



**HAL**  
open science

# Interoperability and computational framework for simulating open channel hydraulics: Application to sensitivity analysis and calibration of Gironde Estuary model

Cédric Goeury, Yoann Audouin, Fabrice Zaoui

## ► To cite this version:

Cédric Goeury, Yoann Audouin, Fabrice Zaoui. Interoperability and computational framework for simulating open channel hydraulics: Application to sensitivity analysis and calibration of Gironde Estuary model. *Environmental Modelling and Software*, 2022, 148, pp.105243. 10.1016/j.envsoft.2021.105243 . hal-04147544

**HAL Id: hal-04147544**

**<https://edf.hal.science/hal-04147544v1>**

Submitted on 3 Aug 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Interoperability and computational framework for simulating open channel hydraulics: application to sensitivity analysis and calibration of Gironde Estuary model

Cédric Goeury, Yoann Audouin, and Fabrice Zaoui

Électricité de France (EDF), Research and Development Division, National Laboratory for Hydraulics and Environment (LNHE), 6 Quai Watier, 78400 Chatou, France

July 2, 2021

## Abstract

Water resource management is of crucial societal and economic importance, requiring a strong capacity for anticipating environmental change. Progress in physical process knowledge, numerical methods and computational power, allows us to address hydro-environmental problems of growing complexity. Modeling of river and marine flows is no exception. With the increase in IT resources, environmental modeling is evolving to meet the challenges of complex real-world problems. This paper presents a new distributed Application Programming Interface (API) of the open source TELEMAC-MASCARET SYSTEM to run hydro-environmental simulations with the help of the interoperability concept. Use of the API encourages and facilitates the combination of worldwide reference environmental libraries with the hydro-informatic system. Consequently, the objective of the paper is to promote the interoperability concept for studies dealing with such issues as uncertainty propagation, global sensitivity analysis, optimization, multi-physics or multi-dimensional coupling. To illustrate the capability of the API, an operational problem for improving the navigation capacity of the Gironde Estuary is presented. The API potential is demonstrated in a re-calibration context. The API is used for a multivariate sensitivity analysis to quickly reveal the most influential parameters which can then be optimally calibrated with the help of a data assimilation technique.

## Software availability

The Application Programming Interface (API) framework described in this article (Fortran APIs and its Python wrapper) is available for download in TELEMAC-MASCARET SYSTEM ([www.opentelemac.org](http://www.opentelemac.org)). TELEMAC-MASCARET is an integrated suite of solvers for use in the field of free-surface flow, available under the GNU General Public License version 3.

## 1 Introduction

Water availability and quality are vital to human health. Increasing global scarcity makes anticipating the evolution of this limited natural resource even more essential. This ability relies on the understanding and prediction of hydrodynamic flows. Hydraulic simulation codes (such as HEC-RAS [22], MOHID [48], DELFT-3D [20], TELEMAC-MASCARET [34]) have been developed for many years and applied to a wide range of hydro-environmental cases, such as prevention of flood risks [71], development of management plans [63] or design schemes for flood alleviation [55].

Hydro-environmental modeling of river and marine flows is increasingly complex and requires a better knowledge of physical and biochemical processes, advanced numerical methods and high-performance computing resources. For many applications, engineers have to deal with complex “modeling systems” for the representation of subparts of a physical system [8, 32]. Increasing a numerical model conceptual complexity in terms of how many aspects of the underlying system are included in it has the potential to improve the accuracy of the results in term of descriptive and predictive capacities. However, adding each new aspect to the model can induce an increase of the error brought about by the uncertainty in the description of the features accumulates and increases the overall uncertainty in the output [14, 44]. This trade-off between “model completeness” and “propagation error” [66] is known as the

conjecture of O’Neill [56]. In maritime or fluvial cases, numerical modeling tools allow replicating the past as well as predicting the future with an uncertainty range that is strongly linked to the approximate description of hydraulic parameters [59], meteorological data [4] and geographical data [1]. In spite of the significant improvement in computational resources and the accuracy of numerical models, the turbulent nature of fluid mechanics equations limits the performance of hydraulic models. The need for realistic hydrodynamic flow simulation is beyond the abilities of deterministic forecast. As a consequence, model uncertainties have to be quantified. Recent advances in simulation and optimization of environmental systems have relied on increasingly detailed models, running many scenarios, in order to quantify these uncertainties [33, 71].

According to Solomatine and Ostfeld [68], significant advances in the means of observing continental waters, especially through orbital equipment (SWOT satellite mission [51]; Sentinel satellite missions (Copernicus program) [46]), allows use of various types of data, with their associated analysis, for water resources management [40]. Thus, environmental modeling is increasingly complemented by data-driven models [37]. Moreover, by taking into account advances in information processing and data management, it becomes possible to reduce uncertainties and better understand the modeling process. Data-driven modeling is a paradigm shift for addressing many problems in science and engineering [49].

To summarize, with the increase in computer resources, environmental modeling is evolving with the need to solve increasingly complex real-world problems involving the environment and its relationship to human systems and activities. If it is to break down research silos, and bring scientists, stakeholders and decision-makers together to solve social [17], economic [31] and environmental problems, an environmental modeling system should support IT tasks, such as aggregation of model components into functional units, component interaction and communication, parallel computing and cross-language interoperability [19, 43]. These technical requirements are closely related to interoperability aspects, namely the capacity of the software to run and share information with other codes. Environmental system development requires both scientific understanding of the phenomena involved and software development proficiency.

This paper presents a new Application Programming Interface (API) of the open source and interoperable TELEMAC-MASCARET SYSTEM (<http://svn.opentelemac.org/svn/opentelemac/>) to solve environmental problems [28, 41]. Since 2017, the TELEMAC-MASCARET SYSTEM has included in its distribution a generic API that allows it to expand its range of applications [5]. With this newly implemented feature, the TELEMAC-MASCARET SYSTEM can be easily coupled or integrated into higher-level platforms to model, study or simulate complex problems [27].

To our knowledge, APIs for an environmental modeling system (with hydraulic (MASCARET 1D, TELEMAC-2D, TELEMAC-3D), water quality (TRACER, WAQTEL), sediment transport (COURLIS, GAIA), and wave hydrodynamics (TOMAWAC and ARTEMIS)) are not yet widespread in the environmental community and are an original innovation. The present work uses as an example France’s Gironde Estuary to demonstrate the operational applicability and outcomes of APIs. The port of Bordeaux in the Gironde Estuary faces many development challenges [38]. To satisfy demand for larger ships, while ensuring navigational safety, the evolution of water depth over time in the estuary needs to be predicted with a high degree of accuracy. To address this issue, the numerical results must be consistent with past observational data. Among other things, this process relies on calibration to determine “empirical adequacy” [57]. In particular, the calibration aims at simulating a series of reference events by adjusting some uncertain physically-based parameters until the comparison is as accurate as possible [21]. The objective of this work is to implement an efficient calibration algorithm, capable of processing measurement optimally, in order to estimate the partially known or missing parameters [53]. Often performed, calibrating a hydrodynamic model is nevertheless a difficult task owing to the complexity of the flows and their interaction with the shoreline, the bathymetry, islands, etc. It is essential to understand, in depth, the relationship between the calibration of modeling parameters and the simulated state variables which are compared to the observations. In this context, a multivariate sensitivity analysis [23, 42] succeeded in identifying the most influential input parameters on the model results. Sensitivity analysis is associated to define the parameters to be calibrated using a physically based data-driven technique [12].

This paper is organized as follows. Section 2 includes a description of the main concepts of the TELEMAC-MASCARET SYSTEM API. Section 3 presents the architecture of implemented APIs. Section 4 describes the application case. Section 5 deals with the calibration problem. Section 6 is the discussion and Section 7 is the conclusion.

## 2 Method

The TELEMAC-MASCARET hydro-informatic system, created in 1987, is an open source ([www.opentelemac.org](http://www.opentelemac.org)) integrated suite of solvers for use in the field of free-surface flows [34]. As displayed on Figure 1, it can carry out simulations of flows (MASCARET, TELEMAC-2D and TELEMAC-3D, gray circle), sediment transports (COURLIS

and GAIA, orange circle), waves (ARTEMIS and TOMAWAC, green circle) and water quality (TRACER and WAQTEL, purple circle). Historically, from the original software solving physics equations (Saint-Venant: MASCARET and TELEMAC-2D, Navier-Stokes: TELEMAC-3D, elliptic mild slope equation: ARTEMIS, simplified equation for the spectro-angular density of wave action: TOMAWAC, Exner: COURLIS and GAIA, water quality processes: TRACER and WAQTEL), the system has gradually evolved to the notion of hydro-informatics with a set of solvers dealing with large heterogeneous data and solving complex dependent problems (overlapping physical components in Figure 1). The various simulation components use high-end algorithms based on the finite element or finite volume method. The ability to study global, regional, or local scale problems by using the same system, makes the TELEMAC-MASCARET a useful tool for assessing the environmental state in the sea, estuaries, coastline and rivers.

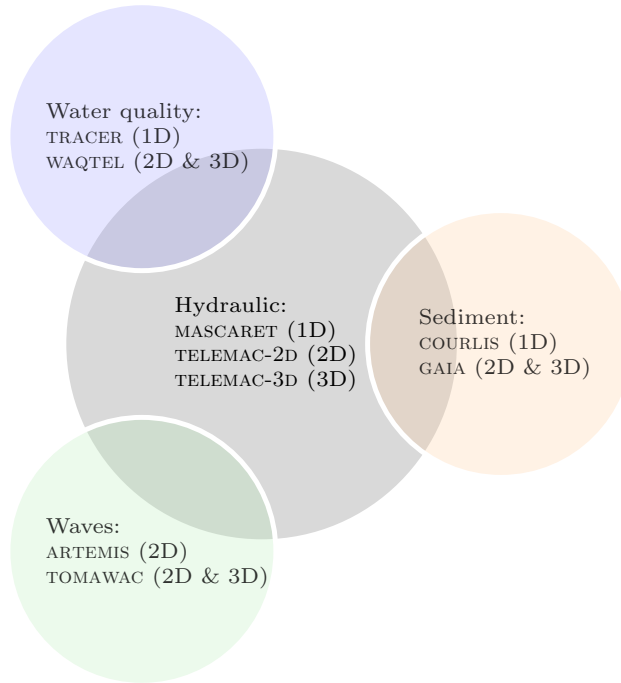


Figure 1: TELEMAC-MASCARET hydro-informatic system

According to Laniak et al. [43], Integrated Environmental Modeling (IEM) highlights the importance of software development and sharing in their role as pluggable components of a larger ecosystem. The work presented here is part of this trend. It shows the implementation procedure for the component-based solver interoperability of the TELEMAC-MASCARET SYSTEM. The API architecture, described in this paper, is directly applicable for any software regardless of its language since it is an easy task to build a wrapper in other programming language as presented in the following for Fortran and Python. The main difficult part is to split the software main program. When a solver is an interoperable component, it can be part of an assembly of elements in permanent interaction that work in a coordinated way. To be considered as an elementary reusable component, a computation model must:

- possess APIs, which can be viewed as new functionalities or entry points in the computational model to provide new services on demand;
- have a resolution algorithm comprising three main functionalities: problem initialization, simulation and finalization of the computation;
- allow instantiation of several clearly separated and identifiable problems;
- be usable in dynamic library form.

In the following, the four points listed above are addressed.

## 2.1 Design

A computer programming model like the object-oriented programming (OOP) aims to organize the software design around data grouped into objects with attributes and methods. This helps the developer to focus on the data rather than the logic required to manipulate them. Most of the OOP concepts (encapsulation, interfaces, polymorphism, modularity, instantiation, ...) are derived from software engineering patterns but relate to the real world. As shown in Figure 2, the first concept used here is Encapsulation with a classic Design Pattern “Facade” [24]. Since the TELEMAC-MASCARET SYSTEM is written in Fortran 77 and 90, the design pattern has been adapted to this non object-oriented language as OOP is well-suited for the design of large and complex programs.

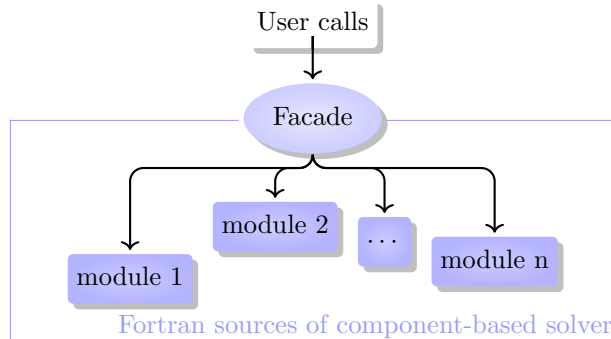


Figure 2: Design Pattern “Facade”

The main benefits of this approach are that it is easier to use and understand, limits the library user dependencies, includes integrity constraints and validation rules and avoids some unnecessary calculations [41].

Internal data has been accessed in order to expand the capabilities of the API. It is possible to examine all the information embedded in the calculation code and modify its structure (specifically the values, some metadata and properties) at runtime. Pointers were used to access the simulation data. As Fortran lacks the pointer definition of other high-level languages such as C or C++, a combination of the `POINTER` and `TARGET` Fortran keywords with adequate updates were employed. For instance, a manually update must be done if a pointer on an `ALLOCATABLE` variable is set before the dynamic allocation since the pointer is no longer valid. This ensures the correct state of the controlled variable. A component-based approach was adopted by having the same structure as all of the TELEMAC-MASCARET physical-dedicated solver, to allow them to be viewed as linkable. The execution of a simulation was partitioned to allow modification of a given parameter at runtime (for example, each time step). The flexibility that comes with interoperability allows deployment of the TELEMAC-MASCARET SYSTEM in specific industrial applications and distribution of TELEMAC-MASCARET solvers as a part of complex modeling system (Applications, blue part in Figure 3). As mentioned by Gil et al. [26], geosciences “will lead to a new generation of knowledge-rich intelligent systems that contain rich-knowledge and context in addition to data, enabling fundamentally new forms of reasoning, autonomy, learning, and interaction”. As presented in Figure 3, the TELEMAC-MASCARET physical-dedicated components can be easily connected together to build multi-physics or multi-dimensional interactions [32] (Coupling, orange part), inserted in optimization or uncertainty quantification processes (Studies, yellow part, see Section 5), wrapped in advanced technology such as cloud-based solutions or embedded systems with data management (OpenMI [29] or Pynsim [39], Cloud services, green part) and integrated in platforms (Salome [64] or Palm [9] for example, Numeric computing environment, pink part).

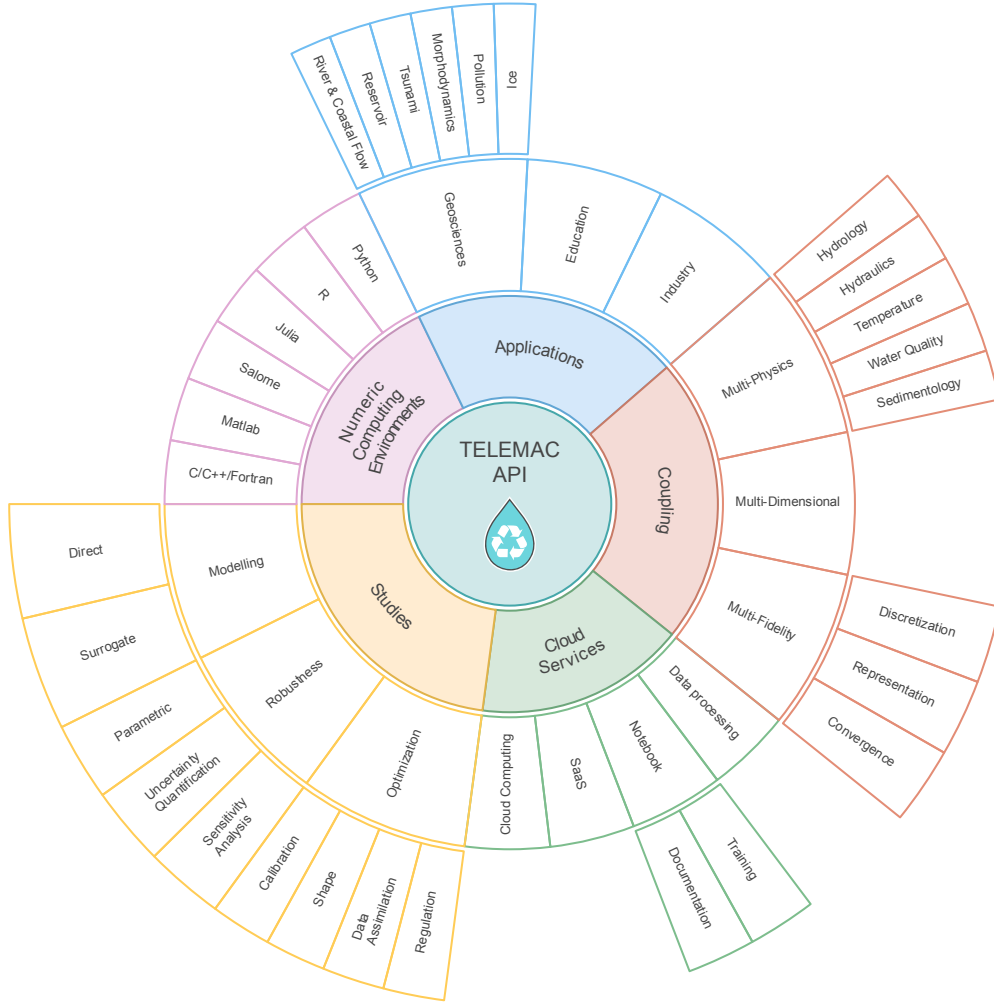


Figure 3: Ecosystem of the TELEMAC-MASCARET API

## 2.2 Technology

To maintain efficient execution, a native code compiler (for example, GNU Compiler Collection, licensed under GNU GPL) running on most systems and hardware architectures, and supporting Fortran, is used and recommended. Moreover, the calculation code of the TELEMAC-MASCARET SYSTEM numerically solves a model based on a system of partial differential equations (PDE). The computational solution relies on an implicit time scheme, removing stability constraints on the time frame, but leading to a linear-system solution of considerable size. The resulting increased complexity of the model structures poses strong limitations in terms of practical implementation and computational requirements. Consequently, High-Performance Computing (HPC) is necessary to tackle realistic applications. The solution deployed in the hydro-informatic system is to decompose the domain of computation. This idea of domain decomposition is to assign to each processor one part of the global domain over which it solves the fluid mechanics problem. The results of the other processors help in defining artificial boundary conditions arising from the partition. The assignments given to each processor are the same, but the data differ. Therefore, the TELEMAC-MASCARET SYSTEM requires the use of an MPI library for communication between processors [16].

It is possible to use MPI technology for the instantiation of several clearly separated and identifiable problems. This is because MPI is adapted for distributed memory systems. This method has been followed for the learning step of sensitivity analysis presented in Subsection 5.1.

### 3 Architecture of API

The architecture is generic for all TELEMAC-MASCARET SYSTEM components (Figure 1). For the sake of clarity, only an API example with the TELEMAC-2D component is shown in the following sections.

#### 3.1 Fortran structure

API's main goal is to control the simulation while running a case. For example, it must allow the user to suspend the simulation at any time step, retrieve some variable values, and possibly change them before resuming the calculation. To make this possible, a Fortran structure called “instance” is used in the API. The instance structure gives direct access to the physical memory of variables, and therefore allows control of the variable. Furthermore, based on decomposition of the main TELEMAC-MASCARET SYSTEM subroutines in three main functionalities (initialization, simulation of one time step and finalization), it is possible to run a hydraulic case for an indefinite sequence of time steps.

Figure 4 describes an example of API workflow where  $\mathbf{X}$  are input parameters and  $O(t)$  is information extracted at given times. For instance, in the framework of the Gironde Estuary application case,  $\mathbf{X}$  represents parameters to be estimated / calibrated (friction coefficients and tidal boundary condition parameters) and  $O(t)$  is the free surface elevation every minute at the measurement stations (see Section 4).

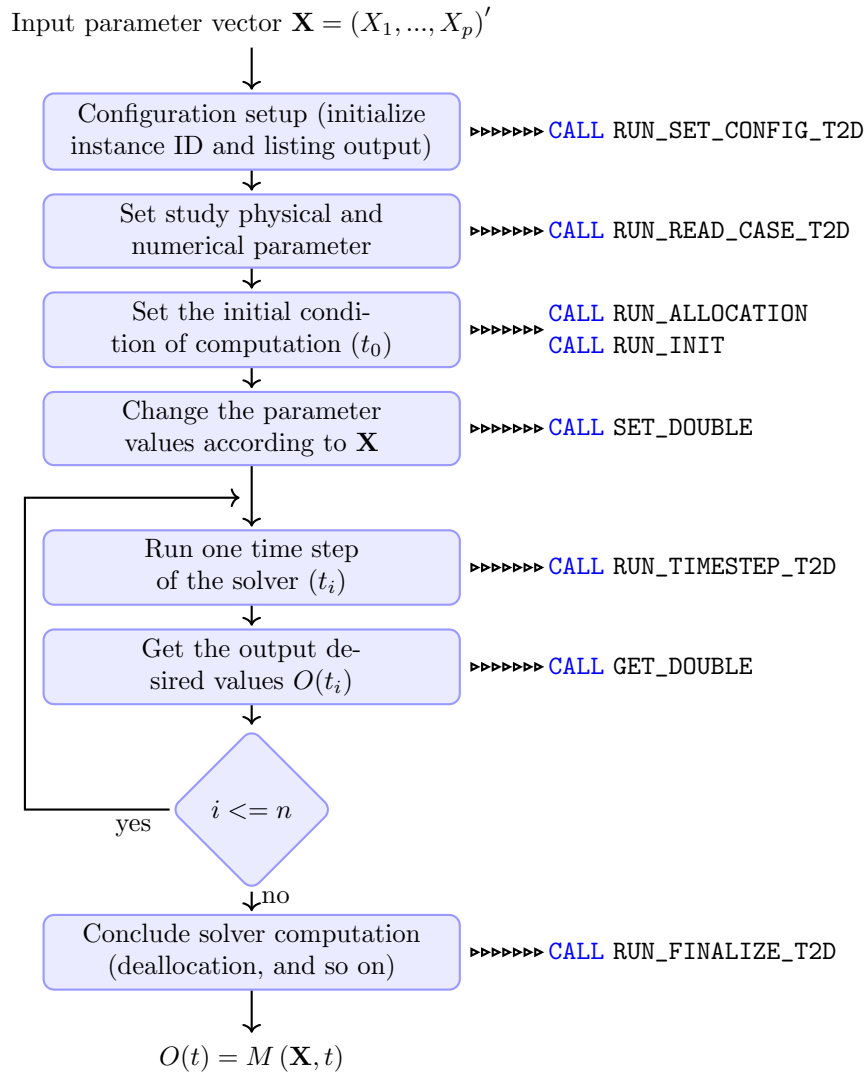


Figure 4: Diagram of study workflow (Function  $M(\mathbf{X}, t)$  in Figure 7) and corresponding Fortran API calls

For each action defined above, the identity number of the instance is used as an input argument allowing

all computation variables to be linked with the corresponding instance pointers. These actions must be done in chronological sequence in order to ensure proper execution of the computation in the API main program. In Fortran and for the shallow water code TELEMAC-2D, it will begin with the call of an initialization subroutine:

```
CALL RUN_SET_CONFIG_T2D(ID, LU, LNG, COMM, IERR)
```

where all parameters are scalars of type integer and ID is an output value giving the instantiation number. The input parameter LU is used to redirect the standard output, LNG has only two possible values for the choice of the english or french language, and COMM is the MPI communicator for the distributed-memory parallelism. The output IERR has to be considered with care as a non-null value stands for the occurrence of errors. A systematic test on this return value is recommended, for instance one may want to stop the program accordingly:

```
IF (IERR.NE.0) EXIT IERR
```

After the initialization phase, the model has to be read from files with the following instruction:

```
CALL RUN_READ_CASE_T2D(ID, CAS_FILE, DICO_FILE, INIT, IERR)
```

The name of the steering file is given with CAS\_FILE as a character string, it stores several input data of the model with key-value pairs. Each possible key is set with a dictionary file whose name is a character string hold by DICO\_FILE. A true value for the logical input INIT will initialize a common computational kernel of all the modules. After this step, a dynamic allocation of arrays in memory is necessary followed by a state initialization of the physics:

```
CALL RUN_ALLOCATION_T2D(ID, IERR)
CALL RUN_INIT_T2D(ID, IERR)
```

All accessible parameters are listed in predefined keyword lists (see Appendix A for TELEMAC-2D). For example, the number of time steps in the simulation is named MODEL.NTIMESTEPS and its value is obtained and stored in the variable NBSTEPS with the following call:

```
CALL GET_INTEGER(ID, "T2D", "MODEL.NTIMESTEPS", NBSTEPS, 0, 0, 0, IERR)
```

Input parameter can be defined here using the following call:

```
CALL SET_DOUBLE(ID, "T2D", "MODEL.SEALEVEL", SEALEVEL, 0, 0, 0, IERR)
```

where the sea level value in the computation (accessed through the keyword MODEL.SEALEVEL) is set to the value of the variable SEALEVEL.

Now the simulation can be done with a loop on the number of time steps while storing all the results of an evolution of the water depth:

```
DO I = 1, NBSTEPS
  CALL RUN_TIMESTEP_T2D(ID, IERR)
  CALL GET_DOUBLE(ID, "T2D", "MODEL.WATERDEPTH", WATERDEPTH(I), 42, 0, 0, IERR)
ENDDO
```

WATERDEPTH is a one-dimensional array of size NBSTEPS. It is used here to store the water depth values at node number 42 of the TELEMAC-2D triangular mesh. Finally, the Fortran program can end with the instance deletion thus freeing the memory used:

```
CALL RUN_FINALIZE_T2D(ID, IERR)
```

Additional functions are available to handle parallelism (see Subsection 2.2 for more details on parallelism in the TELEMAC-MASCARET SYSTEM). The data structure in the current version (official version 8.2) does not yet allow for multiple instantiations, as all the instances point to the same memory area. This constitutes a future development to improve API capacity. However, instantiation associated with the processor communication function overcomes this limitation by ensuring the possibility of several clearly separated and identifiable problems.

## 3.2 Python wrapping

The use of API is not limited to Fortran programming but can also be used by a scripting language. To this end, a Python wrapper is also available. It is relatively easy to use the Fortran API routines directly in Python using the f2py tool of the Python Scipy library [60]. This tool will make it possible to compile Fortran code for use in Python. The only limitations are on the type of arguments of the functions to wrap. Python is a portable, dynamic,



extensible, free language, which allows (without requiring) a modular approach and object-oriented programming [73]. In addition to the benefits of this programming language, Python offers a large number of interoperable libraries for scientific computing, image processing, data processing, machine learning and deep learning. The link between various interoperable libraries with TELEMAC-MASCARET SYSTEM APIs allows the creation of an ever-more efficient calculation chain capable of responding finely to various complex problems such as the application case presented in this article (see Section 4). Moreover, the Python scripting language makes it possible to implement a wrapper in order to provide user-friendly functioning of the Fortran API. Thus, a Python overlay was developed to encapsulate and simplify the different API calls.

For instance, a Python “get” function can encapsulate all API Fortran type-dedicated function `get_type` (where `type` can be replaced by Double, Integer, Boolean and String). File management is also made easier with Python (copy, move, . . . , and reducing the number of arguments). This Python wrapping of TELEMAC-MASCARET SYSTEM API constitutes a package called TelApY distributed in the official version of the hydro-informatic system (Figure 5).

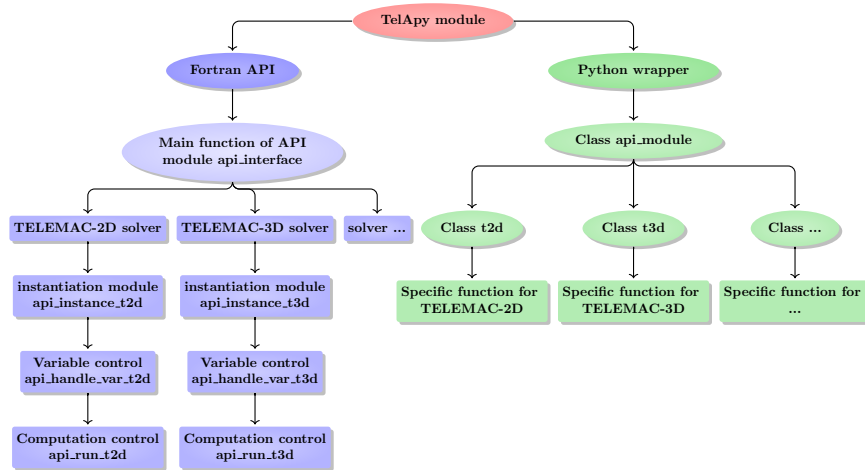


Figure 5: Overview of TelApY module

The TelApY package is provided with a tutorial for people who want to run the TELEMAC-MASCARET SYSTEM physical components in an interactive mode with the help of the Python scripting language. The script corresponding to the application case of Section 4 is presented in Appendix B.

### 3.3 Summary

A single coherent Fortran structure has been deployed in TELEMAC-MASCARET in order to ensure interoperability of the system for each of its solvers. This is based on three main functionalities: instantiation, variable values and computation control. Firstly, instantiation associated with the processor communication function ensures the possibility of a simultaneous existence of several clearly separated and identifiable problems. Secondly, based on variable and computation control, memory access is permitted during simulation, allowing new services to be provided on demand. Finally, the technology allows dynamic compilation of each physical-based component model. All of these features enable the TELEMAC-MASCARET SYSTEM to operate with other existing or future codes without restricting access or implementation. It therefore becomes natural to drive these APIs using Python scripting language. As mentioned by Knox et al. [39], this scripting language offers numerical and scientific libraries and is already a recognized tool in environmental resource modeling.

Finally, the Python wrapping allows use of the TELEMAC-MASCARET SYSTEM for a wide range of studies, such as optimization, coupling and uncertainty quantification.

The Application Programming Interface (API) framework described in this work (Fortran APIs and its Python wrapper) is available for download and distributed in TELEMAC-MASCARET SYSTEM ([www.opentelemac.org](http://www.opentelemac.org)).

## 4 Gironde Estuary application case

The hydrodynamic model TELEMAC-2D was used to solve the shallow water equations (see Section 4.2) in the real case of the Gironde Estuary in southwestern France (Figure 6). This important large-scale estuary involves many

economic and environmental considerations. Several hydrodynamic and morphodynamic TELEMAC-2D studies of the estuary have been carried out for the last decade [35, 74]. In these studies, the importance of the calibration phase was systematically emphasized to achieve operational performance.

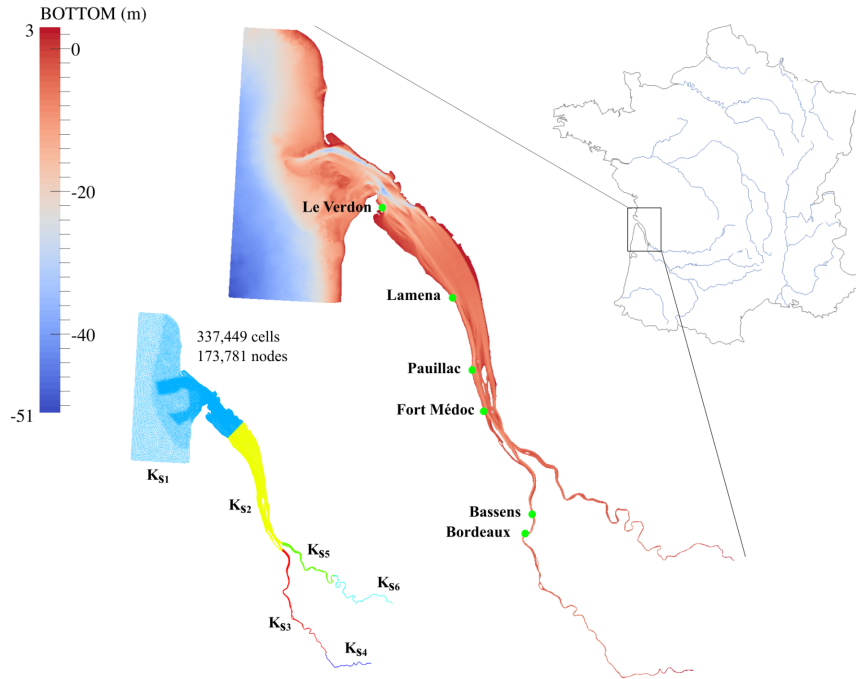


Figure 6: Gironde Estuary. Location of observation stations (●), bathymetry, friction zones ( $K_s$  denotes Strickler coefficient) and mesh size [76]

Model calibration aims at simulating a series of reference events by adjusting some uncertain physically based parameters in order to minimize deviation between measured and computed values of variables of interest. This process of parameter estimation, a subset of the so-called inverse problems, consists in evaluating the underlying input data of a problem from its solution. For free-surface flow hydraulics, parameters that are often unknown or difficult to assess include initial state, bathymetry, bed friction and model boundary conditions. In this work, the bathymetry and initial state are considered as known. In fact, a spin-up of 12 hours was used to generate realistic initial condition. Bathymetry uncertainty was not considered in this study and could constitute an approximation in particular on current speed [13]. A primary challenge of integrating bathymetry in the calibration process is the spatial structure characterization of its uncertainty in relation to channel morphology [45]. This was not studied here and will be the topic of a future study. Consequently, both bed roughness and model boundary conditions are considered in the calibration process. As mentioned by Williams and Esteves [75], the hydrodynamic bed friction is a primary calibration variable for all coastal and estuarine models. It is also essential for modeling other processes accurately such as sediment transport and wave attenuation. Concerning the model boundary conditions, only offshore boundary condition is considered in the calibration process as it is derived from a coarser-scale model and upstream boundary conditions is coming from measurement. The transfer of information between a large scale model and the boundaries of a more local model generally requires empirical adequacy: sea level could need to be corrected to account for seasonal variability (effect of thermal expansion, salinity variations, air pressure, etc.) in addition to long-term sea level rise resulting from climate change and differences in tidal amplitude could be attributable to meteorological effects (storm and surge (atmospheric and wave setup)) [36]. Although this paper focuses on friction and tidal parameters, all TELEMAC-2D variables can be changed by the API, as presented in Table A.1.

#### 4.1 Numerical configuration and available data

The Gironde is a navigable estuary in southwest France formed by the confluence of the Garonne and Dordogne rivers just downstream of the center of Bordeaux. The Gironde Estuary is the largest estuary in western Europe. The hydraulic model covers approximately 195 km between the fluvial upstream and the maritime downstream

boundary conditions, for an area of around 635 km<sup>2</sup>. The finite element mesh is composed of 173,781 nodes (Figure 6). The mesh size varies from 40 m within the area of interest, the navigation channel, to 750 m offshore (western and northern sectors of the model). As shown in Figure 6, six friction areas are considered in the hydraulic model.

The boundary condition along the marine border of the model was set using depth-averaged velocities and water levels from the Legos numerical model TUGO dataset (46 harmonic constants). Surge data, describing the difference between the tidal signal and the observed water level, are taken into account using a data file generated from the Hycom2D model of the SHOM [15]. Surface wind data are also considered to simulate the flow under windy conditions. Time-evolution flow discharge hydrographs are imposed upstream of the Gironde Estuary model on the Garonne and Dordogne rivers. The ability of the numerical model to reproduce complex physical processes such as sediment transport is presented in [58].

The free surface flow was continuously measured (every minute) at the Verdon, Richard, Lamena, Pauillac, Fort Médoc, Ambes, The Marquis, Bassens and Bordeaux locations (Figure 6). For this study, observation results are used over a 36 hour period from August 12 to 14, 2015.

## 4.2 Hydrodynamic model

The TELEMAC-2D code solves the 2D depth-averaged free surface flow equations, also known as shallow water equations (Eqs. 1-3).

$$\frac{\partial h}{\partial t} + \frac{\partial}{\partial x}(hu) + \frac{\partial}{\partial y}(hv) = 0 \quad (1)$$

$$\frac{\partial hu}{\partial t} + \frac{\partial}{\partial x}(huu) + \frac{\partial}{\partial y}(huv) = -gh\frac{\partial Z_s}{\partial x} + hF_x + \nabla \cdot (h\nu_e \nabla(u)) \quad (2)$$

$$\frac{\partial hv}{\partial t} + \frac{\partial}{\partial x}(huv) + \frac{\partial}{\partial y}(hvv) = -gh\frac{\partial Z_s}{\partial y} + hF_y + \nabla \cdot (h\nu_e \nabla(v)) \quad (3)$$

where  $x$  and  $y$  are the horizontal Cartesian coordinates,  $t$  the time,  $u$  and  $v$  the components of the depth-averaged velocity,  $h$  the water depth,  $\nu_e$  an effective diffusion representing depth-averaged turbulent viscosity  $\nu_t$  and dispersion,  $Z_s$  the free surface elevation,  $g$  the gravitational acceleration,  $F_x$  and  $F_y$  refer to the friction force (see Section 4.2.1).

TELEMAC-2D solves the previous equation system using the finite element method on a triangular element mesh. The main results at each node of the computational mesh are the water depth and the depth-averaged velocity components. TELEMAC-2D can take into account propagation of long waves, including non-linear effects, bed friction, effect of the Coriolis force, effects of meteorological phenomena (e.g., atmospheric pressure, rain or evaporation, and wind), turbulence, supercritical and subcritical flows, influence of horizontal temperature and salinity gradients on density and dry area in the computational domain, amongst other processes [34].

### 4.2.1 Friction coefficient

The friction term in the momentum part of the shallow water equations is treated in a semi-implicit form within TELEMAC-2D. The two components of friction force are given in Eq. 4.

$$\begin{cases} F_x = -\frac{u}{2h}C_f\sqrt{u^2 + v^2} \\ F_y = -\frac{v}{2h}C_f\sqrt{u^2 + v^2} \end{cases} \quad (4)$$

where  $C_f$  is a dimensionless friction coefficient.

The roughness coefficient often takes into account friction caused by walls on the fluid or other phenomena such as turbulence. Empirical or semi-empirical formulas are used for calculating  $C_f$  [52]. In the present work,  $C_f$  is given by the widely used Strickler formulas (Eq. 5). The Strickler coefficient ( $\text{m}^{1/3}\text{s}^{-1}$ ) is a calibration parameter of the modeling system, to be adjusted according to field data (e.g., usual measured water levels).

$$C_f = \frac{2g}{h^{1/3}K_s^2} \quad (5)$$

Generally, the friction coefficient is contained in an interval bounded by physical values, as it results from different contributions (e.g., skin friction, bed form dissipation, etc.). In this study, the Strickler bounds are taken as large as possible. In fact, the lower bound of the Strickler coefficient is set to 5 according to U.S.F.H.A [72] (rough bed surface) and the upper one is set to 115 (extreme smooth bed surface).

In the present API context, the friction coefficient is called `MODEL.CHESTR` and in the study corresponds to the Strickler coefficient  $K_{si}$  (where  $i$  denotes the considered friction area).

#### 4.2.2 Tidal amplification parameter

Tidal characteristics are imposed using a database of harmonic constituents to force the open boundary conditions. For each harmonic constituent, the water depth  $h$  and components of velocity  $u$  and  $v$  are calculated, at point  $P$  and time  $t$  by Eq. 6.

$$\begin{cases} f(P, t) = \sum_i f_i(P, t) \\ f_i(P, t) = f_i(t) A_{f_i}(P) \cos \left[ \frac{2\pi t}{T_i} - \varphi_{f_i}(P) + \varphi_i^0 + l_i(t) \right] \end{cases} \quad (6)$$

where  $f$  is either the water depth  $h$  or one of the components of velocity  $u$  or  $v$ ,  $i$  refers to the considered constituent,  $T_i$  is the period of the constituent,  $A_{f_i}$  is the amplitude of water depth or one of the horizontal components of velocity,  $\varphi_{f_i}$  is the phase,  $f_i(t)$  and  $l_i(t)$  are the nodal factors and  $\varphi_i^0$  is the phase at the original time of the simulation. The water level and velocities of each constituent are then submitted to obtain the water depths and velocities for the open boundary conditions (Eq. 7).

$$\begin{cases} h = \alpha \sum_i h_i - Z_b + Z_{ref} - \gamma \\ u = \beta \sum_i u_i \\ v = \beta \sum_i v_i \end{cases} \quad (7)$$

where  $Z_b$  is the bottom elevation and  $Z_{ref}$  the mean reference level. In Eq. 7, the tidal amplitude multiplier coefficients of tidal range and velocity, respectively  $\alpha$  and  $\beta$ , at boundary locations and the sea level correction  $\gamma$  are assumed to be the tidal calibration parameters [61].

The sea level correction parameter is assumed, following expert opinion, to be contained in an interval of plus or minus a meter. The interval of the weighting coefficients of tidal range and velocity is set to  $[0.8, 1.2]$  corresponding to a 20% margin of the initial value.

In the API framework, the sea level correction variable  $\gamma$  is designated as `MODEL.SEALEVEL`, and the weighting coefficients of tidal range  $\alpha$  and velocity  $\beta$  are denoted `MODEL.TIDALRANGE` and `MODEL.VELOCITYRANGE`, respectively.

### 4.3 Summary of parameter range variation

All the model input parameters and their associated probability distribution are summarized in Table 1.

Variable name	Nature	Variation interval	Probability distribution
Fiction coefficient per area ( $m^{1/3}s^{-1}$ )	Real scalar	[5., 115.]	Uniform
Sea level correction parameter ( $m$ )	Real scalar	[-1., 1.]	Uniform
Tidal range coefficient (-)	Real scalar	[0.8, 1.2]	Uniform
Velocity range coefficient (-)	Real scalar	[0.8, 1.2]	Uniform

Table 1: Input parameters and associated probabilistic model for the Gironde Estuary application

## 5 Calibration problem

The inverse problem of calibration can be understood as the computation of the *posterior* distribution  $\pi(\mathbf{X}|\mathbf{Y})$  (“Model Calibration Algorithm step”, Figure 7, as presented in appendix C), where model parameters constitute the p-components of the parameter control vector  $\mathbf{X} = (X_1, \dots, X_p)' \in \mathbb{R}^p$  composed of independent variables defined on some probability space  $(\Omega, \mathcal{A}, \mathbb{P})$  (with  $\Omega$  as a sample space,  $\mathcal{A}$  the  $\sigma$ -algebra of events and  $\mathbb{P}$  the probability measure),  $\mathbf{Y} \in \mathbb{R}^m$  is the observation vector, also defined on a probability space, around the unknown parameter vector  $\mathbf{X}$  and  $\mathbb{R}^m$  is the observation space defined by Eq. 8.

$$\mathbf{Y} = G(\mathbf{X}) + \epsilon_m \quad (8)$$

where  $G: \mathbb{R}^p \rightarrow \mathbb{R}^m$  is a vector-valued function of vector  $\mathbf{X}$  and  $\epsilon_m \in \mathbb{R}^m$  is an observable measurement noise such as  $\mathbb{E}(\epsilon_m) = 0$  and  $\mathcal{R} = cov(\epsilon_m) = \mathbb{E}(\epsilon_m \epsilon_m') \in \mathbb{R}^{m \times m}$  and identified as a multivariate normal distribution,  $\epsilon_m \sim \mathbb{N}(0, \mathcal{R})$ . As a reminder, in the Gironde Estuary application case, the parameter control vector  $\mathbf{X}$  is composed

of the friction and tidal parameters and the observation parameter  $\mathbf{Y}$  is composed of water levels measured every minute at an observation station. Note that the operator  $G$  enabling the passage from the parameter space (where vector  $\mathbf{X}$  lives) to the observation space (where vector  $\mathbf{Y}$  lives) represents a call to the API study workflow defined in Figure 4. As TELEMAC-2D is a discrete time dynamic model, the output  $\mathbf{O}$  is composed of scalar output (water elevation at observation stations interpolated from TELEMAC-2D computational nodes given at  $t \in \mathbb{T} = [1, \dots, T]$ ).

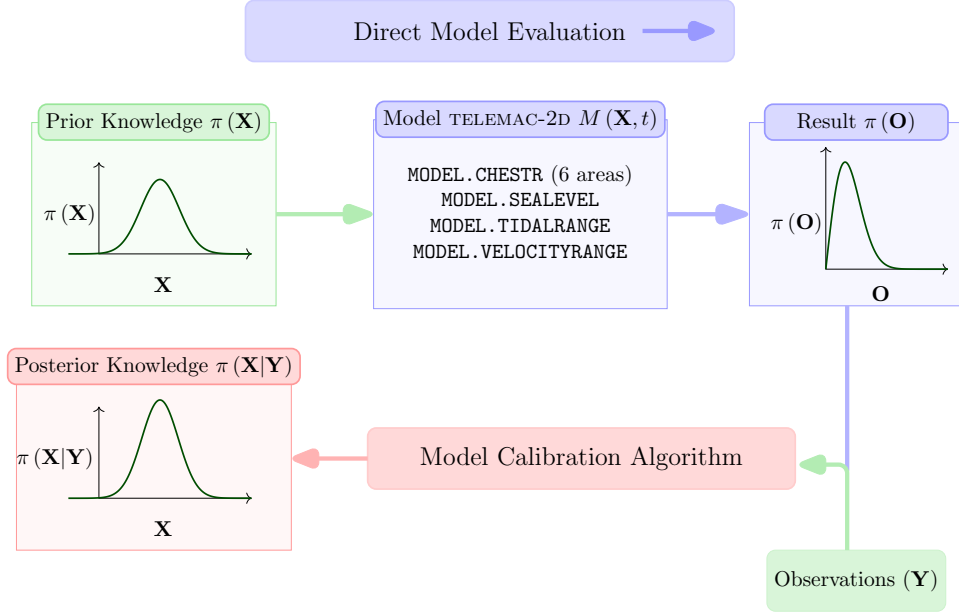


Figure 7: Model calibration under uncertainty using observations framework

Finally, the maximum *a posteriori* is equivalent to the optimal search for the control vector minimizing the objective function  $J(\mathbf{X}) = \frac{1}{2} [\mathbf{X} - \mathbf{X}_0] \mathcal{B}^{-1} [\mathbf{X} - \mathbf{X}_0]' + \frac{1}{2} [\mathbf{Y} - G(\mathbf{X})] \mathcal{R}^{-1} [\mathbf{Y} - G(\mathbf{X})]'$  (where  $\mathcal{B}$  is the *prior* covariance matrix). This is known as the traditional variational data assimilation cost function, called 3D-VAR [12].

Mathematical methods can be used to solve optimization problems. The former can vary significantly according to the form of the cost function (convex, quadratic, nonlinear, etc.), its regularity, and the dimension of the space. Many deterministic optimization methods are known as gradient descent methods, among which is the Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton method used in this work [50]. Using this constrained optimization method makes it possible to impose boundaries during the research process of the model parameters, guaranteeing their physical values. Because the inverse problem is often ill-posed and unstable with available data corresponding to more than one solution, small changes in model results can lead to very different estimates for the input (calibration) parameters. These problems are related to the issue of “parameter identifiability” [54]. Still, the chosen optimization method involves computing the adjoint of the observation operator  $G$  (or the partial derivatives of the operator with respect to its input parameters). In this work, the partial derivatives are approximated using a classical finite difference method. This is a simple solution, numerically sensitive and computationally time-consuming, but the observation operator can be written to make use of parallel computing, thus providing a fast automatic calibration algorithm relevant to real-world applications.

As the inverse problem is defined, a relevant question is what are the effects of the modeling calibration parameters  $\mathbf{X}$  on the simulated state variables  $O(t) = M(\mathbf{X}, t)$  which are compared to the observations? This question can be addressed by multivariate sensitivity analysis.

## 5.1 Multivariate Sensitivity Analysis

The sensitivity analysis aims at quantifying the relative importance of each input model parameter. The variance-based methods aim at decomposing the variance of the output to quantify the participation of variables considered as independent. Conventional approaches to Global Sensitivity Analysis (GSA) compute sensitivity indices called Sobol’ indices and assume that the output variable is scalar. The definition of Sobol’ indices is a result of the ANOVA (ANalysis Of VAriance) variance decomposition [65, 67]. However, as reported by many authors [11, 25, 42], when this approach is applied to each variable of a functional output, it leads to a high degree of redundancy because of the

strong relationship between responses from one time step to another. Moreover, it also misses important features of the output dynamic because many features cannot be efficiently detected through single-time measurements. Thus, the methodology used in this work to compute GSA is a multivariate sensitivity analysis computing ‘‘Generalized Sensitivity Indices’’ [42] which are similar to ‘‘Aggregated Sensitivity Indices’’ [23], as demonstrated in [25, 47].

### 5.1.1 Analysis of Covariance

To perform sensitivity analysis, some parameters and input variables are identified as uncertain, while the others are fixed at given nominal values. TELEMAC-2D denoted  $M$  thereafter ensures the relationship between the vector of model inputs  $\mathbf{X}$  and the output quantity of interest  $O(t)$  whereby  $O(t) = M(\mathbf{X}, t)$  (‘‘Direct Model Evaluation’’, Figure 7). Since TELEMAC-2D is a discrete time dynamic model, for one interest point, the model output  $\mathbf{O}$  is composed of scalar output given at  $t \in \mathbb{T} = [1, \dots, T]$  such as  $\mathbf{O} = (O_1, \dots, O_T)$  (with  $O_i = O(t_i) = M(\mathbf{X}, t_i)$ ). The covariance of the model vector output  $\mathbf{O} = (O_1, \dots, O_T)$  can be partitioned into summands of increasing dimensions (single, pairs, triplets and so on) of input variables, as shown in Eq. 9.

$$Cov(O_1, \dots, O_T) = \sum_{i=1}^p Cov_i(O_1, \dots, O_T) + \sum_{1 \leq i < j \leq p} Cov_{i,j}(O_1, \dots, O_T) + \dots + Cov_{1, \dots, p}(O_1, \dots, O_T) \quad (9)$$

From this equation, some sensitivity indices of multivariate case can be defined by projection of the covariance matrix into scalar using the trace operator. Gamboa et al. [23] present the following aggregated indices:

$$\begin{cases} S1_i(O_1, \dots, O_T) = \frac{Tr[Cov_i(O_1, \dots, O_T)]}{Tr[Cov(O_1, \dots, O_T)]} \\ ST_i(O_1, \dots, O_T) = \frac{Tr[Cov_i(O_1, \dots, O_T)] + \sum_{1 \leq i < j \leq p} Tr[Cov_{i,j}(O_1, \dots, O_T)] + \dots + Tr[Cov_{1, \dots, p}(O_1, \dots, O_T)]}{Tr[Cov(O_1, \dots, O_T)]} \end{cases} \quad (10)$$

### 5.1.2 Computation of the multivariate sensitivity indices

To handle dynamic system behavior under parameter uncertainty, a sample set of configurations is generated using Latin Hypercube Sampling (LHS). In the sampling procedure, the parameter uncertainties are taken from uniform distribution whose limits are defined by the parameter bounds (see Section 4.3). For a fixed number of desired configurations  $n$ ,  $n^p$  LHSs are feasible. It is therefore possible to select the plan optimizing a uniformity criterion (discrepancy, ...) or a statistical criterion (entropy, ...). In this study, the uniformity criterion relying on  $L^2$ -discrepancies was considered and optimized, as described in [18]. In fact, according to these authors, the  $L^2$ -discrepancy optimized LHS have shown some good properties to 2D projections in high dimension. The full set of output dynamics over the complete Latin Hypercube Sampling gives a matrix of size  $n \times T$ .

$$\mathcal{O} = (\mathbf{O}_1, \dots, \mathbf{O}_T) = \begin{pmatrix} o_{1,1} & o_{1,2} & \dots & o_{1,T} \\ o_{2,1} & o_{2,2} & \dots & o_{2,T} \\ \vdots & \vdots & \ddots & \vdots \\ o_{n,1} & o_{n,2} & \dots & o_{n,T} \end{pmatrix} \quad (11)$$

Each row of the matrix  $\mathcal{O}$  corresponds to a temporal trajectory for a fixed scenario of uncertain input parameter. As proposed by Lamboni et al. [42], a Principal Component Analysis (PCA) can be carried out on the matrix  $\mathcal{O}_c$  which is the matrix obtained by centering each column of  $\mathcal{O}$ . The PCA decomposition consists in a small number determination of variable  $(\mathbf{H}_1, \dots, \mathbf{H}_T)$  to transcribe the maximum information, in the sense of variance, contained in the initial variables  $\mathcal{O}_c = (\mathbf{O}_1^c, \dots, \mathbf{O}_T^c)$  (with  $\mathbf{O}_i^c = \mathbf{O}_i - \mathbb{E}(\mathbf{O}_i)$ , where  $\mathbb{E}(\cdot)$  is the mean operator and  $\mathbf{O}_i$  the discrete form of  $O_i$ ) in order to remove redundancies with the least possible loss of information. The variables  $\mathbf{H}_i$  are constructed to be uncorrelated from each other, based on an iterative process that adds complementary information not contained in the previously constructed variables. This methodology can be summarized by resolution of the system described by Eq. 12.

$$\begin{cases} \mathbf{H}_i = \mathcal{O}_c \mathbf{w}_i, i = 1, \dots, T \\ \text{with } Var(\mathbf{H}_1) \geq Var(\mathbf{H}_2) \dots Var(\mathbf{H}_T) > 0 \\ \text{under the constraint } \|\mathbf{w}_i\|^2 = 1 \end{cases} \quad (12)$$

In practice,  $(\mathbf{w}_1, \dots, \mathbf{w}_T)$  is a set of normalised and mutually orthogonal eigenvectors associated with eigenvalues  $(\lambda_1, \dots, \lambda_T)$  of the covariance matrix  $\Sigma = \frac{1}{n} \mathbb{E}[\mathcal{O}_c^T \mathcal{O}_c]$ . Once the eigenvectors resulting from the diagonalization of  $\Sigma$  are determined, the matrix  $\mathcal{O}_c$  can be decomposed as written in Eq. 13.

$$\mathcal{O}_c = \sum_{i=1}^T \mathbf{H}_i \mathbf{w}'_i \quad (13)$$

$\mathbf{H}_i$  variables are uncorrelated, and the principal component matrix  $\mathcal{H} = (\mathbf{H}_1, \dots, \mathbf{H}_T)$  has the same total inertia as  $\mathcal{O}$ , but is mostly concentrated in its first columns. Thus, the expression of covariance decomposition can be expressed as a sum of variance of  $\mathbf{H}_i$  on which the Sobol' decomposition into summands of increasing dimension can be carried out (Eq. 14).

$$\begin{cases} Tr [Cov(O_1, \dots, O_m)] = \sum_{k=1}^T \left( \sum_{i=1}^p V_{(i),k} + \sum_{1 \leq i < j \leq p} V_{(i,j),k} + \dots + V_{(1,\dots,p),k} \right) \\ \text{with } V_{(\epsilon),k} = Var [\mathbb{E}(\mathbf{H}_k | \mathbf{X}_\epsilon)] - \sum_{w, w \in \epsilon} V_{w,k} \end{cases} \quad (14)$$

where  $Var [\mathbb{E}(\mathbf{H}_k | \mathbf{X}_\epsilon)]$  is the variance of the conditional expectation of  $\mathbf{H}_k$  to the input variable  $\mathbf{X}_\epsilon$

Thus, based on an identification process between the formulation that is described in Eq. 10 and Eq. 14, the expression of multivariate sensitivity indices can be given, as described in Eq. 15.

$$\begin{cases} S1_i(O_1, \dots, O_m) = \frac{\sum_{k=1}^T S_{(i),k}^{ob} * \lambda_k}{\sum_{k=1}^T \lambda_k} \\ ST_i(O_1, \dots, O_m) = \frac{\sum_{k=1}^T S^{ob} T_{(i),k} * \lambda_k}{\sum_{k=1}^T \lambda_k} \end{cases} \quad (15)$$

where  $S_{(\epsilon),k}^{ob}$  denotes Sobol' sensitivity indices such as  $S_{(\epsilon),k}^{ob} = \frac{V_{(\epsilon),k}}{Var(\mathbf{H}_k)}$  and  $S^{ob} T_{(i),k}$  represents the total Sobol' sensitivity index such as  $S^{ob} T_{(i),k} = \frac{\sum_{i \in \epsilon} V_{(\epsilon),k}}{Var(\mathbf{H}_k)}$ .

As expressed by Eq. 15, the multivariate sensitivity analysis indices involve Sobol' index computation. The next paragraph is dedicated to the computation process of this mathematical expression based on a combined PCA and Polynomial Chaos Expansion (PCE) emulator.

### 5.1.3 Reduced Order Model PCA-PCE

As already mentioned, Python is a language that can be used in many scientific contexts and can be adapted to any type of use based on dedicated libraries. The multivariate sensitivity analysis explained in the previous section was carried out based on an open-source library for uncertainty treatment named "OpenTURNS", standing for "Open source initiative to Treat Uncertainties, Risks'N Statistics" ([www.openturns.org](http://www.openturns.org)) [6]. A Design of Experiment (DoE) of size 1000 ( $n = 1000$ ) is constructed for the PCA-PCE learning step. This number of model evaluation was determined based on a convergence study carried on the sensitivity analysis (see Section 5.1.4). The MPI library was used for launching and managing the TELEMAC-2D solver computations. At the end of the calculation, a TELEMAC-2D result file was composed of 288 time records corresponding to a physical variable record every 10 minutes ( $T = 288$ ). In the present study, the discrete version of the PCA (singular value decomposition (SVD)) of the matrix  $\mathcal{O}_c$  is considered. This means that  $\mathbf{H}_i$  considered is expressed as  $\mathbf{H}_i = \sqrt{n \lambda_i} \mathbf{A}_i$  with  $\mathbf{A}_i$  orthonormal. In order to have an efficient sensitivity index computation in terms of computational cost, the  $k$  main eigenvectors which explained more than 99.95% of the original variance are considered ( $k \sim 10 \ll T$ ). It is important to note that, owing to truncation, the equivalence between "Aggregated Sensitivity Indices" [23] and "Generalized Sensitivity indices" [42] no longer exists. The Generalized Sensitivity indices represent an approximation of the Aggregated Sensitivity Indices. For each eigenvector, a learning sample of realizations of the decomposition coefficient  $\mathbf{H}_i$  is available. As proposed by Garcia-Cabrejo and Valocchi [25], a polynomial Chaos Expansion can be used as learning step of each decomposition coefficient. Thus, the Sobol' indices needed in Eq. 15 are obtained based on post-treatment of each mode PCE as proposed in [70]. Here, the construction of the PCE is carried out based on Least Angle Regression Stagewise method (LARS) in order to construct an adaptive sparse PCE. In this approach, a collection of possible PCEs, ordered by sparsity, is provided and an optimum can be chosen with an accuracy estimate. It was performed in this study using corrected leave-one-out error [7].

At this stage, a reduced order PCA-PCE model is produced for each observation station of the Gironde Estuary (Figure 6). The validation of these emulators was carried out based on a validation sample generated with the same procedure (optimized LHS) as for the learning step. This sample was composed of 100 validation points ( $n_{val} = 100$ ) and allows, after the TELEMAC-2D unit computing, a temporal predictivity criterion defined by Eq. 16.

$$Q^2(t) = 1 - \frac{\sum_{j=1}^{n_{val}} [\hat{M}_k(\mathbf{X}_j, t) - M(\mathbf{X}_j, t)]^2}{\sum_{j=1}^{n_{val}} [\hat{M}_k(\mathbf{X}_j, t) - \frac{1}{n_{val}} \sum_{i=1}^{n_{val}} M(\mathbf{X}_j, t)]^2} \quad (16)$$

where  $\hat{M}_k(\mathbf{X}_j, t)$  is the estimated PCA-PCE evolution.

Figure 8 shows (from top to bottom) a comparison on a validation configuration between an estimated result from the reduced order PCA-PCE model and the TELEMAC-2D computation. The difference between these two water elevations is then given, and finally, the predictivity criterion is presented.

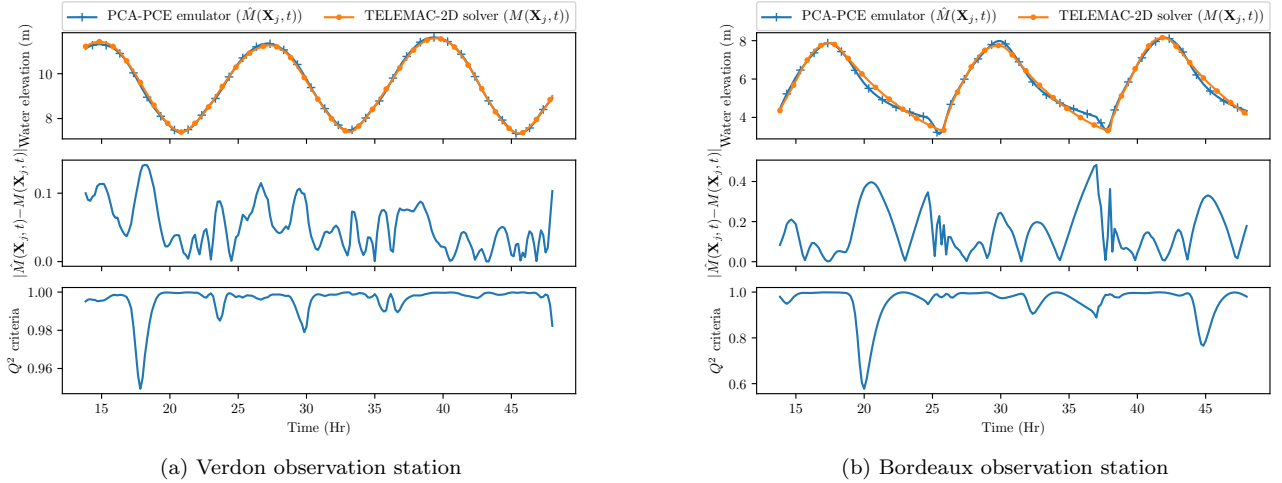


Figure 8: Validation of the reduced order PCA-PCE model

The reduced order model is validated according to the performance displayed in Figure 8. Indeed, the results provided by the emulator and the TELEMAC-2D solver are fairly close. The deviation between these results varies respectively from 10 to 20 centimeters at Verdon and Bordeaux observation stations. This is confirmed by the average predictivity criterion close to 1 for Verdon and Bordeaux during the study period. A coefficient close to 1 shows a good fit between the validation database and the result estimated by the reduced order model. However, a less satisfactory performance of  $Q^2$  criteria can be noticed at the Bordeaux observation station, compared to the Verdon station. The confluence of the Dordogne and Garonne rivers occurs just downstream from Bordeaux and the influence of the hydrological forcing of these two rivers is not studied here. Consequently, the upstream fluvial discharges are not considered as parameters in construction of the PCA-PCE emulator. The missing interactions can explain the observed performance gap in the predictivity criteria.

To conclude, the reduced PCA-PCE model shows good agreement with the TELEMAC-2D computation. The high score of predictivity criteria allows confidence in the sensitivity results post-treated from the reduced order model.

#### 5.1.4 Multivariate sensitivity analysis results

A reduced order model based on PCA-PCE is used to identify influential input parameters. As reported by Pianosi et al. [62], when applying sampling-based sensitivity analysis, sensitivity indices are not computed exactly but they are approximated from the available samples. The robustness and convergence of such sensitivity estimates should therefore be assessed. Thus, three elements are of interest in the following analysis: (i) the convergence of sensitivity indices, (ii) the robustness of the estimates and (iii) relevant sensitivity analysis visualization.

##### Convergence:

To handle this issue, the convergence rates of Generalized Sensitivity Indices are assessed as the sample size increases. Figure 9 presents the evolution of sensitivity indices obtained at Bordeaux observation station.

As displayed on Figure 9, the number of samples needed to reach stable sensitivity estimates can vary from one input factor to another. However, from a sample size of 1000, the sensitivity indices are stabilized. Thus, this number of model evaluation is considered in the following investigations.



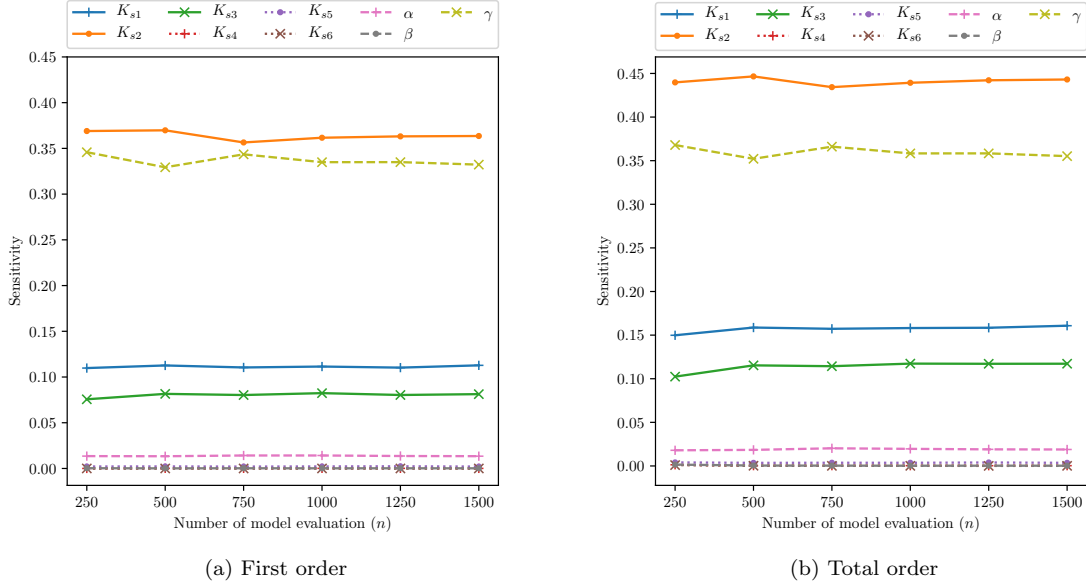


Figure 9: Generalized Sensitivity Indices estimated using an increasing sample size at Bordeaux observation station

### Robustness:

A robustness analysis is carried out in order to evaluate the sensitivity of the estimates to the Design of Experiment. First, the robustness of the sensitivity indices is analysed through confidence intervals. They are obtained by repeating 30 times the methodology described above (see Section 5.1.3). Table 2 presents the result of these confidence intervals, which have finally required  $N = 30 \times n = 3 \times 10^4$  model calls.

Inputs	Verdon		Bordeaux	
	First Order	Total Order	First Order	Total Order
$K_{s1}$	[0.123, 0.131]	[0.131, 0.139]	[0.108, 0.114]	[0.153, 0.162]
$K_{s2}$	[0.00969, 0.0103]	[0.0148, 0.0157]	[0.352, 0.367]	[0.43, 0.448]
$K_{s3}$	$[1.03, 2.98] \times 10^{-5}$	$[4.37, 7.93] \times 10^{-5}$	[0.0790, 0.0827]	[0.115, 0.119]
$K_{s4}$	$[0., 1.79 \times 10^{-6}]$	$[0.275, 2.84] \times 10^{-5}$	$[0.561, 1.59] \times 10^{-4}$	$[2.58, 4.21] \times 10^{-4}$
$K_{s5}$	$[1.32, 3.12] \times 10^{-5}$	$[4.62, 7.35] \times 10^{-5}$	$[1.96, 2.40] \times 10^{-3}$	$[3.46, 4.01] \times 10^{-3}$
$K_{s6}$	$[0., 1.46 \times 10^{-7}]$	$[0.241, 1.41] \times 10^{-5}$	$[0., 8.49] \times 10^{-5}$	$[0.772, 2.17] \times 10^{-4}$
$\alpha$	[0.0452, 0.0481]	[0.0479, 0.0509]	[0.0131, 0.0145]	[0.0184, 0.0203]
$\beta$	$[4.35, 5.15] \times 10^{-4}$	$[4.55, 5.40] \times 10^{-4}$	$[1.25, 2.37] \times 10^{-4}$	$[2.45, 4.16] \times 10^{-4}$
$\gamma$	[0.802, 0.813]	[0.803, 0.813]	[0.326, 0.352]	[0.350, 0.374]

Table 2: Min-max confidence intervals of Generalized Sensitivity Indices (GSI) obtained from 30 repetitions of the computational process at Bordeaux and Verdon observation stations

As shown in Table 2, the bounds of the min-max confidence intervals are relatively close and demonstrate the capacity of the reduced order PCA-PCE model, constructed from a  $L^2$ -discrepancy optimized LHS sample, to produce precise estimates. An alternative sampling method based on low discrepancy sequence of Sobol is also considered. The obtained results are presented in Table D.2. Most of the time, the sensitivity indices obtained with the low discrepancy sequence are included in the min-max confidence intervals obtained from  $L^2$ -discrepancy optimized LHS sample. When out of range, values stay close to the bounds. Consequently, the estimates can be considered satisfactory in terms of accuracy and robustness.

### Results visualization:

Visualization can significantly improve the interpretation of the sensitivity analysis results. For this purpose, radial convergence diagrams, also called chord graphs, are used to simultaneously visualize some computed Generalized Sensitivity Indices (Figure 10). These diagrams plot the main effect of each input variable (first-order multivariate sensitivity analysis proportional to the size of the inner circle); its total influence, including interactions (proportional to the size of the outer circle), existence and extent of second-order effects (second-order multivariate sensitivity indices proportional to width) [10].

Figure 10 displays sensitivity analysis estimates from one  $L^2$ -discrepancy optimized LHS sample used to compute the min-max confidence intervals presented in Table 2. As shown in the Figure, the output variance at each observation station is mainly explained by the upstream friction coefficients (areas one and two of the Gironde Estuary model) and the sea water level. As expected, downstream from the estuary (Bordeaux station), the friction coefficient of area 3 containing the Bordeaux station has a more significant influence. It is noticeable, from the visualization of variable interactions, that even if the tidal range variable contribution is not considerable, its interaction with other parameters should not be neglected. This enhances the utility of simultaneously visualized interactions with main and total effects of sensitivity analysis.

To conclude, the calibration problem initially composed of nine input variables can be reduced to five (the first three being friction area, sea level correction and tidal range coefficients) after considering the results of multivariate sensitivity analysis.

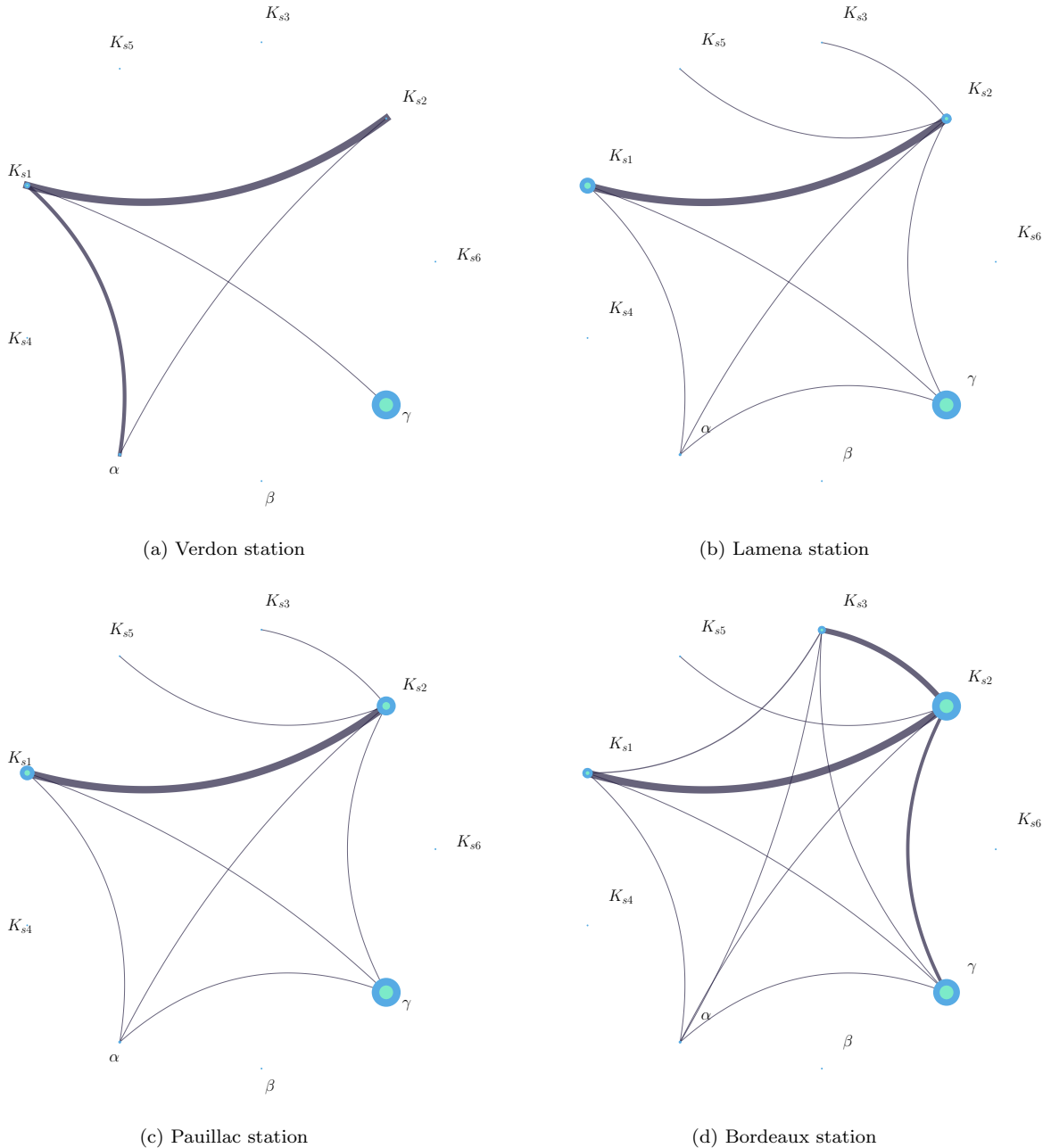


Figure 10: Chord graph of Generalized Sensitivity analysis

## 5.2 Calibration Results

The calibration algorithm (Figure 7) is performed by coupling TELEMAC-2D and the data assimilation library ADAO in Python context through the component TelApy of the TELEMAC-MASCARET SYSTEM. ADAO, A module for Data Assimilation and Optimization, provides modular data assimilation and optimization features in Python (<https://pypi.org/project/adao>) [3]. It can be coupled with other modules or external codes while providing a number of standard and advanced data assimilation or optimization methods. The ADAO library also covers a wide variety of practical applications, from real engineering to experimental methodologies. Its architecture and numerical scalability adapt to the field of application. As shown by Eq. 18, the optimal search for the control vector  $\mathbf{X}$  takes a minimization form of an objective or cost function (given by the expression inside the exponential term) which must satisfy the background error statistics (prior term) and the equivalent observation error (likelihood term). This minimization process, equivalent to the maximum *a posteriori* search, is carried out using the 3D-VAR algorithm. The control parameter is composed of the five most influential variables identified by the multivariate sensitivity analysis performed previously. The initial guess  $\mathbf{X}_0$  is set to random values inside the constrained search space ( $\mathbf{X}_0 = (K_{s1} = 73.76, K_{s2} = 83.62, K_{s3} = 83.62, \alpha = 0.9729, \gamma = 0.8611)'$ ). The observation vector  $\mathbf{Y}$  is the free surface flow evolution extracted every 60 seconds at the Verdon, Lamena, Pauillac, Fort Médoc, Bassens and Bordeaux locations from noon August 12 to midnight August 14, 2015. The chosen optimization method involves computing the partial derivatives of the observation operator  $G$  with respect to  $\mathbf{X}$ , a classical finite differences method with a differential increment set to  $10^{-4}$ . The error background and observation covariance matrices respectively identified by  $\mathcal{B}$  and  $\mathcal{R}$  are token diagonals, meaning they have no error correlations. A small variance value for  $\mathcal{R}$ , justifying great confidence in the observation value, is considered, such as  $\sigma_m^2 = 0.1 * \mathbf{Y}$ . On the contrary, little confidence is given to a prior part such as  $\sigma_b^2 = 10. * \mathbf{X}_0$ . As shown by Figures 11-12, the automatic calibration algorithm finds an optimal solution in about 30 iterations with the following set of parameters  $\mathbf{X}_{MAP} = (K_{s1} = 47.99, K_{s2} = 59.63, K_{s3} = 67.485, \alpha = 0.9114, \gamma = 0.5344)'$ . A rapid decrease in cost function can be observed for the first algorithm iterations.

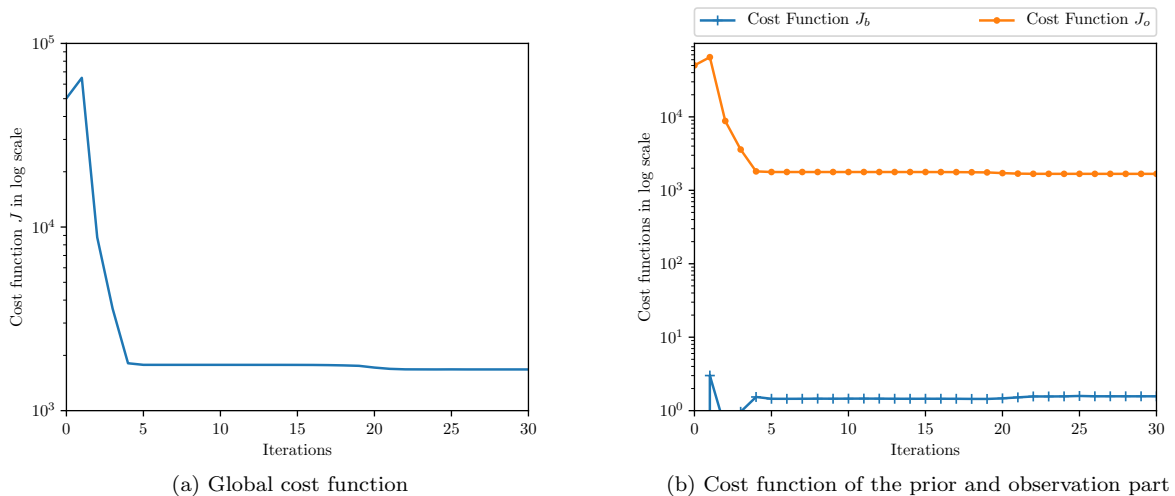


Figure 11: Value of cost functions according to number of algorithmic calibration iterations

The cost function curve behavior is smooth until iteration number 20, where a slight decrease can be observed. As expected, a similar behavior is observed for the parameter to be calibrated.

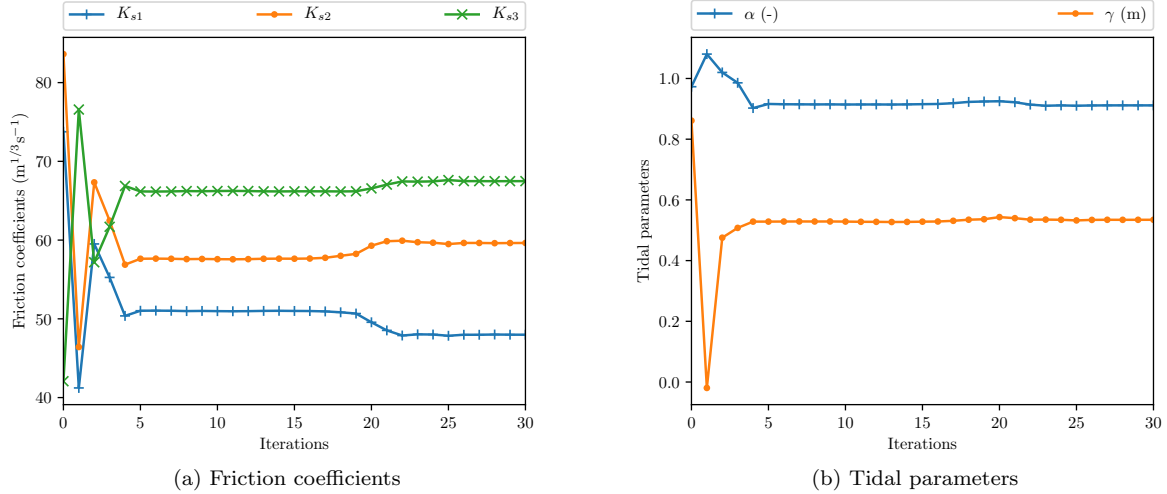


Figure 12: Value of the parameters according to number of algorithmic calibration iterations

Figure 13 displays the results of automatic calibration over the computation period. As expected, the water surface profiles, calculated from the calibrated parameter configuration, are much closer to measurements than the ones computed from the background knowledge parameters. The final results emphasize the efficiency of the automatic calibration tool in the framework of a real configuration. For most of the studied period, the difference between observations and the calibrated configuration is less than 20 cm (about 5% of the tidal range) at the Verdon and Bordeaux stations. However, at the Bordeaux observation station, the calibrated configuration presents some error peaks at low tide. This phenomenon can be induced by a greater effect of the fluvial part of the estuary, which is not well represented in the model, where discharge is expressed as an hourly average.

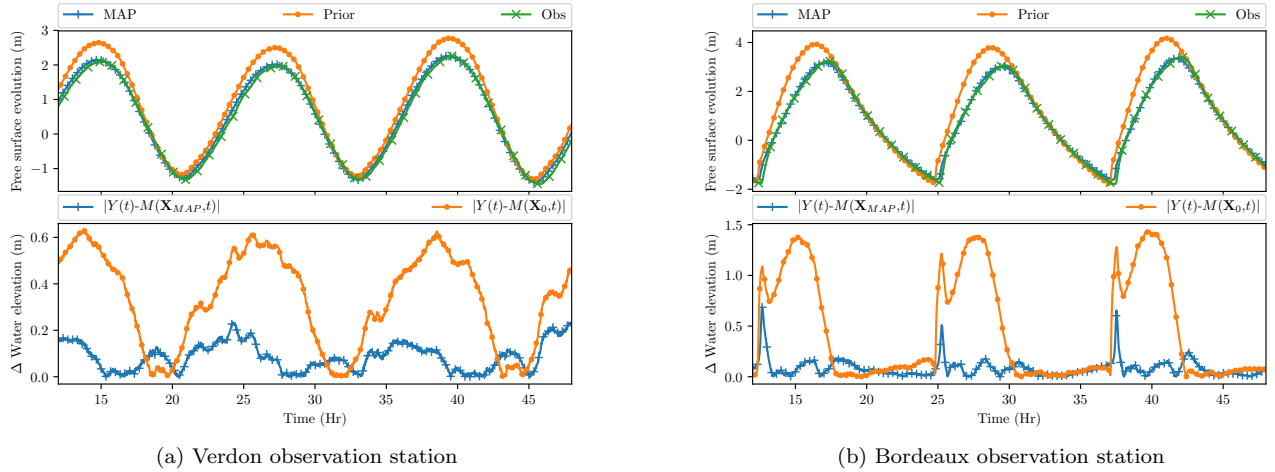


Figure 13: Comparison of the water depth evolution with and without calibration

## 6 Discussion

The concept of interoperability is a generic solution for gathering and exchanging information from various multidisciplinary knowledge. The application, presented in this article, is a case of 2D hydrodynamics and is not representative of all the possibilities offered by the interoperability of TELEMAC-MASCARET SYSTEM. For example, the estuarine sediment transport could be taken into account in order to better model and understand the evolution

of the bed with the API of the GAIA module. More generally the approach presented here can be easily applied to different geoscience problems where TELEMAC-MASCARET is relevant. All APIs of the different modules are freely available as they are part of the TELEMAC-MASCARET SYSTEM.

Applying interoperability criteria on an old and large code is not an easy task because of the transformation effort it requires. For the TELEMAC-MASCARET SYSTEM the transformation towards the concept of interoperability required several years of development. A recommendation is to take into account these criteria at the early stage of development considering the fact that the code will probably have to interact with the outside world. The implementation of APIs makes it possible to extend the scope of the software by facilitating the use for different types of applications. For instance, based on fluid exchange of information, APIs can be used to couple the TELEMAC-MASCARET SYSTEM with a Computational Fluid Dynamics (CFD)-type software (for example Code\_Saturne [2]; [www.code-saturne.org](http://www.code-saturne.org)) to take into account atmospheric and groundwater flows.

The development of standards for publishing interoperable softwares in forms suitable for community interactions remains a major issue [43]. The present work allows the TELEMAC-MASCARET SYSTEM to be integrated in different environments. However, the wrapper for each specific environment must be maintained. A lean standard has been proposed with the OpenMI environment [29]. The integration of the TELEMAC-MASCARET SYSTEM in this environment could constitute an outlook to this work. Since Python is really easy to pick up and learn, a wrapper for this language is distributed in the TELEMAC-MASCARET SYSTEM official version (TelAPy). This choice aims to favorise the dissemination of the hydro-informatic system as an environmental modeling tool. However, the system compilation is a cumbersome process which can slow down its dissemination. To overcome this issue, and based on the Python Package Index (PyPI, <https://pypi.org>), a compiled version of TelAPy might be envisaged to provide all dynamic libraries needed to run TELEMAC-MASCARET SYSTEM APIs.

To demonstrate that TelAPy is a functional system, an example of hydraulic model calibration is presented. This case deals with a series of reference events by adjusting some uncertain physically based parameters until the comparison with observations achieves sufficient accuracy. If performed manually, the model calibration is time-consuming. Fortunately, the process can be largely automated to significantly reduce human workload, as shown in this paper.

A reduced order model based on PCA-PCE is used to identify influential input parameters. This emulator has been created and validated on the basis of learning and validation solver computations. A major issue arising from this methodology concerns the optimization of the number of computational runs needed for sufficient results in the sensitivity analysis. Approximation with the surrogate models must become sufficiently accurate with just limited data available for the learning step. There have been some recent advances in this area, based on adaptive sampling [69].

After most influential parameters have been identified, they are then calibrated using a data-driven technique. The chosen algorithm is based on a minimization process requiring derivatives of the TELEMAC-2D solver with respect to the parameters to be calibrated. Several options exist to obtain the derivatives. The Finite Difference method, used in this work, is easily implemented but returns approximate derivatives whose poor accuracy can degrade the performance of the complete application. A much better option is to create a new program that computes the exact analytical derivatives of the model. Algorithmic Differentiation [30] is a way of automating creation of the derivative program, thus providing accurate derivatives for a minimal development effort. Gradient-based methods are very useful to find local extreme values within a reasonable time, but they cannot pretend to find a global solution in the search domain. On the other hand, metaheuristic optimization algorithms are useful in finding the area of a global solution (minimum or maximum), but the convergence becomes much slower due to the large number of required simulations. One way to improve the efficiency of the calibration would be to combine a derivative-free algorithm (like a metaheuristic) with a gradient-based method (like BFGS) to obtain better solutions within a reasonable time (hybridization).

## 7 Conclusion

The Application Programming Interface (API) of the open source TELEMAC-MASCARET SYSTEM was developed to convert a heterogeneous set of open-source, user-contributed models into a suite of plug-and-play modeling components that can be reused in many different contexts. The APIs provide a user-friendly development framework that can be easily understood, allowing seamless integration of base codes, and do not invalidate existing institutional software development practices. The API development framework does not compel the use of, or supply, specific environmental modeling standards, as its services standardize at a more basic level of internal communication. This gives researchers and model developers more freedom to customize services for the problems they are facing.

The TELEMAC-MASCARET SYSTEM API seeks to assist the hydrodynamic community in resolving new challenges,

such as uncertainties, real-time data assimilation, and multi-physics simulations. Many model developers have limited skills in software development and architecture. Recognizing this and seeking to promote the widest possible adoption of the TELEMAC-MASCARET SYSTEM API by the user community, a scripting feature of modeling is privileged with the Python language, as recommended by Knox et al. [39]. This forms a new module of the system called TelApy. Work performed with Python can be transposed to other scripting languages such as R and Julia. Python is a language that can be used in many contexts and adapted to any type of program based on dedicated libraries. However, it is particularly used to automate tedious tasks, such as the calibration process of a numerical model, saving a significant amount of time in realizing a project. Use of Python is widespread in the scientific community, and it has many libraries optimized and intended for numerical computation.

This flexibility was demonstrated in the Gironde Estuary case. First, a sensitivity analysis was carried out to identify sensitive parameters to calibrate. The most influential variables on water depth variation were the upstream friction coefficients ( $K_{s1}$ ,  $K_{s2}$ ,  $K_{s3}$ ), the tidal amplitudes multiplier coefficient of tidal range  $\alpha$ , and the sea level correction  $\gamma$ . This was achieved by linking a TELEMAC-2D physical-based component to an uncertainty quantification library. To achieve better model accuracy, a calibration process was realized through a physical based data-driven technique using a data assimilation library. To promote and facilitate the dissemination of the deployed approach to the TELEMAC-MASCARET community, the development of a Graphical User Interface (GUI) can be useful. The major benefits are user-friendliness, efficiency and enhancing the quality of hydraulic studies.

The calibration process deployed here can be extended to other solvers of the TELEMAC-MASCARET SYSTEM (water quality, sediment transport, wave propagation and so on). Moreover, there are key potentially available sources of information on continental water bodies (in situ and remote sensing data, for instance). Data assimilation algorithms for integrating observation data into real cases are now increasingly applied to hydraulic problems with two main objectives: optimizing model parameters and improving stream-flow simulation and forecasting. Therefore, the ability to exchange data as computational results has become a growing need, facilitated by interoperability.

## Acknowledgements

The authors gratefully acknowledge contributions from the open-source community, especially that of OpenTURN (Open source initiative for the Treatment of Uncertainties, Risks'N Statistics) and ADAO (a module for Data Assimilation and Optimization). In particular, we would like to thank Angelique Ponçot and Jean-Philippe Argaud from EDF R&D for their constructive discussion on data assimilation, Regis Lebrun from Airbus for the POD construction based on OpenTURN and all of the EDF R&D OpenTURN team. The authors also would like to address special thanks to Kamal El Kadi Abderrezzak for his support and deep reading that greatly enriched this paper. Finally, we would like to thank the anonymous reviewer and Arnald Puy, whose comments and suggestions helped improve the manuscript.

## References

- [1] M. Abily, O. Delestre, N. Bertrand, C.-M. Duluc, and P. Gourbesville. High-resolution modeling with bi-dimensional shallow water equations-based codes – high-resolution topographic data use for flood hazard assessment over urban and industrial environments. *Procedia Eng.*, 154:853 – 860, 2016.
- [2] F. Archambeau, N. Méchitoua, and M. Sakiz. Code Saturne: A finite volume code for the computation of turbulent incompressible flows - industrial applications. *International Journal on Finite Volumes*, 1(1), 2004.
- [3] J.-P. Argaud. User documentation, in the salome 9.3 platform, of the ADAO module for "data assimilation and optimization", 2019.
- [4] G. T. Aronica, F. Franza, P. D. Bates, and J. C. Neal. Probabilistic evaluation of flood hazard in urban areas using Monte Carlo simulation. *Hydrol. Process.*, 26(26):3962 – 3972, 2012.
- [5] Y. Audouin, C. Goeury, F. Zaoui, R. Ata, S. El Idrissi Essebtey, A. Torossian, and D. Rouge. Interoperability applications of TELEMAC-MASCARET SYSTEM. In *Proc. TELEMAC & MASCARET User Conf.*, pages 69 – 76, 2017.
- [6] M. Baudin, R. Lebrun, B. Iooss, and A.-L. Popelin. Openturns: An industrial software for uncertainty quantification in simulation. *Handbook of Uncertainty Quantification*, pages 2001 – 2038, 2017.

- [7] G. Blatman and B. Sudret. Adaptive sparse polynomial chaos expansion based on least angle regression. *J. Comput. Phys.*, 230(6):2345 – 2367, 2011.
- [8] F. Braunschweig, P. C. Leitao, L. Fernandes, P. Pina, and R. J. J. Nevesn. The object-oriented design of the integrated water modeling system MOHID. In *Computational Methods in Water Resources*, volume 55 of *Developments in Water Science*, pages 1079 – 1090. Elsevier, 2004.
- [9] S. Buis, A. Piacentini, and D. Déclat. Palm: a computational framework for assembling high-performance computing applications. *Concurr. Comput.*, 18(2):231 – 245, 2006.
- [10] M. P. Butler, P. M. Reed, K. Fisher-Vanden, K. Keller, and T. Wagener. Identifying parametric controls and dependencies in integrated assessment models using global sensitivity analysis. *Environ. Model. Softw.*, 59:10 – 29, 2014.
- [11] K. Campbell, M. D. McKay, and B. J. Williams. Sensitivity analysis when model outputs are functions. *Reliab. Eng. Syst. Saf.*, 91(10):1468 – 1472, 2006.
- [12] A. Carrassi, M. Bocquet, L. Bertino, and G. Evensen. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *Wiley Interdiscip. Rev. Clim. Change*, 9(5):e535, 2018.
- [13] L. Cea and J. R. French. Bathymetric error estimation for the calibration and validation of estuarine hydrodynamic models. *Estuar. Coast. Shelf Sci.*, 100:124–132, 2012.
- [14] L. Cea, M. Bermúdez, and J. Puertas. Uncertainty and sensitivity analysis of a depth-averaged water quality model for evaluation of escherichia coli concentration in shallow estuaries. *Environ. Model. Softw.*, 12:1526 – 1539, 2011.
- [15] E. P. Chassignet, H. E. Hurlburt, O. M. Smedstad, G. R. Halliwell, P. J. Hogan, A. J. Wallcraft, R. Baraille, and R. Bleck. The HYCOM (HYbrid Coordinate Ocean Model) data assimilative system. *J. Mar. Syst.*, 65 (1):60 – 83, 2007.
- [16] L. Clarke, I. Glendinning, and R. Hempel. The MPI Message Passing Interface Standard. In *Proc. Programming Environments for Massively Parallel Distributed Systems*, pages 213 – 218, 1994.
- [17] I. Daloglu, J. I. Nassauer, R. Riolo, and D. Scavia. An integrated social and ecological modeling framework—impacts of agricultural conservation practices on water quality. *Ecol. Soc.*, 19(3), 2014.
- [18] G. Damblin, M. Couplet, and B. Iooss. Numerical studies of space filling designs: optimization of latin hypercube samples and subprojection properties. arXiv, 2013.
- [19] O. David, J. C. Ascough, W. Lloyd, T. R. Green, K. W. Rojas, G. H. Leavesley, and L. Ahuja. A software engineering perspective on environmental modeling framework design: The object modeling system. *Environ. Model. Softw.*, 39:201 – 213, 2013.
- [20] DELTARES. Delft3d-flow user’s manual: Simulation of multi-dimensional hydrodynamic flows and transport phenomena, including sediments, 2014.
- [21] N. V. Dung, B. Merz, A. Bárdossy, T. D. Thang, and H. Apel. Multi-objective automatic calibration of hydrodynamic models utilizing inundation maps and gauge data. *Hydrol. Earth Syst. Sci.*, 15(4):1339 – 1354, 2011.
- [22] G. Dyhouse, J. Hatchett, and J. Benn. *Floodplain Modeling Using HEC-RAS*. Haestad Methods water resources modeling collection. Bentley Institute Press, 2007. ISBN 9781934493021.
- [23] F. Gamboa, A. Janon, T. Klein, and A. Lagnoux. Sensitivity analysis for multidimensional and functional outputs. arXiv, 2013.
- [24] E. Gamma, R. Helm, R. Johnson, and J. M. Vlissides. *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley Professional, 1994.
- [25] O. Garcia-Cabrejo and A. Valocchi. Global sensitivity analysis for multivariate output using polynomial chaos expansion. *Reliab. Eng. Syst. Saf.*, 126:25 – 36, 2014.

- [26] Y. Gil, S. A. Pierce, H. Babaie, A. Banerjee, K. Borne, G. Bust, M. Cheatham, I. Ebert-Uphoff, C. Gomes, M. Hill, J. Horel, L. Hsu, J. Kinter, C. Knoblock, D. Krum, V. Kumar, P. Lermusiaux, Y. Liu, C. North, V. Pankratius, S. Peters, B. Plale, A. Pope, S. Ravela, J. Restrepo, A. Ridley, H. Samet, S. Shekhar, K. Skinner, P. Smyth, B. Tikoff, L. Yarmey, and J. Zhang. Intelligent systems for geosciences: An essential research agenda. *Commun. ACM*, 62(1):76 – 84, 2018.
- [27] C. Goeury, T. David, R. Ata, S. Boyaval, Y. Audouin, N. Goutal, A.-L. Popelin, M. Couplet, M. Baudin, and R. Barate. Uncertainty quantification on a real case with TELEMAT-2D. In *Proc. TELEMAT & MASCARET User Conf.*, pages 44 – 51, 2015.
- [28] C. Goeury, Y. Audouin, and F. Zaoui. Telapy user’s manual, 2017.
- [29] J. B. Gregersen, P. J. A. Gijbbers, and S. J. P. Westen. OpenMI: Open Modeling Interface. *J. Hydroinformatics*, 9(3):175 – 191, 2007.
- [30] A. Griewank and A. Walther. *Evaluating Derivatives*. Society for Industrial and Applied Mathematics, second edition, 2008.
- [31] J. J. Harou, M. Pulido-Velazquez, D. E. Rosenberg, J. Medellín-Azuara, J. R. Lund, and R. E. Howitt. Hydro-economic models: Concepts, design, applications, and future prospects. *J. Hydrol.*, 375(3):627 – 643, 2009.
- [32] Q. Harpham and E. Danovaro. Towards standard metadata to support models and interfaces in a hydro-meteorological model chain. *J. Hydroinformatics*, 17(2):260 – 274, 2014.
- [33] J. D. Herman, M. P. Reed, , H. B. Zeff, and G. W. Characklis. How should robustness be defined for water systems planning under change? *J. Water Resour. Plan. Manag.*, 141:04015012, 2015.
- [34] J.-M. Hervouet. *Hydrodynamics of Free Surface Flows : Modeling with the finite element method*. Wiley, 2007. ISBN 978-0470035580.
- [35] N. Huybrechts, C. Villaret, and F. Lyard. Optimized predictive two-dimensional hydrodynamic model of the Gironde Estuary in France. *J. Waterw. Port. Coast.*, 138(4):312–322, 2012.
- [36] D. Idier, X. Bertin, P. Thompson, and M. Pickering. Interactions between mean sea level, tide, surge, waves and flooding: Mechanisms and contributions to sea level variations at the coast. *Surv. Geophys.*, 40:1603–1630, 11 2019.
- [37] M. Kim, C. P. Gerba, and C. Y. Choi. Assessment of physically-based and data-driven models to predict microbial water quality in open channels. *J. Environ. Sci*, 22(6):851 – 857, 2010.
- [38] F. Klein, A. Fort, A. Sottolichio, A. Beudin, G. Mattarolo, C. Goeury, A. Ponçot, J.-P. Argaud, S. Orseau, P. Tassi, N. Huybrechts, S. Cai, H. Smaoui, R. Gasset, Y. Nedelec, V. Laborie, R. Leroux, S. Barthelemy, M. Ali, and K. S. Gironde XL : an example of how estuarine scientific research can improve dredging and navigation operations in ports. In *VII Congreso Nacional de la ATPYC*, 2018.
- [39] S. Knox, P. Meier, J. Yoon, and J. J. Harou. A Python framework for multi-agent simulation of networked resource systems. *Environ. Model. Softw.*, 103:16 – 28, 2018.
- [40] S. Knox, J. Tomlinson, J. J. Harou, P. Meier, D. E. Rosenberg, J. R. Lund, and D. E. Rheinheimer. An open-source data manager for network models. *Environ. Model. Softw.*, 122:104538, 2019.
- [41] J.-M. Lacombe, F. Zaoui, and N. Goutal. Mascaret – an open souce 1-d simulation code for flow hydrodynamic with a generic application programming interface. *Open Water Journal*, 2(21), 2013.
- [42] M. Lamboni, H. Monod, and D. Makowski. Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. *Reliab. Eng. Syst. Saf.*, 96(4):450 – 459, 2011.
- [43] G. F. Laniak, G. Olchin, J. Goodall, A. Voinov, M. Hill, P. Glynn, G. Whelan, G. Geller, N. Quinn, M. Blind, S. Peckham, S. Reaney, N. Gaber, R. Kennedy, and A. Hughes. Integrated environmental modeling: A vision and roadmap for the future. *Environ. Model. Softw.*, 39:3 – 23, 2013.



- [44] M. J. Lees, L. A. Camacho, and S. Chapra. On the relationship of transient storage and aggregated dead zone models of longitudinal solute transport in streams. *Water Resour. Res.*, 36:213 – 224, 2000.
- [45] C. J. Legleiter, P. C. Kyriakidis, R. R. McDonald, and J. M. Nelson. Effects of uncertain topographic input data on two-dimensional flow modeling in a gravel-bed river. *Water Resour. Res.*, 47(3), 2011.
- [46] Z. Malenovsky, H. Rott, J. Cihlar, M. E. Schaepman, G. García-Santos, R. Fernandes, and M. Berger. Sentinels for science: Potential of Sentinel-1, -2, and -3 missions for scientific observations of ocean, cryosphere, and land. *Remote Sens. Environ.*, 1120:91 – 101, 2012.
- [47] A. Marrel, N. Saint-Geours, and M. De Lozzo. Sensitivity analysis of spatial and/or temporal phenomena. *Handbook of Uncertainty Quantification*, pages 1327 – 1357, 2017.
- [48] M. Mateus and R. Neves. *Ocean modeling for coastal management - Case studies with MOHID*. IST Press, 2013.
- [49] F. J. Montáns, F. Chinesta, R. Gómez-Bombarelli, and J. N. Kutz. Data-driven modeling and learning in science and engineering. *Comptes Rendus Mécanique*, 347:845 – 855, 2019.
- [50] J. L. Morales and J. Nocedal. Remark on “Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-Scale Bound Constrained Optimization”. *ACM Trans. Math. Softw.*, 38(1), 2011.
- [51] R. Morrow, L.-L. Fu, F. Ardhuin, M. Benkiran, B. Chapron, E. Cosme, F. d’Ovidio, J. T. Farrar, S. T. Gille, G. Lapeyre, P.-Y. Le Traon, A. Pascual, A. Ponte, B. Qiu, N. Rasche, C. Ubelmann, J. Wang, and E. D. Zaron. Global Observations of Fine-Scale Ocean Surface Topography With the Surface Water and Ocean Topography (SWOT) Mission. *Front. Mar. Sci.*, 6:232, 2019.
- [52] H. Morvan, D. Knight, N. Wright, X. Tang, and A. Crossley. The concept of roughness in fluvial hydraulics and its formulation in 1d, 2d and 3d numerical simulation models. *J. Hydraul. Res.*, 46(2):191 – 208, 2008.
- [53] J. B. Nagel, J. Rieckermann, and B. Sudret. Principal component analysis and sparse polynomial chaos expansions for global sensitivity analysis and model calibration: Application to urban drainage simulation. *Reliab. Eng. Syst. Saf.*, 195:106737, 2020.
- [54] I. M. Navon. Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography. *Dyn. Atmospheres Oceans*, 27(1):55 – 79, 1998.
- [55] R. I. Ogie, C. Adam, and P. Perez. A review of structural approach to flood management in coastal megacities of developing nations: current research and future directions. *J. Environ. Plan. Manag.*, 63:127 – 147, 2020.
- [56] R. V. O’Neill and B. Rust. Aggregation error in ecological models. *Ecol. Model.*, 7(2):91–105, 1979. ISSN 0304-3800.
- [57] N. Oreskes, K. Shrader-Frechette, and K. Belitz. Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, 263(5147):641–646, 1994.
- [58] S. Orseau, N. Huybrechts, P. Tassi, D. Pham Van Bang, and F. Klein. Two-dimensional modeling of fine sediment transport with mixed sediment and consolidation: Application to the Gironde Estuary, France. *Int. J. Sediment Res.*, 2020.
- [59] F. Pappenberger, K. J. Beven, M. Ratto, and P. Matgen. Multi-method global sensitivity analysis of flood inundation models. *Adv. Water. Resour.*, 31(1):1 – 14, 2008.
- [60] P. Peterson. F2py: A tool for connecting Fortran and Python programs. *Int. J. Comput. Sci. Eng.*, 4(4):296 – 305, 2009.
- [61] C.-T. Pham and F. Lyard. Use of tidal harmonic constants databases to force open boundary conditions in TELEMAC. In *Proc. TELEMAC & MASCARET User Conf.*, pages 165 – 172, 2012.
- [62] F. Pianosi, K. Beven, J. Freer, J. W. Hall, J. Rougier, D. B. Stephenson, and T. Wagener. Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environ. Model. Softw.*, 79:214 – 232, 2016.

- [63] J. Ran and Z. Nedovic-Budic. Integrating spatial planning and flood risk management: A new conceptual framework for the spatially integrated policy infrastructure. *Comput. Environ. Urban Syst.*, 57:68 – 79, 2016.
- [64] A. Ribes and C. Caremoli. Salome platform component model for numerical simulation. In *Proc. International Computer Software and Applications Conference*, volume 2, page 553 – 564, 2007.
- [65] A. Saltelli. Making best use of model evaluations to compute sensitivity indices. *Comput. Phys. Commun.*, 145(2):280 – 297, 2002.
- [66] A. Saltelli. A short comment on statistical versus mathematical modelling. *Nat. Commun.*, 10(1):3870, August 2019. ISSN 2041-1723.
- [67] I. M. Sobol’. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simul.*, 55(1):271 – 280, 2001.
- [68] D. P. Solomatine and A. Ostfeld. Data-driven modeling: some past experiences and new approaches. *J. Hydroinformatics*, 10(1):3 – 22, 2008.
- [69] M. Steiner, J.-M. Bourinet, and T. Lahmer. An adaptive sampling method for global sensitivity analysis based on least-squares support vector regression. *Reliab. Eng. Syst. Saf.*, 183:323 – 340, 2019.
- [70] B. Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliab. Eng. Syst. Saf.*, 93(7):964 – 979, 2008.
- [71] J. Teng, A. Jakeman, J. Vaze, B. F. W. Croke, D. Dutta, and S. Kim. Flood inundation modeling: A review of methods, recent advances and uncertainty analysis. *Environ. Model. Softw.*, 90:201 – 216, 2017.
- [72] United States Federal Highway Administration. *Guide for Selecting Manning’s Roughness Coefficients for Natural Channels and Flood Plains*. U.S. Department of Transportation, Federal Highway Administration, 1984.
- [73] G. Van Rossum and F. L. Drake. *Python 3 Reference Manual*. CreateSpace, 2009. ISBN 1441412697.
- [74] C. Villaret, L. Van, N. Huybrechts, D. Pham van bang, and O. Boucher. Consolidation effects on morphodynamics modeling: application to the Gironde Estuary. *La Houille Blanche*, 6:15 – 24, 2010.
- [75] J. Williams and L. Esteves. Guidance on setup, calibration, and validation of hydrodynamic, wave, and sediment models for shelf seas and estuaries. *Adv. Civ. Eng.*, 2017:1–25, 12 2017.
- [76] F. Zaoui, C. Goeury, and Y. Audouin. A metamodel of the TELEMAC errors. arXiv, 2019.

## A telemac-2d list of API variables

Variable name	Definition
MODEL.AT	Current time
MODEL.DT	Time step
MODEL.BCFILE	Boundary condition file name
MODEL.BND_TIDE	Options for tidal boundary condition
MODEL.BOTTOMELEVATION	Bottom elevation
MODEL.CHESTR	Roughness coefficient
MODEL.FAIR	Wind influence coefficient
MODEL.COTE	Water level on boundary conditions
MODEL.CPL_PERIOD	Coupling period with TELEMAC-MASCARET solvers
MODEL.DEBIT	Flowrate discharge on boundary conditions
MODEL.DEBUG	Activation debug mode keyword
MODEL.FLUX_BOUNDARIES	Flux on the boundary conditions
MODEL.GEOMETRYFILE	Geometry file name
MODEL.METEOFIL	Binary atmospheric file name
MODEL.FO2FILE	Formatted data file name

MODEL.LIQBCFILE	Liquid boundary file name
MODEL.PREFILE	Previous computation file name
MODEL.GRAPH_PERIOD	Graphical output period
MODEL.HBOR	Value of prescribed water depth on boundary conditions
MODEL.IKLE	Mesh connectivity table
MODEL.NACHB	Number of processors sharing a given node
MODEL.KNOLG	Pointer to get the global numbering of a node in the initial mesh
MODEL.KP1BOR	Boundary node connectivity
MODEL.LIHBOR	Boundary type for the water depth
MODEL.LISTIN_PERIOD	Listing output period
MODEL.LIUBOR	Boundary type for horizontal velocity
MODEL.LIVBOR	Boundary type for vertical velocity
MODEL.LT	Current time step
MODEL.COMPLEO	Graphical output counter
MODEL.PTINIG	Number of the first graphical printout time step
MODEL.NPTIR	Number of border nodes for subdomains
MODEL.NBOR	Boundary node global number
MODEL.NELEM	Number of elements in the discretization mesh
MODEL.NELMAX	Maximum possible number of elements
MODEL.NPOIN	Number of nodes in the computational mesh
MODEL.NPTFR	Number of boundary nodes
MODEL.NTIMESTEPS	Number of time steps
MODEL.NUMLIQ	Liquid boundary number
MODEL.POROSITY	Porosity
MODEL.RESULTFILE	Result file name
MODEL.SEALEVEL	Sea level calibration coefficient
MODEL.TIDALRANGE	Tidal range calibration coefficient
MODEL.VELOCITYRANGE	Velocity range calibration coefficient
MODEL.UBOR	Value of prescribed horizontal velocity on boundary conditions
MODEL.VBOR	Value of prescribed vertical velocity on boundary conditions
MODEL.VELOCITYU	Horizontal velocity
MODEL.VELOCITYV	Vertical velocity
MODEL.WATERDEPTH	Water depth
MODEL.X	Horizontal coordinates of mesh points
MODEL.XNEBOR	Horizontal component of the normal at boundary nodes
MODEL.Y	Vertical coordinates of mesh nodes
MODEL.YNEBOR	Vertical component of the normal at boundary nodes
MODEL.EQUATION	Name of solved equation
MODEL.AK	Turbulent kinetic energy term of $k - \epsilon$ turbulence model
MODEL.EP	Dissipation term of $k - \epsilon$ turbulence model
MODEL.ITURB	Turbulence model
MODEL.INIT_DEPTH	Initial value of water depth
MODEL.TRACER	Tracer value on mesh nodes
MODEL.NTRAC	Number of tracers
MODEL.FLOWRATEQ	Solid transport flowrate
MODEL.DCLA	Median grain size value
MODEL.SHIELDS	Critical Shields parameter
MODEL.XWC	Settling velocity
MODEL.Z	Free surface elevation
MODEL.QBOR	Value of prescribed discharge flowrate on boundary nodes
MODEL.EBOR	Value of prescribed turbulence dissipation term on boundary nodes
MODEL.FLBOR	Value of prescribed bottom elevation on boundary nodes
MODEL.TOB	Shear stress
MODEL.LIQBOR	Boundary type for the discharge
MODEL.NSICLA	Number of bed material size classes
MODEL.NOMBLAY	Number of layers in the bed

MODEL.CONCENTRATION	Concentration at the current time step
MODEL.EVOLUTION	Evolution of bed
MODEL.PARTHENIADES	Parthenaides erosion coefficient of each bed layer
MODEL.VOLU2D	Basis integral

Table A.1: TELEMAC-2D variables accessible through API

## B Programming the Gironde estuary case in Python

```

# Configuration step
# 'cas_file': telemac steering file
# 'for_file': User fortran folder
# 'comm': MPI communicator
t2d = Telemac2d(cas_file, user_fortran=for_file, comm=comm)

# Reading of TELEMAC-2D steering file and run memory allocation
t2d.set_case()

# Initialization of the TELEMAC-2D model
t2d.init_state_default()

# Six friction coefficient values are used
CHESTR = [CF1, CF2, CF3, CF4, CF5, CF6]
for k in range(len(CHESTR)):
    # Set the corresponding value to each node according to the area location
    for i in AREA[k]:
        t2d.set("MODEL.CHESTR", CHESTR[k], i=i)

# Set the tidal range value with the 'Hrange' variable
t2d.set("MODEL.TIDALRANGE", Hrange)
# Set the tidal velocity value with the 'Vrange' variable
t2d.set("MODEL.VELOCITYRANGE", Vrange)
# Set the sea level value with the 'SeaLevel' variable
t2d.set("MODEL.SEALEVEL", SeaLevel)

# MPI barrier
comm.Barrier()

# Initialize a list 'waterLevels' to store results in certain coordinate points
waterLevels = []
for j in xrange(len(self.observation_points)):
    waterLevels.append([])

# Time loop
ntimestep = self.t2d.get("MODEL.NTIMESTEPS")
for i in xrange(ntimestep):
    # Running one time step
    t2d.run_one_time_step()
    # Extract results if the time corresponds to observations
    if i >= debObs and i <= finObs and i % freqObs == 0:
        # Getting stations water level values
        for j in range(len(self.observation_points)):
            # 'H' is the water depth and 'Zf' the bathymetric level
            H, _, _, Zf = getStationValues(self.observation_points[j])
            waterLevels[j].append(H+Zf)

# Finalize
comm.Barrier()
t2d.finalize()
del self.t2d

```

## C Bayesian inference formulation of calibration problem

The *posterior* distribution  $\pi(\mathbf{X}|\mathbf{Y})$  can be determined through the well-known Bayes rule (Eq. 17).

$$\pi(\mathbf{X}|\mathbf{Y}) = \frac{\pi(\mathbf{Y}|\mathbf{X})\pi(\mathbf{X})}{\int \pi(\mathbf{Y}|\mathbf{X})\pi(\mathbf{X})d\mathbf{X}} \quad (17)$$

The term  $\pi(\mathbf{Y}|\mathbf{X})$ , called the likelihood, can be interpreted as the probability density function of the observed data, conditional upon a set of parameter values (considered as random variables), and as a measure of the information provided by the observations on the parameter values. From Eq. 8, the likelihood is expressed as  $\pi(\mathbf{Y}|\mathbf{X}) \propto \exp(-\frac{1}{2}[\mathbf{Y} - G(\mathbf{X})]\mathcal{R}^{-1}[\mathbf{Y} - G(\mathbf{X})]')$ .

The term  $\pi(\mathbf{X})$  represents *a priori* knowledge of the unknown parameters  $\mathbf{X}$ . This term is classically taken as a multivariate normal distribution with known mean  $\mathbf{X}_0$  (derived from measurement data or previous computation) and covariance matrix  $\mathcal{B} \in \mathbb{R}^{p \times p}$  positive definite such as  $\pi(\mathbf{X}) \propto \exp(-\frac{1}{2}[\mathbf{X} - \mathbf{X}_0]\mathcal{B}^{-1}[\mathbf{X} - \mathbf{X}_0]')$ .

From the previous expressions of *a priori* and likelihood, the *posterior* distribution is given by the following equation:

$$\pi(\mathbf{X}|\mathbf{Y}) \propto \exp\left(-\frac{1}{2}[\mathbf{Y} - G(\mathbf{X})]\mathcal{R}^{-1}[\mathbf{Y} - G(\mathbf{X})]' - \frac{1}{2}[\mathbf{X} - \mathbf{X}_0]\mathcal{B}^{-1}[\mathbf{X} - \mathbf{X}_0]'\right) \quad (18)$$

The maximum *a posteriori* (MAP) is equivalent to the formulation of the optimal search of control vector  $\mathbf{X}$ , which must satisfy the *a priori* error statistics  $J_b = \frac{1}{2}[\mathbf{X} - \mathbf{X}_0]\mathcal{B}^{-1}[\mathbf{X} - \mathbf{X}_0]'$  and the equivalent observation error statistics  $J_{obs} = \frac{1}{2}[\mathbf{Y} - G(\mathbf{X})]\mathcal{R}^{-1}[\mathbf{Y} - G(\mathbf{X})]'$ .

## D Generalized Sensitivity Analysis based on low discrepancy sequence of Sobol

Inputs	GSI		Verdon		Bordeaux	
	First Order	Total Order	First Order	Total Order	First Order	Total Order
$K_{s1}$	0.126	0.133	0.110	0.156		
$K_{s2}$	0.01	0.0153	0.358	0.437		
$K_{s3}$	$2.29 \times 10^{-5}$	$7.45 \times 10^{-5}$	0.0816	0.118		
$K_{s4}$	0.	$1.33 \times 10^{-5}$	$9.69 \times 10^{-5}$	$3.44 \times 10^{-4}$		
$K_{s5}$	$2.53 \times 10^{-5}$	$6.50 \times 10^{-5}$	$2.06 \times 10^{-3}$	$3.54 \times 10^{-3}$		
$K_{s6}$	$1.65 \times 10^{-7}$	$8.13 \times 10^{-6}$	$4.17 \times 10^{-5}$	$1.86 \times 10^{-4}$		
$\alpha$	0.0464	0.0492	0.0135	0.019		
$\beta$	$4.82 \times 10^{-4}$	$5.07 \times 10^{-4}$	$1.81 \times 10^{-4}$	$2.70 \times 10^{-4}$		
$\gamma$	0.809	0.809	0.341	0.364		

Table D.2: Generalized Sensitivity Indices (GSI) based on 1024 ( $2^{10}$ ) low discrepancy Sobol points at Bordeaux and Verdon observation stations