Mathematical modelling and parameter estimation of the Serra da Mesa basin

C. Mocenni\textsuperscript{a}, E. Sparacino\textsuperscript{a}, A. Vicino\textsuperscript{a}, J.P. Zubelli\textsuperscript{b,∗}

\textsuperscript{a}Dipartimento di Ingegneria dell’Informazione and Centro per lo Studio dei Sistemi Complessi, Università di Siena, Via Roma 56, 53100 Siena, Italy
\textsuperscript{b}Instituto Nacional de Matemática Pura e Aplicada, Estrada Dona Castorina 110, 22460-320 Rio de Janeiro, Brazil

Received 11 January 2007; accepted 21 June 2007

Abstract

This work concerns the development and calibration of several classes of mathematical models describing ecological and bio-geochemical aspects of aquatic systems. We focus our experimental analysis on the Serra da Mesa lake in Brazil, from which the biological information is extracted by real online measurements provided by the SIMA monitoring program of the Brazilian Institute for Space Research (INPE).

A preliminary analysis is carried out so as to define the input–output data to be accounted for by the models. Furthermore, several classes of mathematical models are considered for fitting real data of biological processes. In order to do that, a two-step parameter identification/validation procedure is applied: the first step uses the integrals of the differential equations to reduce the nonlinear estimation problem to a linear least squares one. The parameter vector resulting from the first step is used for initializing a nonlinear minimization procedure. The results are discussed to assess the fitting performances of the physical and black-box models proposed in the paper. Several simulations are presented that could be used for developing scenario analysis and managing the real system.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Inverse problems; Biomathematics; Ecological modelling; Biophysical phenomena; Environmental studies

0. Introduction

Mathematical modelling of ecosystems plays a crucial role in the study and management of natural resources (see \cite{1,3–5} for engineering and ecological aspects of environmental modelling and \cite{16,21,22} for applications of the models and resource management in the Mediterranean Sea). In the particular case of the Amazon region, due to its size and peculiarities, one needs to develop models using the available observed satellite data as much as possible (for the modelling aspects of the Amazon region see \cite{26}). The research reported herein concerns the mathematical

---

\textsuperscript{*}Corresponding author.
E-mail address: zubelli@impa.br (J.P. Zubelli).

0895-7177/S - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.mcm.2007.06.010
modelling of large basins, such as the Serra da Mesa lake (see the models of this watershed developed by the authors in [15,27]) that is located in the Brazilian Amazon region. The goal is to model, calibrate, and simulate some of the relevant ecological processes, forecast the future evolution and derive the critical components of the systems. In order to accomplish this task, we have to build simplified models of complex systems and match them with noisy/incomplete data.

The long term objective of the research described here is to understand, describe and predict qualitatively as well as quantitatively the dynamics of complex aquatic systems of the Amazon region. We start with concentrated and relatively simple models. A wide literature is available on ecological models of complex ecosystems: see for example [6] for simple microalgal dynamical models, [10,11] for nutrient related models and [25,13] for complex bio-geochemical models of lagoons. Concerning the mathematical analysis of ecological nonlinear models and qualitative analysis of complex systems refer to [9,14]. More specifically, in this work we present the results obtained by means of parameter identification of physical models, which are based on the main biochemical and physical processes, as well as of black-box models. For identification theory, refer to [2], and for applications of identification procedures to physical models, see [7–9].

The resulting models may thus be used for prediction of the relevant variables in different time scales so as to support management, protection and control of ecosystems. The models under consideration could also be used for environment modelling since they trade off the simplicity of population models and the complexity of bio-geochemical ones.

We also confront some of the mathematical issues related to parameter estimation (calibration) in the context of noisy data and model mismatch. Our approach is a two-step identification process which might be of interest in other situations too.

Besides the afore-mentioned performances in terms of simulation and prediction, the models could be useful for managing and, if necessary, controlling the ecosystems. In fact, due to their simple formulation, they could be easily integrated into scenario analysis and decision support systems tools.

The paper is organized as follows: Section 1 describes the site under consideration and the monitoring program providing the measured data. Section 2 is focused on the description of the mathematical models used in the work. In Section 3 the identification procedure is explained. Section 4 presents the results in terms of the fitting performance, sensitivity to parameters and model comparison analysis. We close in Section 5 with some final remarks and suggestions for further research.

1. The Serra da Mesa lake and monitoring data

Handling aquatic ecosystems requires systematic monitoring of physical, chemical, and biological parameters. To this purpose the development of monitoring programs and analysis techniques for water quality management plays a crucial role. INPE keeps eight sites in the north of Brazil constantly under control: Corumbá, Curuai, Itumbiara, Manso, Serra Da Mesa and Tucurui. Thanks to an online data collection system we can monitor the main variables for the water quality analysis. In particular, this work is focused on two sets of data containing measurements collected during summer and winter seasons into Serra da Mesa lake, which is one of the biggest artificial lakes in Brasil. It is formed by the Tocantins river in the Minaçu (GO) plateau (460 m over the see level), north of Brasilia. It has a water volume of 54,4 billion cubic meters and an area of 1.784 square kilometers. Besides representing a tourist and fishing attraction, it gives power to waterwheels of an important hydroelectric plant generating 1275 MW. In 2001 the water level decreased by 9 m causing an energy crisis and the failure of several companies.

The exogenous measured data and the state variables used in the models are reported in Table 1.\textsuperscript{1}

 Both sets of data (summer and winter) consist of 144 samples and show missing data due to satellite transmission problems. Moreover, the shallow water causes considerable errors in chlorophyll-a measurements, which include outliers and negative values during the summer period. For these reasons a preliminary data recovering procedure was applied. The \( \hat{v}(i) \) missing measurement is thus replaced by the estimate \( \hat{v}(i) \):

\[
\hat{v}(i) = \frac{1}{2N} \sum_{j=1}^{N} (v(i+j) + v(i-j))
\]

\textsuperscript{1} In this work, we use the Chlorophyll-a data as a representative estimation of the phytoplankton biomass [3].
where $N = 12$.

Furthermore, during the winter period we do not have nutrient measurements; for this reason the relative set of data is used only for phytoplankton identification.

A preliminary correlation analysis shows a significant daily periodicity in the measurements of oxygen concentration, solar radiation and water temperature.

2. The models

In this section the mathematical models of phytoplankton dynamics used in the work are presented. In particular, two different dynamical models are described by a set of ordinary differential equations: the first one, the *PZ model*, is the extended version of a model previously developed by two of the authors [9,14]; the second one, the *Wampum model*, is taken from the literature as a benchmark [6]. The last alternative is a stochastic transfer function model derived by the structure of the previous ones.

A set of equations developed for water quality dynamics is also proposed and coupled with phytoplankton models.

2.1. Modelling microalgae dynamics

The phytoplankton variable is a population of microalgae representing the producers, i.e., the set of vegetable species performing carbon fixation, measured in terms of biomass. In this work, the constants and equations reflect the behavior of “prototypical” entities (in forms of Diatoms, Peridenes and Microflagelates).

Composition and abundance of phytoplankton are both related to the physical and chemical properties of the ecosystem, so normally microalgae are considered a reliable indicator for the trophic status of the ecosystem.

In the following, the dynamical models of phytoplankton are presented.

2.1.1. The PZ model

The *PZ model* is an extension of a simple model for the phytoplankton dynamics (see [9,14] for details on the equations and applications of the model and [23] for simple models of phytoplankton):

$$\dot{v}_1 = k_{1,1} f_2(u_2) f_3(u_3) v_1 - k_{1,2} v_1^2,$$

where

$$f_2(u_2) = 1.09 \frac{u_2 - T_{OPT}}{T_w},$$

and

$$f_3(u_3) = 0.9 u_3 e^{-E_z}.$$  

The function $f_2(u_2)$ represents the temperature effect on photosynthesis [12] and the function $f_3(u_3)$ represents the light attenuation [19]. The constants of the environmental exogenous inputs are reported in the first four rows of Table 2.

The first term in Eq. (2) accounts for the photosynthetic activity, which produces oxygen leading to an increase of phytoplankton biomass. The process is influenced by the temperature (3) and the light intensity (4). The second
term on the RHS of the equation represents the natural mortality that is assumed to be proportional to the square of phytoplankton biomass itself. The formulation of this structure is based on the logistic equation [4], which is one of the best known models of population dynamics for the vegetation microorganisms. The first two rows of Table 3 report the parameters of the model.

A more complex model is proposed in this paper for the phytoplankton dynamics as follows:

\[ \dot{v}_1 = k_{1,1} f_2(u_2) f_3(u_3) v_1 - k_{1,2} v_1^2 - k_{1,3} v_1 v_4 + k_{1,4} f_s(t) \]

(5)

\[ \dot{v}_4 = k_{4,1} v_1 v_4 - k_{4,2} v_4 \]

(6)

where \( f_2(u_2) \) and \( f_3(u_3) \) are given in (3), (4) and

\[ f_s(t) = \left( 1 + \mu \sin \left( \frac{\pi}{12} t \right) \right). \]

(7)

The function \( f_s(t) \) reproduces the effects of periodic forcing related to the photoperiod. The meaning and value of the constant of the environmental exogenous input is reported in the fifth row of Table 2, while the parameters of the model are reported in Table 3.

The variable \( v_4 \) introduced in Eq. (5) is the biomass of herbivore zooplankton consumers. These are mainly copepods as some species of Acartia. The dynamics of this variable is regulated by a growth due to grazing on phytoplankton and by the losses for natural mortality. The phytoplankton–zooplankton model is based on a logistic predator–prey system with Holling II type response [4], with linear mortality in the zooplankton equation [3].

2.1.2. The Wampum model

The Wampum model has been developed by Romanowicz et al. [6,17] for the control of the phytoplankton biomass in the Elbe River (Germany). The equations of the model are:

\[ \dot{v}_1 = k_{1,1} q_1(u_2) f_1(u_3, v_1) v_1 - k_{1,2} q_2(u_2) v_1, \]

(8)

where

\[ q_1(u_2) = 1.014^{u_2 - T_{OPT}}, \]

(9)

\[ q_2(u_2) = 1.02^{u_2 - T_{OPT}}. \]

(10)
The functions $q_1(u_2)$ and $q_2(u_2)$ describe the effect of the temperature in the algal growth and mortality, respectively, whereas

$$f_i(u_3, v_1) = \frac{u_3 e^{-\lambda(v_1) z}}{\sqrt{K_i^2 + u_3^2 e^{-2\lambda(v_1) z}}}$$

represents the light limitation factor resulting from vertically averaging the so-called “Smith formula” [18]. According to the well known Beer’s law on the light attenuation (see [23] for a description of the mathematical formulation of the law), the light intensity at depth $z$ below the water surface is $I e^{-\lambda z}$, with $I$ denoting the radiation intensity at the water surface and $\lambda(v_1) = \lambda_m + \lambda_s v_1$ the total light attenuation due to mineral compounds ($\lambda_m$) and algal self shading ($\lambda_s$). The constants of the environmental exogenous inputs are described in Table 2.

The first term of Eq. (8) accounts for the phytoplankton growth. The light climate is one of the most important factors influencing the evolution of algae populations, as pointed out in [17]. The second term of the equation represents the natural mortality due to loss and respiration. Table 4 reports the parameters of the model.

### 2.2.1. The Oxy-model

The oxygen concentration represents the link between population ecological processes (related to phytoplankton) and biophysical phenomena (related to exogenous inputs and nutrient dynamics). In this paper the oxygen dynamics is firstly analyzed as a single kinetic equation and then coupled with the phytoplankton and nutrient

---

**Table 4**

Parameters of the Wampum model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Process description</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{1,1}$</td>
<td>Phytoplankton growth</td>
<td>$(^\circ C)^{-1} (t)^{-1}$</td>
</tr>
<tr>
<td>$k_{1,2}$</td>
<td>Phytoplankton natural mortality</td>
<td>$(^\circ C)^{-1} (t)^{-1}$</td>
</tr>
</tbody>
</table>
and accounts for the aerobic production of nutrients by mineralization of organic matter. Table 5 reports the meaning of the parameters of the model.

2.2.2. The nutrient model

The nutrient variable refers to the nitrogen compounds in water and sediments. There is evidence from field experiments [25] that often the main source of nutrients in shallow water systems for phytoplankton growth must come from recycling due to bacterial activity and sediment release, while the losses are due to the photosynthetic activity, outgoing flows of water and material and retaining from sediment.

Nutrient dynamics is regulated by the following equation [10,11]:

\[ \dot{v}_3 = k_{3,1} f_M(u_2) f_a(v_2) - k_{3,2} f_2(u_2) f_3(u_3) v_1 v_3 + k_{3,3}(f - v_3), \]  

where \( f_2(u_2), \ f_3(u_3), \ f_M(u_2) \) and \( f_a(v_2) \) are given in (3), (4), (16) and (15).

The first term of the RHS of Eq. (17) accounts for the aerobic production of nutrients by mineralization of organic matter. The second one represents the consumption due to the photosynthetic activity of phytoplankton species. The last one shows the nutrient quantity that is exchanged with external sources. Table 6 reports the meaning of the model parameters.

The two physical models PZ (5) and Wampum (8) describing the phytoplankton dynamics have been coupled with the oxygen (13) and nutrient (17) models. Numerical results of the identification of the coupled models are reported in Section 4.3.

3. Model identification

The parameter identification for the physical models is based on the minimization of a cost function (see [2] for identification theory and fundamentals), representing the mean square error between simulated and experimental data.
A two-step procedure to determine an initial condition for a nonlinear minimization is applied to the phytoplankton equation (see [7,8] for a detailed description and applications of the method), as described in the following.

3.1. Setting the initial parameter vector for nonlinear estimation

Consider the simplified version of the phytoplankton dynamical model (2) described by a nonlinear non-autonomous ordinary differential equation. Rewrite it as

$$\frac{\dot{v}_1}{v_1} = k_{1,1} M(t) - k_{1,2} v_1,$$

(18)

where $M(t) = f_2(u_2(t)) f_3(u_3(t))$. Then, integrating over a time interval $[t_i, t_{i+1}]$ gives

$$\ln(v_1(t_{i+1})/v_1(t_i)) = k_{1,1} \int_{t_i}^{t_{i+1}} M(\tau)d\tau - k_{1,2} \int_{t_i}^{t_{i+1}} v_1(\tau)d\tau.$$  

(19)

Considering Eq. (19) for $i = 1, \ldots, N - 1$, we obtain a linear system of equations in the variable $\theta := (k_{1,1}, k_{1,2})'$ of the form

$$Y = U\theta,$$

(20)

where

$$Y = (\ln(v_1(t_{i+1})/v_1(t_i)))_{i=1,...,N-1},$$

and

$$U = \left(\int_{t_i}^{t_{i+1}} M(\tau)d\tau, \int_{t_i}^{t_{i+1}} v_1(\tau)d\tau\right)_{i=1,...,N-1}.$$

If we replace $v_1(\cdot), u_2(\cdot), u_3(\cdot)$ by measurements, and approximate the integrals in (19) by numerical quadrature, then Eq. (20) becomes

$$Y = \hat{U}\theta + e,$$

(21)

where $e$ is an error caused by noise and numerical quadrature, and $\hat{U}$ is the approximate value of $U$.

We now choose as initial parameter estimates for the model (21) a least squares estimate of $\theta$ in (21) given by

$$\theta_{LS} = (\hat{U}'\hat{U})^{-1}\hat{U}'Y.$$  

(22)

3.2. Nonlinear estimation

The initial guess $\theta_{LS}$ obtained above would most likely not coincide with the correct value even in the absence of noise in the measurements and model imperfections. In order to improve on such an estimate we now perform a nonlinear estimation as described in the following.

We consider the cost function $F(\theta)$, representing the mean square error between simulated and experimental data [2]

$$F(\theta) = \frac{1}{N} \sum_{i=1}^{N} e^2(t_i) = \frac{1}{N} \sum_{i=1}^{N} (\phi(\theta, t_i) - \bar{\phi}(t_i))^TW(\phi(\theta, t_i) - \bar{\phi}(t_i)),$$

(23)
Table 7
Identification results for the phytoplankton ($v_1$) dynamical models

<table>
<thead>
<tr>
<th>Model</th>
<th>Fit</th>
<th>MSE</th>
<th>$\bar{e}_r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PZ-summer</td>
<td>23.1892</td>
<td>0.9666</td>
<td>0.074718</td>
</tr>
<tr>
<td>PZ-winter</td>
<td>−8.3209</td>
<td>0.33986</td>
<td>0.0919</td>
</tr>
<tr>
<td>WP-summer</td>
<td>8.3049</td>
<td>1.3775</td>
<td>0.16589</td>
</tr>
<tr>
<td>WP-winter</td>
<td>5.505</td>
<td>0.25864</td>
<td>0.12939</td>
</tr>
<tr>
<td>OE-summer</td>
<td>57.0698</td>
<td>0.30195</td>
<td>3.99e−05</td>
</tr>
<tr>
<td>OE-winter</td>
<td>32.1203</td>
<td>0.13346</td>
<td>0.06509</td>
</tr>
</tbody>
</table>

where $\phi(t_i)$ is the measurement vector at time $t_i$, $\phi(\theta, t_i)$ is the vector of corresponding values provided by the model at time $t_i$, $\theta$ is the vector of model parameters, and $W$ is a suitable weight matrix.

The parametric identification of the microalgae dynamical models (PZ, Wampum and OE) is performed using the phytoplankton data set ($v_1$). The parametric identification of the water quality model has been based alternatively on the oxygen ($v_2$) and nutrients ($v_3$) concentrations.

The identification of the coupled models is performed by minimizing the cost function (23), where $\theta$ is the vector of all the parameters related to the three equations involved in the procedure. The vector $\phi(t_i) = (v_1(t_i), v_2(t_i), v_3(t_i))^T$ includes measurements $\phi(\theta, t_i) = (v_1(t_i), v_2(t_i), v_3(t_i))^T$ is the vector of corresponding values provided by the model at time $t_i$ and $W = \text{diag}([\sigma^{-2}(v_j)], j = 1, \ldots, 3)$, where $\sigma^{-2}(v_j)$ is the sample variance of the data $v_j$.

Nonlinear optimization has been performed by means of a quasi-Newton algorithm implemented in the NAG Toolbox for Matlab.

3.3. Identification of the black-box models

The parameter estimation of the black-box models is performed by minimizing the one-step ahead prediction error implemented in the System Identification Toolbox for Matlab [2]:

$$ J(\theta) = \frac{1}{N} \sum_{i=1}^{N} (y(t) - \hat{y}(t|t - 1; \theta))^2, $$

(24)

where $y(t)$ is the measurement value and $\hat{y}(t|t - 1; \theta)$ is the model predicted output at time $t$.

4. Results

In this section the numerical results of the parameter identification of the models are presented. The results are organized similarly to those of Sections 2 and 3. Comparisons and benchmark exercises between models are also interpreted.

Here, the following fit indicator is used:

$$ \text{Fit} = \left(1 - \frac{\|y(t) - \hat{y}(t|t - 1)\|}{\|y(t) - \text{mean}(y(t))\|}\right) \cdot 100 $$

(25)

which represents the percentage of real data variance captured by the models. Moreover, the mean square error (MSE) and the sample mean ($\bar{e}_r$) of the residual error signal will be computed and displayed. See [2] for a detailed description.

4.1. Microalgae dynamical models

As far as the PZ model is concerned, estimation of the parameters of these models has been performed through a two-step identification procedure. In fact, the structure of this model is such that it is possible to compute an initial estimate of the parameters by the integral minimization procedure described in Section 3. This fact allowed us to reduce consistently uncertainty on the parameter vector initialization.

A comparison between the models PZ and Wampum on the numerical results obtained shows that the former model performs better on the summer data, while the latter provides slightly better performance on winter data (see Table 7).
Table 8
Estimated parameter values for the PZ model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Initial value (winter)</th>
<th>Final value (winter)</th>
<th>Initial value (summer)</th>
<th>Final value (summer)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{1,1}$</td>
<td>$3.868 \times 10^{-4}$</td>
<td>$8.311 \times 10^{-4}$</td>
<td>$-1.278 \times 10^{-2}$</td>
<td>$2.668 \times 10^{-4}$</td>
</tr>
<tr>
<td>$k_{1,2}$</td>
<td>$2.201 \times 10^{-4}$</td>
<td>$4.298 \times 10^{-4}$</td>
<td>$-1.183 \times 10^{-2}$</td>
<td>$4.056 \times 10^{-3}$</td>
</tr>
<tr>
<td>$k_{1,3}$</td>
<td>$1 \times 3$</td>
<td>$1.144 \times 10^{-4}$</td>
<td>$1 \times 3$</td>
<td>$1.169 \times 10^{-6}$</td>
</tr>
<tr>
<td>$k_{1,4}$</td>
<td>$1 \times 3$</td>
<td>$1.387 \times 10^{-3}$</td>
<td>$1 \times 3$</td>
<td>$4.646 \times 10^{-2}$</td>
</tr>
<tr>
<td>$k_{4,1}$</td>
<td>$1 \times 1$</td>
<td>$1.144 \times 10^{-4}$</td>
<td>$1 \times 1$</td>
<td>$1.257 \times 10^{-1}$</td>
</tr>
<tr>
<td>$k_{4,2}$</td>
<td>$1 \times 1$</td>
<td>$1.053 \times 10^{-2}$</td>
<td>$1 \times 1$</td>
<td>$1.876 \times 10^{-1}$</td>
</tr>
</tbody>
</table>

Table 9
Estimated parameter values for the Wampum model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Initial value (winter)</th>
<th>Final value (winter)</th>
<th>Initial value (summer)</th>
<th>Final value (summer)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{1,1}$</td>
<td>$1.593 \times 10^{-2}$</td>
<td>$1.614 \times 10^{-2}$</td>
<td>$-2.365 \times 10^{-1}$</td>
<td>$-2.338 \times 10^{-1}$</td>
</tr>
<tr>
<td>$k_{1,2}$</td>
<td>$4.845 \times 10^{-3}$</td>
<td>$4.531 \times 10^{-3}$</td>
<td>$-4.706 \times 10^{-2}$</td>
<td>$-4.622 \times 10^{-2}$</td>
</tr>
</tbody>
</table>

Fig. 1. Fitting results of PZ model: simulation (solid line) and measurements (dashed line). Summer data.

The values of the estimated parameters are reported in Tables 8 and 9 and simulation results of the estimated models are reported in Figs. 1–4.

Comparing the numerical results of physical models with the stochastic OE model, it turns out that OE provides better results than the PZ and the Wampum models. This fact is more evident on the summer dataset (see Table 7 and Figs. 5 and 6). Anyway, it should be stressed that the OE model involves the estimation of 8 parameters, while PZ and Wampum models require 6 and 2 parameters, respectively. The estimated values of the OE model parameters are reported in Table 10.

4.2. Water quality models

Water quality models have been estimated on the summer dataset only, because of the lack of nutrient data in the winter period. Numerical results obtained are reported in Table 11, Figs. 7 and 8. Table 12 shows the values of the estimated parameters. Notice that the dissolved oxygen model shows excellent performance on the real data.

4.3. Coupled models

In this subsection the results of the identification of the coupled models are presented.
The microalgae physical models are coupled with the water quality ones and all the parameters are estimated. The previously estimated parameters for the microalgae and quality models are used for the initialization of the augmented parameter vector of the coupled model.
Table 10
Estimated parameter values for the OE model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Final value (winter)</th>
<th>Final value (summer)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_{1,1}$</td>
<td>$-1.952$</td>
<td>$-1.889$</td>
</tr>
<tr>
<td>$a_{1,2}$</td>
<td>$9.629e-1$</td>
<td>$8.919e-1$</td>
</tr>
<tr>
<td>$a_{2,1}$</td>
<td>$-1.721$</td>
<td>$-1.909$</td>
</tr>
<tr>
<td>$a_{2,2}$</td>
<td>$7.278e-1$</td>
<td>$9.316e-1$</td>
</tr>
<tr>
<td>$b_{1,1}$</td>
<td>$4.946e-1$</td>
<td>$6.319e-1$</td>
</tr>
<tr>
<td>$b_{1,2}$</td>
<td>$-1.157$</td>
<td>$-6.316e-1$</td>
</tr>
<tr>
<td>$b_{1,3}$</td>
<td>$6.691e-1$</td>
<td>$0$</td>
</tr>
<tr>
<td>$b_{2,1}$</td>
<td>$7.603e-5$</td>
<td>$-3.753e-4$</td>
</tr>
<tr>
<td>$b_{2,2}$</td>
<td>$9.773e-5$</td>
<td>$2.149e-4$</td>
</tr>
<tr>
<td>$b_{2,3}$</td>
<td>$-1.98e-4$</td>
<td>$0$</td>
</tr>
</tbody>
</table>

The results of the identification procedure corresponding to all the models considered are reported in Table 13. Plots of real and fitted data of nutrients are reported in Fig. 9(a)–(c). As it can be observed from Table 13 and Fig. 9, a remarkable improvement of the performance of the coupled PZ model is obtained. In fact, in this case the coupled PZ model provides the best results both in terms of the Fit and MSE criteria.
Table 11
Identification results for the water quality models

<table>
<thead>
<tr>
<th>Model</th>
<th>Fit</th>
<th>MSE</th>
<th>( \bar{e}_r )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oxy ((v_2))</td>
<td>90.1653</td>
<td>0.00142</td>
<td>0.00346</td>
</tr>
<tr>
<td>Nut ((v_3))</td>
<td>37.3527</td>
<td>0.03063</td>
<td>0.031663</td>
</tr>
</tbody>
</table>

Fig. 7. Fitting results of Oxygen model: simulation (solid line) and measurements (dashed line). Summer data.

Fig. 8. Fitting results of Nutrient model: simulation (solid line) and measurements (dashed line). Summer data.

Table 12
Estimated parameter values for the Oxy and Nut models

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Initial value</th>
<th>Final value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k_{2,1} )</td>
<td>1e−4</td>
<td>8.679e−5</td>
</tr>
<tr>
<td>( k_{2,2} )</td>
<td>1e−4</td>
<td>4.450e−2</td>
</tr>
<tr>
<td>( k_{2,3} )</td>
<td>1e−4</td>
<td>1e−9</td>
</tr>
<tr>
<td>( k_{2,4} )</td>
<td>1e−4</td>
<td>3.341e−7</td>
</tr>
<tr>
<td>( k_{3,1} )</td>
<td>1e−4</td>
<td>1.147e−3</td>
</tr>
<tr>
<td>( k_{3,2} )</td>
<td>1e−4</td>
<td>7.054e−4</td>
</tr>
<tr>
<td>( k_{3,3} )</td>
<td>1e−4</td>
<td>1.940e−3</td>
</tr>
<tr>
<td>( f )</td>
<td>1e−4</td>
<td>2.141e−4</td>
</tr>
</tbody>
</table>

Summer data.
Table 13
Identification results for the coupled models

<table>
<thead>
<tr>
<th>Model</th>
<th>Fit</th>
<th>MSE</th>
<th>$\bar{\epsilon}_r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PZ-coupled-phyto</td>
<td>22.8704</td>
<td>0.97464</td>
<td>0.12606</td>
</tr>
<tr>
<td>PZ-coupled-oxy</td>
<td>90.3727</td>
<td>0.00136</td>
<td>0.00435</td>
</tr>
<tr>
<td>PZ-coupled-nut</td>
<td>41.443</td>
<td>0.02676</td>
<td>0.02458</td>
</tr>
<tr>
<td>WP-coupled-phyto</td>
<td>8.2382</td>
<td>1.3795</td>
<td>0.13521</td>
</tr>
<tr>
<td>WP-coupled-oxy</td>
<td>90.4364</td>
<td>0.00134</td>
<td>0.00322</td>
</tr>
<tr>
<td>WP-coupled-nut</td>
<td>26.5657</td>
<td>0.04209</td>
<td>-0.00367</td>
</tr>
<tr>
<td>OE-coupled-oxy</td>
<td>90.507</td>
<td>0.00132</td>
<td>0.00579</td>
</tr>
<tr>
<td>OE-coupled-nut</td>
<td>36.9703</td>
<td>0.0310</td>
<td>0.023278</td>
</tr>
</tbody>
</table>

Fig. 9. Nutrient dynamics in summer. Identification results of PZ-coupled (a), Wampum-coupled (b) and OE-coupled (c) models. Simulations are represented by solid lines and measurements by dashed lines.

4.4. Sensitivity analysis

In this subsection we discuss the sensitivity of PZ and Wampum models with respect to the reconstructed parameters. In order to do that, we perform two studies:

1. The change of the cost function with respect to a variation of the parameters when each parameter is varied in a range of $\pm 10\%$.
2. The logarithmic derivative of the simulated values given by the model with respect to the parameters according to the formula...
Fig. 10. Sensitivity analysis of PZ and Wampum models. MSE values for ±10% parameter variations in winter and summer datasets.

\[
\text{MSE}_{\text{SENS}} = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{\partial v(t_i)}{\partial k} / v(t_i) \right)^2
\]

(26)

where \( k \) denotes any of the parameters \( k_{i,j} \) in the models.

With reference to the first indicator, numerical results are reported in Fig. 10. We remark that the Wampum cost variations are smoother than the PZ ones. In any case, the PZ model provides much better performances on the summer data than in the winter. In fact, the minimum cost is consistently lower than that provided by the Wampum model during the same period.

With reference to the second indicator (26), Table 14 reports the numerical results obtained for the PZ and Wampum models. The improvement of the PZ model performances, reported in Table 7, may be due to the introduction of the predation term \( k_{1,3} \). This fact is confirmed by a higher sensitivity of the model output to this parameter in the summer period. On the other hand, the higher complexity of the model may increase the sensitivity with respect to the initial conditions.

The above comments allow us to conclude that while the Wampum model is more robust to initial conditions and parameter variations, it seems to be quite unable to capture seasonal dynamics such as summer–winter climatic changes. Analogous considerations hold when comparing the performances of the PZ and Wampum models coupled with the water quality ones (see Table 13).
Table 14
Sensitivity analysis of PZ model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MSESENS (winter)</th>
<th>MSESENS (summer)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{1,1}$</td>
<td>1.849e+8</td>
<td>1.831e+7</td>
</tr>
<tr>
<td>$k_{1,2}$</td>
<td>1.077e+6</td>
<td>3.534e+4</td>
</tr>
<tr>
<td>$k_{1,3}$</td>
<td>3.227</td>
<td>5.290e+9</td>
</tr>
<tr>
<td>$k_{1,4}$</td>
<td>1.845e+2</td>
<td>1.984e+4</td>
</tr>
</tbody>
</table>

Mean square value of $\frac{\partial v}{\partial k_i}/v$ in the winter and summer.

Table 15
Sensitivity analysis of the Wampum model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MSESENS (winter)</th>
<th>MSESENS (summer)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{1,1}$</td>
<td>6.692e+2</td>
<td>2.528e+2</td>
</tr>
<tr>
<td>$k_{1,2}$</td>
<td>6.623e+3</td>
<td>7.919e+3</td>
</tr>
</tbody>
</table>

Mean square value of $\frac{\partial v}{\partial k_i}/v$ in the winter and summer.

5. Conclusions and suggestions for further research

In this article, physical and black-box models describing the microalgae and the water quality dynamics in the Serra da Mesa basin in the Brazilian Amazon region are proposed. The variables involved in the models are: phytoplankton, which is considered as a population of microalgae, oxygen and nutrients, which are the most important indicators for the water quality. Two different physical models and a stochastic transfer function one are developed for the phytoplankton dynamics, while two physical models account for the oxygen and nutrient dynamics. The microalgae physical models are then coupled with the water quality ones for integrating ecological and physico-chemical processes.

By using a nonlinear parametric optimization procedure, the parameters of the developed models have been estimated on the real data. A novel two-step parameter identification procedure for identifying the microalgae dynamic models has been adopted. Several simulations are presented and the results are discussed to evaluate the fitting performances and sensitivities of the models.

The analysis of sensitivity and the values shown in Tables 14 and 15 indicate a substantial discrepancy in the orders of magnitude for the relative sensitivity of the parameters. These values, together with numerical experiments we performed by varying the initial guesses for the parameters, suggest that further improvements could be obtained by introducing suitable regularization terms in the cost function.

Acknowledgments

C.M. and J.P.Z would like to thank Arcilan T. Assireu, Claudio Barbosa, Gilberto Camara and Evlyn M. L. Moraes Novo (INPE) for their hospitality and for the help with questions concerning the SIMA project. J.P.Z. would like to thank the hospitality of the University of Siena and the C.S.C., where part of this research was developed. J.P.Z. acknowledges financial support from CNPq through grants 302161 and 474085.

References


