# CHAPTER 1

# APPLICATION TO THE NUMERICAL COASTAL MODEL CROCO

In this chapter, we will study the problem of calibration of a specific model under uncertainties. We will first address the problem of the high dimensionality of the input space by segmentating it in arbitrary way. Then, after applying some of the methods introduced in the previous chapter.

[? ]

## 1.1 The ocean modelling system CROCO

CROCO [1] is a new oceanic modeling system built upon ROMS_AGRIF and the non-hydrostatic kernel of SNH (under testing), gradually including algorithms from MARS3D (sediments) and HYCOM (vertical coordinates). An important objective for CROCO is to resolve very fine scales (especially in the coastal area), and their interactions with larger scales. It is the oceanic component of a complex coupled system including various components, e.g., atmosphere, surface waves, marine sediments, biogeochemistry and ecosystems. [? ? ]

### 1.1.1 Numerical setting of the model

The model used in this thesis is roughly the same as [? ]. The spatial domain ranges from 9°W to 1°E and from 43°N to 51°N, and spans most of the Bay of Biscay, the English Channel and the eastern part of the Celtic Sea. The resolution is 1/14°, which

[1] CROCO and CROCO_TOOLS are provided by https://www.croco-ocean.org

leads to a mesh size between 5 km and 6 km. The bathymetry map is shown Fig. 1.1. We can notice roughly two regions based on the depth map: the region near the coasts which correspond to the continental shelf, where the water depth is less than 200 m, and the offshore region of the Bay of Biscay, where the depth is closer to 5000 m. . CROCO can

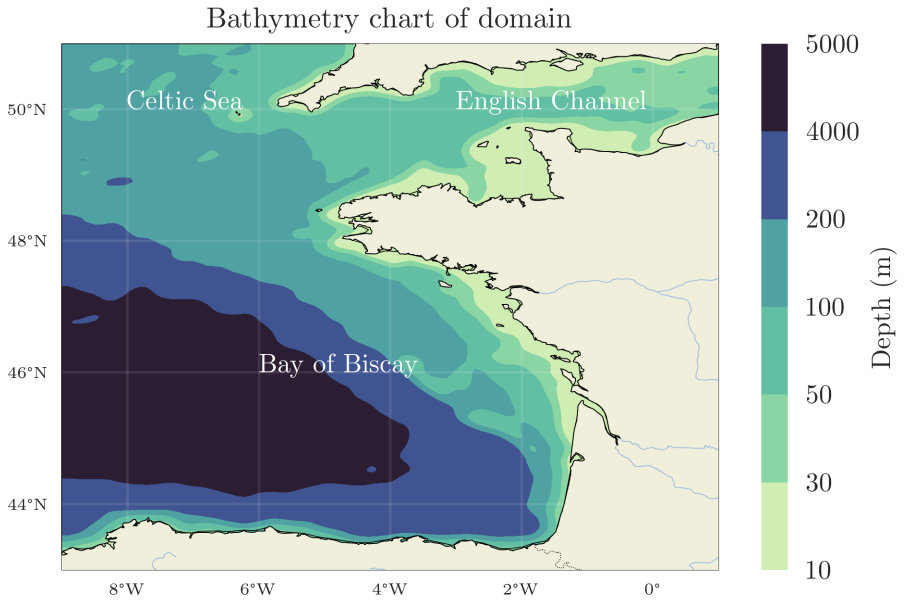


Figure 1.1: Bathymetry used in CROCO , and geographical landmarks. The continental shelf correspond roughly to the area with depth less than 200 m (green hue), while the abyssal plain has a depth larger than 4000 m (blue hue)

solve 3D fluid motions equations, but in this configuration is used only with one vertical level and thus solves numerically the shallow water equations:

$$\begin{cases} \frac{\partial u}{\partial t} + \nabla \cdot (\vec{v}u) - fv &= -\frac{\partial \phi}{\partial x} + \mathcal{F}_u + \mathcal{D}_u \\ \frac{\partial v}{\partial t} + \nabla \cdot (\vec{v}v) + fv &= -\frac{\partial \phi}{\partial y} + \mathcal{F}_v + \mathcal{D}_v \end{cases} \tag{1.1}$$

### 1.1.2 Modelling of the bottom friction

In CROCO , the bottom friction is modelled using a quadratic friction drag coefficient, as described Eqs. (1.2) and (1.3), expression which can be derived by assuming a logarithmic profile of the velocity at the bottom.

$$\boldsymbol{\tau}_b = C_d \|\mathbf{u}_b\| \mathbf{u}_b \tag{1.2}$$

$$C_d = \left( \frac{\kappa}{\log\left(\frac{H}{z_{0,b}}\right)} \right)^2 \text{ for } C_d \in [C_d^{\min}, C_d^{\max}] \tag{1.3}$$

where $\kappa$ is the Von Kármán constant, equal to 0.41.

Based on the documentation of the SHOM, the Fig. 1.2 shows a simplified version of the map of the repartition of tthe different types of sediments. Most of the ocean floor is here composed of sand. There is also very few areas where Siltic soil is listed.

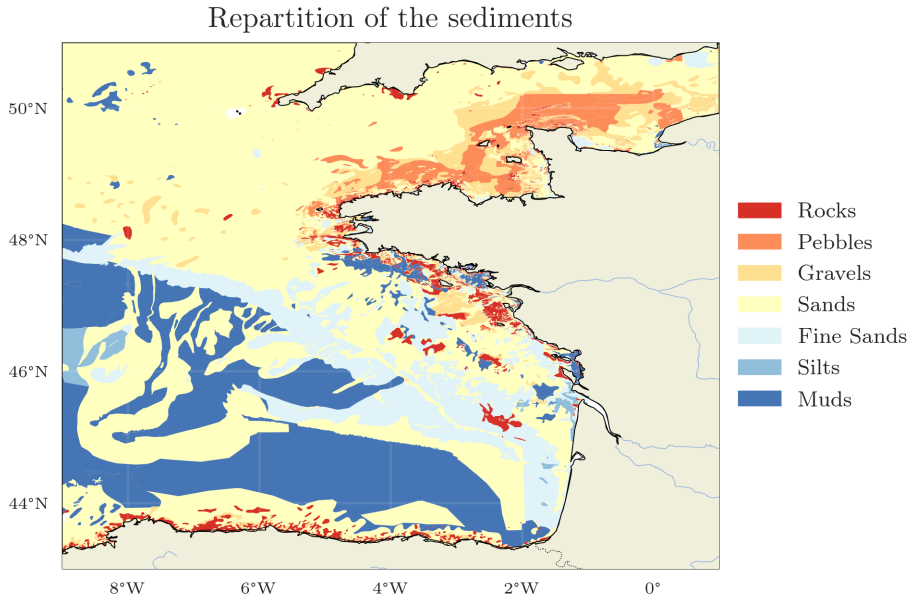


Figure 1.2: Repartition of the sediments on the ocean floor.

| Code | Description | Size of the majority of particles | $z_{0,b}^{\text{truth}}$ |
|---|---|---|---|
| Roche | Rock | Larger | 50 mm |
| C | Pebble | >20 mm | 25 mm |
| G | Gravel | [20 mm, 2 mm] | 7 mm |
| S | Sand | [2 mm, 0.5 mm] | 1 mm |
| SF | Fine Sand | [0.5 mm, 0.05 mm] | $1.5 \times 10^{-1}$ mm |
| Si | Silt | [0.05 mm, 0.01 mm] | $2 \times 10^{-2}$ mm |
| V | Muds | < 0.05 mm | $2 \times 10^{-2}$ mm |

Table 1.1: Type of sediments and size of the majority of particles for each type of sediment

Figure 1.2 shows that the largest sediments are rocks but are mostly located in the Bay of Biscay, near the boundary of the continental shelf. Pebbles however are mostly located in the shallow region in the English Channel, thus it may be expected that controlling the size of pebbles is highly influencal in a calibration context.

On Fig. 1.3 is shown the drag coefficient $C_d$ as a function of the roughness $z_{0,b}$ of the ocean floor, for different height of the water column $H$. We can see that the higher the water column height, the less variation appears when adjusting the bottom roughness $z_{0,b}$. Considering the physical properties of the bottom friction and the types of sediments,

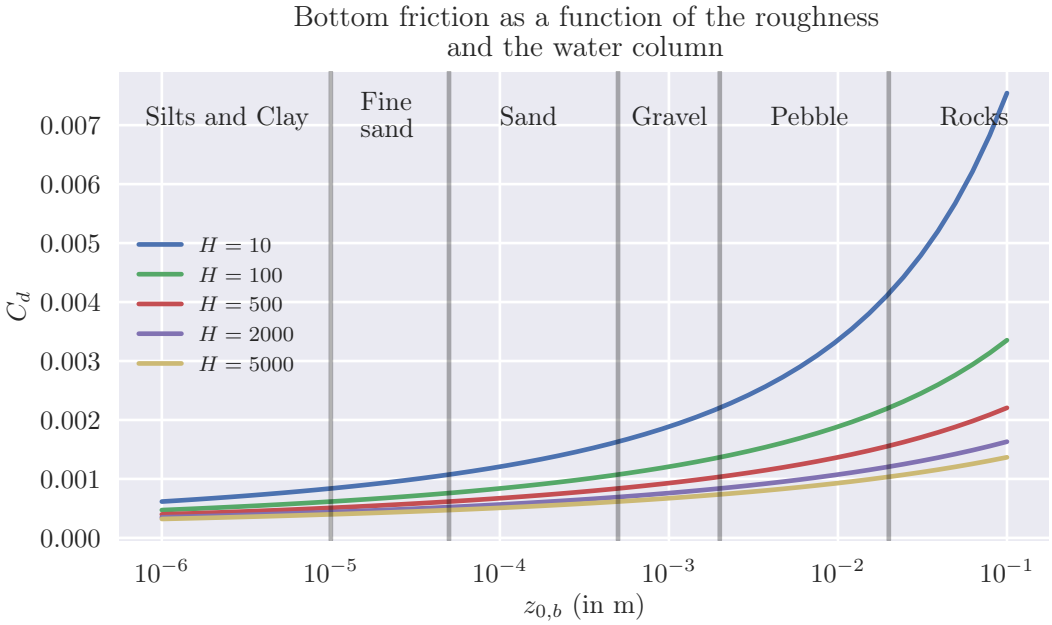


Figure 1.3: Drag coefficient $C_d$ as a function of the column water height and the roughness at the bottom

it can be expected that the English Channel, and at a lesser extent the rest of the continental shelf are the areas which are the most influential for the calibration.

## 1.2 Deterministic calibration of the bottom friction

### 1.2.1 Twin experiments setup

In a twin experiment setup, the observation $y$ are generated using the numerical model, and thus can be compared directly

### 1.2.2 Cost function definition

Let $y \in \mathbb{R}^{N_{\text{mesh}} \cdot N_{\text{time}}}$ be the observations, generated using a particular configuration of the model. We define $J$ as

$$J(\theta) = \sum_{t=1}^{N_{\text{time}}} \sum_{i=1}^{N_{\text{mesh}}} (\eta_{t,i}(\theta) - y_{t,i})^2 \tag{1.4}$$

$$= \|\mathcal{M}(\theta) - y\|_2^2 \tag{1.5}$$

Equivalently, as mentioned **??**, by assuming that the distribution of the (random) observation vector is known and $Y \mid \theta \sim \mathcal{N}(\mathcal{M}(\theta), I)$ where $I$ is the identity matrix, $J$ is proportional to the negative log-likelihood of the data.

### 1.2.3 Optimisation without uncertainties

The optimisation is first carried using M1QN3, a version of a gradient-descent procedure, as described in [**?** ]. We can first look to control $z_{0,b}$ at every cell of the mesh: $\theta = (\theta_1, \cdots, \theta_p)$ where $\theta_i = z^i_{0,b}$ and $p = 15\,684$. Due to the large number of points whose friction can be controlled., a finite difference method to get the gradient is unfeasible. Instead, Tapenade [**?** ], an Automatic Differentiation tool has been used in order to get the gradient of the cost function $J$ using the adjoint method, as described **??**. The optimisation procedure is stopped after 400 iterations, and the estimated controlled parameter is shown Fig. 1.4. Figure 1.5 shows the evolution of the cost function and the squared norm of the gradient during the optimisation procedure.

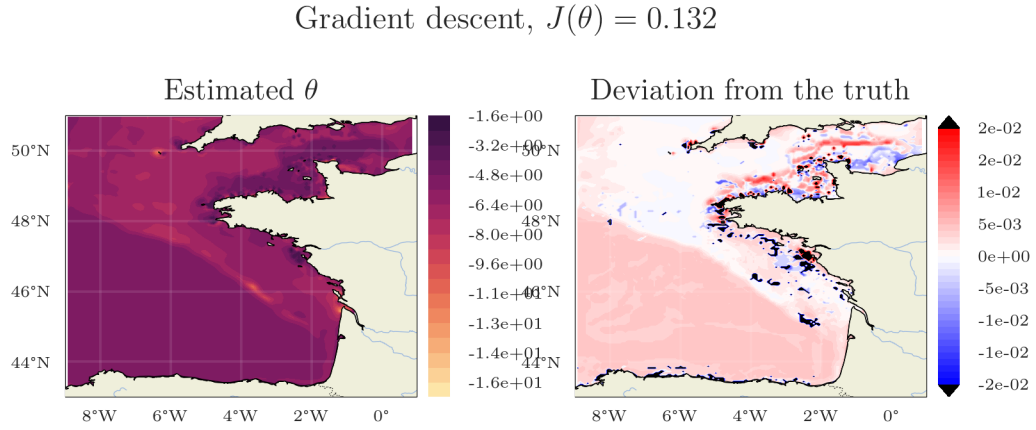


Figure 1.4: Optimization of $z_{0,b}$ on the whole space using gradient obtained via adjoint method, after 400 iterations TODO: Taille et colorbar

By comparing Fig. 1.4 with Fig. 1.2 and Fig. 1.1, we can have a first overview on the links between sediments type, depth, and *identifiability*, which can be understood as the ability of the optimisation procedure to retrieve a truth value.

On a first look, we can see that the abyssal plain (the deep region off the Bay of Biscay) remains mostly untouched by the optimisation, while the continental shelf, except for the English Channel, is retrieved. In terms of sediments, the bottom of the continental shelf is largely composed of sand. On Fig. 1.6, we can see that indeed, points of the mesh corresponding to sands have seen their $z_{0,b}$ value shifted toward the truth, while for silts and muds, the procedure has not been able to identify their roughness. This is probably due to the fact that those sediments lay at great depth, and thus have little effect on the circulation per Eq. (1.3).

In the English Channel, similar conclusions can be drawn: the size of the pebbles is well retrieved, but the control variable of points mapped to gravel do seem to compensate: on the northern part of the channel the size of the gravel is overestimated, while it is underestimated on the southern part. Finally the rocks appear to be hard to capture:

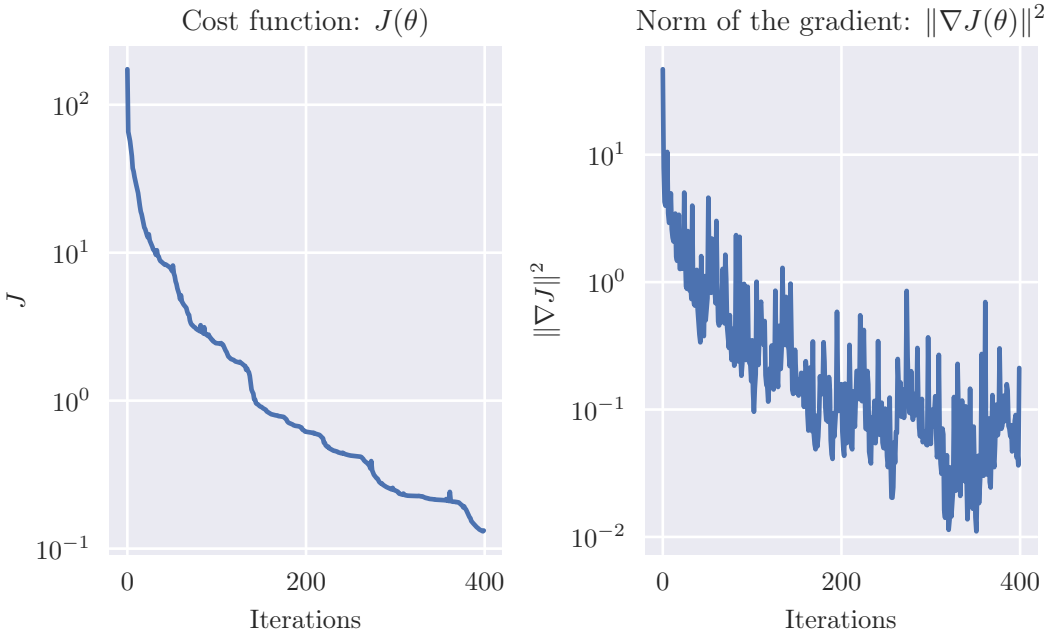


Figure 1.5: Evolution of the cost function and the squared normed of the gradient

their assumed size, *i.e.* their truth value is significantly larger than the rest of the sediments, and they are sparsely distributed.

### 1.2.4 Tidal modelling

[**?** ] TPX model of tides

| Darwin Symbol | Period (h) | Species |
|---:|---:|---:|
| $M_2$ | 12.4206 | Principal Lunar Semidiurnal |
| $S_2$ | 12 | Principal solar Semidiurnal |
| $N_2$ | 12.65834751 | Larger Lunar Elliptic Semidiurnal |
| $K_2$ | 11.96723606 | Lunisolar Semidiurnal |
| $K_1$ | 23.93447213 | Lunar Diurnal |
| $O_1$ | 25.81933871 | Lunar Diurnal |
| $P_1$ | 24.06588766 | Solar Diurnal |
| $Q_1$ | 26.868350 | Larger Lunar Elliptic Diurnal |
| $M_f$ | 13.660830779 | Lunisolar Fortnightly |
| $M_m$ | 27.554631896 | Lunar Monthly |

Table 1.2: Tide components

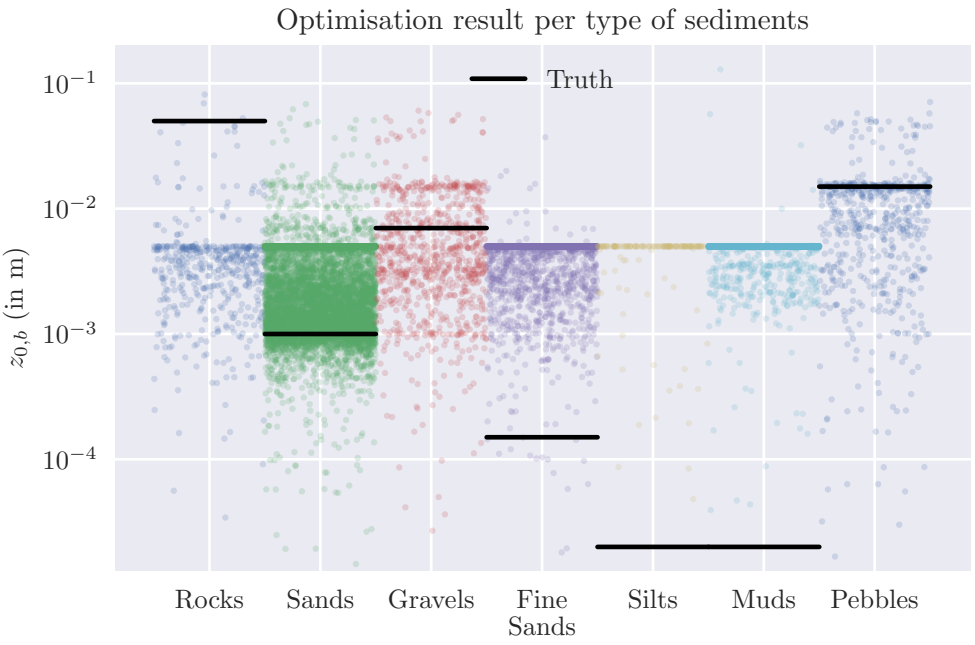


Figure 1.6: Results of the optimisation procedure, depending on the type of sediments. The initial value is $5 \times 10^{-3}$ m

The uncertainties in this configuration represent an error on the amplitude of the tide:

$$\tilde{A}_i(u_i) = A_i(1 + 0.01(2u - 1)) \tag{1.6}$$

so $\tilde{A}_i(0) = 0.99A_i$, $\tilde{A}_i(0.5) = A_i$ and $\tilde{A}_i(1) = 1.01A_i$

## 1.3 Sensitivity analysis

Sensitivity analysis (often abbreviated as *SA*), aims at quantififying the effect of the variation of some input variable to the output of the model [**?** **?** ]. Intuitively, *SA* ties the variation of the input to the variation of the output. It can then be approached at two different scales: around a nominal value, using the gradient, and at a global scale, by considering the inputs as random variable, and by measuring the variance of the output.

### 1.3.1 Methods of Sensitivity analysis

#### 1.3.1.a Local sensitivity analysis

Local sensitivity analysis [**?** ] refers to the study of how a small perturbation $\delta\theta$ of a nominal value $\theta$ affects the output of the numerical model. As we assume that the numerical model is accessible through the cost function $J$, a straightforward way to

quantify this perturbation is to consider the partial derivative of $J$, with respect to each component of the control variable $\theta = (\theta_1, \dots, \theta_p)$:

$$\frac{\partial J}{\partial \theta_i}(\theta) \tag{1.7}$$

The normalized local sensitivity at $\theta$ associated with the $i$-th component is then

$$\frac{\Delta J / J}{\Delta \theta_i / \theta_i} = \frac{\theta_i}{J(\theta)} \frac{\partial J}{\partial \theta_i} \tag{1.8}$$

#### 1.3.1.b Global Sensitivity Analysis: Sobol' indices

[**?** ] The $i$-th Sobol' indice of order 1 is defined as

$$S_i = \frac{\mathbb{V}\mathrm{ar}_{X_i}\left[\mathbb{E}_Y\left[Y \mid X_i\right]\right]}{\mathbb{V}\mathrm{ar}_Y\left[Y\right]} \tag{1.9}$$

where $Y = J(X)$ is the random variable. These are computed using a replicated method [**? ?** ], in order to get bootstrap confidence interval for the first and second order effects, and total effects

$$S_{T_i} = 1 - \frac{\mathbb{V}\mathrm{ar}_{X_{-i}}\left[\mathbb{E}_Y\left[Y \mid X_{-i}\right]\right]}{\mathbb{V}\mathrm{ar}_Y\left[Y\right]} \tag{1.10}$$

where $X_{-i} = (X_1, \dots X_{i-1}, X_{i+1}, \dots, X_p)$ is random vector of $p-1$ components.

### 1.3.2 Application to CROCO

The bottom friction affects the ocean circulation through two factors, as shown Eq. (1.3), first by the bottom roughness, $z_{0,b}$, and by the ocean depth. We are first going to perfrom a sensitivity analysis, in order to quantify the role of each sediment based region, without incorporating the knowledge on the typical size of the sediment there.

#### 1.3.2.a SA on the class of sediments

#### 1.3.2.b SA on the tide components

Fig. 1.7 shows the Sobol indices of order 1 (left), 2 (right), and the total effect indices, along with bootstrap confidence intervals. The variation of the variable associated with the $M_2$ component of the tide has the most impact on the cost function.

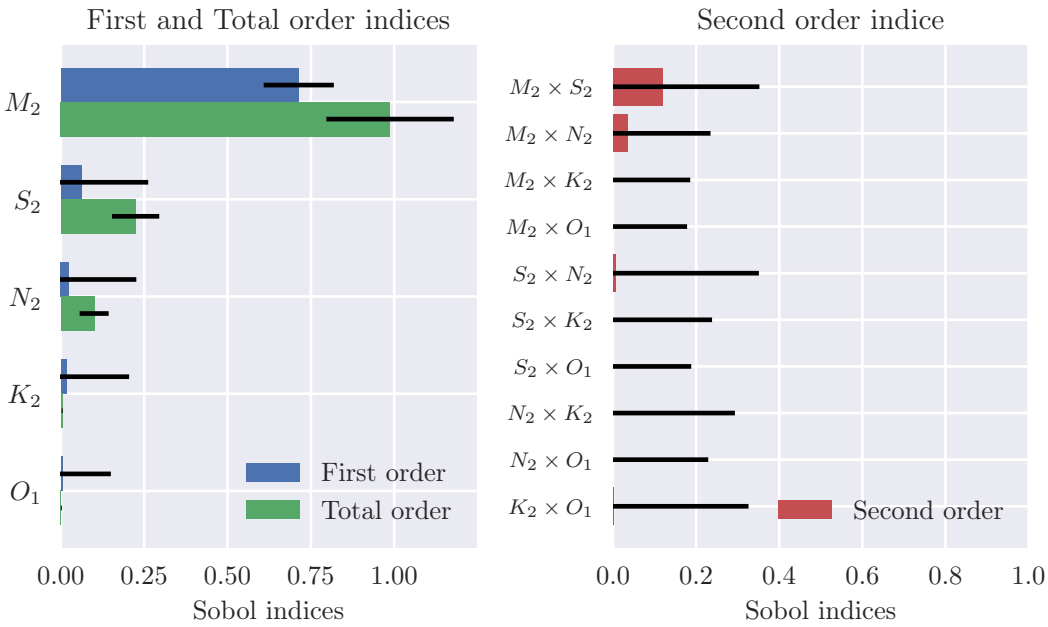


Figure 1.7: Global SA on the different components of the tide