

# Benchmark MODECOGeL

## A global sensitivity analysis approach for marine biogeochemical modeling

C. Prieur<sup>a</sup>, L. Viry<sup>a</sup>, E. Blayo<sup>a</sup>, J.-M Brankart<sup>b</sup>

<sup>a</sup>Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP\*, LJK, 38000 Grenoble, France

<sup>b</sup>Univ. Grenoble Alpes, CNRS, Grenoble INP\*, IGE, 38000 Grenoble, France

\* Institute of Engineering Univ. Grenoble Alpes

---

### Abstract

Marine biogeochemical models are now commonly included as modules in complex ocean circulation modeling systems. They are thus increasingly used for many applications. For such applications, *sensitivity analysis* (SA), i.e. methods that aim to quantify the relative influence of the inputs on some given outputs in a complex system like a numerical model, are a valuable tool. Indeed, they can help better understand the model itself, and identify which parameters are most influential and should be calibrated carefully.

---

**These methods may be divided in two main categories.** *local* sensitivity methods that consider the behavior of the solution with respect to small parameter variations and *global* sensitivity methods that determine the behavior of the solution under parameter perturbations of arbitrary magnitude.

**Benchmark objectives.** In [7], the authors propose a global sensitivity analysis (GSA) approach for marine biogeochemical modeling. One aim of their paper is to demonstrate that, although demanding in terms of computation, a variance-based GSA in the full-dimension is becoming affordable, thanks to the development of distributed computing environments. They apply their approach to the MODECOGeL model, for which the input space is of dimension  $d = 74$ , and consider different quantities of interest (QoI). The MODECOGeL model is briefly described in Section 1.1. The  $d$  uncertain input parameters of the model are described in Section 1.2 and the QoIs are listed in Section 1.3. This study required a very large number of model evaluations that the authors decided to provide to the MASCOT NUM community, and more generally to the community of uncertainty quantification. The inputs/outputs were obtained from specific schemes, corresponding to different strategies for estimating sensitivity indices. In order to provide a self-contained document, a very brief description of the GSA strategies used in [7] is given in Section 1.4. Then, in Section 2, the different datasets that we provide are described.

## 1. Description of the ocean biogeochemical model

### 1.1. The MODECOGeL model

The model used is MODECOGeL<sup>1</sup>. It was developed in [1] for investigating the biogeochemical activity in the Ligurian sea by coupling a 1D hydrodynamic model of the mixed layer to a 12-component ecosystem model.

The hydrodynamic model is a 1D version of the GHER primitive equations model [6]. The state variables are the horizontal velocity, the potential temperature, the salinity, and the turbulent kinetic energy. A full description of the model can be found in [3] or [2], where it is applied to simulate the behavior of

---

<sup>1</sup>MODECOGeL: MODèle d'ECOsystème du GHER (GeoHydrodynamics and Environment Research) et du LOV (Laboratoire d'Océanographie de Villefranche-sur-Mer)

the system during the **FRONTAL** oceanographic campaigns from 1984 to 1988. In the present study, the model is applied to years 2006–2007, and the atmospheric dataset is extracted mainly from the Côte d’Azur meteorological buoy located at the **DYFAMED** station (**BOUSSOLE** project) at hourly frequency [5].

The ecosystem model provides a 12-component description of the ecosystem of the Ligurian Sea (see state variables in Table 1). A detailed description of this model can be found in [2]. The time evolution of each state variable is governed by the equation:

$$\frac{\partial C_i}{\partial t} = \text{ADV}_i + \text{DIFF}_i + \text{SMS}_i \quad \text{with} \quad \text{SMS}_i = \sum_{j \neq i} \text{FLUX}(C_j \rightarrow C_i) \quad (1)$$

where  $\text{ADV}_i$  and  $\text{DIFF}_i$  are advection and diffusion terms (governed by the hydrodynamic model), and  $\text{SMS}_i$  is the “source minus sink” term summing up the fluxes ( $\text{FLUX}(C_j \rightarrow C_i)$ ) between the various components of the ecosystem (conservation of course imposes that  $\text{FLUX}(C_j \rightarrow C_i) = -\text{FLUX}(C_i \rightarrow C_j)$ ). These fluxes can be sorted into several categories: primary production, secondary production, mortality, exudation, excretion, growth of bacteria, decomposition of particulate organic matter, and nitrification. We refer to Appendix A for a detailed description of these different processes.

Equation (1) is solved numerically between the sea surface and 405 m depth using a constant vertical discretization (1 m) and a constant time step (6 minutes). Outputs are saved daily at all depths.

It is important to note that the mathematical expression of these flux terms contains numerous parameters whose values are not known precisely. In the following, consistent with the objective of this study, we provide only a brief overview of these model parameters, with a specific focus on the assumptions that we make regarding their respective uncertainties.

Variable	Acronym	Name
$C_1$	<b>NO3</b>	Nitrate
$C_2$	<b>NH4</b>	Ammonium
$C_3$	<b>PicP</b>	Picophytoplankton
$C_4$	<b>NanP</b>	Nanophytoplankton
$C_5$	<b>MicP</b>	Microphytoplankton
$C_6$	<b>NanZ</b>	Nanozooplankton
$C_7$	<b>MicZ</b>	Microzooplankton
$C_8$	<b>MesZ</b>	Mesozooplankton
$C_9$	<b>BAC</b>	Bacteria
$C_{10}$	<b>DON</b>	Dissolved organic nitrogen
$C_{11}$	<b>POM1</b>	Particulate organic matter (size 1)
$C_{12}$	<b>POM2</b>	Particulate organic matter (size 2)

Table 1: Model state variables.

### 1.2. Model parameters

The biogeochemical fluxes ( $C_j \rightarrow C_i$ ) parameterized in **MODECOGeL** are summarized in Table A.3 (Appendix). Each flux depends on several of the parameters listed in Table B.4 (Appendix). Moreover, the parameterization of each process as a joint function of the model state and parameters is often complex and nonlinear. As a result, it is usually impossible for the user to know the sensitivity of the whole system to the parameters. A systematic computational approach is thus needed.

To apply the SA method, a probability distribution must be specified for each input parameter. This has been done here using the following guidelines:

- In the absence of any reliable information about possible correlations, the uncertainties of the various parameters are assumed independent.
- Most parameters are constrained to be either positive or negative, for consistency with the formulations used to parameterize the processes. They are assumed to follow a Gamma distribution.

- Parameters constrained between 0 and 1 are assumed to follow a Beta distribution.
- Some parameters are constrained to be larger than 1. Their logarithm is assumed to follow a Gamma distribution.
- Some parameters are not constrained, and are assumed to follow a Gaussian distribution.
- Three different values for standard deviations are used (5%, 20%, 50% of the expected value) according to the confidence we have in the parameters. These values were provided by biogeochemical modelers using a priori knowledge.

The resulting probability distributions are given in Table B.4 (Appendix).

### 1.3. Quantities of interest

The quantities of interest (QoI), i.e. the output values  $y$ , must be defined according to the main scientific objectives of the sensitivity study. In the present case, for this example, we have chosen to focus on characterizing the simulation of phytoplankton (concentrations  $C_3, C_4, C_5$  in Table 1), which is at the base of the marine food web.

As an additional quantity, we also introduce chlorophyll concentration (noted  $C_0$ ), which is what is observed by ocean color data, and which can be approximately computed from phytoplankton concentrations using a constant chlorophyll to nitrogen ratio ( $\alpha$ ):

$$C_0 = \alpha(C_3 + C_4 + C_5) \quad (2)$$

Since sensitivity analysis applies to scalar output quantities, to apply the SA method we need to reduce the time and space variations of  $C_0, C_3, C_4, C_5$  to some scalar indicators. Note that the cost of the method is almost independent of the number of these indicators. However they must be defined before running the sensitivity study. To illustrate the method, we thus decided to introduce a range of different QoI characterizing  $C_0, C_3, C_4$  (model outputs for microphytoplankton concentration are not retained), without limiting our choice to simple linear diagnostics.

Table 2 summarizes the QoI  $Y_{ij}$  that we will use in our application. The second index  $j$  corresponds to the computed diagnostic while the first index  $i$  corresponds to the concentration ( $i = 0, 3, 4, 5$ ) to which it is applied. This set of five diagnostics is meant to characterize (i) the maximum intensity of the phytoplankton spring bloom, (at the surface and as a vertical average), (ii) the time at which it occurs, and (iii) the overall average over the whole simulation.

Index $j$	Diagnostic	Definition
1	surface maximum	$\max_t C_i(0, t)$
2	time of surface maximum	$\operatorname{argmax}_t C_i(0, t)$
3	maximum of vertical average	$\max_t \frac{1}{Z} \int_0^Z C_i(z, t) dz$
4	time of maximum of vertical average	$\operatorname{argmax}_t \frac{1}{Z} \int_0^Z C_i(z, t) dz$
5	time and vertical average	$\frac{1}{ZT} \int_0^T \int_0^Z C_i(z, t) dz dt$

Table 2: Quantities of interest  $Y_{ij}$ . The maximum depth for averaging is  $Z = 40$  m and  $T$  is the total duration of the experiment.

### 1.4. Different strategies for estimating indices

In the case for complex models, one evaluates the model for selected values of the input parameters, and the resulting output values are used to estimate sensitivity indices of interest.

In our study, the different strategies we apply are:

- The replication procedure introduced in [4] (and further studied in [9]) to estimate all first-order Sobol’ indices with only **two replicated  $d$ -dimensional Latin hypercube samples of size  $n$** , that is with only  $2n$  model evaluations.
- The replication procedure introduced in [9] to estimate all closed second-order Sobol’ indices with only **two replicated  $d$ -dimensional randomized orthogonal arrays of strength 2 and size  $n$** , that is again with only  $2n$  model evaluations. Due to constraints in the construction of orthogonal arrays of strength 2,  $n$  must be chosen as  $q^2$ , with  $q$  a prime number greater than, or equal to,  $d-1$ .
- We apply the procedure introduced in [8, Theorem 1] to estimate all first-order and total Sobol’ indices with a cost of only  $(d+2)n$  model evaluations. This procedure is based on combinatorial arguments.
- It is interesting to compare the results we have just obtained using the global Sobol’ indices method with equivalent results obtained using local methods applied to the same MODECOGeL model.

We have thus computed the gradients of our QoIs with regard to the 74 input parameters. The partial derivatives are approximated by

$$\frac{\partial Y}{\partial X_j}(x_1, \dots, x_d) \simeq \frac{Y(x_1, \dots, x_{j-1}, x_j + \alpha, x_{j+1}, \dots, x_d) - Y(x_1, \dots, x_d)}{\alpha} \quad (3)$$

where  $\alpha$  is a small value ensuring very good accuracy of the estimate (in our case, we checked that any value of  $\alpha$  between  $10^{-3} \sqrt{\text{Var}(X_j)}$  and  $10^{-6} \sqrt{\text{Var}(X_j)}$  is relevant and leads to the same value for the gradient).  $d+1$  (i.e. 75 in our case) model runs are required. The local character of this gradient appears clearly in (3): the gradient is computed for a particular value  $(X_1, \dots, X_d) = (x_1, \dots, x_d)$ . In our case, we choose to compute it for the mean values of the 74 parameters (listed in Table B.4), i.e. for  $(x_1, \dots, x_d) = (\text{E}(X_1), \dots, \text{E}(X_d))$ .

## 2. Description of the data for each strategy

To compute the Sobol’ indices, we used the R `sensitivity` package. The choice of the designs of experiments was driven by the functions of this package that we used for the estimation of first-, second-order and total Sobol’ indices, as already mentioned in Section 1.4. A small percentage of the model evaluations failed or led to completely spurious results, due to unrealistic combinations of the input parameters. The corresponding QoIs were then set to NA and treated as missing values in the study presented in [7].

### 2.1. First-order Sobol’ indices with the replication procedure

- **Two replicated designs** (Latin hypercubes) **of size  $N$**  for a total cost of  $2N$  model evaluations.

We considered 4 different values for the size  $N$  of each design:  $10^3$ ,  $10^4$ ,  $10^5$  and  $10^6$ .

The information needed to compute the first-order Sobol’ indices from the model outputs is contained in a R dataset, one dataset for each size.

```
inputRLHS.74_1e+3.RData ,
inputRLHS.74_1e+4.RData ,
inputRLHS.74_1e+5.RData ,
inputRLHS.74_1e+6.RData
```

which contain:

- **d**: number of parameters
- **parClass**:  $\frac{\sigma}{\bar{x}}$  for each parameter (see Appendix)

- **parNames**: names of parameters
  - **parCent**: mean for each parameter
  - **N**: an integer giving the size of each replicated design
  - **X<sub>1</sub>**: a matrix containing the first replicated design. The field separator character is "white space", the first column of the table contains the row names.
  - **X<sub>2</sub>**: a matrix containing the second replicated design with the same format
  - **perm1**, **perm2**: table of permutations needed to compute **X<sub>2</sub>** from **X<sub>1</sub>**. **perm1** and **perm2** are both  $N \times d$  matrices whose columns are permutations of  $\{1, \dots, N\}$ . The  $j^{\text{th}}$  column of **X<sub>2</sub>** is obtained from the  $j^{\text{th}}$  column of **X<sub>1</sub>** by applying  $\pi_2^j \circ (\pi_1^j)^{-1}$ , with  $\pi_1^j$  the permutation of  $\{1, \dots, N\}$  encoded in the  $j^{\text{th}}$  column of **perm1** and  $\pi_2^j$  the permutation of  $\{1, \dots, N\}$  encoded in the  $j^{\text{th}}$  column of **perm2**.
- **The quantities of interest** For each **N**, the 15 QoIs computed on **X<sub>1</sub>** and **X<sub>2</sub>** are stored in two files, each of size  $(1 + N) \times 20$  in table format (csv file):

QI\_X1\_RLHS\_NA\_1e+3.csv , QI\_X2\_RLHS\_NA\_1e+3.csv ,  
 QI\_X1\_RLHS\_NA\_1e+4.csv , QI\_X2\_RLHS\_NA\_1e+4.csv ,  
 QI\_X1\_RLHS\_NA\_1e+5.csv , QI\_X2\_RLHS\_NA\_1e+5.csv ,  
 QI\_X1\_RLHS\_NA\_1e+6.csv , QI\_X2\_RLHS\_NA\_1e+6.csv.

The first line of this file contains the names of these quantities of interest.  
 To read these files in R, you can use the command **read.csv** as follows:

```
qi ← read.csv(file="QI_X1_RLHS_NA_1e+3.csv",header=TRUE,sep=" ",row.names=1)
```

## 2.2. Second-order Sobol' indices with the replication procedure

- **Two replicated designs** (randomized orthogonal arrays) **of size N** for a total cost of  $2N$  model evaluations.

For the second-order,  $N = q^2$  where  $q \geq d - 1$  is a prime number corresponding to the number of levels of the orthogonal array used. We considered 3 different values for  $q$ : 73 ( $N=5\ 329$ ), 227 ( $N=51\ 529$ ) and 709 ( $N=506\ 681$ ).

The information needed to compute the second-order Sobol' indices from the model outputs is contained in a R dataset, one dataset for each sample size **N**.

inputRLHS2\_74\_73.RData,  
 inputRLHS2\_74\_227.RData,  
 inputRLHS2\_74\_709.RData

which contain:

- **d**: number of parameters
- **parClass**:  $\frac{\sigma}{x}$  for each parameter (see Appendix)
- **parNames**: names of parameters
- **parCent**: mean for each parameter
- **N**: an integer giving the size of each orthogonal arrays
- **X<sub>1</sub>**: a matrix containing the first orthogonal array. The field separator character is "white space", the first column of the table contains the row names.
- **X<sub>2</sub>**: a matrix containing the second orthogonal array with the same format

- **perm1**, **perm2**: table of permutations needed to compute  $\mathbf{X}_2$  from  $\mathbf{X}_1$ . As previously, **perm1** and **perm2** are both  $\mathbf{N} \times \mathbf{d}$  matrices whose columns are permutations of  $\{1, \dots, \mathbf{N}\}$ . The  $j^{\text{th}}$  column of  $\mathbf{X}_2$  is obtained from the  $j^{\text{th}}$  column of  $\mathbf{X}_1$  by applying  $\pi_2^j \circ (\pi_1^j)^{-1}$ , with  $\pi_1^j$  the permutation of  $\{1, \dots, \mathbf{N}\}$  encoded in the  $j^{\text{th}}$  column of **perm1** and  $\pi_2^j$  the permutation of  $\{1, \dots, \mathbf{N}\}$  encoded in the  $j^{\text{th}}$  column of **perm2**.
- **The quantities of interest**: For each  $\mathbf{q}$  ( $\mathbf{N} = q^2$ ), the 15 QoIs computed on  $\mathbf{X}_1$  and  $\mathbf{X}_2$  are stored in two files of size  $(\mathbf{N} + 1) \times 20$  in table format (csv file):

QI\_X1\_RLHS2\_NA\_Q73.csv , QI\_X2\_RLHS2\_NA\_Q73.csv ,  
 QI\_X1\_RLHS2\_NA\_Q227.csv , QI\_X2\_RLHS2\_NA\_Q227.csv ,  
 QI\_X1\_RLHS2\_NA\_Q709.csv , QI\_X2\_RLHS2\_NA\_Q709.csv.

The first line of this file contains the names of the 15 QoIs.

To read these files in R, you can use the command **read.csv** as previously:

```
qi ← read.csv(file="QI_X1_RLHS2_NA_Q73.csv",header=TRUE,sep=" ",row.names=1)
```

### 2.3. Monte Carlo estimation of total Sobol' indices introduced by Saltelli

- **Two random samples**  $\mathbf{X}_1$  and  $\mathbf{X}_2$  of size  $\mathbf{N}$  for a total cost of  $(2 + \mathbf{d})\mathbf{N}$  model evaluations where  $\mathbf{d}$  is the number of parameters.

To compute the total indices, we used an experimental design  $\mathbf{X}$  from  $\mathbf{X}_1$  and  $\mathbf{X}_2$  of size  $(2 + \mathbf{d})\mathbf{N}$ . We considered one value for  $\mathbf{N} = 10^5$ .

The information needed to compute the total Sobol' indices from the model outputs is contained in a R dataset, one dataset for each size.

```
inputSalt_74_1e+5.RData
```

which contain:

- **d**: number of parameters
- **parClass**:  $\frac{q}{x}$  for each parameter (see Appendix)
- **parNames**: names of parameters
- **parCent**: mean for each parameter
- **N**: an integer giving the size of each replicated design
- **X<sub>1</sub>**: a  $\mathbf{N} \times \mathbf{d}$  matrix containing the first i.i.d. sample of input parameters. The field separator character is "white space", the first column of the table contains the row names.
- **X<sub>2</sub>**: a  $\mathbf{N} \times \mathbf{d}$  matrix containing the second i.i.d. sample of input parameters with the same format
- **X**: a  $(\mathbf{d} + 2)\mathbf{N} \times \mathbf{d}$  matrix containing the design of experiments

Note that the design  $\mathbf{X}$  is obtained from  $\mathbf{X}_1$  and  $\mathbf{X}_2$  as follows:  $\mathbf{X}$  is divided in  $1 + \mathbf{d} + 1 = 2 + \mathbf{d}$  matrices, denoted by  $\mathbf{X}_1, \mathbf{C}_1, \dots, \mathbf{C}_d, \mathbf{X}_2$ , each of size  $\mathbf{N} \times \mathbf{d}$ . These matrices are concatenated vertically. For any  $1 \leq i \leq \mathbf{d}$ ,  $\mathbf{C}_i$  is a matrix with  $\mathbf{N}$  lines and  $\mathbf{d}$  columns. Its  $i^{\text{th}}$  column is the one of  $\mathbf{X}_1$  while all its other  $\mathbf{d} - 1$  columns are the ones of  $\mathbf{X}_2$ . Such a design  $\mathbf{X}$  is called a pick-freeze scheme. The scheme is summarized below:

$$\mathbf{X}_1 = \begin{pmatrix} x_{1,1}^1 & \dots & x_{d,1}^1 \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ x_{1,n}^1 & \dots & x_{d,n}^1 \end{pmatrix} \quad \mathbf{X}_2 = \begin{pmatrix} x_{1,1}^2 & \dots & x_{d,1}^2 \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ x_{1,N}^2 & \dots & x_{d,N}^2 \end{pmatrix}$$

$$\mathbf{C}_i = \begin{pmatrix} x_{1,1}^2 & \cdots & x_{i,1}^1 & \cdots & x_{d,1}^2 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ x_{1,N}^2 & \cdots & x_{i,N}^1 & \cdots & x_{d,N}^2 \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{C}_1 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{C}_d \\ \mathbf{X}_2 \end{pmatrix}$$

- **The quantities of interest:** For each value of  $\mathbf{N}$ , a file of size  $(1 + (2 + \mathbf{d})\mathbf{N}) \times 15$  in table format (csv file) that contains the 15 quantities of interest computed on  $\mathbf{X}$ .

QI\_Salt\_NA\_1e+5.csv

The first line of this file contains the names of the QoIs.

To read these file in R, you can use the command **read.csv** :

```
qi <- read.csv(file="QI_Salt_NA_3e+4.csv",header=TRUE,sep=" ",row.names=1)
```

#### 2.4. Local method with gradient calculation

The information needed to compute the local indices from the model outputs is contained in a R dataset:

#### inputLocal\_74\_2.RData

which contains:

- **nparam:** number of parameters
- **nrowplan** : an integer giving the size of the experimental design
- **alpha:** 2-dimensional vector of values for  $\alpha$  (see Equation (3)) to approximate the gradients of our quantities of interest
- **parCent** : mean for each parameter
- **parClass:**  $\frac{\sigma}{\bar{x}}$  for each parameter (see Appendix)
- **parNames:** names of parameters
- **planSimul:** a matrix  $\mathbf{X}$  of size  $(1 + 2\mathbf{d})N \times \mathbf{d}$  containing the experimental design
- **qiNames:** names of quantities of interest
- **tabqi:** a matrix containing the 15 quantities of interest retained

More precisely, if  $(x_1, \dots, x_d) = (E(X_1), \dots, E(X_d))$ ,

$$\mathbf{X} = \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_{d-1} & x_d \\ x_1 + \alpha_1 & x_2 & x_3 & \cdots & x_{d-1} & x_d \\ x_1 + \alpha_2 & x_2 & x_3 & \cdots & x_{d-1} & x_d \\ x_1 & x_2 + \alpha_1 & x_3 & \cdots & x_{d-1} & x_d \\ x_1 & x_2 + \alpha_2 & x_3 & \cdots & x_{d-1} & x_d \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ x_1 & x_2 & x_3 & \cdots & x_{d-1} & x_d + \alpha_1 \\ x_1 & x_2 & x_3 & \cdots & x_{d-1} & x_d + \alpha_2 \end{pmatrix}.$$

## Appendix A. Description of biogeochemical fluxes

The biogeochemical fluxes ( $C_j \rightarrow C_i$ ) parameterized in MODECOGeL are summarized in Table A.3. Each flux depends on several parameters, which are indicated by referring to the parameter list in Table B.4 in Appendix B. To give an idea of the role of each parameter in MODECOGeL, the biogeochemical fluxes are organized into several categories using different colors:

- **Primary production (green)** is the growth of phytoplankton by photosynthesis. In Table A.3, this corresponds to all fluxes from nutrients ( $C_1, C_2$ ) to phytoplankton ( $C_3, C_4, C_5$ ). Parameters govern the maximum growth rate (1–3), and the dependence on nutrient concentrations (4–10), to solar irradiance (11–16), and to temperature (18–20).
- **Secondary production (blue)** is the growth of zooplankton by grazing of phytoplankton or by assimilation of bacteria and particulate organic matters. In Table A.3, this corresponds to all fluxes to zooplankton ( $C_6, C_7, C_8$ ). Parameters govern the ingestion rate (28–33), the dependence on prey concentration (34–36), the efficiency according to the type of prey (37–43), and the fraction actually used for growth (55–59).
- **Mortality (red)** of living species, including a parameterization of predation by higher trophic levels. In Table A.3, this corresponds to all fluxes from phytoplankton or zooplankton or bacteria ( $C_3$  to  $C_9$ ) to particulate organic matter ( $C_{11}$  or  $C_{12}$ ). Parameters govern mortality rates (44–50) and predation (51–53).
- **Exudation (magenta)** by phytoplankton. In Table A.3, this corresponds to all fluxes from phytoplankton ( $C_3, C_4, C_5$ ) to dissolved organic nitrogen ( $C_{10}$ ). Parameters are exudation rates (25–27).
- **Excretion (pink)** by zooplankton and bacteria. In Table A.3, this corresponds to all fluxes from zooplankton or bacteria ( $C_6$  to  $C_9$ ) to ammonium and dissolved organic nitrogen ( $C_2$  and  $C_{10}$ ). Parameters govern excretion rates (60–63), the dependence on temperature (64–67), the tradeoff between ammonium and dissolved organic matter (68), and the excreted fraction of predation (54).
- **Growth of bacteria (yellow)** from ammonium and dissolved organic matter. In Table A.3, this corresponds to all fluxes to bacteria ( $C_9$ ). Parameters govern the growth rate (23–24).
- **Decomposition of particulate organic matter (orange)**. In Table A.3, this corresponds to all fluxes from particulate organic matter ( $C_{11}$  or  $C_{12}$ ) to dissolved organic nitrogen ( $C_{10}$ ). Parameters are decomposition rates (69–70).
- **Nitrification (brown)**. In Table A.3, this corresponds to the flux from ammonium ( $C_2$ ) to nitrate ( $C_1$ ). The parameter is the nitrification rate (72).



	Nutrients		Phytoplanktons			Zooplanktons			BAC	DON & POM		
	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$	$C_6$	$C_7$	$C_8$	$C_9$	$C_{10}$	$C_{11}$	$C_{12}$
$C_1$			1,4,5, 11,14, 17,20	2,4,6, 12,15, 18,21	3,4,7, 13,16, 19,22							
$C_2$	72	72	1,8, 11,14, 17,20	2,9, 12,15, 18,21	3,10, 13,16, 19,22				23,24			
$C_3$			1,5,8, 11,14, 17,20, 25,44			28,31, 34,55				25	44	
$C_4$				2,6,9, 12,15, 18,21, 26,45			29,32, 35,56			26	45	
$C_5$					3,7,10, 13,16, 19,22, 27,46			30,33, 36,37, 57		27	46	
$C_6$						28,31, 34,47, 55,60, 64	29,32, 35,39, 56			60,64, 68	47	
$C_7$							29,32, 35,48, 56,61, 65	30,33, 36,40, 57		61,65, 68	48	
$C_8$								30,33, 36,49, 52,53, 54,57, 62,66		54,62, 66,68		49,51, 52,53
$C_9$						28,31, 34,38, 55	29,32, 35,56		23,24, 50,63, 67		50	
$C_{10}$									23,24			
$C_{11}$							29,32, 35,41, 58	30,33, 36,42, 59		69	69	
$C_{12}$								30,33, 36,43, 59		70		70

Table A.3: Biogeochemical fluxes from variable  $C_i$  (line  $i$ ) to variable  $C_j$  (column  $j$ ). Numbers in the boxes refer to parameter indices, given in Table B.4.

## AppendixB. Model parameters

This table describes the different probability distributions chosen for input parameters.

Index	Name	Unit	Pdf	Mean	Std	Std/Mean
1	<b>PicP</b> max growth rate	$t^{-1}$	$\Gamma(25, 0.12)$	3.	0.6	20%
2	<b>NanP</b> max growth rate	$t^{-1}$	$\Gamma(25, 0.1)$	2.5	0.5	20%
3	<b>MicP</b> max growth rate	$t^{-1}$	$\Gamma(25, 0.08)$	2.	0.4	20%
4	dependence of <b>NO3</b> limitation to <b>NH4</b>	$C^{-1}$	$\Gamma(400, 0.00365)$	1.46	0.073	5%
5	<b>NO3</b> semisaturation for <b>PicP</b>	$C$	$\Gamma(4, 0.125)$	0.5	0.25	50%
6	<b>NO3</b> semisaturation for <b>NanP</b>	$C$	$\Gamma(4, 0.175)$	0.7	0.35	50%
7	<b>NO3</b> semisaturation for <b>MicP</b>	$C$	$\Gamma(4, 0.25)$	1.0	0.5	50%
8	<b>NH4</b> semisaturation for <b>PicP</b>	$C$	$\Gamma(4, 0.075)$	0.3	0.15	50%
9	<b>NH4</b> semisaturation for <b>NanP</b>	$C$	$\Gamma(4, 0.125)$	0.5	0.25	50%
10	<b>NH4</b> semisaturation for <b>MicP</b>	$C$	$\Gamma(4, 0.175)$	0.7	0.35	50%
11	optimal PAR for <b>PicP</b>	$I$	$\Gamma(25, 0.4)$	10.	2.	20%
12	optimal PAR for <b>NanP</b>	$I$	$\Gamma(25, 0.6)$	15.	3.	20%
13	optimal PAR for <b>MicP</b>	$I$	$\Gamma(25, 0.8)$	20.	4.	20%
14	variation of light limitation for <b>PicP</b>	–	$-\Gamma(4, 0.2)$	-0.8	0.4	50%
15	variation of light limitation for <b>NanP</b>	–	$-\Gamma(4, 0.175)$	-0.7	0.35	50%
16	variation of light limitation for <b>MicP</b>	–	$-\Gamma(4, 0.15)$	-0.6	0.3	50%
17	optimal temperature for <b>PicP</b>	$T$	$\mathcal{N}(15, 3^2)$	15.	3.	20%
18	optimal temperature for <b>NanP</b>	$T$	$\mathcal{N}(15, 3^2)$	15.	3.	20%
19	optimal temperature for <b>MicP</b>	$T$	$\mathcal{N}(16, 3.2^2)$	16.	3.2	20%
20	variation of temp. limitation for <b>PicP</b>	–	$-\Gamma(4, 0.125)$	-0.5	0.25	50%
21	variation of temp. limitation for <b>NanP</b>	–	$-\Gamma(4, 0.125)$	-0.5	0.25	50%
22	variation of temp. limitation for <b>MicP</b>	–	$-\Gamma(4, 0.1375)$	-0.55	0.275	50%
23	bacteria growth limitation	–	$\Gamma(4, 0.15)$	0.6	0.3	50%
24	semisaturation for <b>BAC</b> growth	$C$	$\Gamma(4, 0.125)$	0.5	0.25	50%
25	exudation ratio for <b>PicP</b>	–	$\Gamma(4, 0.015)$	0.06	0.03	50%
26	exudation ratio for <b>NanP</b>	–	$\Gamma(4, 0.0125)$	0.05	0.025	50%
27	exudation ratio for <b>MicP</b>	–	$\Gamma(4, 0.01)$	0.04	0.02	50%
28	max ingestion rate for <b>NanZ</b>	$t^{-1}$	$\Gamma(25, 0.12)$	3.	0.6	20%
29	max ingestion rate for <b>MicZ</b>	$t^{-1}$	$\Gamma(25, 0.08)$	2.	0.4	20%
30	max ingestion rate for <b>MesZ</b>	$t^{-1}$	$\Gamma(25, 0.06)$	1.5	0.3	20%
31	threshold ingestion for <b>NanZ</b>	$C$	$\Gamma(4, 0.0125)$	0.05	0.025	50%
32	threshold ingestion for <b>MicZ</b>	$C$	$\Gamma(4, 0.0075)$	0.03	0.015	50%
33	threshold ingestion for <b>MesZ</b>	$C$	$\Gamma(4, 0.0025)$	0.01	0.005	50%
34	semisaturation for ingestion by <b>NanZ</b>	$C$	$\Gamma(4, 0.125)$	0.5	0.25	50%
35	semisaturation for ingestion by <b>MicZ</b>	$C$	$\Gamma(4, 0.1875)$	0.75	0.375	50%
36	semisaturation for ingestion by <b>MesZ</b>	$C$	$\Gamma(4, 0.25)$	1.	0.5	50%
37	efficiency of <b>MesZ</b> on <b>MicP</b>	–	$\beta(4.2, 1.05)$	0.8	0.16	20%
38	efficiency of <b>NanZ</b> on <b>BAC</b>	–	$\beta(4.2, 1.05)$	0.8	0.16	20%
39	efficiency of <b>MicZ</b> on <b>NanZ</b>	–	$\beta(4.2, 1.05)$	0.8	0.16	20%
40	efficiency of <b>MesZ</b> on <b>MicZ</b>	–	$\beta(4.2, 1.05)$	0.8	0.16	20%
41	efficiency of <b>MicZ</b> on <b>MOP1</b>	–	$\beta(19.8, 79.2)$	0.2	0.04	20%
42	efficiency of <b>MesZ</b> on <b>MOP1</b>	–	$\beta(19.8, 79.2)$	0.2	0.04	20%
43	efficiency of <b>MesZ</b> on <b>MOP2</b>	–	$\beta(19.8, 79.2)$	0.2	0.04	20%
44	mortality rate for <b>PicP</b>	$t^{-1}$	$\Gamma(4, 0.015)$	0.06	0.03	50%
45	mortality rate for <b>NanP</b>	$t^{-1}$	$\Gamma(4, 0.0125)$	0.05	0.025	50%
46	mortality rate for <b>MicP</b>	$t^{-1}$	$\Gamma(4, 0.01)$	0.04	0.02	50%
47	mortality rate for <b>NanZ</b>	$t^{-1}$	$\Gamma(4, 0.015)$	0.06	0.03	50%
48	mortality rate for <b>MicZ</b>	$t^{-1}$	$\Gamma(4, 0.0125)$	0.05	0.025	50%
49	mortality rate for <b>MesZ</b>	$t^{-1}$	$\Gamma(4, 0.0075)$	0.03	0.015	50%
50	mortality rate for <b>BAC</b>	$t^{-1}$	$\Gamma(4, 0.015)$	0.06	0.03	50%
51	threshold for predation	$C$	$\Gamma(4, 0.005)$	0.02	0.01	50%
52	maximum predation rate on <b>MesZ</b>	$t^{-1}$	$\Gamma(4, 0.25)$	1.	0.5	50%
53	semisaturation for predation on <b>MesZ</b>	$C$	$\Gamma(4, 0.25)$	1.	0.5	50%
54	excreted fraction of predation on <b>MesZ</b>	–	$\beta(2.33, 4.67)$	0.333	0.167	50%
55	fraction of grazing used for growth of <b>NanZ</b>	–	$\beta(4.2, 1.05)$	0.8	0.16	20%
56	fraction of grazing used for growth of <b>MicZ</b>	–	$\beta(4.2, 1.05)$	0.8	0.16	20%
57	fraction of grazing used for growth of <b>MesZ</b>	–	$\beta(4.2, 1.05)$	0.8	0.16	20%
58	fraction of POM used for growth of <b>MicZ</b>	–	$\beta(12, 12)$	0.5	0.1	20%
59	fraction of POM used for growth of <b>MesZ</b>	–	$\beta(12, 12)$	0.5	0.1	20%
60	excretion rate for <b>NanZ</b>	$t^{-1}$	$\Gamma(4, 0.0375)$	0.15	0.075	50%
61	excretion rate for <b>MicZ</b>	$t^{-1}$	$\Gamma(4, 0.025)$	0.1	0.05	50%
62	excretion rate for <b>MesZ</b>	$t^{-1}$	$\Gamma(4, 0.0125)$	0.05	0.025	50%
63	excretion rate for <b>BAC</b>	$t^{-1}$	$\Gamma(4, 0.0375)$	0.15	0.075	50%
64	temperature variation of excretion for <b>NanZ</b>	–	LogGamma	1.05	0.0525	5%
65	temperature variation of excretion for <b>MicZ</b>	–	LogGamma	1.05	0.0525	5%
66	temperature variation of excretion for <b>MesZ</b>	–	LogGamma	1.02	0.051	5%
67	temperature variation of excretion for <b>BAC</b>	–	LogGamma	1.04	0.052	5%
68	fraction of excretion as <b>DOM</b>	–	$\beta(2.75, 8.25)$	0.25	0.125	50%
69	POM1 decomposition rate	$t^{-1}$	$\Gamma(4, 0.01625)$	0.065	0.0325	50%

Index	Name	Unit	Pdf	Mean	Std	Std/Mean
70	POM2 decomposition rate	$t^{-1}$	$\Gamma(4, 0.015)$	0.06	0.03	50%
71	sedimentation velocity for <b>MicP</b>	$V$	$\Gamma(4, 0.25)$	1.	0.5	50%
72	nitrification rate	$t^{-1}$	$\Gamma(4, 0.0075)$	0.03	0.015	50%
73	light attenuation coefficient in sea water	–	$\Gamma(25, 0.0016)$	0.04	0.008	20%
74	fraction of photosynthetically active radiation	–	$\Gamma(25, 0.02)$	0.5	0.1	20%

Table B.4: Model parameters  $X_i$ . Units are: time  $t$  in days, concentration  $C$  in mmolN/m<sup>3</sup>, irradiance  $I$  in W/m<sup>2</sup>, and velocity  $V$  in m/day. The notation  $-\Gamma(.,.)$  means that the parameter is negative and that its opposite follows a  $\Gamma(.,.)$  distribution.

- [1] G. Lacroix. *Simulation de l'écosystème pélagique de la mer Ligure à l'aide d'un modèle unidimensionnel. Etude du bilan de matière et de la variabilité saisonnière, interannuelle et spatiale*. PhD thesis, Université Pierre et Marie Curie, Paris, France, 1998.
- [2] G. Lacroix and M. Grégoire. Revisited ecosystem model (MODECOGeL) of the Ligurian sea: seasonal and interannual variability due to atmospheric forcing. *J. Marine Syst.*, 37(4):229–258, 2002.
- [3] G. Lacroix and P. Nival. Influence of meteorological variability on primary production dynamics in the Ligurian Sea (NW Mediterranean sea) with a 1D hydrodynamic/biological model. *J. Marine Syst.*, 16(1-2):23–50, 1998.
- [4] T. Mara and O. Rakoto Joseph. Comparison of some efficient methods to evaluate the main effect of computer model factors. *Journal of Statistical Computation and Simulation*, 78(2):167–178, 2008.
- [5] J.C. Marty and J. Chiavérini. Hydrological changes in the ligurian sea (nw mediterranean, dyfamed site) during 1995–2007 and biogeochemical consequences. *Biogeosciences*, 7(7):2117–2128, 2010.
- [6] J. C. J. Nihoul and S. Djenidi. Perspectives in three-dimensional modelling of the marine system. In J.C.J. Nihoul and B.M. Jamart, editors, *Three-Dimensional Models of Marine and Estuarine Dynamics*, pages 1–34. Elsevier, Amsterdam, 1987.
- [7] C. Prieur, L. Viry, E. Blayo, and J.-M. Brankart. A global sensitivity analysis approach for marine biogeochemical modeling. *Ocean Modelling*, 139, 101402, 2019.
- [8] A. Saltelli. Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145(2):280–297, 2002.
- [9] J.-Y. Tissot and C. Prieur. A randomized orthogonal array-based procedure for the estimation of first- and second-order Sobol' indices. *Journal of Statistical Computation and Simulation*, 85(7):1358–1381, 2015.