Mathematical Modelling and Parameter Estimation of the Serra da Mesa Basin

C. Mocenni[a], E. Sparacino[a], A. Vicino[a] and J. P. Zubelli[b]

[a]Dipartimento di Ingegneria dell'Informazione, Università di Siena,
Via Roma 56, 53100 Siena, Italy

[b]Instituto Nacional de Matemática Pura e Aplicada,
Estrada Dona Castorina 110, 22460-320 Rio de Janeiro, Brasil

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. [1]

Keywords: Inverse problems, biomathematics, ecological modelling, biophysical phenomena, environmental studies

INTRODUCTION

Mathematical modelling of ecosystems plays a crucial role in the study and management of natural resources (see [1,3–5] for engineering and ecological aspects of environmental modelling and [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 Amazon region see [26]). The research reported herein concerns the mathematical modelling of large basins, such as the Serra

[1]The research reported here is part of an international cooperation among the Instituto Nacional de Matemática Pura e Aplicada (Br), the Laboratório Nacional de Computação Científica (Br), and the Centro per lo Studio dei Sistemi Complessi of the University of Siena (It).

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 microalgae 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 most relevant biochemical and physical processes, as well as on black-box models. For identification theory, refer to [2], and for applications of identification procedure 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 represent a compromise between 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 would 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, Curuaí, Itumbiara, Manso, Serra Da Mesa* and *Tucuruí.* 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 squared 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 meters causing an energetic crisis and the failure of several companies.

The exogenous measured data and the state variables used in the models are reported in Table 1[2].

Table 1
Exogenous inputs and state variables.

Name	Var.	Units
Wind velocity	u_1	[m/s]
Water temperature	u_2	[°C]
Solar Radiation	u_3	[W/m^2]
Percentage of oxygen saturation	u_4	[%]
Chlorophyll-a concentration	v_1	[μg/l]
Oxygen concentration	v_2	[mg/l]
Nutrient Concentration	v_3	[mg/l]

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 some errors in chlorophyll-a measurements. These include outliers and negative values during the summer period. For those reasons a preliminary data recovering procedure was applied. The i-th missing measurement is thus replaced by the estimate $\hat{v}(i)$:

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

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 models of the 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 one is a stochastic transfer

[2]In this work, we use the Chlorophyll-a data as a representative estimation of the phytoplankton biomass [3].

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 sequel, the dynamical models of phytoplankton are introduced.

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):

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

where

$$f_2(u_2) = 1.09^{\frac{u_2 - T_{OPT}}{T_W}}, \tag{3}$$

and

$$f_3(u_3) = 0.9 u_3 e^{-Ez}. \tag{4}$$

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 the Table 2.

Table 2
Constants of the models.

Parameter	Biological Meaning	Value	Units
T_{OPT}	Optimal temperature	29	°C
T_W	Temperature width	1	°C
E	Albedo coefficient	3	m^{-1}
z	Water depth	1	m
μ	Amplitude of the photoperiod forcing function	1.47	-
K_l	Light intensity (71% max value of algae growth rate)	200	$W\,m^{-2}$
λ_m	Algae mineral compounds coefficient	0.2	-
λ_s	Algae self shading coefficient	0.02	-

The first term in Equation (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 model of population dynamics for the vegetation microorganisms. The first two rows of Table 3 report the parameters of the model.

Table 3
Parameters of the PZ-model.

Par.	Biological Meaning	Units
$k_{1,1}$	Phytoplankton growth rate	$[\mathrm{Wm}^{-2}]^{-1}[t]^{-1}$
$k_{1,2}$	Phytoplankton natural mortality	$[t]^{-1}$
$k_{1,3}$	Phytoplankton losses for grazing	$[\mathrm{mg\,l}^{-1}]^{-1}[t]^{-1}$
$k_{1,4}$	External phytoplankton input/output	$[\mu\mathrm{g\,l}^{-1}][t]^{-1}$
$k_{4,1}$	Zooplankton grow rate	$[\mu\mathrm{g\,l}^{-1}][t]^{-1}$
$k_{4,2}$	Zooplankton natural mortality	$[t]^{-1}$

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

$$\dot{v}_1 = k_{1,1} f_2(u_2) f_3(u_3) v_1 - k_{1,2} v_1^2 - k_{1,3} v_1 v_4 + k_{1,4} f_s(t) \tag{5}$$

$$\dot{v}_4 = k_{4,1} v_1 v_4 - k_{4,2} v_4 \tag{6}$$

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

$$f_s(t) = \left(1 + \mu \sin\left(\frac{\pi}{12} t\right)\right). \tag{7}$$

The function $f_s(t)$ reproduces the effects of periodic forcing related to the photoperiod. The constants of the environmental exogenous inputs are reported in the fifth and sixth row of the Table 2, while the parameters of the model are reported in Table 3.

The variable v_4 introduced in Equation (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].

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:

$$\dot{v}_1 = k_{1,1} q_1(u_2) f_l(u_3, v_1) v_1 - k_{1,2} q_2(u_2) v_1, \tag{8}$$

where

$$q_1(u_2) = 1.014^{u_2 - T_{OPT}}, \tag{9}$$

$$q_2(u_2) = 1.02^{u_2 - T_{OPT}}. \tag{10}$$

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

$$f_l(u_3, v_1) = \frac{u_3 e^{-\lambda(v_1)z}}{\sqrt{K_l^2 + u_3^2 e^{-2\lambda(v_1)z}}} \tag{11}$$

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 $Ie^{-\lambda z}$, 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 (λ_m) and algal self shading (λ_s). The constants of the environmental exogenous inputs are described in Table 2.

The first term of equation (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.

Table 4
Parameters of the Wampum-model.

	Description	Units
$k_{1,1}$	Phytoplankton grow rate	$[°\mathrm{C}]^{-1}[t]^{-1}$
$k_{1,2}$	Phytoplankton natural mortality	$[°\mathrm{C}]^{-1}[t]^{-1}$

THE OUTPUT ERROR (OE) MODEL

OE model is a Black-Box Multiple Input Single Output (MISO) Stochastic Transfer Function (STF) model. The structure of the model is derived from the Wampum one [6]. The mathematical formulation of the model is

$$v_1(t) = \sum_{i=1}^{i=M} \frac{B_i(z^{-1})}{A_i(z^{-1})} u_i(t - \delta_i) + \xi(t) \ , \tag{12}$$

where $v_1(t)$ is the algae concentration; $u_i(t - \delta_i)$ is the vector of input variables at sample time t; δ_i denotes the time delay for the i-th input; and $\xi(t)$ represents a Gaussian white noise.

The polynomials $A_i(z^{-1})$ and $B_i(z^{-1})$ are defined as $A_i(z^{-1}) = 1 + a_{i,1}z^{-1} + \ldots + a_{i,n}z^{-n}$ and $B_i(z^{-1}) = b_{i,0} + b_{i,1}z^{-1} + \ldots + b_{i,m}z^{-m}$, where a_i and b_j are the model parameters and the operator z^{-i} denotes an i-step backward shift in time.

2.2. MODELLING WATER QUALITY

This section concerns the water quality modelling process. For this purpose the oxygen and nutrient dynamics need be taken into account and their dynamic models will be described in the sequel. In fact, these two components are the most important indicators for eutrophication and pollution from external sources, being involved in the photosynthesis and mineralization of organic carbon in water and sediment layers.

The Oxy-model. Oxygen is an ecological variable of extreme importance for the overall functionality of the system. It is produced by the photosynthetic activity, integrated by physical re-aeration due to wind regime, consumed for respiration by all species of the living community and for biochemical reactions attending organic matter degradation. The oxygen quantity in water and sediments is an important indicator of anoxic crises due to excessive growth of microalgae (leading to important release of nutrients from the sediments by bacterial mineralization activity), high temperature and scarce re-aeration.

For these reasons 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 equations. For the coupling of ecological models with physico-chemical aspects, see [13,14,24], while for the effects of nutrients on phytoplankton growth refer to [20].

In the present work, the oxygen dynamics will be modelled by the following equation:

$$\dot{v}_2 = k_{2,1} f_2(u_2) f_3(u_3) v_1 v_3 + k_{2,2} f_W(u_1)(\frac{v_2}{u_4} - v_2) - k_{2,3} f_M(u_2) f_\alpha(v_2) - k_{2,4} v_1 v_2, \tag{13}$$

where

$$f_W(u_1) = 0.641 + 0.0256 \left(\frac{u_1}{0.0447}\right)^2 , \tag{14}$$

$$f_\alpha(v_2) = \frac{v_2}{k_{AE} + v_2} , \tag{15}$$

and

$$f_M(u_2) = e^{0.07 u_2}. \tag{16}$$

The functions $f_2(u_2)$ and $f_3(u_3)$ are given in (3), (4); the functions $f_W(u_1)$ and $f_\alpha(v_2)$ represent the re-aeration rate and aerobic oxygen consumption (with the limiting factor $K_{AE} = 0.5[\mathrm{mg\,l}^{-1}]^2$), respectively, and the function $f_M(u_2)$ describes the organic matter degradation as function of temperature.

The first production term in Equation (13) represents the primary production performed by phytoplankton photosynthetic process. The second one, represents the equilibrium physico-chemical reaction between gaseous oxygen and dissolved oxygen. The third one, accounts for aerobic bacterial respiration consumption, while the fourth term represents phytoplankton respiration losses. Table 5 reports the parameters of the model.

The Nutrient Model. The nutrient variable refers to the nitrogen compounds in water and sediments. There is evidence from field experiments [25] that often in shallow water systems the main source of nutrients for phytoplankton growth must come from

Table 5
Parameters of the Oxy-model.

	Description	Units
$k_{2,1}$	Oxygen produced by photosynthesis	$[\mathrm{Wm}^{-2}]^{-1}[\mu\mathrm{g\,l}^{-1}]^{-1}[t]^{-1}$
$k_{2,2}$	Oxygen exchange with atmosphere	$[\mathrm{m\,s}^{-1}]^{-1}[t]^{-1}$
$k_{2,3}$	Oxygen consumption by aerobic activity	$[°C]^{-1}[\mathrm{mg\,l}^{-1}][t]^{-1}$
$k_{2,4}$	Oxygen consumption by phytoplankton	$[\mu g\, l^{-1}]^{-1}[t]^{-1}$

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]:

$$\dot{v}_3 = k_{3,1} f_M(u_2) f_\alpha(v_2) - k_{3,2} f_2(u_2) f_3(u_3) v_1 v_3 + k_{3,3}(f - v_3) \ , \tag{17}$$

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

The first term of the RHS of Equation (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 parameter of the model.

Table 6
Parameters of the Nut-model.

	Description	Units
$k_{3,1}$	Nutrient production by abiotic processes	$[°\mathrm{C}]^{-1}[t]^{-1}$
$k_{3,2}$	Nutrient consumption by photosynthesis	$[\mu\mathrm{g\,l}^{-1}]^{-1}[\mathrm{W\,m}^{2}]^{-1}[t]^{-1}$
$k_{3,3}$	External nutrient input/output	$[t]^{-1}$
f	External nutrient input	$[\mathrm{mg\,l}^{-1}]$

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 Subsection 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 sequel.

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

$$\frac{\dot{v}_1}{v_1} = k_{1,1}M(t) - k_{1,2}v_1, \tag{18}$$

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

$$\ln(v_1(t_{i+1}))/v_1(t_i)) = k_{1,1} \int_{t_i}^{t_{i+1}} M(\tau)\ d\tau - k_{1,2} \int_{t_i}^{t_{i+1}} v_1(\tau)\ d\tau. \tag{19}$$

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

$$Y \ = \ U\ \theta, \tag{20}$$

where

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

and

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

If we replace $v_1(\cdot)$, $u_2(\cdot)$, $u_3(\cdot)$ by measurements, and approximate the integrals in (19) by numerical quadrature, then equation (3) becomes

$$Y \ = \ \widehat{U}\ \theta \ + \ e\ , \tag{21}$$

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

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

$$\theta_{LS} = (\widehat{U}'\ \widehat{U})^{-1}\ \widehat{U}'Y. \tag{22}$$

NONLINEAR ESTIMATION

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

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

$$F(\theta) = \frac{1}{N}\sum_{i=1}^{N} e^2(t_i) = \frac{1}{N}\sum_{i=1}^{N}(\phi(\theta, t_i) - \overline{\phi}(t_i))^T W(\phi(\theta, t_i) - \overline{\phi}(t_i)), \tag{23}$$

where $\overline{\phi}(t_i)$ 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, θ 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 ($\overline{v}_1$). The parametric identification of the water quality model has been based alternatively on the oxygen ($\overline{v}_2$) and ($\overline{v}_3$) concentrations.

The identification of the coupled models is performed by minimizing the cost function (23), where θ is the vector of all the parameters related to the three equations involved in the procedure, the data set is the vector $\overline{\phi}(t_i) = (\overline{v}_1(t_i), \overline{v}_2(t_i), \overline{v}_3(t_i))'$, $\phi(\theta, t_i) = (v_1(t_i), v_2(t_i), v_3(t_i))'$ is the vector of corresponding values provided by the model at time t_i and $W = \text{diag}\ \{\overline{\sigma}^{-2}(\overline{v}_j), j = 1, \ldots, 3\}$, where $\overline{\sigma}^{-2}(\overline{v}_j)$ is the sample variance of the data $\overline{v}_j$.

The nonlinear minimization in this work makes use of a quasi-Newton algorithm with simple bounds and was implemented using NAG foundation toolbox for Matlab subroutine *e04jaf*.

IDENTIFICATION OF THE BLACK-BOX MODELS

The parameter estimation of the black-box models is performed by minimizing the one-step ahead prediction error [2]:

$$J(\theta) = \frac{1}{N} \sum_{t=1}^{N} (y(t) - \hat{y}(t|t-1;\theta))^2, \tag{24}$$

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:

$$Fit = \left(1 - \frac{||y(t) - \hat{y}(t|t-1)||}{||y(t) - \text{mean}(y(t))||}\right) \cdot 100 \tag{25}$$

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.

As it may be observed from the numerical results obtained, the model performs better on the winter data than on the summer data. See MSE in Table 7.

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

Model	Fit	MSE	$\bar{e}_r$
PZ-Summer	23.1892	0.9666	0.074718
PZ-Winter	-8.3209	0.33986	0.0919
WP-Summer	8.3049	1.3775	0.16589
WP-Winter	5.505	0.25864	0.12939
OE-Summer	57.0698	0.30195	3.99e-05
OE-Winter	32.1203	0.13346	0.06509

This fact can be at least partially explained by the fact that the summer database has a larger amount of missing data. Further explanation is the presence of ecological or physical processes affecting the system in summer that are not taken into account by the adopted model. In fact, differences in water level and hydrodynamics conditions in the two seasons are quite relevant to the ecosystem behavior.

Concerning the comparison between the models *PZ* and *Wampum*, the first one shows better fitting results in summer and slightly worse in winter (see MSE in Table 7).

The values of the estimated parameters are reported in Tables 8 and 9 and simulation results of the estimated models are reported in Figures 1, 2, 3 and 4.

Table 8
Estimated parameter values for the PZ model.

Parameter	Initial Value (Winter)	Final Value (Winter)	Initial Value (Summer)	Final Value (Summer)
$k_{1,1}$	3.868e-4	8.311e-4	-1.278e-2	2.668e-4
$k_{1,2}$	2.201e-4	4.298e-4	-1.183e-2	4.056e-3
$k_{1,3}$	1e-3	1.144e-4	1e-3	1.169e-6
$k_{1,4}$	1e-3	1.387e-3	1e-3	4.646e-2
$k_{4,1}$	1e-1	1.144e-4	1e-1	1.257e-1
$k_{4,2}$	1e-1	1.053e-2	1e-1	1.876e-1

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 Figures 5 and 6). Anyway, it should be stressed that the *OE* model involves the estimation of 8 parameters, while *PZ* and *Wampum* models require of 6 and 2 parameters respectively. The estimated values of the parameters are reported in Table 10.

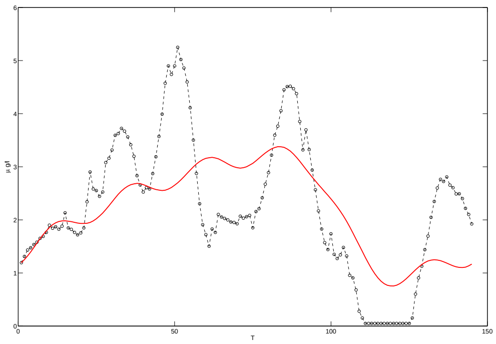

Figure 1. Fitting results of *PZ* model: simulation (solid line) and measurements (dashed line). Summer data.

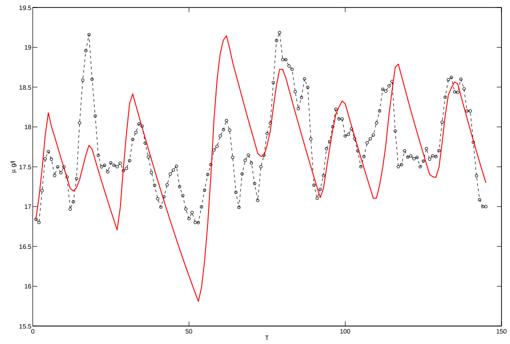

Figure 2. Fitting results of *PZ* model: simulation (solid line) and measurements (dashed line). Winter data.

Table 9
Estimated parameter values for the Wampum model.

Parameter	Initial Value (Winter)	Final Value (Winter)	Initial Value (Summer)	Final Value (Summer)
$k_{1,1}$	1.593e-2	1.614e-2	-2.365e-1	-2.338e-1
$k_{1,2}$	4.845e-3	4.531e-3	-4.706e-2	-4.622e-2

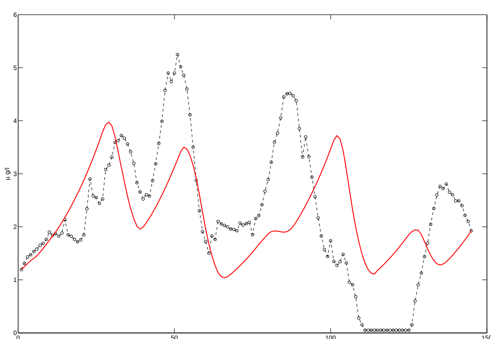

Figure 3. Fitting results of *Wampum* model: simulation (solid line) and measurements (dashed line). Summer data.

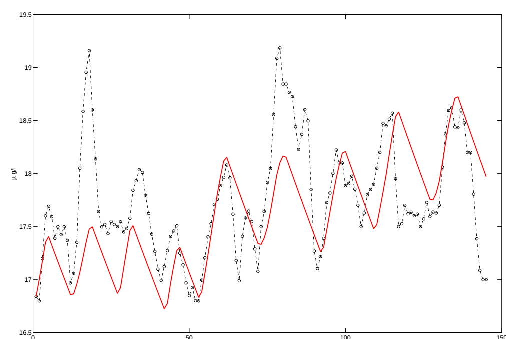

Figure 4. Fitting results of *Wampum* model: simulation (solid line) and measurements (dashed line). Winter data.

Table 10
Estimated parameter values for the OE-Model.

Parameter	Final Value Winter	Final Value Summer
$a_{1,1}$	- 1.952	-1.889
$a_{1,2}$	9.629e-1	8.919e-1
$a_{2,1}$	- 1.721	-1.909
$a_{2,2}$	7.278e-1	9.316e-1
$b_{1,1}$	4.946e-1	6.319e-1
$b_{1,2}$	- 1.157	-6.316e-1
$b_{1,3}$	6.691e-1	0
$b_{2,1}$	7.603e-5	-3.753e-4
$b_{2,2}$	9.773e-5	2.149e-4
$b_{2,3}$	- 1.98e-4	0

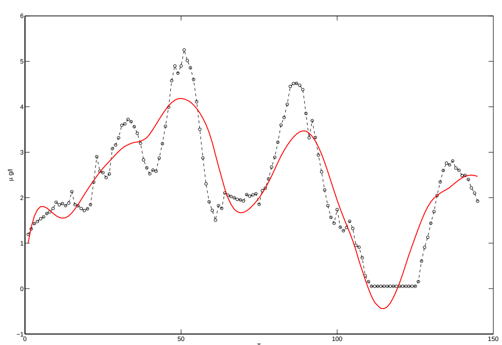

Figure 5. Fitting results of *OE* model: simulation (solid line) and measurements (dashed line). Summer data.

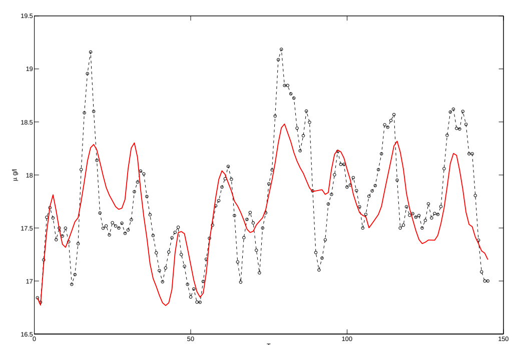

Figure 6. Fitting results of *OE* model: simulation (solid line) and measurements (dashed line). Winter data.

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 and Figures 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.

Table 11
Identification results for the water quality models.

Model	Fit	MSE	$\bar{e}_r$
Oxy (v_2)	90.1653	0.00142	0.00346
Nut (v_3)	37.3527	0.03063	0.031663

Table 12
Estimated parameter values for the Oxy and Nut models. Summer data.

Parameter	Initial Value	Final Value
$k_{2,1}$	1e-4	8.679e-5
$k_{2,2}$	1e-4	4.450e-2
$k_{2,3}$	1e-4	1e-9
$k_{2,4}$	1e-4	3.341e-7
$k_{3,1}$	1e-4	1.147e-3
$k_{3,2}$	1e-4	7.054e-4
$k_{3,3}$	1e-4	1.940e-3
f	1e-4	2.141e-4

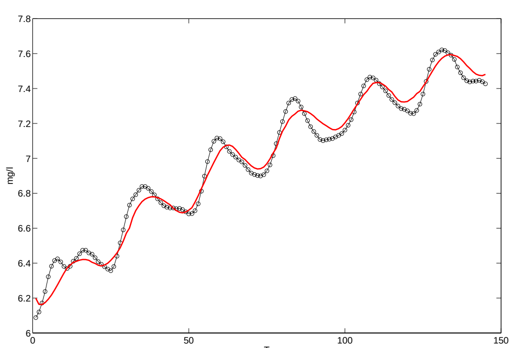

Figure 7. Fitting results of *Oxygen* model: simulation (solid line) and measurements (dashed line). Summer data.

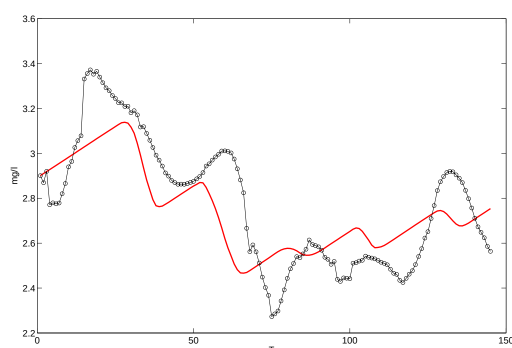

Figure 8. Fitting results of *Nutrient* model: simulation (solid line) and measurements (dashed line). Summer data.

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.

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 Figures 9 (a), (b) and (c). As it can be observed from Table 13 and Figure 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 13
Identification results for the coupled models.

Model	Fit	MSE	$\bar{e}_r$
PZ-Coupled-phyto	22.8704	0.97464	0.12606
PZ-Coupled-oxy	90.3727	0.00136	0.00435
PZ-Coupled-nut	41.443	0.02676	0.02458
WP-Coupled-phyto	8.2382	1.3795	0.13521
WP-Coupled-oxy	90.4364	0.00134	0.00322
WP-Coupled-nut	26.5657	0.04209	-0.00367
OE-Coupled-oxy	90.507	0.00132	0.00579
OE-Coupled-nut	36.9703	0.0310	0.023278

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 ±10%.

2. The logarithmic derivative of the simulated values given by the model with respect to the parameters according to the formula

$$MSE_{SENS} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{\partial v(t_i)}{\partial k} / v(t_i) \right)^2 , \tag{26}$$

   where k here denotes any of the parameters $k_{i,j}$ in the models.

In the case of the cost function, the corresponding results are reported in Figure 10. We remark that the *Wampum* cost variations are smoother than the *PZ* ones. This calls for nonsmooth optimization, which is a natural continuation of the research developed here. In any case, the *PZ* model provides much better performances on the summer data

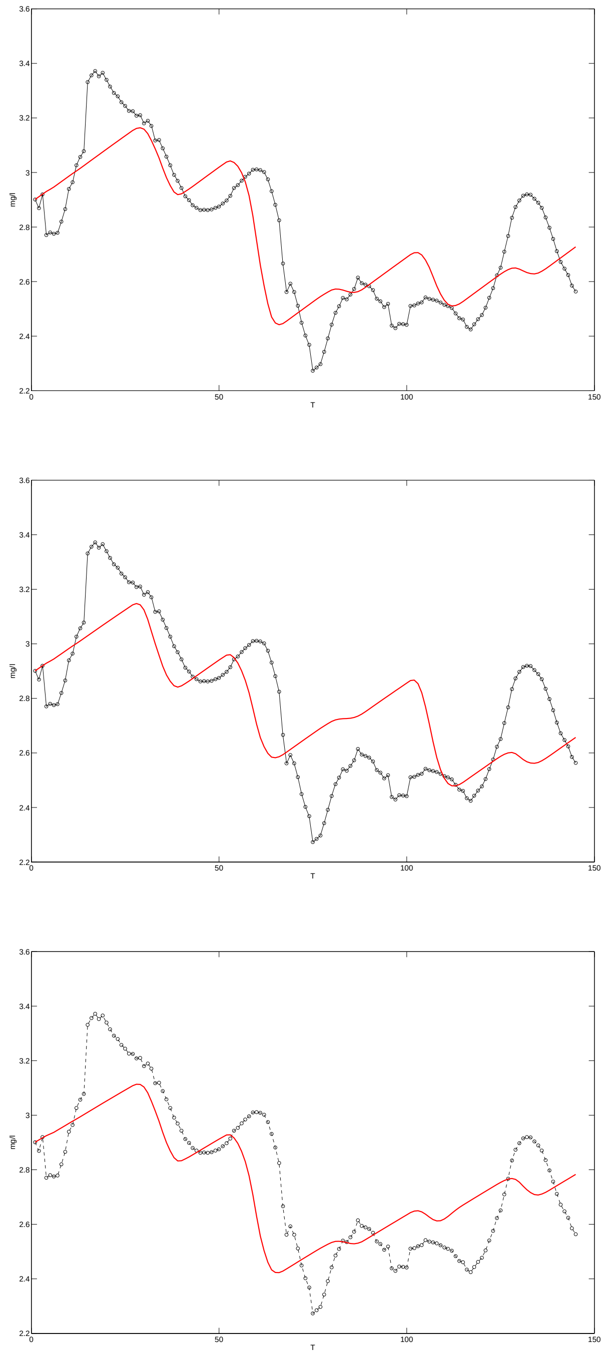

Figure 9. Nutrient dynamics in summer. Identification results of *PZ-coupled* (a), *Wampum-coupled* (b) and *OE-coupled* (c) models. Simulations are drawn with solid lines and measurements with dashed lines.

than in the winter. In fact, the minimum cost is consistently lower than that provided by the *Wampum* model during the same period.

In the case of the sensitivity of the state variable v in equations (5) and (8) with respect to the parameters: We computed $\partial v/\partial k_i$ for each model. Table 14 shows the squared relative averages of $\partial v/\partial k_i$ for the winter and summer data. 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 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. Mean square value of $\frac{\partial v}{\partial k_i}/v$ in the winter and summer.

Parameter	MSE_{SENS} Value (Winter)	MSE_{SENS} Value (Summer)
$k_{1,1}$	1.849e+8	1.831e+7
$k_{1,2}$	1.077e+6	3.534e+4
$k_{1,3}$	3.227	5.290e+9
$k_{1,4}$	1.845e+2	1.984e+4

Table 15
Sensitivity analysis of the *Wampum* model. Mean square value of $\frac{\partial v}{\partial k_i}/v$ in the winter and summer.

Parameter	MSE_{SENS} Value (Winter)	MSE_{SENS} Value (Summer)
$k_{1,1}$	6.692e+2	2.528e+2
$k_{1,2}$	6.623e+3	7.919e+3

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

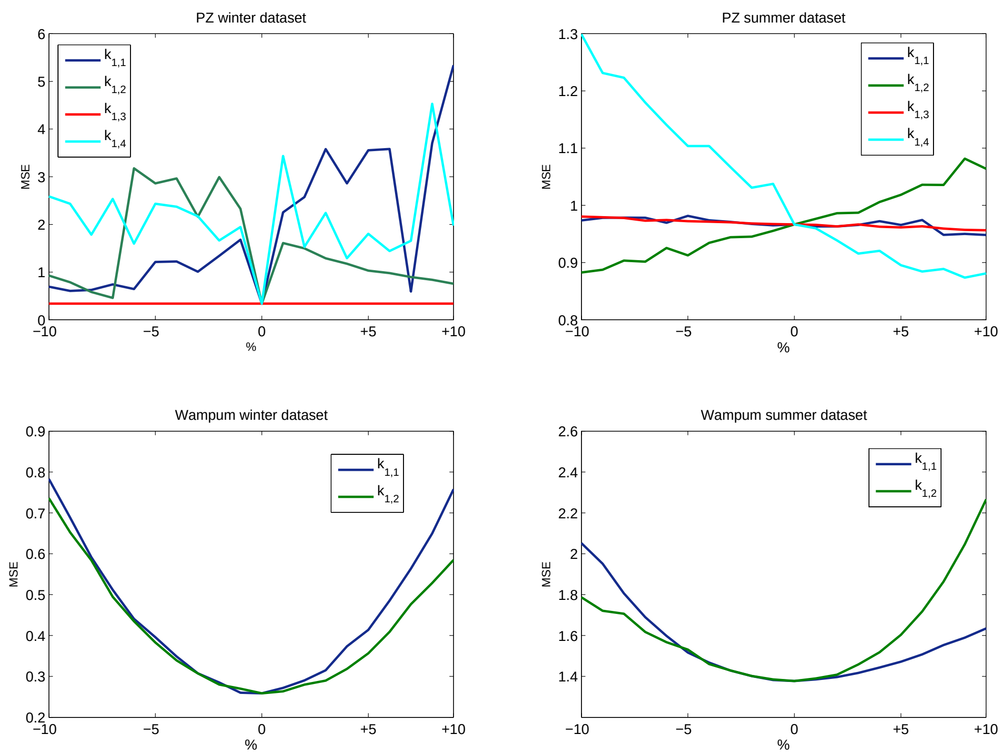

Figure 10. Sensitivity analysis of *PZ* and *Wampum* models. MSE values for ±10% parameter variations in winter and summer datasets.

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.

Using a nonlinear parametric optimization procedure the most important parameters of the previously developed models have been selected and estimated on the basis of the real data. A novel two-step parameter identification/validation procedure for identifying the microalgae dynamic models has been applied. Several simulation 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 in further research one may want to use some regularization terms on 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

1. G. M.Masters, *Introduction to environmental engineering and science*, Prentice-Hall.
2. L. Ljung, *System Identification. Theory for the User*, Prentice-Hall PTR, 1999.
3. M. Scheffer, *Ecology of shallow lakes*, Chapman Hall, 1998.
4. J. D. Murray, *Mathematical Biology*, Springer-Verlag, 1993.
5. W. S. C. Gurney, R. M. Nisbet, *Ecological Dynamics*, Oxford University Press, 1998.
6. R. J. Romanowicz, U. Callies, P. C. Young, Water quality modelling in rivers with limited observational data: river Elbe case study, in*Proceeding of the Int. Conf.: Complexity and Integrated Resource Management*, 14-17 June 2004, Germany.
7. C. E. Hann, J. G. Chase, J. Lin, T. Lotz, C. V. Doran, G. M. Shaw, Integral-based parameter identification for long-term dynamic verification of a glucose-insulin system model, in*Computer Methods and Programs in Biomedicine*, vol. 77 259-270, 2005.
8. C. E. Hann, J. G. Chase, J. Lin, T. Lotz, K. A. McAuley, J. Wong, S. Andreassen, Integral-based identification of a physiological insulin and glucose model on euglycaemic clamp and IVGTT trials, in*Proceeding of* 14^{th} *IFAC Symposium on System Identification*, Newcastle, Australia, 2006.
9. A. Garulli, C. Mocenni, A. Vicino, Integrating identification and qualitative analysis for the dynamic model of a lagoon, in*International Journal of Bifurcation and Chaos*, volume 13 (2)357-374, 2003.
10. A. Chapelle, A preliminary model of nutrient cycling in sediments of a Mediterranean lagoon, in *Ecological Modelling 80*, 131-147 , 1995.
11. A. Chapelle,A. Mè nesguen, J.M. Deslous-Paoli,P. Souchu, N. Mazouni, A. Vaquer,B. Millet, Modelling nitrogen, primary production and oxygen in a Mediterranean lagoon. Impact of oysters farming and inputs from the watershed, in *Ecological Modelling 127*, 161-181, 2000.
12. J. Norberg, D. DeAngelis, Temperature effects on stocks and stability of a phytoplankton-zooplankton model and the dependence on light and nutrients, in *Ecological Modelling 95*, 75-86, 1997.
13. J.M. Zaldivar, E. Cattaneo, M. Plus, C.N. Murray, G. Giordani, P. Viaroli, Long-term simulation of main bio-geochemical events in a coastal lagoon: Sacca Di Goro (Northern Adriatic Coast, Italy), in *Continental Shelf Research 23*, 1847-1875, 2003.
14. C. Mocenni and A. Vicino, Modeling the ecological competition between seaweed and seagrass in an aquatic environment: a case study, *Proceedings of the IFAC SYSID 2006*, Newcastle, March 29-31, 2006.
15. C. Mocenni, E. Sparacino, A. Vicino and J. Zubelli, Modelling and Parameter Identification of an Aquatic Ecosystem: the Serra da Mesa Example, *Workshop on Math-*

ematical Methods and Modeling of Biophysical Phenomena, Angra dos Reis, Rio de Janeiro, March 5-11, 2006.
16. D. Marinov, J.M. Zaldivar, A. Norro, G. Giordani, P. Viaroli, *Integrated modelling in coastal lagoons part B: lagoon model Sacca Di Goro case study (Italian Adriatic sea shoreline)*, European Communities, 2005.
17. C. S. Reynolds, J.P. Descy, J. Padisàk, Are phytoplankton dynamics in rivers so different from those in shallow lakes? *Hydrobiologia* 289,1-7. 1994.
18. E. L. Smith, Photosynthesis in Relation to Light and Carbon Dioxide, *Proc. Nat. Acad. Sci.* 22, 504-510. 1936.
19. P.H. C.Eilers, J.C.H. Peeters, A model for the relationship between light intensity and the rate of photosynthesis in phytoplankton, in *Ecological Modelling 42*, 199-215, 1988.
20. R.W. Eppley, J.B. Rogers, J.J. MC Carthy, A. Sournia, Light/dark periodicity in nitrogen assimilation of the marine phytoplankters Skeletonema costatum and Coccolithus huxley, in *N-limitant chemostat culture*, 150-154, 1971.
21. J.M. Zaldivar, F. Somma, F. Bouraoui, M. Austoni, G. Giordani, P. Viaroli, M. Plus, *Preliminary study of spatio-temporal data analysis tools for the management of coastal lagoons*, European Communities, 2004.
22. D. Marinov, J.M. Zaldìvar, L. Galbiati, F. Bouraoui, F. Sena, F.J. Elorza, A. Norro, G. Giordani, P. Viaroli, *Integrated modelling in coastal lagoons part C: coupling watershed-lagoon models Sacca Di Goro case study*, European Communities, 2005.
23. Huisman, J., and F. J. Weissing. Light-limited growth and competition for light in well-mixed aquatic environments: an elementary model. *Ecology* 75:507520, 1994.
24. E. H. S. Van Duin and L. Lijklema, Modelling photosynthesis and oxygen in a shallow, hypertrophic lake, *Ecological Modelling 45-4*, 243-260, 1989.
25. V. Hull, C. Mocenni, M. Falcucci and N. Marchettini, A trophodynamic model for the lagoon of Fogliano (Italy) with ecological modifying parameters, *Ecological Modelling 134*, 153-167, 2000.
26. CT Brasil Ministério da Ciência e Tecnologia, Rede Temática de Pesquisa em Modelagem Ambiental da Amazônia (GEOMA), *www.geoma.lncc.br*, 2002.
27. E. Sparacino, *Modellistica, identificazione ed analisi dei dati di monitoraggio ambientale per zone umide della regione Amazzonica*, Master Thesis in Information Engineering, Università di Siena, 2004-2005.