncertainty in the structure of a model. This representation suggests a natural way of quantifying and accounting for model structure error/uncertainty in the making of forecasts of future behavior, through an extended set of model parameters treated as a set of stochastic processes — hence the label TVP (Time-Varying Parameters) for our proposed approach. Extension of the set of model parameters is important in its own right, and derives from the way in which the innovations representation accounts for stochastic disturbances of the model's state-variable dynamics.

When compared with three other, more familiar approaches to accounting for the predictive propagation of MSEU, i.e., use of a model fitting error, expansion of parametric uncertainty, and Bayesian model averaging, the TVP approach performs well. Future work is accordingly in progress to extend the preliminary numerical results of the present paper to more challenging areas, in particular, to hypothetical situations where uncertainty in the {acknowledged unknown} is significant, to future conditions in which the input forcing functions of the system are more radically different from those observed in the past, and to a case study with actual field data. We are also assessing the implications of the present work with respect to ways of reducing MSEU in the process of model structure identification, again using the innovations representation and the Recursive Prediction Error algorithm (Lin and Beck, 2007).

REFERENCES (RÉFERENCES)

Beck, M.B., 2002. Environmental Foresight and Models: A Manifesto. Elsevier, New York.

Beck, M.B., 2007. Environment: How best to look forward? Science 316(5822), 202-203 (13 April).

Beck, M.B., Gupta, H., Rastetter, E., Shoemaker, C., Tarboton, D., Butler, R., Edelson, D., Graber, H., Gross, L., Harmon, T., McLaughlin, D., Paola, C., Peters, D., Scavia, D., Schnoor, J.L., Weber, L., 2009. Grand Challenges of the Future for Environmental Modeling, White Paper, National Science Foundation, Arlington, Virginia (ISBN: 978-1-61584-248-3).

Beck, M.B., Halfon, E., 1991. Uncertainty, identifiability and the propagation of prediction errors: A case study of Lake Ontario. J. Forecast. 10, 135-161.

Beck, M.B., van Straten, G. 1983. Uncertainty and Forecasting of Water Quality. Springer-Verlag, New York.

Beck, M.B., Young, P.C. 1976. Systematic identification of DO-BOD model structure. J. Environ. Eng. Div. 102, 902-927.

Beven, K.J., 2002. Uncertainty and the detection of structural change in models of environmental systems. In: Beck, M.B. (Ed.), Environmental Foresight and Models: A Manifesto. Elsevier, New York, pp. 227-250.

Beven, K.J., 2005. On the concept of model structure error. Water Sci. and Technol. 52(6), 167-175.

Brockwell, P.J., Davis, R.A., 2002. Introduction to Time Series and Forecasting, Second Ed. Springer, New York.

Bulygina, N., Gupta, H. 2009. Estimating the uncertain mathematical structure of a water balance model via Bayesian data assimilation. Water Resour. Res. 45, W00B13, doi:10.1029/2007WR006749.

Butts, M.B., Payne, J.T., Kristensen, M., Madsen, H., 2004. An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. J. Hydrol. 298, 242-266.

Chen, J., Beck, M.B., 2002. Detecting and forecasting growth in the seeds of change. In: Beck, M.B. (Ed.), Environmental Foresight and Models: A Manifesto. Elsevier, New York, pp. 351-374.

Diks, C.G.H., Vrugt, J.A., 2010. Comparison of point forecast accuracy of model averaging methods in hydrologic applications. Stochastic Environ. Res. Risk Assess. 24, 809-820.

Draper, D., 1995. Assessment and propagation of model uncertainty. J. Royal Stat. Soc., B 57(1), 45-97.

Gaganis, P., Smith, L., 2008. Accounting for model error in risk assessments: Alternatives to adopting a bias towards conservative risk estimates in decision models. Adv. Water Resour. 31, 1074-1086.

Hendrix, E.M.T, G.-Tóth, B., 2010. Introduction to Nonlinear and Global Optimization. Springer, New York.

Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian model averaging: A tutorial. Stat. Sci. 14(4), 382-417.

Jakeman, A.J., Letcher, R.A., Norton, J.P., 2006. Ten iterative steps in development and evaluation of environmental

models. Environ. Model. Softw. 21 (5), 602-614. doi:10.1016/j.envsoft.2006.01.004.

Kavetski, D., Kuczera, G., Franks, S.W., 2006a. Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. Water Resour. Res. 42, W03407, doi:10.1029/2005WR004368.

Kavetski, D., Kuczera, G., Franks, S.W., 2006b. Bayesian analysis of input uncertainty in hydrological modeling: 2. Application. Water Resour. Res. 42, W03408, doi:10.1029/2005WR004376.

Lin, Z., Beck, M.B., 2007. On the identification of model structure in hydrological and environmental systems. Water Resour. Res. 43, W02402, doi:10.1029/2005WR004796.

Ljung, L., 1987. System Identification: Theory for the User. Prentice-Hall, Englewood Cliffs, NJ.

O'Neill, R.V., 1973. Error analysis of ecological models. In: Nelson, D.J. (Ed.), Radionuclides in Ecosystems. Conference Publication 71051, NTIS, Springfield, Virginia.

Oppenheimer, M., O'Neill, B.C., Webster, M., Agrawala, S., 2007. The Limits of Consensus. Science 317, 1505-1506 (14 September).

Osidele, O.O., Beck, M.B., 2004. Food web modelling for investigating ecosystem behaviour in large reservoirs of the south-eastern United States: Lessons from Lake Lanier, Georgia. Ecol. Model. 173, 129-158.

Osidele, O.O., Zeng, W., Beck, M.B., 2006. A random search methodology for examining parametric uncertainty in water quality models. Water Sci. Technol. 53, 33-40.

Petersen, A.C., 2006. Simulating Nature: A Philosophical Study of Computer-Simulation Uncertainties and Their Role in Climate Science and Policy Advice. Het Spinhuis, Apeldoorn, The Netherlands.

Poeter, E., Anderson, D., 2005. Multimodel ranking and inference in ground water modeling. Ground Water 43(4), 597-605.

Refsgaard, J.C., van der Sluijs, J.P., Brown, J., van der Keur, P., 2006. A framework for dealing with uncertainty due to model structure error. Adv. Water Resour. 29, 1586-1597.

Rencher, A.C., 2000. Linear Models in Statistics. John Wiley & Sons, New York.

Rojas, R., Feyen, L., Dassargues, A., 2008. Conceptual model uncertainty in groundwater modeling: Combining generalized likelihood uncertainty estimation and Bayesian model averaging. Water Resour. Res. 44:W12418, doi:10.1029/2008WR006908.

Tsai, F.T.-C., 2010. Bayesian model averaging assessment on groundwater management under model structure uncertainty. Stochastic Environ. Res. Risk Assess. 24, 845-861.

van Straten, G., Keesman, K.J., 1991. Uncertainty propagation and speculation in projective forecasts of environmental change: A lake eutrophication example. J. Forecast. 10, 163-190.

Wadsworth, H.M., 1998. Handbook of Statistical Methods for Engineers and Scientists, Second Ed., McGraw-Hill, New York.

Wagener, T., McIntyre, N., Lees, M.J., Wheater, H.S., Gupta, H.V., 2003. Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis. Hydrol. Processes 17, 455-476.

Warmink, J.J., Janssen, J.A.E.B., Booij., M.J., Krol, M.S., 2010. Identification and classification of uncertainties in the application of environmental models. Environ. Model. Softw. 25, 1518-1527.