INTEGRATING IDENTIFICATION AND QUALITATIVE ANALYSIS FOR THE DYNAMIC MODEL OF A LAGOON

A. GARULLI*, C. MOCENNI† and A. VICINO‡

Dipartimento di Ingegneria dell’Informazione,
Centro per lo Studio dei Sistemi Complessi,
Università di Siena, Via Roma 56, 53100 Siena, Italy
*garulli@ing.unisi.it
†mocenni@ing.unisi.it
‡vicino@ing.unisi.it

A. TESI

Dipartimento di Sistemi e Informatica, Università di Firenze
Via di S. Marta 3, 50139 Firenze, Italy
atesi@dsi.unifi.it

Received July 11, 2001; Revised February 13, 2002

This paper deals with the identification and the qualitative analysis of a dynamic model of a shallow lagoon. The model describes the relations between biotic (phytoplankton, zooplankton) and abiotic (oxygen, nutrients) components of a lagoon. The first step of the paper is to derive estimates for the model parameters through an identification procedure using real data. The second step is to perform a qualitative analysis of the model dynamics, via the introduction of a parameterized reduced order model. The main contribution of the paper is to make an effort in the direction of synergizing the identification stage with the qualitative analysis of the model dynamics, in order to gain a better understanding of the system behavior and obtain more reliable estimates for the model parameters and exogenous inputs.

Keywords: Biological processes; lagoon; dynamic model; identification; qualitative analysis.

1. Introduction

The dynamic model presented in this paper describes biotic and abiotic processes in a shallow saltmarsh coastal lagoon. Several models have been proposed for phytoplankton-zooplankton-nutrients dynamics [Childers & McKellar, 1987; Franks et al., 1986; Kremer & Nixon, 1978] and dissolved oxygen concentration in marine and freshwater environment [Ginot & Hervé, 1994; Van Duin & Lijklema, 1989]. The model considered in the paper concerns two trophic levels and it stems from a more general lagoon model introduced in [Hull et al., 2000; Mocenni et al., 1999]. These levels refer to a producer (phytoplankton) and a herbivore consumer (zooplankton). Two other components enrich the picture of the ecosystem: dissolved inorganic nutrients in water column involved in photosynthesis and regenerated by bacterial populations, and dissolved oxygen content in water column. The model includes also the effect of exogenous environmental factors acting on the ecosystem processes, such as light [Childers & McKellar, 1987; Eilers & Peeters, 1988], temperature [Kremer & Nixon, 1978; Norberg & Angelis, 1997] and wind [Van Duin & Lijklema, 1989]. These driving inputs play a key
role in the model analysis and identification developed in the paper.

The aim of the paper is to derive estimates for the nonlinear model parameters through an identification procedure using real data. In this respect, a lot of work has been devoted to the identification of black-box nonlinear models for the analysis of asymptotic properties of real data, when physical models are not available [Aguirre & Billings, 1994; Coca & Billings, 1997]. This approach has been successfully applied for modeling biological processes (see, e.g. [Coca et al., 2000]). On the other hand, qualitative analysis of ecological systems is a powerful tool for the investigation and prediction of the behavior of systems subject to exogenous inputs (see e.g. [Kuznetsov, 1995; Kuznetsov et al., 1992; Strogatz, 1994]).

The main contribution of this paper is that of synergizing the identification stage with the qualitative analysis of the dynamics of the model, in order to gain a better understanding of the system behavior and obtain more reliable estimates for the model parameters and exogenous inputs. In particular, the complete model is reduced to a second order model, which depends on two biologically meaningful parameters. These parameters are used to perform the bifurcation analysis on the reduced model. The analysis of the system qualitative behavior in the two parameter plane allows one to improve the results of the parameter estimation procedure. This approach may prove useful in contexts where the a priori information on the system structure and the available data do not allow for sufficiently accurate model identification.

The paper is organized as follows. Section 2 introduces the model describing the main biochemical and biological processes as well as the exogenous inputs acting on the system. In Sec. 3 an appropriate tuning procedure based on least squares optimization for the estimation of the model parameters is performed, with reference to the specific case study of the Caprolace lagoon, located in the Parco Nazionale del Circeo (Italy). In Sec. 4 the essence of the nonlinear dynamics of the model is captured in the formulation of a reduced order model. A qualitative analysis of the dynamical behavior of the system is performed, via a suitable parameterization of the reduced order model. In Sec. 5 the information provided by the qualitative analysis is exploited to improve the performance of the identification procedure, through a proper selection of the starting point in the least squares estimation optimization procedure. Finally, some concluding remarks are drawn in Sec. 6.

2. Model Formulation

2.1. Driving environmental functions

The exogenous inputs to the lagoon system described in this subsection are light, wind and temperature. Such inputs usually show an annual periodicity, inducing seasonal phenomena in ecosystems [Kremer & Nixon, 1978]. The month is assumed as sampling time for our analysis.

Based on references [Childers & McKellar, 1987; Eilers & Peeters, 1988; Hull et al., 1991; Kremer & Nixon, 1978; Van Duin & Lijklema, 1989], we assume the following equations for describing the amount of light time per month $u_1(t)$, the temperature $u_2(t)$ and the average wind intensity $u_3(t)$:

$$u_1(t) = \mu_{11} + \mu_{12} \sin\left(\frac{2\pi}{12}(t + \mu_{13})\right),$$

$$u_2(t) = \mu_{21} + \mu_{22} \sin\left(\frac{2\pi}{12}(t + \mu_{23})\right),$$

$$u_3(t) = \mu_{31} + \mu_{32} \cos\left(\frac{2\pi}{12}(t + \mu_{33})\right),$$

where $\mu_{ij}$, $i, j = 1, 2, 3$ are parameters to be fixed on the basis of the available data (see Sec. 3.1).

An additional variable playing an important role in the model is the limiting factor of temperature on the growth of phytoplankton biomass. According to the analysis proposed in [Kremer & Nixon, 1978], this limiting factor $u_4$ can be expressed by the following equation:

$$u_4(t) = \gamma e^{\delta u_2(t)}.$$  

The values of the constants $\gamma$ and $\delta$ are reported in Table 2.

2.2. State variable equations

In this subsection, the dynamic state equations representing the mathematical model will be introduced. The model parameters, denoted by $k_i$, $i = 0, 1, \ldots, 16$, $k_P$, $k_T$, $k_X$, $k_{AE}$ and $k_{AN}$, are all
Table 1. Definition of state variables of the model.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Biological Meaning</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>phytoplankton biomass</td>
<td>mg m$^{-3}$</td>
</tr>
<tr>
<td>$x_2$</td>
<td>zooplankton biomass</td>
<td>mg m$^{-3}$</td>
</tr>
<tr>
<td>$x_3$</td>
<td>dissolved oxygen concentration</td>
<td>mg l$^{-1}$</td>
</tr>
<tr>
<td>$x_4$</td>
<td>nutrients</td>
<td>mg m$^{-3}$</td>
</tr>
</tbody>
</table>

non-negative. Table 1 reports the state variables needed to construct the mathematical model.

**Phytoplankton.** This population represents the producer (in forms of Diatoms, Peridenes and Microflagelates), i.e. the set of vegetal species performing carbon fixation, measured in terms of biomass. The dynamic equation describing the evolution of the phytoplankton $x_1$ is

$$\dot{x}_1 = k_1 u_1 u_4 x_1 \frac{x_4}{k_0 + x_4} - k_2 x_1^2 - k_3 x_2 \frac{x_1}{k_P + x_1}. \tag{5}$$

The first term in Eq. (5) accounts for the photosynthetic activity, which produces oxygen leading to an increase of phytoplankton biomass. The process is influenced by light intensity and limited by temperature and nutrient concentration. The second term of the equation represents the natural mortality that is assumed to be proportional to the square of phytoplankton biomass itself. The formulation of the first two terms is based on the logistic equation [Murray, 1993], which is the best known model of population dynamics for the vegetation microorganisms. The third term accounts for losses due to grazing of zooplankton.

Note that $k_P$ represents the half saturation value, since if $x_1 = k_P$, then $x_1/(k_P + x_1) = 0.5$, thus producing a reduction of 50% of the grazing and growth of zooplankton (see Eq. (6) below).

**Zooplankton.** The variable $x_2$ is the biomass of herbivore consumers (mainly copepods as some species of Acartia). The zooplankton evolution equation is

$$\dot{x}_2 = k_4 x_2 \frac{x_1}{(k_P + x_1)} - k_5 x_2. \tag{6}$$

The dynamics of this variable is regulated by a growth due to grazing on phytoplankton and by the losses for natural mortality. Here, it is assumed that there is no interspecific competition between zooplankters, as it can be checked in the mortality term.

**Dissolved oxygen.** The equation describing the dynamics of dissolved oxygen $x_3$ is

$$\dot{x}_3 = k_6 u_3 + k_7 u_1 u_4 x_1 \frac{x_4}{k_0 + x_4} - k_8 f_\alpha(x_3) - k_9 f_\beta(x_3) + k_10 \frac{C_S(u_2) - x_3}{k_T + u_2} - k_{11} x_1 x_3 - k_{12} x_2 x_3, \tag{7}$$

where

$$f_\alpha(x_3) = \frac{x_3^2}{k_{AE} + x_3^2}, \quad f_\beta(x_3) = \frac{x_3}{k_{AN} + x_3^2}. \tag{8}$$

The first two terms of Eq. (7) account for wind reaeration and photosynthesis oxygen production. The third and fourth terms represent the losses due to the bacterial activity, while the last two terms show the decrease of oxygen content due to the phytoplankton and zooplankton respiration. The fifth term accounts for the equilibrium physical-chemical reaction between gaseous oxygen and dissolved oxygen. The rate of this process is globally influenced by water temperature: the higher the temperature, the lower is the probability that the

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_{11}$</td>
<td>0.5</td>
<td>t (month)</td>
</tr>
<tr>
<td>$\mu_{12}$</td>
<td>0.125</td>
<td>t (month)</td>
</tr>
<tr>
<td>$\mu_{13}$</td>
<td>5.7</td>
<td>t (month)</td>
</tr>
<tr>
<td>$\mu_{14}$</td>
<td>17.5</td>
<td>°C</td>
</tr>
<tr>
<td>$\mu_{24}$</td>
<td>11</td>
<td>°C</td>
</tr>
<tr>
<td>$\mu_{33}$</td>
<td>3</td>
<td>t (month)</td>
</tr>
<tr>
<td>$\mu_{34}$</td>
<td>1.8</td>
<td>m s$^{-1}$</td>
</tr>
<tr>
<td>$\mu_{35}$</td>
<td>0.8</td>
<td>m s$^{-1}$</td>
</tr>
<tr>
<td>$\mu_{36}$</td>
<td>10</td>
<td>t (month)</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.59</td>
<td>[t (month)]$^{-1}$</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.0633</td>
<td>[°C]$^{-1}$</td>
</tr>
<tr>
<td>$c_1$</td>
<td>14.6</td>
<td>mg l$^{-1}$</td>
</tr>
<tr>
<td>$c_2$</td>
<td>-0.4</td>
<td>mg l$^{-1}$ [°C]$^{-1}$</td>
</tr>
<tr>
<td>$c_3$</td>
<td>0.008</td>
<td>mg l$^{-1}$ [°C]$^{-2}$</td>
</tr>
</tbody>
</table>
exchange occurs from air to water. The parameter $k_T$ represents the temperature limiting factor. Furthermore, since the rate $k_{10}$ is always positive, the direction of the chemical reaction $[C_S(u_2) - x_3]$ is related to the over/under saturated level of oxygen concentration $x_3$ in the water. For temperature ranges between 7°C and 28°C, the saturation value $C_S(u_2)$ changes according to the following equation [Van Duin & Lijklema, 1989] (see Table 2 for the values of the constants $c_i$, $i = 1, 2, 3$):

$$C_S(u_2) = c_1 + c_2u_2 + c_3u_2^2.$$  

(9)

Finally, the functions $f_\alpha$ and $f_\beta$, described in Eq. (8), represent the dependence of aerobic and anaerobic activities on dissolved oxygen concentration, respectively. Indeed, the degradation of organic matter performed by anaerobic bacterial pool (in water and sediments) is a very important process responsible for about 50% of the mineralization of organic matter [Izzo & Hull, 1991]. This process always occurs under environmental conditions that cause the amount of dissolved oxygen to be insufficient for complete mineralization. If the matter undergoes anaerobic breakdown, the result is fermentation leading to sulphate reduction [Jørgensen, 1977]. Bacterial sulphate reduction indirectly consumes the oxygen available in the water column, since the hydrogen sulﬁde produced as a reaction to anaerobic respiration is reoxidized chemically using a double amount of oxygen compared to the requirements of the aerobic process, thus exacerbating the oxygen shortage.

Note that $k_{AE}$ and $k_{AN}$ define the shape of $f_\alpha$ and $f_\beta$, respectively, and they will be fixed on the basis of availability of oxygen in the considered lagoon.

**Nutrients.** The nutrient variable $x_4$ refers to the nitrogen and phosphorus compounds in water and sediments. There is evidence from field experiments [Hull et al., 1991] that in coastal shallow water lagoons (and in particular in the Caprolace lagoon) the main source of nutrients 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. We model nutrients dynamics according to the following equation:

$$\dot{x}_4 = k_{13}f_\alpha(x_3) + k_{14}f_\beta(x_3) - k_{15}u_1u_4x_1 \frac{x_4}{k_0 + x_4} - k_{16} \frac{x_4}{k_X + x_4},$$  

(10)

where $f_\alpha(x_3)$ and $f_\beta(x_3)$ are given in (8) and $k_X$ represents the nutrient fixation half saturation constant.

The first two terms of Eq. (10) account for the aerobic and anaerobic production of nutrients by mineralization of organic matter, respectively. The consumption terms are due to the photosynthetic activity of phytoplankton species and the fixation in the sediment.

### 3. Experimental Results

#### 3.1. Estimation techniques and experimental data

The parametric identification has been performed for the model introduced in the previous section. The identification technique is based on the minimization of the cost function $F(\theta)$, representing the mean square error between simulated and experimental data [Ljung, 1999; Garulli et al., 1999]

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

(11)

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 $W_i(\bar{\phi}(t_i))$ is a suitable weighting matrix (see the next subsection for more details).

The set of real data used for parameter identification refers to experiments performed by the Laboratorio Centrale di Idrobiologia of Rome [Hull et al., 1991] and ENEA [1995]. These experimental studies concern the research about two saltmarsh coastal ecosystems situated in the Parco Nazionale del Circeo (Italy): the Fogliano and Caprolace lagoons.

The data collected consists of

1. Average monthly values of the biomasses of the most important vegetal and animal species and of the dissolved oxygen, over a period of three years. In particular, as usually done, the data related to the phytoplankton biomass are assumed proportional to the Chlorophyll “a” contained in the water.
2. Average daily measurements of environmental factors like photoperiod, water temperature, and wind stress. These data have been used to estimate the parameters of the driving functions $u_1$, $u_2$, $u_3$ in (1)–(3) as reported in Table 2.

### 3.2. Numerical results

Parametric identification of the model has been performed using three sets of data: phytoplankton ($\overline{X}_1$) and zooplankton ($\overline{X}_2$) biomasses and oxygen concentration ($\overline{X}_3$). Following the notation adopted for the cost function (11), we have

$$\theta = (k_0, \ldots, k_{12})', \quad \phi(t_i) = (\overline{X}_1(t_i), \overline{X}_2(t_i), \overline{X}_3(t_i))',$$

$$W_i = \text{diag}\{\overline{X}_j^{-2}(t_i)\}, \quad j = 1, 2, 3.$$  

Note that the parameter vector $\theta$ only contains a subset of the model parameters. The parameters $k_{13}, \ldots, k_{16}, k_P, k_T, k_X, k_{AE}, k_{AN}$ are assumed to be known constants. This choice is suggested by the lack of data on nutrients and the knowledge of typical admissible values for some of these parameters from the literature. The assumed values, reported in Table 3, have been adjusted on the basis of field observations about the considered ecosystem. For example, the values of $k_{AE}$ and $k_{AN}$ in (8) are based on the observation that the activation/deactivation of aerobic and anaerobic bacterial activity falls in the range $[3 \div 4]$ mg l$^{-1}$ oxygen.

Minimization of the cost function (11) and (12), performed by using the NAG package [NAG, 1998], leads to the estimate $\hat{\theta} = (\hat{k}_0, \ldots, \hat{k}_{12})'$ of the parameter vector reported in Table 4. The units of some parameters in Tables 3 and 4 follow from the fact that these parameters represent the aggregate effect of more than one physical quantity of interest, such as rates of variation, concentrations, etc. Figures 1(a)–1(c) show the model fitting of the real data. It is clear that the identification procedure does not provide a satisfactory fitting of the data; this may be due to the fact that the estimated parameter $\hat{\theta}$ represents a local minimum of the cost function $F(\theta)$.

To look for the global minimum of $F(\theta)$, one typically performs multiple runs of the optimization procedure for different starting points. However, since no guidelines are available on how to select the initial points in the parameter space spanned by $k_0, \ldots, k_{12}$, this approach can become prohibitive from a computational point of view.

Motivated by this fact, in Sec. 5 we will provide a different approach to deal with this problem, which is based on integrating information of model dynamics in the identification procedure.

### 4. Reduced Order Model and Qualitative Analysis

The investigation of the dynamics of the periodically forced nonlinear model (5)–(10) is of fundamental importance to understand the ecological consequences on the lagoon system behavior. Indeed, to achieve a complete understanding of the long term behavior of the model, a qualitative analysis is necessary to determine the existing dynamical solutions (equilibria, periodic solutions, \ldots) and their stability properties.

<table>
<thead>
<tr>
<th>Par.</th>
<th>Biological Meaning</th>
<th>Value</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{13}$</td>
<td>nutrient production by aerobic activity</td>
<td>5</td>
<td>mg m$^{-3}$[t]$^{-1}$</td>
</tr>
<tr>
<td>$k_{14}$</td>
<td>nutrient production by anaerobic activity</td>
<td>$10^{-6}$</td>
<td>mg m$^{-3}$[mg l$^{-1}$][t]$^{-1}$</td>
</tr>
<tr>
<td>$k_{15}$</td>
<td>nutrient consumption by photosynthesis</td>
<td>$10^{-6}$</td>
<td>[t]$^{-1}$</td>
</tr>
<tr>
<td>$k_{16}$</td>
<td>sediment retained nutrients</td>
<td>3.8</td>
<td>mg m$^{-3}$[t]$^{-1}$</td>
</tr>
<tr>
<td>$k_P$</td>
<td>half saturation constant for phytoplankton</td>
<td>100</td>
<td>mg m$^{-3}$</td>
</tr>
<tr>
<td>$k_T$</td>
<td>limiting factor in temperature</td>
<td>1</td>
<td>°C</td>
</tr>
<tr>
<td>$k_X$</td>
<td>half saturation constant of nutrients</td>
<td>1</td>
<td>mg m$^{-3}$</td>
</tr>
<tr>
<td>$k_{AE}$</td>
<td>limiting factor in aerobic process</td>
<td>25</td>
<td>[mg l$^{-1}$]$^2$</td>
</tr>
<tr>
<td>$k_{AN}$</td>
<td>limiting factor in anaerobic process</td>
<td>0.4</td>
<td>[mg l$^{-1}$]$^2$</td>
</tr>
</tbody>
</table>
This section is devoted to this qualitative analysis, whose aim is to gain knowledge on the behavior of systems whose parameters are different from the estimated ones. Such an investigation of “nearby” systems is motivated by two main reasons. The first one is related to the uncertainty affecting the numerical parameters in the expression of the exogenous inputs, which have been estimated from the available data (see Sec. 3.1). The second one, most important in the present context, concerns the case when there is evidence that the identified model corresponds to an unsatisfactory local minimum of the cost function (11). In this case, qualitative analysis of nearby systems may help in selecting the most appropriate values of the parameters. This fact is particularly useful in view of the large number of parameters of the model to be estimated and the limited amount of available real data, and it will be exploited in a systematic way in Sec. 5.

The exact analysis of the dynamics of the estimated model is a prohibitive task, since such a model is a fourth order nonlinear system with a large number of parameters. Therefore, a different approach based on two successive steps will be followed. In the first step, a reduced order model of the dynamics is built catching the most significant biological aspects of the system. In particular, a second order model involving only two free parameters is looked for, because in this case the nonlinear dynamics can be analyzed quite efficiently from a computational point of view. In the second step, a qualitative analysis of the obtained reduced model with respect to the two parameters is performed. These two steps are described in the forthcoming Secs. 4.1 and 4.2.

4.1. Reduced order model

The physical motivation for the considered reduced order model lies in the fact that, while the oxygen dynamics depends directly on phytoplankton and zooplankton biomasses, the feedback of the variable $x_3$ on $x_1$ and $x_2$ occurs through the nutrients concentration variable $x_4$ (see (5)–(10)). More specifically, the global effect of the exogenous variables ($u_1$, $u_4$) and nutrients ($x_4$) on phytoplankton growth in Eq. (5) is represented by the term

$$u_1(t)u_4(t)\frac{x_4(t)}{k_0 + x_4(t)},$$

which, from a biological point of view, represents the growth rate parameter of the phytoplankton cells due to the effect of light, temperature and nutrients. This term can be interpreted as an auxiliary exogenous input. Moreover, since the model (5)–(10) is forced periodically, the nutrient concentration typically exhibits a periodic behavior (see [Farkas, 1994] for a complete discussion of the dynamical properties of periodically forced nonlinear systems).
Integrating Identification and Qualitative Analysis for the Dynamic Model of a Lagoon

systems), and therefore the auxiliary exogenous input (13) can be assumed as a periodic signal. As a consequence, the reduced order model becomes:

\[
\dot{x}_1 = \lambda_1 s(t)x_1 - \lambda_2 x_1^2 - \lambda_3 x_2 \left( \frac{x_1}{k_P + x_1} \right),
\]

(14)

\[
\dot{x}_2 = \lambda_4 x_2 \left( \frac{x_1}{k_P + x_1} \right) - \lambda_5 x_2.
\]

(15)

where the auxiliary exogenous input is:

\[
s(t) = M + N \sin \left( \frac{2\pi}{12} t + 1 \right).
\]

(16)

The constants \(M\) and \(N\) in (16) are additional parameters of the reduced order model. Since \(u_1, u_4,\) and \(x_4\) take non-negative values, we impose the constraint \(N \leq M\).

Note that, if the nutrient concentration \(x_4(t)\) of the complete model has exactly an annual periodicity and the terms of the Fourier series of the periodic function (13) are all negligible, except for the first two ones, then the reduced order model closely reproduces the dynamical behavior of the complete model, by selecting \(i_t = k_i, i = 1, \ldots, 5\) and \(M, N\) as the amplitude of the first two terms of the Fourier series of (13).
4.2. Qualitative analysis

Let us consider Eqs. (13) and (16), representing the cumulative effect of light ($u_1$), temperature ($u_4$) and nutrient concentration ($x_4$) on the reduced model (14) and (15). The introduction of the additional variables $M$ and $N$ allows for taking into account two sources of model uncertainty: the simplified expression used for describing the exogenous inputs (1)–(4); the a priori knowledge exploited for fixing the coefficients of the nutrient Eq. (10).

Several identification trials of the reduced order model for different values of $M$ and $N$ have been performed, thus obtaining the estimates of the corresponding parameters $\lambda_i$, $i = 1, \ldots, 5$ in Eqs. (14) and (15). Figure 2 reports the residual cost function computed in the identification trials of the reduced order model performed for $0 \leq M \leq 2$, $0 \leq N \leq 2$, $N \leq M$. Observe that the cost function looks rather flat near the diagonal $M = N$.

In the sequel a qualitative analysis is performed to investigate the possible dynamical behaviors predicted by the model, as well as their stability properties as a function of the two parameters $M$ and $N$. In this respect, instead of analyzing the dynamical behavior of each identified reduced model, it turns out more appropriate, from the numerical accuracy viewpoint, to provide a dynamic model whose coefficients $\lambda_i$ are continuous functions of $M$ and $N$. Indeed, such parametric dynamic model can be obtained via a least squares interpolation of the parameter estimates $\hat{\lambda}_i$, $i = 1, \ldots, 5$ through a second order polynomial in $M$ and $N$, denoted by $\hat{\lambda}_i = \hat{\lambda}_i(M,N)$. The resulting $(M,N)$-parameterized dynamic model is:

$$\dot{x}_1 = \hat{\lambda}_1(M,N)x_1 - \hat{\lambda}_2(M,N)x_1^2 - \hat{\lambda}_3(M,N)x_2 \left( \frac{x_1}{k_p + x_1} \right),$$

$$\dot{x}_2 = \hat{\lambda}_4(M,N)x_2 \left( \frac{x_1}{k_p + x_1} \right) - \hat{\lambda}_5(M,N)x_2.$$  

(17)  

(18)

It is easy to verify that, for $N = 0$, the model (17) and (18) has three equilibrium points. Two of them are trivial: the origin $(0,0)$, which is a saddle point, and an additional saddle point $E_1 = (\lambda_1(M,0)x_1/\hat{\lambda}_2(M,0),0)$. The third equilibrium is the point of interest $E_2 = (x_1^*, x_2^*)$ whose explicit expression can be easily obtained from (17) and (18).
The following considerations can be made on the properties of the equilibrium point $E_2$.

$N = 0$. The equilibrium point $E_2$ is stable for $M > 0.5$; for $M = 0.5$ one of the two eigenvalues vanishes; for $M < 0.5$ the model shows a saddle point.

$N > 0$. For small values of $N$, the periodically forced model retains the stability properties of the unforced system. In particular, the stable equilibria give rise to stable periodic solutions of period $T = 12$ months (see [Farkas, 1994] for a theoretical explanation). As $N$ increases, such periodic solution may undergo some bifurcation, thus generating different solutions [Kuznetsov, 1995; Strogatz, 1994]. Clearly, the detection of bifurcations is important for understanding the system behavior especially for larger values of $N$ (see also [Kremer & Nixon, 1978]). In the following, $(F)$ denotes a flip bifurcation, $(N - S)$ a Neimark–Sacker bifurcation and $(T)$ a tangent or fold bifurcation.

The bifurcations can be studied via the Poincaré map [Kuznetsov, 1995; Strogatz, 1994] which can be computed using suitable software packages (in our case, LOCBIF (Interactive Local Bifurcation Analyzer [Khibnik et al., 1992]) and WINPP (The Dynamical Systems Tool [Ermentrout, 1999]) have been used). These packages perform both simulations of dynamical systems and numerical bifurcation analysis.

Figure 3 shows the complete picture in the $(M, N)$-plane of the bifurcations of the fixed points of the Poincaré map associated to the equilibrium studied. Recall that the half plane $M < N$ (below the diagonal) does not correspond to physically meaningful system behaviors. For any $N > 0$, the system shows at least one T-periodic solution. Its long term behavior is clear in Fig. 3, where stable T-periodic solutions are present in the gray regions only. Several bifurcation curves have been obtained. Specifically, $F_1$ and $F_1^{-1}$ represent the curves of the period T flip bifurcations, $T_1$ the fold bifurcations, and $N-S_1$ the curve of the Neimark–Sacker bifurcations. Furthermore, $F_2$ and $F_2^{-1}$ represent the curves of the period 2T flip bifurcations, $T_2$ and $T_2^{-1}$ the fold bifurcations, and $N-S_2$ the curve of the Neimark–Sacker bifurcations.

![Bifurcation diagram in $(M, N)$ plane. Gray: region of stable T-periodic solutions.](image-url)
Fig. 4. Bifurcation diagram in \((M,N)\) plane. Gray: region of stable 2T-periodic solutions.

Fig. 5. The two coexisting stable periodic solutions corresponding to the point C1 of Fig. 4 \((M = 1.57 \text{ and } N = 0.61)\): (a) 2T-periodic solution generated via a fold bifurcation (curve T2); (b) 2T-periodic solution born via a flip bifurcation (curve F1). The unstable T-periodic solution generated via this flip bifurcation (dotted line) and the related 2T-periodic stable one (solid line) are drawn in (c).
By the analysis of such bifurcation curves several conclusions on the nonlinear dynamics can be drawn. For instance, the flip bifurcation curves $F_1$ and $F_1^{-1}$ give rise to periodic solutions of period $2T$. Such solutions are stable for all $M$ and $N$ belonging to the gray regions in Fig. 4. Also, we observe that there are regions in the plane $(M,N)$ where there exist more than one stable solution. For example, the two stable trajectories of period $2T$ corresponding to point $C_1$ of Fig. 4 ($M = 1.57, N = 0.61$) are shown in Fig. 5. Instead, one stable solution of period $T$ coexists with one stable solution of period $2T$ (see Fig. 6) for the point $C_2$ of Fig. 3 ($M = 1.38, N = 0.38$).

Moreover, trajectories different from $T$- or $2T$-periodic solutions exist. Indeed, the Neimark-Sacker bifurcations (see N-S curves in Figs. 3 and 4) give rise to torus stable attractors. One of these attractors is depicted in Fig. 7.

It is also expected that chaotic attractors are very likely to exist in the considered system. For example, two different cascades of flip bifurcations...
Fig. 7. Toroidal solution born via a Neimark–Sacker bifurcation of a stable T-periodic solution. The simulation is obtained for a point close to the curve N-S1 in Figs. 3 and 4 ($M = 0.1, N = 0.3$).

Fig. 8. Some attractors obtained for $M = 1.07$ and different values of $N$ in the region A shown in Figs. 3 and 4. The 2T, 4T, and 8T-periodic stable solutions generated via successive flip bifurcations for $N = 0.41$, $N = 0.51$, and $N = 0.54$ are shown in (a)–(c), respectively. Such sequence of flip bifurcations leads to the chaotic attractor depicted in (d) for $N = 0.88$. 
are observed in the regions A and B. The first one is shown in Fig. 8 where three stable solutions of period 2T, 4T and 8T are depicted in (a)–(c), respectively. They represent the first terms of a cascade of period doubling bifurcations leading to the chaotic attractor drawn in (d). The second cascade of flip bifurcations arises by decreasing $N$, starting from the curve $F1_1$, through the curve $F2_1$, and so on. Figure 9 reports a 2T-periodic stable solution (a) and a chaotic attractor (b).

Summing up, it is clear that the parameterized system (17) and (18) offers a wide range of different behaviors as a function of $M$ and $N$. This analysis may be particularly useful to select most appropriate parameter estimates in domains of the parameter space where the cost function is flat.

5. An Iterative Procedure for the Selection of Model Parameters
In this section, we show how the information provided by qualitative analysis of the reduced order model can be exploited for the identification of the complete model. The key idea is to select
the parameters \(\lambda_1, \ldots, \lambda_5\) in (14) and (15) taking into account the analysis performed in Sec. 4, and then use the obtained values when initializing the minimization of the identification cost function \(F(\theta)\) in (11) and (12) for the complete model.

Since \(F(\theta)\) generally exhibits several local minima, the choice of the optimization starting point is crucial. The insight given by qualitative analysis is used to choose the “region of interest” in the parameter space, where the dynamics of the system is most likely to lie. Clearly, qualitative analysis of the reduced order model (14) and (15) provides information only on the subspace of the complete model parameter space spanned by \(k_1, \ldots, k_5\). The initial values of the remaining parameters \((k_0, \ldots, k_{12}\) and \(k_0)\) can be set to the values obtained from a previous identification of the complete model, such as the one described in Sec. 3.

The above strategy can be iterated by alternatively performing identification of the complete model and selection of \(M, N\) together with the corresponding values of \(\lambda_i(M, N), i = 1, \ldots, 5\), in the reduced order one \((17)\) and \((18)\). The iterative procedure can be described as follows.

**Step 0.** Let \(\theta(0)\) be given by a preliminary identification of the complete model. Set \(h = 0\) and use qualitative analysis to select \(M(h)\) and \(N(h)\) in Eq. (16), according to the values of the residual cost function with respect to \(M\) and \(N\) (see Figs. 3 and 4) and the bifurcation \((M, N)\)-plane (see Figs. 3 and 4).

**Step 1.** For given \(M(h)\) and \(N(h)\), select the values of \(\lambda_i(h) = \hat{\lambda}_i(M(h), N(h)), i = 1, \ldots, 5\), for the parameterized dynamic model \((17)\) and \((18)\).

**Step 2.** Choose \(\theta_{in} = [k_0(h), \lambda_1(h), \ldots, \lambda_5(h), \hat{k}_0(h), \ldots, \hat{k}_{12}(h)]\) as the initial parameter vector in the numerical minimization of the cost function \(F(\theta)\), and denote by \(\theta(h+1)\) the parameter vector provided by the minimization procedure.

**Step 3.** Perform Fourier analysis of the term

\[
u_1(t)u_4(t) \frac{x_4(h+1)(t)}{k_0(h+1) + x_4(h+1)(t)} \quad (19)
\]

where \(x_4(h+1)(t)\) is obtained from simulation of the complete model with parameters \(\theta(h+1)\), and compute the new values of \(M(h+1), N(h+1)\) for the exogenous input (16) of the reduced order model.

**Step 4.** Set \(h = h + 1\) and go back to Step 1.

The iterative procedure can be stopped when \(M\) and \(N\) do not change significantly during one iteration. Several simulation studies have shown that convergence is usually achieved within few iterations.

The above procedure has been applied to the case study considered in the previous sections. As a starting parameter vector \(\theta(0)\), the result of identification performed in Sec. 3.2 has been used. Recall that this led to quite unsatisfactory fitting results, corresponding to a cost value \(F(\theta) = 59.4\). Notice that Fourier analysis on the term in (19) gives values \(M = 0.16\) and \(N = 0.12\). However, exploiting qualitative analysis, we select \(M(0) = 0.5\), \(N(0) = 0.5\) and correspondingly the parameter values for the reduced order model reported in Table 5, row \(I - 1\).

This choice is motivated by the value of the residual cost function of the reduced order model, which achieves lower values near the diagonal \(M = N\) (see Fig. 2), and by the fact that a richer dynamic behavior can be found in the region A of the \((M, N)\)-plane (see Figs. 3 and 4).

Then, Step 2 gives the new parameter vector \(\theta(1)\) (Table 5, row \(I - 2\)), which corresponds to the cost \(F(\theta(1)) = 21.9\). The values \(M(1) = 1.067\), \(N(1) = 0.87\) are then obtained in Step 3.

As shown in Table 5, the next iteration (rows \(II\)) does not change significantly the values of \(M\) and \(N\); hence, \(\theta(2)\) can be kept as final identified parameter vector for the complete model. Figure 10 shows the behavior of the four state variables, for this model. It can be seen that fitting is remarkably improved, with respect to Fig. 1.

It is worth noting that the values of \(M\) and \(N\) provided by the iterative procedure correspond to a point lying in region A of Figs. 3 and 4. Qualitative analysis associates this point to a chaotic behavior of the reduced order model, similar to that drawn in Fig. 8(d). This is not surprising, as we are dealing with identification of a model showing a wide variety of dynamic behaviors, based on a very small set of data. Loosely speaking, we may say that it is not possible to predict the long term behavior of the system, on the basis of the available data.

Nevertheless, the contribution provided by qualitative analysis is still useful. In fact, different initial values of \(M\) and \(N\) have been tried to initialize the procedure in Step 0. If the values corresponding to the result of “blind” identification performed in Sec. 3 are chosen \((M = 0.16,\)
Table 5. Iterative procedure.

<table>
<thead>
<tr>
<th>Step</th>
<th>$k_0$</th>
<th>$k_1$</th>
<th>$k_2$</th>
<th>$k_3$</th>
<th>$k_4$</th>
<th>$k_5$</th>
<th>$k_6$</th>
<th>$k_7$</th>
<th>$k_8$</th>
<th>$k_9$</th>
<th>$k_{10}$</th>
<th>$k_{11}$</th>
<th>$k_{12}$</th>
<th>Cost $\left( M, N \right)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>I − 0</td>
<td>40.8</td>
<td>1.29</td>
<td>0.0002</td>
<td>0.93</td>
<td>6.9</td>
<td>5.8</td>
<td>0</td>
<td>0</td>
<td>0.007</td>
<td>1.21</td>
<td>0</td>
<td>0.0001</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I − 1</td>
<td>0.3</td>
<td>0.0004</td>
<td>1.75</td>
<td>12.65</td>
<td>10.73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.5, 0.5)</td>
</tr>
<tr>
<td>I − 2</td>
<td>0.0088</td>
<td>0.16</td>
<td>0.00007</td>
<td>1.8</td>
<td>12.6</td>
<td>10.76</td>
<td>0</td>
<td>0</td>
<td>0.006</td>
<td>0</td>
<td>0</td>
<td>0.038</td>
<td>2 $\cdot 10^{-6}$</td>
<td>0.001</td>
</tr>
<tr>
<td>I − 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(1.067, 0.87)</td>
</tr>
<tr>
<td>II − 1</td>
<td>0.19</td>
<td>0.0001</td>
<td>1.29</td>
<td>12.5</td>
<td>10.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II − 2</td>
<td>0</td>
<td>0.16</td>
<td>0.0001</td>
<td>1.8</td>
<td>13.1</td>
<td>11.2</td>
<td>0.35</td>
<td>0.007</td>
<td>0.66</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0009</td>
<td>21.8</td>
</tr>
<tr>
<td>II − 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(1.067, 0.87)</td>
</tr>
</tbody>
</table>
Fig. 10. Identification results obtained by the application of the iterative method based on qualitative analysis (see Sec. 5): model prediction (solid line) and data (○).

For the complete model. The resulting identified complete model is the one showing the best trade-off between low value of the identification cost and good agreement with the reduced order dynamics, which may be seen as a validation of the obtained complete model based on qualitative analysis of the reduced order one.

Clearly, another possibility is to constrain the search in the iterative procedure to a certain subset of the qualitative plane, thus obtaining the best complete model (in terms of the identification cost value) among those that exhibit a desired qualitative behavior.
6. Conclusions

In this paper a procedure for identifying the dynamic model of a lagoon by using real data has been proposed. The new approach relies on the systematic use of the information on the model dynamics in the identification procedure. More specifically, in order to understand the dynamical behavior of the complete model on the basis of real data, a suitable reduced order model is introduced and its qualitative analysis is performed as a function of two variables. These variables are instrumental for the iterative procedure designed for parameter estimation of the complete model. Indeed, the qualitative analysis of the reduced order model provides useful information for picking the starting parameter vector in the estimation nonconvex optimization program.

From a biological point of view, it has been observed that a model describing the dynamics of a lagoon exhibits a richness of possible behaviors, depending on the model parameters, in agreement with theoretical and experimental biological knowledge. In this respect, the application to the Caprolace lagoon provides satisfactory results.

The proposed identification procedure may also be useful when more complicated models are considered, which is often the case when modeling environmental systems. Further testing of the whole identification scheme in more general contexts is under investigation and it will be the object of future work.

References


