

Sensitivity of tidal modelling in coastal configurations: an uncertainty study based on field-measurement reduction

Rem-Sophia Mouradi, Cédric Goeury, Pablo Tassi, Fabrice Zaoui, Olivier
Thual

► To cite this version:

Rem-Sophia Mouradi, Cédric Goeury, Pablo Tassi, Fabrice Zaoui, Olivier Thual. Sensitivity of tidal modelling in coastal configurations: an uncertainty study based on field-measurement reduction. TELEMATICS USER CONFERENCE (TUC) 2020, Oct 2020, Anvers (en ligne), Belgium. hal-03106245

HAL Id: hal-03106245

<https://hal.archives-ouvertes.fr/hal-03106245>

Submitted on 11 Jan 2021

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.

Sensitivity of tidal modelling in coastal configurations: an uncertainty study based on field-measurement reduction

Rem-Sophia Mouradi^{1,2}, Cédric Goeury¹, Pablo Tassi¹, Fabrice Zaoui¹

¹National Laboratory for Hydraulics and Environment (LNHE)
Electricité De France (EDF) R&D
6 Quai Watier, Chatou, France
rem sophia.mouradi@gmail.com

Olivier Thual²

²Climate, Environment, Coupling and Uncertainties
research unit (CECI)
Université de Toulouse, CERFACS/CNRS
Toulouse, France

Abstract—Hydrodynamic models are increasingly used in operational industrial contexts for prediction and analysis purposes, including risk assessment and design optimization. In the modelling phase, expertly made choices can have consequences on either computational cost (i.e. mesh resolution, domain size, etc.), or closures to unknowns and their parameters (friction, turbulence, etc.). These choices are examples of epistemic uncertainty in hydrodynamic applications. Simultaneously, in the last few years, field measurements have become more accessible, giving the opportunity to deepen validation processes, and to challenge standard modelling practices based on expert judgement.

In this study, a sensitivity analysis to modelling choices such as domain extent and friction closure formulas is assessed in the context of tidal forcing in a coastal area. Comparison to field measurements of free surface and velocity components is performed using Dimensionality Reduction (DR), by means of Proper Orthogonal Decomposition (POD).

Firstly, numerical simulations with small, medium and large scale computational domains and two different friction parameterizations (i.e. Strickler and Colebrook-White), are performed and statistically compared. For this purpose, Monte Carlo (MC) simulations are launched for each modelling configuration, using samples of uncertain friction parameters and tidal boundary conditions. The different behaviors are therefore analyzed using comparative statistics on the available measurement points (average, confidence intervals, etc.).

Secondly, using POD, the behavior of each configuration, over the space of all possible events, is reduced to few representative components, commonly called modes or patterns. These modes are mutually compared for the different configurations, as well as to POD modes deduced from a distinct analysis on field measurements. A sensitivity analysis on POD modes using Sobol' indices shows that the chosen configurations can have different sensitivities to the same uncertain input parameters. This suggests that a calibration procedure might respond to different control parameters depending on the modelling choices. Consequently, the optimal estimates found from one configuration to the other, may differ.

I. INTRODUCTION

Shallow Water Equations (SWE) are commonly used in coastal applications as good compromise between precision and computational cost. However, modelling a real case application is not always trivial. For example, the influence of the computation domain extension on the results is in general poorly evaluated, and the uncertainty implied by source terms closures, for example friction, is often overlooked. In the absence of alternatives, these choices are determined by expert opinion. In particular, calibration of the model on measurements is used to justify the model's capacity to provide physically coherent information, and therefore its predictive capability. The fitted model is then used to analyze the flow between calibration points. However, the equivalence between good-fit and accuracy can be questioned. For example, friction is often calibrated to fit measurements at different stream points, spatially or temporally distributed, as in [4, 10]. It is then assumed that the bed and the flow characteristics are uniformly distributed between the calibration points, and that the model is trusted for the other choices, namely domain size. This of course is a strong assumption.

The goal of the present study is to analyze the uncertainty resulting from common modeling choices: domain size and friction closure. For this purpose, four domain extents are compared, and two friction formulas are confronted. Different challenges are however encountered. Firstly, realistic hydrodynamic cases are high dimensional, i.e. they involve numerous parameters, and the response is spatio-temporal, which makes pointwise analysis difficult. Secondly, measurements are also spatio-temporal fields and are sometimes noisy. Proper Orthogonal Decomposition (POD) is therefore used to reduce the dimensionality of the numerical output on one hand, and to reduce and smoothen the observation on the other hand. This makes the comparison of uncertainties resulting from different model choices easier. Indeed, POD is respectively applied on the numerical model results and on observations. The resulting patterns are confronted. Then, the coefficients associated to the numerical POD patterns are learned using PCE (Polynomial Chaos

Expansion), to provide a meta-model that helps perform low cost Sensitivity Analysis (SA) [13, 14], and can be used to replace the model for fast calibration.

Last, besides patterns comparison, the most influencing parameters for each modeling configuration are optimally calibrated and compared. A 3DVAR algorithm is used [15], even though measurements are distributed in time. This is made possible thanks to POD that reduces the temporal behavior over the simulation window to few scalars instead of a high dimensional time series.

An example of a power plant's cooling intake, located in a coastal area is studied as an application case. The intake provides the plant with water via a pumping system. The study of the intakes hydrodynamics is of industrial interest, but the external currents should be trustworthy. Hydrodynamic uncertainty should therefore be quantified. Five measurement points are available off the coast, and 48 hours simulations are launched for the uncertainties investigation.

This paper is organized as follows. Firstly, theoretical background on hydrodynamic modeling, Uncertainty Quantification (UQ), using PCE, POD, SA and 3DVAR are shortly described in Section II. Secondly, modeling uncertainties are investigated in Section III, using four domain extents and two friction closures. Thirdly, optimal calibration results are compared for the modeling settings in Section IV. Lastly, a conclusion is given in Section IV.

II. MATERIALS

A. Shallow Water Equations

Shallow Water Equations (SWE) are obtained by depth-averaging the three-dimensional Reynolds-averaged free-surface Navier-Stokes equations, allowing the representation of almost-horizontal, two-dimensional (2D), shallow flows [1]. The mass and momentum conservation equations are defined in (1), where the system unknowns are the velocity components $\mathbf{u} = (u, v)^T$ along the Cartesian coordinates (x, y) and the free surface elevation $\eta := h + b$, with h the water depth and b the bottom elevation. The gravitational acceleration g and the water density ρ are considered constant. Vector $\boldsymbol{\tau}_b$ denotes the bottom shear stress, vector \mathbf{F} represents external forces (Coriolis, surface tension, wave radiation, etc.), and ν_e is the effective viscosity accounting for kinematic, eddy and "dispersion" viscosity, the latter resulting from the vertical integration.

$$\begin{cases} \frac{\partial h}{\partial t} + \nabla \cdot (h\mathbf{u}) = 0 \\ \frac{\partial (h\mathbf{u})}{\partial t} + \nabla \cdot (h\mathbf{u} \otimes \mathbf{u}) = -g\nabla\eta - \frac{\boldsymbol{\tau}_b}{\rho} + \frac{h\mathbf{F}}{\rho} + \nabla \cdot (h\nu_e \nabla \mathbf{u}) \end{cases} \quad (1)$$

In this work, the external forces are omitted and the effective viscosity is set to water's kinematic viscosity for simplicity. While this is practical for a first investigation of the model's behaviour in a UQ framework, it is worth mentioning that the omitted terms are physically important and must, in principle, be considered. The bottom shear stress and the hydrodynamic Boundary Conditions (BC) need closure.

Physical parameterizations are generally used, here considered uncertain, and discussed in the following in Subsection B.

B. Uncertain friction and boundary conditions

1) *Friction*: Bed shear stress is unknown and needs closure. It is capital for environmental applications, as it has considerable influence on the flow because of the energy dissipation it induces [2]. Its exact formulation remains unknown, but many formulas can be found in literature, with specific calibration parameters. It is generally expressed as in (2), where C_f is a dimensionless friction coefficient.

$$\boldsymbol{\tau}_b = \frac{1}{2} \rho C_f |\mathbf{u}| \mathbf{u}, \quad (2)$$

Literature formulas for C_f are either empirical or semi-empirical [2]. One of the most widely used empirical formulas is Strickler's model (3) (or Manning-Strickler in the Anglo-Saxon terminology, as explained in [3]). Coefficient B takes values in the range [21.1, 26.613] $m^{1/2}s^{-1}$ and k_s is the bed roughness height, often set to the median bed grain size [3]. The formula is usually written as a function of the so-called Strickler coefficient $K := B/k_s^{1/6}$. It can be noted that equivalence with Nikuradse is assured for the particular value $\frac{k_s}{h} = 0.037$, if and only if $B=26.613$ (developments in [19]).

$$C_f = \frac{2g}{B^2} \left(\frac{k_s}{h} \right)^{1/3}. \quad (3)$$

Semi-empirical formulas express the impact of near-bed turbulence on bed resistance to the flow. Indeed, not only turbulence modifies the currents in the water column, but the turbulence regime also changes the bed resistance [3]. For example, Colebrook-White's implicit formula (4) can be used, where B_1 , B_2 and B_3 are dimensionless, with values respectively in the ranges [2, 2.14], [0, 7.17] and [8.888, 14.83], as reported by Yen [5].

$$C_f = \frac{\lambda}{4} = \frac{1}{4 \left(-B_1 \log \left(\frac{1}{B_2} \left(\frac{k_s}{h} + \frac{B_3}{Re\sqrt{\lambda}} \right) \right)^2 \right)}. \quad (4)$$

For both formulas, k_s must be defined. This variable is often adopted as a calibration parameter, although being physically complex to describe, as it results from different contributions (e.g. skin friction, bed forms dissipation, etc.) [6]. For coastal applications, van Rijn [6] proposes the

formula $k_s = \sqrt{(k_s^c)^2 + (k_s^{mr})^2 + (k_s^d)^2}$, where the total roughness k_s is induced by ripples k_s^c , mega-ripples k_s^{mr} and dunes k_s^d . Using the bounds reported in [6] for each component, the variation interval $k_s \in [0.00064, 1.023] m$ is obtained.

2) *Tidal Boundary Conditions*: Tidal forcing are usually imposed as BC in coastal applications. In this study, the TPXO data-base is used [7], particularly the European Shelf (ES) local model within TELEMAT-2D [8]. The hydrodynamic unknowns at the boundary are modelled as a superposition of harmonic components, as in (5) and (6),

$$F(\mathbf{p}, t) = \sum F_i(\mathbf{p}, t), \quad (5)$$

$$F_i(\mathbf{p}, t) = f_i(t) A_{F_i}(\mathbf{p}) \cos\left(\frac{2\pi t}{T_i} - \phi_{F_i}(\mathbf{p}) + u_i^0 + v_i(t)\right), \quad (6)$$

where the term F at point \mathbf{p} and time t represents the unknown (velocity component and/or water depth), F_i a harmonic component with constant period T_i , amplitude A_{F_i} , phase ϕ_{F_i} , phase at origin of times u_i^0 , and temporal nodal factors $f_i(t)$ and $v_i(t)$. Thompson's method is then used to prescribe BC [9], and three parameters, denoted CTL (Coefficient of Tidal Level), MTL (Mean Tidal Level) and CTV (Coefficient of Tidal Velocity), can be used to calibrate the BC on measurements, as in (7) and (8). For example, MTL allows 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 [17].

$$h(\mathbf{p}, t) = CTL \times \sum h_i(\mathbf{p}, t) - z_f + MTL, \quad (7)$$

$$\mathbf{u}(\mathbf{p}, t) = CTV \times \sum \mathbf{u}_i(\mathbf{p}, t). \quad (8)$$

In this study, MTL variation interval is deduced from measurements (Section III.A) as [4.0 m CM, 6.0 m CM], whereas the non-dimensional parameters CTL and CTV are expertly determined and respectively set to [0.8, 1.2] and [0.8, 3.0]. Using these bounds, the measurements fall within the simulated min-max interval (see Fig. 3). Coefficients MTL , CTL and CTV can be used to compensate the effects of storm and surge (atmospheric and wave setup), as the latter are not modelled and not taken into consideration in the TPXO database.

C. Uncertainty Quantification

The objective of UQ studies is to determine the impact of uncertain inputs or model on the uncertainty of simulated output. Firstly, an identification and quantification of uncertainty sources should be performed, as done in Subsection B. Here, in the absence of other information, Uniform densities are deduced from literature value ranges for each variable (Subsection B). Secondly, Monte Carlo (MC) simulations are realized. A random sample of the uncertain inputs of size 1000 is generated with the Uniform laws, and corresponding calculations are launched. Thirdly, MC outputs are analysed. Statistics such as mean and standard deviation can be calculated, and the overall behaviour can be analysed, for example by reducing the output's space to representative components using Proper Orthogonal Decomposition (POD), described in Subsection D.1. Additionally, Sensitivity Analysis (SA) is performed to rank the uncertain inputs by their influence on the output. For this purpose, Polynomial Chaos Expansion (PCE) can be used to calculate Sobol' indices, as explained in Subsections D.2 and D.3.

Lastly, once the most influencing parameters are identified, they can be used to fit the model on measurements. For example, optimal methods like 3DVAR, shortly described in Subsection D.4, can be used. This gives a best estimation for parameters as well as associated confidence intervals. The latter are generally much narrower than the initial UQ intervals, which reduces uncertainty.

1) *Proper Orthogonal Decomposition*: The goal of POD is to extract patterns of a continuous bi-variate function. These patterns, when added and multiplied by adequate coefficients, explain the main dynamics. Let $\mathbf{u}: \Omega \times \mathbb{T} \rightarrow \mathbb{D} = Im(\mathbf{u})$ be a continuous function of two variables $\epsilon, \delta \in \Omega \times \mathbb{T}$. The following relationships and

properties hold for any $\Omega \times \mathbb{T}$ and Hilbert space \mathbb{D} characterized by its scalar product $(\cdot, \cdot)_{\mathbb{D}}$ and induced norm $\|\cdot\|_{\mathbb{D}}$. For example, this could concern the reduction of a temporal series, controlled by stochastic parameters ($\epsilon = \{t_1, \dots, t_n\}$ and $\delta = \{\theta_1, \dots, \theta_V\}$), where \mathbb{D} is a set of scalar real values or vector real values (e.g. \mathbb{R} or \mathbb{R}^2). POD consists then in an approximation of $\mathbf{u}(\epsilon, \delta)$ at a given order $d \in \mathbb{N}^*$ [11] as in (9),

$$\mathbf{u}(\epsilon, \delta) \approx \sum_{k=1}^d v_k(\delta) \sigma_k \boldsymbol{\varphi}_k(\epsilon), \quad (9)$$

where $\{v_k(\delta)\}_{k=1}^d \subseteq \mathcal{C}(\mathbb{T}, \mathbb{R})$ and $\{\boldsymbol{\varphi}_k(\epsilon)\}_{k=1}^d \subseteq \mathcal{C}(\Omega, \mathbb{D})$, with $\mathcal{C}(\mathbb{A}, \mathbb{B})$ denoting the space of continuous functions defined over \mathbb{A} and arriving at \mathbb{B} , and $\{\sigma_k\}_{k=1}^d \subseteq \mathbb{R}$. The objective is to identify $\{\boldsymbol{\varphi}_k(\cdot)\}_{k=1}^d$ that minimizes the distance of the approximation to the true value $\mathbf{u}(\cdot, \cdot)$, over the whole $\Omega \times \mathbb{T}$ domain, with an orthonormality constraint for $\{\boldsymbol{\varphi}_k(\cdot)\}_{k=1}^d$ using the scalar product $(\cdot, \cdot)_{\mathbb{D}}$. This can be defined, in the least-squares sense, as a minimization problem.

The minimization problem is defined for all orders $d \in \mathbb{N}^*$, so that the members $\boldsymbol{\varphi}_k$ are ordered according to their importance. The family $\{\boldsymbol{\varphi}_k(\epsilon)\}_{k=1}^d$ is called the POD basis. The solution to this problem is well established [11, 12]: the POD basis of \mathbb{D} of order d is the set of orthonormal eigenvectors of an operator $\mathfrak{R}: \mathbb{D} \rightarrow \mathbb{D}$ defined as $\mathfrak{R}(\boldsymbol{\varphi}) = \langle (\mathbf{u}, \boldsymbol{\varphi})_{\mathbb{D}} \times \mathbf{u} \rangle_{\mathbb{T}}$, if the eigenvectors are taken in decreasing order of corresponding eigenvalues, which are $\{\lambda_k = \sigma_k^2\}_{k=1}^d$. An accuracy rate, also called Explained Variance Rate (EVR), can be calculated as $\sum_{k \leq d} \lambda_k / \sum_{k=1}^{+\infty} \lambda_k$, which tends to 1 (perfect approximation) when d tends to $+\infty$. Each λ_k represents the variance proportion carried by mode k . When a given $d \ll \min(\dim(\mathbb{D}), \dim(\mathbb{T}))$ corresponds to a high accuracy rate, we speak of dimensionality reduction.

2) *Polynomial Chaos Expansion*: The idea behind Polynomial Chaos Expansion (PCE) is to formulate an explicit model that links a variable (output) to conditioning parameters (inputs), both living in a probability space. The interest variable denoted \mathbf{Y} and the input parameters denoted $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_V)$ are characterized by a given density. The models response can be approximated as in (10),

$$\mathbf{y} = \mathcal{M}(\boldsymbol{\theta}) = \mathcal{M}_0 + \sum_{i=1}^V \mathcal{M}_i(\theta_i) + \sum_{1 \leq i < j \leq V} \mathcal{M}_{i,j}(\theta_i, \theta_j) + \dots + \mathcal{M}_{1,\dots,V}(\theta_1, \theta_2, \dots, \theta_V), \quad (10)$$

where \mathcal{M}_0 is the mean of \mathbf{y} and $\mathcal{M}_{I \subseteq \{1, \dots, V\}}$ is the common contribution of variables $I \subseteq \{1, \dots, V\}$ to \mathbf{y} . For PCE, these contributions have a polynomial form, as in (11),

$$\mathbf{y} = \mathcal{M}(\boldsymbol{\theta}) = \sum_{|\underline{\alpha}| \leq P} c_{\underline{\alpha}} \zeta_{\underline{\alpha}}(\theta_1, \theta_2, \dots, \theta_V), \quad (11)$$

with $\underline{\alpha} = \{\alpha_1, \dots, \alpha_V\}$ and $|\underline{\alpha}| = \sum_{i=1}^V \alpha_i$, defining a multivariate polynomial basis $\{\zeta_{\underline{\alpha}}, \underline{\alpha} \in \mathbb{N}^V \text{ and } |\underline{\alpha}| \in \{0, \dots, P\}\}$ as $\zeta_{\underline{\alpha}}(\theta_1, \theta_2, \dots, \theta_V) = \prod_{i=1}^V \xi_{\alpha_i}^{(i)}(\theta_i)$. For each parameter θ_i , $\{\xi_{\alpha_i}^{(i)}, \alpha_i \in \{0, \dots, P\}\}$ is an orthonormal polynomial basis, and $P \in \mathbb{N}$ is a chosen polynomial degree. $c_{\underline{\alpha}}$ are deterministic coefficients that can be estimated using different methods. In this work, the Least Angle Regression Stagewise method (LARS) is used to construct an adaptive

sparse PCE. Further details on PCE in general and LARS in particular can be found in [20].

3) *Sensitivity Analysis*: PCE can be used with variance decomposition [13, 14] to analyse the influence of the inputs variances on the output's variance [14]. For the general case of a multivariate model written as in (10), the contribution of the polynomial indexed by $\underline{\alpha}$, to the output \mathbf{y} can be computed in terms of variance, as in (12),

$$S_{\underline{\alpha}} = \frac{\text{var}[c_{\underline{\alpha}}\zeta_{\underline{\alpha}}]}{\text{var}[\mathbf{y}]} = \frac{c_{\underline{\alpha}}^2}{\sum_{|\underline{\beta}| \leq P} c_{\underline{\beta}}^2}, \quad (12)$$

where $S_{\underline{\alpha}}$ are the well-known Sobol' indices [14]. Adding them for all the polynomials that contribute to \mathbf{y} equals 1. They allow to rank all terms by their relative contribution to \mathbf{y} . The contributions can either be: (i) analysed for each polynomial; (ii) used to compute the 1st order contribution of a variable θ_i alone (1st Sobol' index denoted S_i) by adding the monomial contributions only; (iii) or used to compute the total contribution of θ_i (total Sobol' index denoted S_i^T) by adding the contributions of all polynomials involving θ_i .

4) *Optimal Calibration using 3DVAR*: An automatic algorithm is here used for inverse parameters estimation from observations and a numerical model. On the first hand, observations are not perfect. On the other hand, one may have a first guess for the parameters (physical knowledge, previous simulations, etc.), but the latter is uncertain. In order to find the best compromise between measurements errors and parameters first guess errors, the optimization takes the form of a minimization problem, for the function defined in (13),

$$J(\boldsymbol{\theta}) = \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_b)^T \mathbf{B}^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}_b) + \frac{1}{2}(\mathbf{y} - \mathcal{H}(\mathcal{M}(\boldsymbol{\theta})))^T \mathbf{R}^{-1}(\mathbf{y} - \mathcal{H}(\mathcal{M}(\boldsymbol{\theta}))), \quad (13)$$

where \mathbf{y} is the observation, \mathcal{M} the numerical model, \mathcal{H} an operator from the simulation to the observation space, $\boldsymbol{\theta}$ the set of unknown parameters and $\boldsymbol{\theta}_b$ a background knowledge (or first guess) of the parameters. The background and observation errors are represented by error covariance matrices, respectively denoted \mathbf{B} and \mathbf{R} . This is commonly known as a data assimilation variational problem, where J is called cost function, and minimizing it is referred to as 3DVAR [15, 16].

Here, the observation and simulation are the same variables expressed on the same locations and times, i.e. \mathcal{H} is identity. The observation can be POD reduced and approximated as a matrix product $\mathbf{y} \approx \boldsymbol{\Phi}_Y \boldsymbol{\Sigma}_Y \tilde{\mathbf{y}}$, where a finite number of POD modes is stored in $\boldsymbol{\Phi}_Y$, the corresponding square roots of eigenvalues are stored in $\boldsymbol{\Sigma}_Y$ and $\tilde{\mathbf{y}}$ contains the multiplicative coefficients, which are a reduced form of \mathbf{y} . Next, the simulation result $\mathcal{M}(\boldsymbol{\theta})$ can be POD reduced and the coefficients learned using PCE as a function of $\boldsymbol{\theta}$, which is written as $\mathcal{M}(\boldsymbol{\theta}) \approx \boldsymbol{\Phi}_X \boldsymbol{\Sigma}_X \tilde{\mathcal{M}}(\boldsymbol{\theta})$, where $\tilde{\mathcal{M}}(\boldsymbol{\theta})$ is a vector storing PCE models for the reduced version of simulation result. The cost function can then be approximated as in (14).

$$J(\boldsymbol{\theta}) \approx \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_b)^T \mathbf{B}^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}_b) + \frac{1}{2}(\tilde{\mathbf{y}} - \tilde{\mathbf{H}} \tilde{\mathcal{M}}(\boldsymbol{\theta}))^T \tilde{\mathbf{R}}^{-1}(\tilde{\mathbf{y}} - \tilde{\mathbf{H}} \tilde{\mathcal{M}}(\boldsymbol{\theta})), \quad (14)$$

where $\tilde{\mathbf{H}} = (\boldsymbol{\Phi}_Y \boldsymbol{\Sigma}_Y)^{-1} \boldsymbol{\Phi}_X \boldsymbol{\Sigma}_X$ is a new linear operator (matrix) that links the reduced observation to the PCE model of reduced simulation $\tilde{\mathcal{M}}(\boldsymbol{\theta})$, and $\tilde{\mathbf{R}}^{-1} = (\boldsymbol{\Phi}_Y \boldsymbol{\Sigma}_Y)^T \mathbf{R}^{-1} (\boldsymbol{\Phi}_Y \boldsymbol{\Sigma}_Y)$. The new minimization problem is an approximation of the original, but is less costly. Indeed, reduction implies that the dimension of $\tilde{\mathbf{y}}$ is much lower than the full observation's, and that of $\tilde{\mathcal{M}}(\boldsymbol{\theta})$ is much lower than the numerical model's. Additionally, PCE performs in seconds while the numerical model needs hours, making the model evaluation time negligible.

In the following, the uncertainties resulting from modelling choices (domain extension, friction closure) are investigated using the previously described theoretical elements.

III. UNCERTAINTIES RELATED TO MODELING CHOICES

Firstly, a description of the study case and available data is given in Subsection A. The case is used to investigate the differences that come from domain extent choice in Subsection B, and friction formula choice in Subsection C.

A. Case study

The study site is located on the eastern English Channel coast in northern France. The study zone is mega-tidal and dominated by a semi-diurnal circulation, with moderate wave activity. In particular, a power plant's cooling intake is of interest. Its upstream boundary is connected to the sea, and a pumping system ensures the plant's cooling. Hydrodynamic models of different extents, as shown in Fig. 1-a, can be used to represent the flow in the intake. The choice of domain extent and other simulation parameters (Section II) is considered uncertain. Consequently, four domain extensions are compared in Subsection B and two friction formulas confronted in Subsection C. For validation, five measurement points (indicated in Fig. 1-a) of $(u, v)^T$ and η over a two-month period are available. In addition to tidal effects, possible occurrence of storms, surges and resulting non-linear interactions with the tides influence these measurements. However, it is difficult to isolate their effect in the measured quantities as highlighted in [17]. The tidal BC coefficients introduced in Subsection B.2 will therefore be used to calibrate the total signal.

Firstly, domains of different sizes, centred on the cooling intake, are created. Four domains in particular, of sizes 800 m, 2 km, 7 km and 8 km, are compared in Section B. These denominations correspond to the distance from the intake entrance to the offshore, and equal distance on either side of the intake, to the east and to the west alongshore. Secondly, mesh convergence is assessed on the 2 km domain, and results in elements of size 50 m at the sea, 2.5 m at the intake walls and 0.5 m at the intake pumps. This configuration is kept for all domain sizes, and the resulting geometries share the same mesh in the common zones. The mesh is shown in Fig. 1-b, where the intake is coloured in red, and the growth of elements size from the intake to the sea is visible. The corresponding

meshes contain 28 188, 31 814, 74 079 and 87 617 nodes for the four domain extensions respectively. Thirdly, bathymetries are retrieved from different sources: a Digital Elevation Model (DEM) composed of global and local bathymetries [18], an interpolation of topographic beach profiles and a nearshore multi-beam bathymetry inside the intake and at its vicinity. The same bathymetries are applied for all domain sizes, and the resulting geometries share the same bathymetry in the common zones. It should be mentioned that the used data are not necessarily available at measurement date, and are spatially interpolated. This results in an epistemic uncertainty. Additional sources of uncertainty, namely measurement errors should be noted. However, this is not explicitly dealt with. For the sake of simplicity, calibrating friction is considered to compensate, in average, for the uncertainty in the right hand side terms of SWE.

Lastly, for all domain extensions, the BC are directly interpolated from the TPXO data-based. As all domain sizes keep smaller than a TPXO element size, the differences in the BC only result from linear interpolation due to the distance of boundary elements from the TPXO nodes. The idea is to evaluate the sensitivity of the simulations to pure extension of the calculation domains and not to changes in the BC nature. For the same reasons, BC of the smaller domains are not interpolated from larger domains because this would compensate for the difference of extension, which is exactly what we want to analyse.

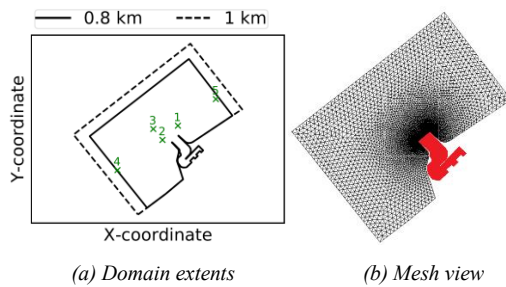


Figure 1- Examples of domain extents with representation of measurement points outside the intake and view of the mesh.

Measurements of tidal periods are extracted and superposed as in Fig. 2.

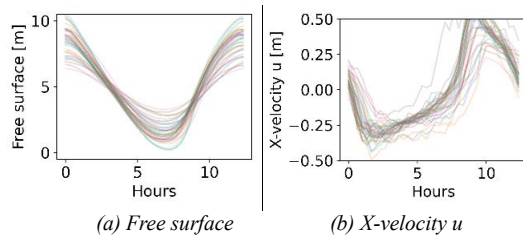


Figure 2- Superposition of field measurements in front of the intake on point 1, for two hydrodynamic variables.

The different periods are considered as realizations of a temporal series, and POD is performed. The behaviour over two months is therefore reduced to a few components. The associated EVR are shown in Fig. 3. Velocity components u and v show the same behaviour. The two-months measurements, from which 38 periods were extracted for each

variable, can therefore be reduced to 2 components, giving 99% of captured variance for free surface, and over 96% for the velocity components.

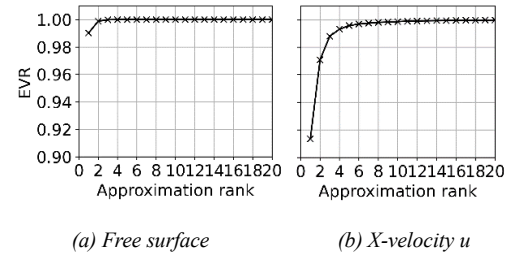


Figure 3- EVR for the free surface and X-velocity u after measurement based POD reduction on Point 1.

B. Domain size implications

As introduced in Section I and represented in Fig. 1, the comparison of hydrodynamic computations with domains of different sizes, centred on the cooling intake, is attempted. The four domains of sizes respectively denoted 800 m, 2 km, 7 km and 8 km, presented in Subsection A, are compared.

In this section the Strickler formula is used for all domain extensions. The following uncertain parameters are considered for UQ: three for BC calibration (MTL , CLT , CTV) and two Strickler coefficients at the intake (K_1 , red zone in Fig. 1-b) and at sea (K_2). The inputs bounds are described in Section II.B. In particular, the exact calculation for K_1 and K_2 using the formula $K = B/k_s^{1/6}$ and the bounds from Section II.B for B and k_s gives the interval $[21.02, 90.66] m^{1/3}s^{-1}$.

A number of 1000 MC simulations is launched on each domain. The min-max values of velocity u and examples of MC realizations are shown in Fig. 4 for two domains. The 2 km domain reaches higher velocities than the 7 km domain, with the same uncertain parameters and bounds. Simulation examples (plots in colours in Fig. 4) show abrupt variations at low tides with the 2 km domain. Conversely, the 7 km one shows a smoother behaviour. For both domains, measurements fall within the modelled interval. Last, whatever the modelling domain, a slight time-lag between measurements and numerical simulations is observed.

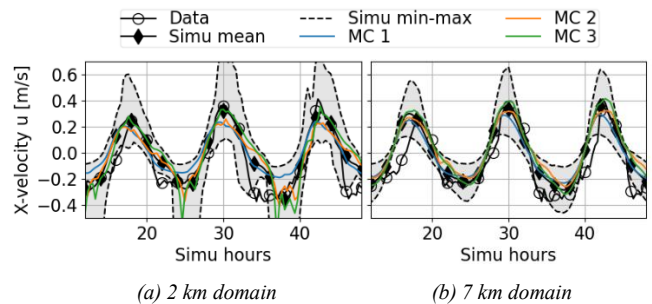


Figure 4- MC min-max envelope of x-velocity u extracted at Point 5, for domains of size 2 km and 7 km, and examples of MC realizations in colours.

Investigation of the 2 km domain shows the presence of tidal flats at BC, which might be the cause of velocity oscillations. Indeed, an extraction of the hydrodynamic variables on the 2 km contour at low tide is performed in Fig. 5. It shows that the velocity extremums are much higher for

the 2 km domain (Fig. 5-b) than for bigger extents. These higher velocities are due to the negligible water depths on the same locations (Fig. 5-a). Indeed, the velocity BC provided by the TPXO data base are not directly $(u, v)^T$, but transport fluxes $(uh, vh)^T$. Division by water depth h causes the velocity BC to be infinite when h is negligible. An often used practical engineering solution is to deepen the bathymetry at BC. Nevertheless, this solution is not used to avoid additional uncertainties and the impact of propagation of this error at BC is studied and compared to other domains.

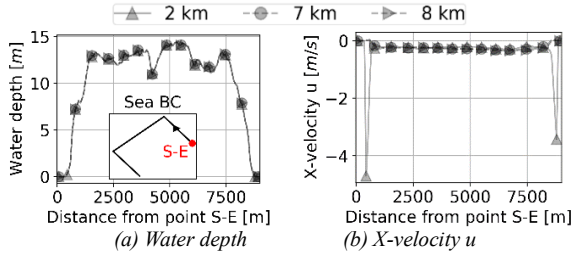


Figure 5- Hydrodynamic variables extracted on the 2 km sea BC at low tide for three different domain extents.

POD is used in order to analyse the full time series variations at a given point. The EVR is shown for example on Point 1 in Fig. 6. The problem is highly reducible. Free surface elevation reaches 99% variance with two modes, for all domain sizes. The relative RMSE (Root Mean Squared Error) between the simulation results and a 2-Mode POD reduction, averaged over the MC sample, is around 0.3%. Velocity components show different behaviours for different domains. For example, the y-velocity v is much less reducible for the 2 km domain, probably linked to the numerical error at the BC. The average relative RMSE, between the simulations and a 2-Mode POD, is around 5.8% with the 8 km domain, while it reaches 8.4% with the 800 m domain. It can also be noticed that the 800 m, 7 km and 8 km domains EVRs are ordered by domain size and 99% of the variance is captured with 2 modes, except for 2 km.

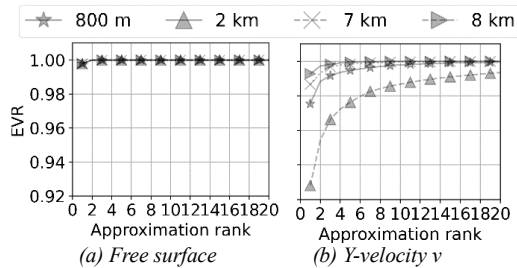


Figure 6- EVR for two hydrodynamic variables after MC simulations POD reduction on Point 1.

Next, free surface modes are shown in Fig. 7 for Point 1, where numerical and measurements modes are compared. Both modes are identical for all domains, and their shapes are comparable to measurement modes. However, the amplitudes of numerical Mode 1 are smaller than measurements. This difference can be explained by the represented information. Measurements Mode 1 is characterized by larger amplitudes, because measurements tidal coefficient varies, whereas the numerical Mode 1 only represents MC stochastic variation around the same tide. Furthermore, Mode 1 is always strictly

positive, be it for the numerical or real mode. When multiplied by a positive coefficient (corresponding to a given MC realization or a given measurement period), it stays positive and translates vertically, i.e. the mean tidal level changes. Mode 2 oscillates from negative to positive, but not symmetrically. Adding it to Mode 1 corrects both tidal range and mean. It can be noticed that the periods of Mode 1 and 2 are approximately semi-diurnal, which can for example be compared to the principal lunar and solar semidiurnal tidal components (M2 and S2). It should be noted however that POD modes have no theoretical reason to recover the tidal harmonics, as would be the case with a Fourier decomposition. This is rather related to the statistical importance of such harmonics in the global variance of the system. Lastly, a phase shift is observed between simulations and reality, for both modes. This phase is more important at ebb than at flood. It may correspond to the un-modelled effect of tide-surge interactions. Indeed, as explained in [17], the latter can lead to more surge at low than at high tide, resulting in a phase lag where the surge precedes the high water by few hours.

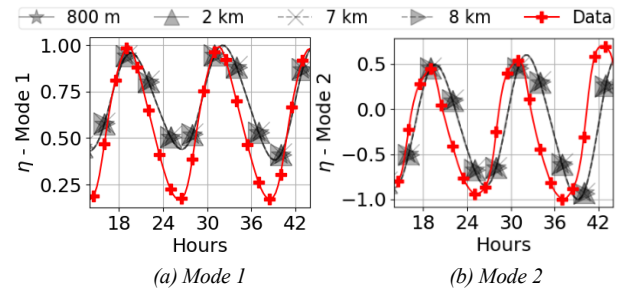


Figure 7- Comparison of the first two modes of free surface to reality, after reduction on point 1, for all tested domain sizes.

Comparison of X-velocity modes is shown in Fig. 8. First, differences between domains can be observed, and are bigger in Mode 2 than in Mode 1. Once again, the domain of size 2 km behaves differently even for Mode 1 (oscillations). This is a numerical artefact, as no oscillations are noticed in the measurements.

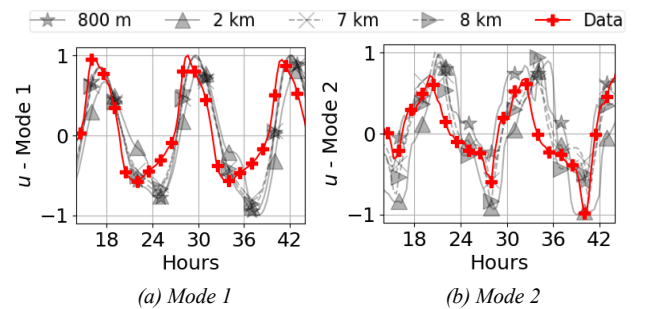


Figure 8- Comparison of the first two modes of x-velocity u to reality, after reduction on point 1, for all tested domain sizes.

Secondly, no model succeeds in reproducing the asymmetry observed at the extremums of measurements for Mode 1. Something is missing in the modelling that cannot be corrected by domain extent (Coriolis force, turbulence, waves, storm and atmospheric surges, non-linear interactions of tides with the latter, more precise bathymetry, etc.). For Mode 2, the overall numerical behaviour is comparable to the real one. It seems however that the domain of size 2 km fails at

reproducing the sharp minimal peaks (for example at Hour 40 in Fig. 8-b) that the other models capture better. The same conclusions hold for Y-velocity v , with oscillations at the minimum noticed with the numerical simulations, and not present in the real mode. For modes of higher rank, no similarity between the numerical results and the measurements is observed.

Next, Sobol' indices of the inputs are compared for all domain sizes. They are calculated using degree 3 PCE models of the POD temporal coefficients, as explained in Section II.C.3. The MC sample is separated to a training set (80%) and a test set (20%) and PCE models are learned on the training set. For example, with the 8 km model, it can be noted that the 90th percentile of the relative RMSE, between the simulations and a 2-Mode POD-PCE meta-model, calculated on the test set, remains below 0.3% for the free surface, and below 6.1% for the velocities. Sobol' indices results for free surface and X-velocity u are shown in Fig. 9. For free surface, the only influencing variable for Mode 1 is MTL (Mean Tidal Level). For Mode 2, the influence of coefficient CTL is dominant, and MTL is in second position. This is coherent with previous interpretation of free surface modes. No differences between domain sizes are noticed. For the velocity components, influences change with domain sizes. For Mode 1, the most influencing variable is CTV for all domain extents, followed by the Strickler coefficient K_2 . This tendency is inverted for Mode 2. However, CTV is more important for the smallest domain of 800 m, whereas K_2 is less important, probably due to the spatial proximity of the BC to the analysed Point 1. For the 2 km domain, K_2 is much more important than for the other extents for both modes. This can be explained by the higher velocity values, as K_2 interacts with velocity in the shear stress formula (2). An interaction between K_2 and CTV is noticed (undashed portion), also explained by the shear stress formula.

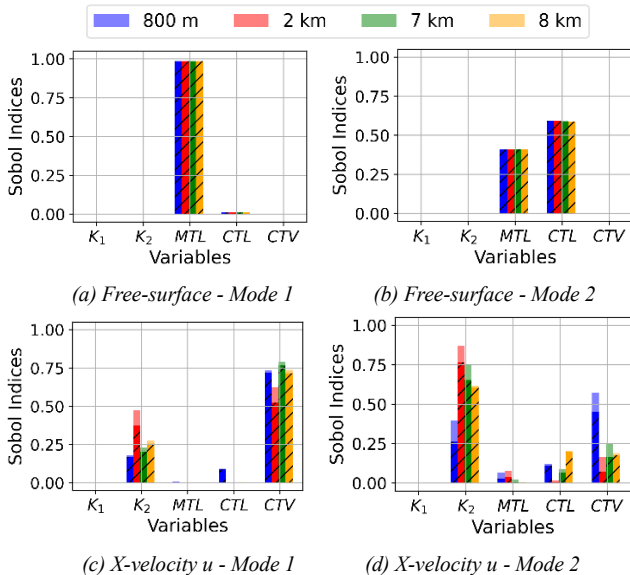


Figure 9- Sobol' indices for the two first modes of free surface and x-velocity u on Point 1, with the different domain sizes. Full bar plot represents total Sobol' indices. The dashed portion corresponds to the 1st order Sobol' index, and the remaining to the interaction with other variables.

Slight differences are observed between the 7 km and 8 km domains, namely more importance in Mode 1 for K_2 and CTV with 7 km, which could be explained by higher proximity of the BC. The differences are larger for Mode 2, but the latter is associated to less variance percentage. Moreover, the Strickler coefficient dominance for Mode 2 means that the observed velocity peaks in Fig. 8 can be moderated by friction, except for the smallest domain, where BC is dominant. The smallest the domain, the more controllable it is with BC only. Last, intake's friction K_1 has no influence outside. Calibration outside can hence be performed without controlling K_1 .

C. Friction closure influence

The same analysis strategy is adopted to study the Strickler and Colebrook friction formulas, introduced in Section II.B. The same domain of size 8 km with the bathymetry and mesh described in Subsection III.A are used. Eight uncertain parameters are used for Colebrook's UQ: three for BC (MTL , CTL , CTV), roughness heights at the intake (k_s^1) and at sea (k_s^2) and three structural uncertainty parameters (B_1 , B_2 and B_3). With the Strickler formula, six uncertain parameters are used: three for BC, two roughness heights, and one structural parameter B . All inputs and corresponding bounds are described in Section II.B. Temporal responses at the five measurement points are reduced using POD. No differences can be observed in terms of EVR, all variables can be reduced to 2 modes for over 99% of variance. The modes shapes are compared, and no differences are observed in the free surface first two modes. Slight differences can be seen in the velocity modes, as shown in Fig. 10, particularly for Mode 2 for which small oscillations appear with Strickler's model and not with Colebrook's. No formula succeeds however in reproducing the observed asymmetry of measurements Mode 1, and a phase can be noticed for both modes. Apparently, this lack of asymmetry representation cannot be corrected with friction closure modification either.

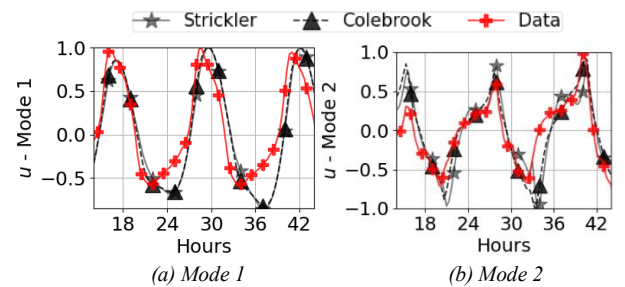


Figure 10- Comparison of the first two modes of x-velocity u to reality, after reduction on point 1, for two friction formulas.

SA is performed and no differences are observed for the free surface modes. Sobol' indices of the X-velocity u modes are shown in Fig. 11. Naturally, Sobol' indices of B are plotted only for Strickler's model, and those of B_1 , B_2 and B_3 are plotted only for Colebrook's. Slight differences are observed in Fig. 11.

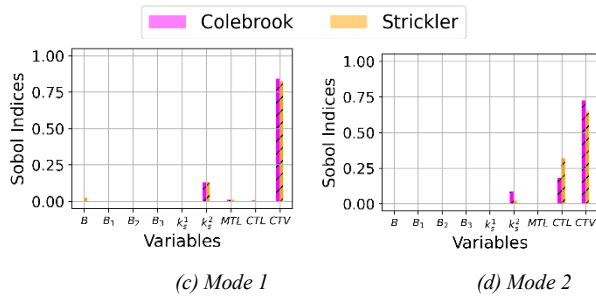


Figure 11- Sobol' indices for the two first modes of free surface and x-velocity u on Point 1, with different friction formulas. Full bar plot represents total Sobol' indices. The dashed portion corresponds to the 1st order Sobol' index, and the remaining to the interaction with other variables.

For Mode 1 there is a slight influence of the structural parameter B for the Strickler model, whereas the parameters of Colebrook's model show no influence. Additionally, the velocity BC correction denoted CTV gains influence with Colebrook, whereas the sea rugosity k_s^2 loses impact. The increase of influence for CTV with Colebrook's model is more important for Mode 2. The ranking of variables is however the same, even though the proportion of influence changed.

IV. IMPLICATIONS OF EXPERTLY MADE CHOICES ON OPTIMAL CALIBRATION

An example of the uncertainty of optimal calibration is shown on the domain extents variation. A 3DVAR algorithm is used, with the same configurations, parameters and bounds, for all domains. The measurement is reduced to 2 modes, and the calibration is performed on the learned 2-modes POD-PCE meta-models for each numerical simulation, by minimizing the cost function in (14). This allows the observation to be simpler (smoothing) and the problem dimension to be lower (2 coefficients per each variable, instead of 48 hours temporal series). The observation's error covariance matrix \mathbf{R} is considered diagonal (no error correlations) and calculated from measurement errors (5 cm for free surface and 1 cm/s for velocity). The background of each input parameter is set to the average of its variation interval, and the background's error covariance matrix \mathbf{B} , also considered diagonal, is estimated from the interval's variance (squared maximum distance from mean). Calibration results are shown in Fig. 12.

Firstly, it can be noticed that while oscillations are slight with the domain of size 2 km at point 4, they become more and more important when approaching the intake, at Point 3 where they gain amplitude and at Point 1 when the response becomes completely uncontrolled. Without specific correction, this domain is therefore of no interest in operational conditions. Secondly, the velocities represented with size 800 m are of lower amplitudes than with the largest domains of sizes 7 and 8 km. For Point 3 for example in Fig. 12-b, the 800 m domain fits better the minimum velocities than the biggest domains, and vice-versa at the maximum velocities. However, it can be seen in Fig. 12-c that the velocity shapes and amplitudes modelled by the biggest domains are more realistic than with the smallest ones. Lastly, even though domains of sizes 7 and 8 km show similar behaviour, slight differences can be observed. For Point 3 for example in Fig. 12-b, the extrema

reached by domain of size 8 km are higher than with the 7 km domain. Hence, even with similar choices for domain extent (7 and 8 km), the best fit could be uncertain. Physical analysis and prediction are therefore also uncertain.

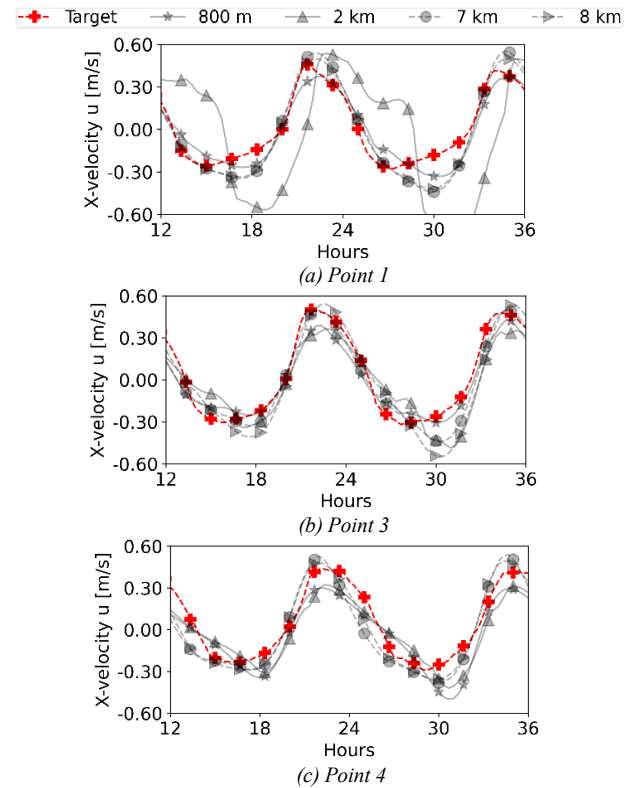


Figure 12- Example of optimal calibration with the tested domain extents, for x-velocity u at points 1, 3 and 4.

In particular, a comparison of X-velocity u profile at the intake's entrance for domains of sizes 7 and 8 km, at half ebb tide, is given in Fig. 13, after 3DVAR calibration. As a reminder, no calibration point is available on this profile. It can be noticed that differences are bigger and analysis is even more uncertain in locations where no measurement is available, even though both domains were calibrated.

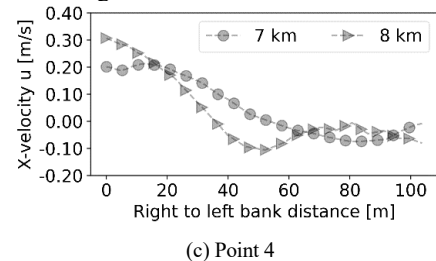


Figure 13- Intake's entrance cross-sectional profile of x-velocity at half ebb tide after 3DVAR calibration.

Last, the optimal parameterizations are different with the four domain sizes. For example, the Strickler coefficient at the sea equals $33.9 \text{ m}^{1/3} \text{ s}^{-1}$ for the domain of size 8 km whereas it is equal to $55.85 \text{ m}^{1/3} \text{ s}^{-1}$ for the domain of size 800 m, when CTV equals 4.16 for the 8 km domain and 4.38 for the 800 m domain. This shows that, in a coastal configuration, the generalization of such parameters, friction coefficient for

example, is hardly possible, without further investigations about their meaning, as optimal values are strongly related to the numerical choices that precede calibration. Additionally, the optimal values for such parameters (e.g. friction coefficient) can be time-varying. Therefore, a dynamic optimal-fitting procedure may be necessary.

V. CONCLUSION

In this study, uncertainties resulting from standard modelling choices in hydrodynamics, i.e. domain size and closure choice for friction were studied.

Firstly, measurements were POD reduced, and their patterns compared to numerical POD patterns resulting from different modelling configurations in a UQ framework. While similarities between observations and simulations are identified, some particularities present in the measurements, such as tidal velocity asymmetry, could not be modelled with the chosen configurations. This emphasizes the need of additional physics to be incorporated into the model, for example the Coriolis force, turbulence, waves, storm and atmospheric surges, non-linear interactions of tides with surge, or more precise bathymetric information. Noticing the lack of accord between the numerical and measurement-based patterns might help saving High Performance Computing resources: it is vain to try calibrating these configurations to capture particularities of the measurements when their modes behave differently, i.e. when the corresponding physical process is missing in the model.

Secondly, SA was performed on each modelling configuration. Differences were noticed, in terms of importance ordering and influence magnitude. It also indicates that the smaller the domain, the more important BC are, which is a coherent conclusion. This could have consequences on calibration processes. Indeed, as a last investigation, a 3DVAR algorithm was tested, using the same parameters, on all modelling configurations. In this context, POD reduction on both measurements and simulations along with PCE meta-models were used to dramatically reduce the computational time required by the 3DVAR algorithm. The results show that the obtained optimal states and associated optimal parameters may differ, even with close modelling choices (e.g. close domain extents), which highlights on uncertainties inherent to common modelling choices.

ACKNOWLEDGEMENT

The authors acknowledge the French National Association of Research and Technology (ANRT) for funding, and are grateful to the TELEMAT-MASCARET (hydrodynamics), OpenTURNS (uncertainty quantification) and ADAO (data assimilation) developer communities.

REFERENCES

- [1] J.-F. Gerbeau and B. Perthame. *Derivation of Viscous Saint-Venant System for Laminar Shallow Water; Numerical Validation*. Research Report RR-4084, INRIA, 2000.
- [2] 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*. Journal of Hydraulic Research, 46: 191-208, 03 2008.
- [3] M. Marriott and R. Jayaratne. *Hydraulic Roughness – links between manning's coefficient, nikuradse's equivalent sand roughness and bed grain size*. Proc. 5th Annual Conf. on Advances in Computing and Technology, 01 2010.
- [4] H. H. Barnes. *Roughness characteristics of natural channels*. Number 1849. US Government Printing Office, 1967.
- [5] B.C. Yen. *Open channel flow resistance*. Journal of hydraulic engineering, 128(1) : 20-39, 2002.
- [6] L. C. van Rijn. *Unified view of sediment transport by currents and waves. I: Initiation of motion, bed roughness, and bed-load transport*. Journal of Hydraulic Engineering, 133(6) : 649-667, 2007.
- [7] G. D. Egbert and S. Y. Erofeeva. *Efficient Inverse Modeling of Barotropic Ocean Tides*. Journal Atmospheric and Oceanic Technology, 19 (2) : 183-204, 02 2002.
- [8] C.-T. Pham and F. Lyard. *Use of tidal harmonic constants databases to force open boundary conditions in telemac*. In Proceedings of the XIXth TELEMAT-MASCARET User Conference 2012.
- [9] J.-M. Hervouet. *Hydrodynamics of Free Surface Flows : Modeling with the Finite Element Method*. John Wiley & Sons, Ltd : Hoboken, NJ, USA, 2007.
- [10] K. Fisher and F. Dawson. *Reducing uncertainty in river flood conveyance: Roughness review*. 07 2003.
- [11] Lumley, J. L. (1967). *The Structure of Inhomogeneous Turbulent Flows*. Proceedings of the International Colloquium on the Fine Scale Structure of the Atmosphere and Its influence on Radio Wave Propagation, edited by A. M. Yaglom, and V. I. Tatarsky, Doklady Akademii Nauk SSSR, Nauka, Moscow.
- [12] L. Sirovich. *Turbulence and the dynamics of coherent structures: I, II and III*. Quarterly Applied Mathematics, 45: 561, 1987.
- [13] Y. Caniou. *Global sensitivity analysis for nested and multiscale modelling*. PhD thesis, 2012.
- [14] B. Sudret. *Global sensitivity analysis using polynomial chaos expansions*. Reliability Engineering & System Safety, 93(7): 964-979, 2008.
- [15] A. Carrassi, M. Bocquet, L. Bertino and G. Evensen. *Data assimilation in the geosciences: An overview of methods, issues and perspectives*. Wiley Interdisciplinary Reviews: Climate Change, 9(5):e535, 2018.
- [16] J.-P. Argaud. *User Documentation in SALOME 7.5 platform of the ADAO module for Data Assimilation and Optimization*. EDF R&D report, 2016.
- [17] Idier, D., Bertin, X., Thompson, P. and Pickering, M.D., 2019. *Interactions between mean sea level, tide, surge, waves and flooding: mechanisms and contributions to sea level variations at the coast*. Surveys in Geophysics, 40(6), pp.1603-1630.
- [18] Shom. *MNT Bathymétrie de façade Atlantique (Projet Homonim)*. 2015. http://dx.doi.org/10.17183/MNT_ATL100m_HOMONIM_WGS84
- [19] Machiels, O., Erpicum, S., Archambeau, P., Dewals, B. and Pirotton, M., 2009. *Bottom friction formulations for free surface flow modeling*. In 8th NCTAM Congress
- [20] Blatman, G. and Sudret, B., 2011. Adaptive sparse polynomial chaos expansion based on least angle regression. *Journal of computational Physics*, 230(6), pp.2345-2367.

