

INVESTIGATING METAMODELING CAPABILITY TO PREDICT WATER LEVELS AND MARINE FLOODING MAPS FOR EARLY-WARNING SYSTEMS: APPLICATION ON THE ARCACHON LAGOON (FRANCE)

Sophie Lecacheux¹, Jeremy Rohmer¹, Eva Membrado¹, Rodrigo Pedreros¹, Andrea Filippini¹, Deborah Idier¹, Servane Gueben-Vénière², Denis Paradis³, Alice Dalphiné³, David Ayache³

¹BRGM, 3 av. C. Guillemin - 45060 Orléans - France

²KEYROS, 3 bis rue Jules Vallès - 75011 Paris - France

³METEO-FRANCE, DIROP/MAR/DAS - Toulouse

Correspondence to: S. Lecacheux (s.lecacheux@brgm.fr)

Abstract - Marine flooding events during storms are expected to occur more frequently due to sea level rise. Hence, early warning systems (EWS) dedicated to marine flooding are expected to develop in the coming years. In this study, we compare three data-driven methodologies to overcome the computational burden of numerical simulations. They are all based on the statistical analysis of pre-calculated databases, to downscale total water levels and to predict marine flooding maps from offshore metocean operational forecasts. While the first one is a simple analog-based research from offshore metocean conditions, the next two both use a machine learning type metamodel to predict total water levels at the coast, and either an analog or a deep-learning approach to predict marine flooding maps. The analysis, carried out with a cross-validation exercise and historical storms on the pilot site of Arcachon lagoon (Southwest of France), reveals that the analog-based approach is a valuable first step to explore the dataset and improve the understanding of flooding phenomena, but lack precision for operational forecast applications. On the other hand, the two metamodel-based approaches are more suitable for fast prediction with a lower prediction error of inland water heights for the deep-learning approach (on the order of 10 cm). Both approaches can then be complementary depending on the type of event, the required level of prediction accuracy to support operational decision making, and the forecast lead time. In this sense, the study also underlines the usefulness of precalculated databases to conduct a preparatory work with crisis managers to determine the type of information and the right level of complexity required to address operational needs.

Keywords: Marine flooding, maps, forecast, metamodels, Arcachon lagoon

1 **1 Introduction**

2 The high rates of population growth and urbanization in coastal regions tend to constantly
3 increase marine flooding risks in low-lying areas. Projections estimate the number of people
4 living in coastal flood-prone areas at more than 200 million by 2100 (Hauer et al., 2021). Marine
5 flooding is a phenomenon resulting from the combination between various processes generated
6 at different time and space scales (atmospheric circulation, waves, atmospheric surge, tide, and
7 sometimes river discharge) and the local configuration of the coast (coastal bathymetry and
8 elevation, protection structures, land use, hydraulic networks, etc.). The numerous variables,
9 scales and sources of uncertainty make marine flooding events very complex to predict several
10 hours or days in advance. However, marine conditions forecasts (Toledano et al., 2022; Lorente
11 et al., 2019) and coastal flood early warning systems (Stansby et al., 2012; Dietrich et al., 2013;
12 Apecechea et al., 2023) have gained a significant impulse in the last decades. Today, there is a
13 growing demand not only to shift towards ever more local marine flooding forecasts, but also
14 to better account for uncertainties, which requires approaches using multi-sources or even
15 probabilistic metocean conditions (see a recent discussion by Turner et al., 2024).

16 Recent improvements in high performance computing enabled numerical weather prediction
17 systems to move from deterministic to probabilistic forecasting using Ensemble Prediction
18 Systems (EPS) (Descamps et al., 2015). While EPS are increasingly used to predict river flows
19 and induced continental floods in several countries (Wu et al., 2020), it is only emerging for
20 marine flooding (Hawkes et al., 2008; Lecacheux et al., 2018; Biolchi et al., 2022). Despite
21 ongoing efforts to develop new generations of high performance oceanographic and phase-
22 resolving models accounting for complex processes (Filippini et al., 2018), the main challenge
23 still remains the computer power required to run multiple simulations with a chain of models
24 of increasing resolutions (from a hundred meters for coastal waves and surges to a few meters
25 for marine flooding). Most of the examples of – deterministic or probabilistic – systems
26 forecasting explicitly marine flooding rely on three solutions to overcome these limitations: (1)
27 High Performance Computing (HPC); (2) reduced process complexity with generalized
28 overflowing models or empirical formula; and (3) statistical analysis of pre-calculated flooding
29 scenarios (Lecacheux et al., 2021; Beuzen et al., 2019). Now, operational needs and constraints
30 of potential users must guide these technical choices. It is about linking top-down and bottom-
31 up approaches by connecting technical capabilities with local users' needs to identify the

1 suitable level of complexity able to produce the right level of information (parameters,
2 resolution, and precision) for decision making (Demerri et al., 2016; Tarchiani et al., 2020).

3 In this study, we focus on methodologies based on pre-calculated databases to downscale total
4 water levels (TWL) and marine flooding maps from offshore metocean conditions as developed
5 at European scale in ECFAS project (Le Gal et al, 2023 and 2024). Databases are often used to
6 work by analogy, based on storm similarity with forecasted conditions, but this solution also
7 provides complementary possibilities to (1) improve the understanding of marine flooding
8 processes and identify key scenarios and action plans to facilitate foresight and anticipation in
9 crisis management (Luesink et al., 2024); (2) apply machine learning (denoted ML) type
10 methods to develop very rapid forecasting models (or metamodels) able to replace numerical
11 models to carry out simulations in real time. ML-based metamodeling techniques have made
12 great progresses in this field of application (ML-based metamodeling techniques have made
13 great progresses for flood predictions (storm surge: e.g., Macdonald et al., 2025; hurricane: e.g.,
14 Irwin et al., 2025; tsunامي: e.g., Ragu Ramalingam et al., 2025; estuarine: e.g., Wang et al.,
15 2025a; river flood: e.g., Wang et al., 2025b) and opens up encouraging perspectives for marine
16 flooding forecasting. Through a statistical analysis of pre-calculated training databases, they
17 can predict key flooding indicators (surge, discharge, water height on land, etc.) at a given
18 spatial location of interest within reasonable time and computing resources while preserving
19 the accuracy of full process models. Yet, some issues remain to push this metamodel-based
20 approach toward operational applications, and more specifically the production of spatialized
21 indicators such as inland water height maps, which is still a matter of active research research
22 (among others, Ma et al., 2022; Spiller et al., 2023; Fraehr et al., 2024; Rohmer et al., 2024a;
23 Jung et al., 2025). Indeed, producing such spatialized indicators requires to overcome the
24 difficulty in learning complex mathematical relationships between the metocean forcing
25 conditions and the generation of flood maps (i.e. output of high dimensional related to the
26 number of pixels, typically of several 10,000 s).

27 The objective of this work is to investigate different approaches, based on pre-calculated
28 databases, to produce information that is relevant for operational forecast applications. This
29 implies analysing not only the added value in terms of accuracy regarding the required
30 implementation effort but also in terms of adequacy according to the needs of crisis managers.
31 We consider three approaches with increasing complexity. The first one is a simple analog-
32 based method, which is based on the analysis of the similarity between the coming metocean

conditions and an individual event in the pre-calculated training database (also named as "look-up table"). The next two are ML-based methods aiming to supplement or substitute the analog-based approach with (1) a regression-type metamodel to predict total water levels in the lagoon; (2) deep learning techniques to predict the marine flooding maps. To support the analysis, we focus on the Arcachon lagoon (Southwest of France, Fig. 1) where a database of a few hundred flood scenarios have been simulated for the Service of Civil Defense in a previous study (Lecacheux et al., 2023) and a participatory work with crisis managers enabled to identify operational needs and critical thresholds associated with graduated categories of flood impacts.

In the following chapters, we present the pilot site, the databases and lessons learnt from their analysis with crisis managers in Section 2, the three approaches for total water levels and marine flooding prediction in Section 3, and the comparison of their skills in Section 4. In the discussion in Section 5, we analyze the pros and cons of the different methods regarding local users' needs while Section 6 summarizes the main conclusions and addresses the perspectives for application of these models for operational forecast.

2 Site, databases setup and analysis

2.1 Study area

Arcachon lagoon is a large low-lying area located in a semi-closed environment and mainly subject to marine overflows driven by offshore storm surge and wave forcing associated with local winds (Fig. 1, A). It is a well-studied site, which benefits from a rich bibliography, covering both its general functioning (Bouchet et al, 1997), the geological and morphosedimentary context (Saltel et al., 2014), the hydrological functioning (IFREMER, 1997) and oceanographic characteristics (Castelle et al., 2017; Dupuis et al., 2016). In recent years the lagoon has been regularly affected by minor to moderate marine flooding notably during storms Klaus (2009), Xynthia (2010), and more recently Céline (2023). As illustrated on Fig. 1 (B), these storms generated total water levels (tide + storm surge) from annual (~3 m NGF, NGF being the official vertical reference system to measure elevations in France) to approximately 10-years (~3.5 m NGF) return periods at the Arcachon-Eyrac tide gauge but there is no historical example of event beyond. Yet, statistical and hazard studies (Mugica et al., 2014; SHOM, 2022) demonstrated that a 100-year return period event could reach up to 3.7 NGF at the same location, then affecting almost ten thousand of persons (SIBA, 2015).

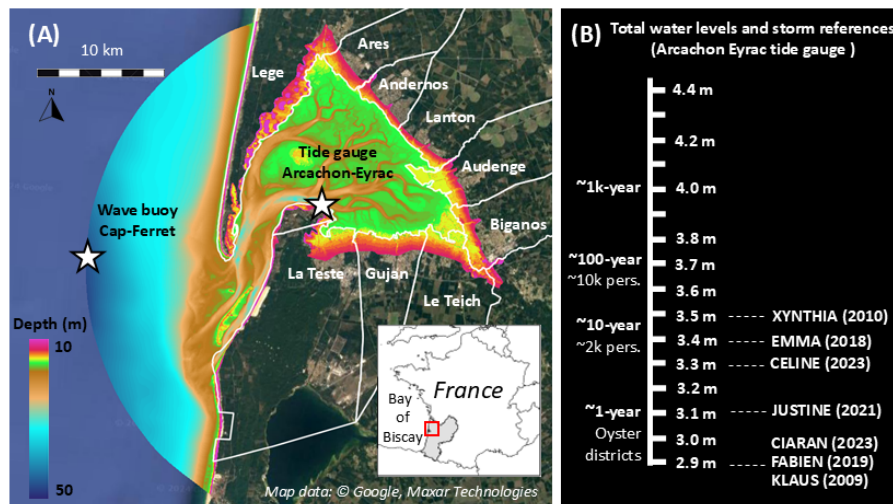


Figure 1: A) Location and bathymetric map of the study site: white stars indicate the location of the tide gauge named Arcachon-Eyrac and the wave buoy named Cap-Ferret. B) Total Water levels (m NGF) with associated return periods and historical storms references. Map data: © Google, Maxar Technologies

Based on this bibliography and historical events, the main elements to consider when modeling coastal floods on this area are:

- Marine flooding by overflowing is dominant on the scale of the entire lagoon even if local sectors can also be exposed to wave overtopping (as the seafront of Andernos, see location in Fig. 1, A).
- Waves have a significant effect on marine flooding through the generation of a homogeneous wave setup inside the lagoon (up to several tens of centimeters) generated by the breaking of the swell at entrance. of the lagoon).
- West/northwest winds are the most frequent and can generate an additional storm surge inside the lagoon particularly at the extreme east (from Biganos to Arès) where total water levels may reach several tens of centimeters higher than in Arcachon in case of strong winds;
- Although river floods are characteristic hazards of wetlands such as the Arcachon lagoon, available studies and measures do not enable to characterize the conjunction between storm surge and river floods on this site (in particular at the Leyre delta between Le Teich and Biganos).

In this study we focus on marine overflowing only, without considering neither wave overtopping (localized on the seafront of Andernos) nor the conjunctions of river flood and storm surge (located in the Leyre Delta). Thus, the study focuses on the simulation of the total steel water level or TWL (including tide, atmospheric surge caused by winds and pressure drop and the wave set-up caused by the breaking of waves in entry of the lagoon) and the resulting

coastal flooding. The dynamic water level (including the swash component) is not taken into account.

2.2 Design of scenarios for learning and validation of the metamodels

The synthetic storm conditions constituting the scenarios of the pre-calculated database (used for the analysis of the physical processes and for setting up the ML-based techniques, see Section 3) are built on the basis of a tri-variate statistical analysis of the extremes of wind, wave and surge conditions at the Cap-Ferret wave buoy located at a depth of 50 m in front of the lagoon (Fig. 2, A). The reference data are extracted at high tide from reanalyses over the period 1979-2009: CFSR (Saha et al., 2010) for wind, ANEMOC 3 (Raoult et al., 2018) for waves and MARS30 (Mugica et al., 2014) for the storm surge. After fitting probability laws for significant wave heights (Hs), skew surge (SPM, defined as the difference between the maximum simulated water level and the predicted high tide level) and wind intensity (U) using the Generalized Pareto law (Coles et al., 2001), the dependence models are adjusted for U, Hs and SPM according to the conditional extreme model of Heffernan and Tawn (2004) while managing the covariates (peak period - Tp, wave directions - Dp and winds - Du). Finally, many random combinations (here chosen at 150,000) with the same statistical characteristics as the input data are generated via Monte Carlo simulations. To constitute the learning database, 50 moderate to strong storm conditions (Hs>4m) are selected via a clustering algorithm (Camus et al., 2011) aimed at maximizing the diversity of scenarios (Fig. 2, A). These 50 conditions are associated with 10 tide levels between coefficients 40 and 120 (i.e. varying from 1 m to 2.8 m NGF at the tide gauge every 20 cm), bringing the total number of scenarios to 500. Each scenario is therefore described by 6 storm parameters at the buoy (U, Du, Hs, Tp, Dp, SPM) and a high tide level (T) at the gauge.

These synthetic events are completed with historical and pseudo-historical events for validation and demonstration purposes. Historical events (Table 1) correspond to 8 storms (Klaus, Xynthia, Emma, Fabien, Justine, Ciaran, Domingos, Céline) that occurred between 2009 and 2023 and whose forcing conditions are extracted from observed winds and waves as well as reanalysed storm surge (from Météo-France Arpège model) at the Cap-Ferret wave buoy location. Table 1 and Fig. 2 (A) show their characteristics at high tide and Fig. 2 (B) an example of parameters time series for storm Xynthia. Although they are well scattered in the synthetic dataset range, these events remained quite limited in terms of total water level (Fig. 1, B) and marine flooding impact (as mentioned in Section 2.1) because the most intense storms (Fabien,

Klaus and Domingos) occurred at low tide coefficients. Thus, pseudo-historical events are added by combining these 8 storms (described by parameters U, Du, Hs, Tp, Dp, SPM) with the same ten high tide levels (T) as the synthetic events, to create marine flooding events of different severity (greater or lower than historical events). This results in 80 additional events hereafter named *Pseudo_StormName_TideLevel(NGF)*.

	XYNTHIA	KLAUS	EMMA	JUSTINE	FABIEN	CELINE	CIARAN	DOMINGOS
SPM (m)	0.44	0.49	0.33	0.21	0.38	0.27	0.3	0.47
Hs (m)	6.7	7	6.46	8	8.57	4.25	7.36	8.78
Tp (s)	12	14	11	16	13.5	14	13	13.5
Dp (°)	266	277	251	273	276	290	280	275
U (m.s ⁻¹)	23.5	28	23.5	19.5	28	11.5	20	25
Du (°)	250	230	250	280	270	190	270	250
Tide (m)	2.74	1.73	2.71	2.46	1.85	2.77	2.07	1.3

Table 1: Storm parameters (U, Du, Hs, Tp, Dp, SPM) at high tide for the 8 historical storms.

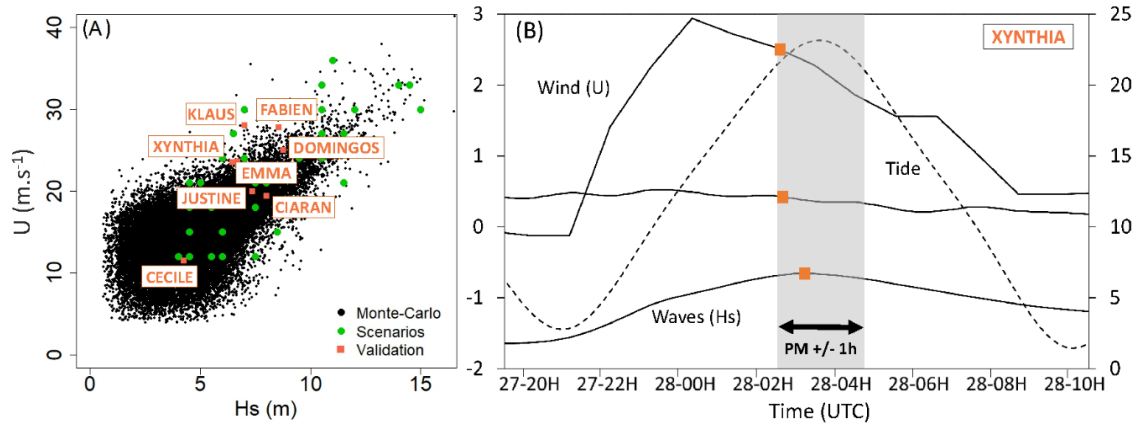


Figure 2: A) Monte-Carlo simulations (black points), synthetic (green points) and historical or validation (orange points) storm conditions at high tide. B) Example of storm parameters time series and extraction of conditions at high tide (orange points) for storm Xynthia (2010).

2.3 Flood numerical model set up and simulation of the scenarios

The marine flooding model set up for the project includes a chaining of WW3 spectral wave model (Tolman, 2014) and UHAINA hydrodynamic models (Filippini et al., 2024) that resolves Saint-Venant equations. The grid is an unstructured mesh with a resolution from kilometric offshore to decametric inside the lagoon (from 20 m to 50m) and on land (~10 m) (Fig. 3). This chain of models is forced by tidal harmonics, atmospheric surge and waves at the western boundary (that passes through the Cap-Ferret buoy) and by the wind on the entire domain. It can simulate (1) the propagation of tide, atmospheric surge and waves inside the lagoon; (2)

the generation and propagation of the additional surge induced by local winds and the breaking of waves and (3) the marine flooding by overflowing. Despite the resolution of the mesh remains limited on land, the model represents the main obstacles for water flows thanks to the line constraints that represent the protection structures (dikes), the little walls, the embankments and the roads. For each point of the grid, an altitude is extracted from a digital terrain model (DTM) built with a combination of bathymetric surveys with resolutions of 100 m offshore and 10 m inside the lagoon as well as LIDAR data from 2016 survey on land. Finally, the land cover is represented by a manning coefficient extracted from friction maps (Mugica et al., 2014).

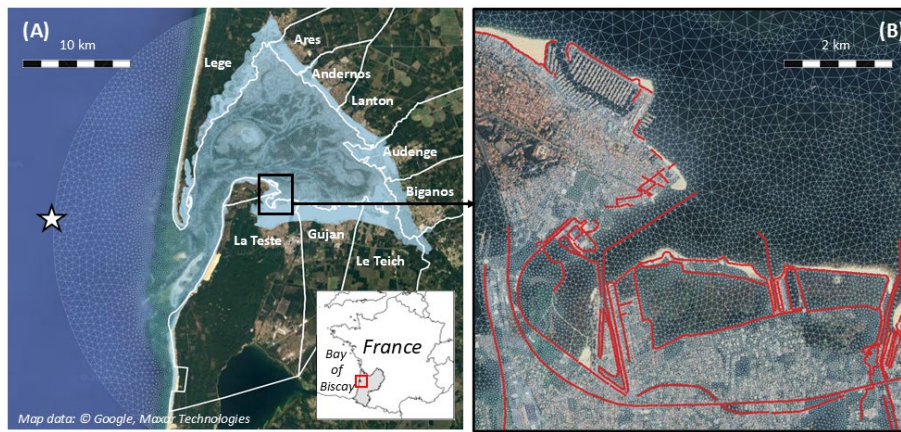


Figure 3: A) Unstructured mesh used in the modelling chain (UHAINA and WW3) at the scale of the lagoon. B) Zoom centred on La-Teste-de-Buch. Map data: © Google, Maxar Technologies

For historical and pseudo-historical events, forcing conditions come from time-varying observations at the Cap-Ferret buoy (for waves and wind) and reanalysed data of Météo-France Arpège model (for storm surge). For synthetic events, it corresponds to the set of parameters at high tide determined with the statistical analysis (U, Du, Hs, Tp, Dp, SPM) that are applied as stationary conditions during the tidal cycle. Thus, a fundamental difference between synthetic cases (for learning of metamodels) and historical or pseudo-historical cases (for validation of metamodels) is the consideration of temporality.

To evaluate the influence of the temporal variations of these conditions before and after high tide, sensitivity tests are carried out by simulating the 8 historical storms of the validation set on a tidal cycle with both dynamic conditions (i.e. with time varying time series) and stationary conditions (i.e. with the conditions extracted at high tide ± 1 h). For the maximum total water level over time (TWLmax), the comparisons between the 8 storm simulations (dynamic and stationary) and tide gauge observations show errors less than 10 cm (Table 2). More generally,

table 3 shows that differences between dynamic and stationary simulations is less than 10 cm from Arcachon to Biganos (on the southwest part of the lagoon) and less than 14 cm from Audenge to Lège (on the northeast part of the lagoon) where the effect of strong winds on the amplification of storm surge is the most important (like for Domingos). Regarding flooded surfaces and maximum water height on land over time (Hmax), the flooded sectors are generally well reproduced and very similar for dynamic and stationary simulations. Figure 4 presents an example for Xynthia (in Gujan-Mestras and Le Teich) and shows that dynamic and stationary simulations result in very similar patterns as observations despite TWLmax differences of 8 cm. The interested reader will find further information about simulations sensitivity tests in Lecacheux et al. (2013). They corroborate that (1) the modeling chain reproduces correctly TWLmax and Hmax on the ten municipalities (2) TWLmax and Hmax are mainly controlled by conditions at high tide, which justifies the use of storm parameters at high tide (U, Du, Hs, Tp, Dp, SPM) in the simulations.

	XYNTHIA	KLAUS	EMMA	JUSTINE	FABIEN	CELINE	CIARAN	DOMINGOS
OBS	3.58	2.92	3.38	3.19	2.92	3.37	2.86	2.46
DYN	3.51	2.9	3.4	3.2	2.85	3.23	2.82	2.36
STAT	3.59	2.86	3.41	3.2	2.9	3.24	2.84	2.39

Table 2: Comparison between TWLmax (in meters) at Arcachon tide gauge for the 8 historical events observed (OBS), modelled with dynamic (DYN) or stationary simulations (STAT).

	LA TESTE	GUJAN	LE TEICH	BIGANOS	AUDENGE	LANTON	ANDERNOS	ARES	LEGE
MIN	0	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
MED	0.04	0.05	0.05	0.05	0.05	0.04	0.04	0.03	0.02
MAX	0.08	0.09	0.1	0.1	0.12	0.13	0.14	0.13	0.12

Table 3: Differences between TWLmax (in meters) simulated with dynamic and stationary conditions for the 8 historical events : minimum (MIN), median (MED) and maximum (MAX).

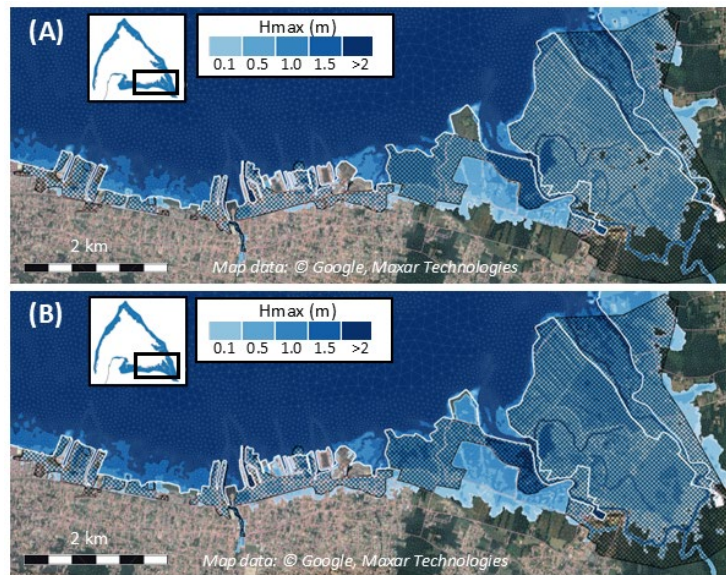


Figure 4: Comparison between observed and dynamic (A) and stationary (B) simulations of flooded areas for Xynthia in Gujan Mestras and Le Teich. Map data: © Google, Maxar Technologies

Scenarios of the learning and validation datasets are thus simulated on a tidal cycle by applying stationary conditions of surge, wind and waves on the entire tidal cycle. For each event, simulations are carried out sequentially, starting from the highest high tide level to the lowest, and stopping at the last overflowing level (corresponding approximately to TWLmax of 3 m NGF inside the lagoon). It induces that strong storm events are simulated with more tide levels than moderate storm events (because moderate events generate marine flooding only for the highest tidal coefficients). To sum up, the learning dataset is composed of 220 simulated scenarios and the validation dataset of 32 simulated scenarios (corresponding to the subset of the initial scenarios that generated marine flooding). They are all described by (1) forcing conditions of 6 storm parameters at the buoy (U , D_u , H_s , T_p , D_p , SPM) and a high tide level (T) at the gauge (2) maps of maximum total water level in the lagoon (TWL_{max}) and water height on land (H_{max}).

2.4 Overview and analysis of the datasets with crisis managers

TWL_{max} obtained with the simulations of learning scenarios vary from 2.6 m to 4.4 m NGF at the tide gauge. For comparison, the annual and centennial levels at the same location are respectively 3 m and 3.7 m NGF (Fig. 1). The learning dataset thus enables to describe a wide range of events from annual to exceptional.

The exposure of the municipalities surrounding the lagoon is quite variable. Figure 5 presents two maps showing the percentage of scenarios generating marine flooding (A) as well as TWLmax and Hmax induced by the most intense one (B). They highlight that:

- Municipalities with wetlands, either limited (La-Teste-de-Buch, Lanton and Ares) or quite extended (Le Teich, Biganos, Audenge) present a greater general exposure with a higher rate of flooding and higher potential Hmax (although it concerns non-urbanized areas).
- Municipalities without wetlands (Arcachon, Gujan-Mestras, Andernos, Lege-Cap-Ferret) are mainly exposed near the seafront with a rapid decrease of flood rate and potential Hmax when moving further inland.

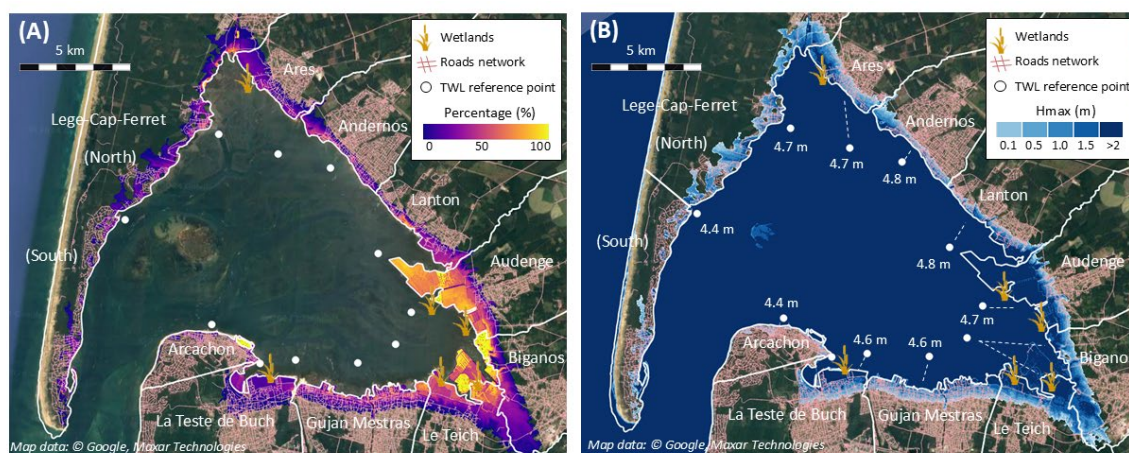


Figure 5: A) Percentage of scenarios (calculated from the learning database) leading to marine flooding on each node of the model mesh. B) Water levels denoted TWLmax on the municipalities reference points and Marine flooding map (Hmax) for the most intense scenario. Map data: © Google, Maxar Technologies

But being exposed does not necessarily mean being vulnerable. The impact of a flooding event depends not only on its severity, but also on the stakes involved and the challenges faced by crisis managers. Crisis managers need aggregated indicators to strengthen their capacity to understand the information and take actions. To address this issue on the Arcachon lagoon, collaborative work was realized with the municipalities to classify the scenarios into 3 categories (CAT) representing graduated levels of impact and related to specific action plans:

- CAT 1 corresponds to limited flooding events concerning only seafronts, port docks or oyster farming districts that are regularly subject to marine flooding and for which the municipalities are well prepared and autonomous to manage the events.

- CAT 2 corresponds to moderate flooding events concerning few residential areas, transport or energy networks, and for which logistical capacity of the municipalities is not exceeded.
- CAT 3 finally corresponds to major events (either by their extension or because critical infrastructures are affected) for which the municipalities need help to manage the events.

As marine flooding patterns are mainly controlled by local TWLmax, these 3 main categories (CAT 1, CAT 2, CAT 3) have been translated into thresholds of TWLmax on 10 reference points in front of each municipality (Fig. 5, B). These thresholds generally correspond to increments of about 30 cm in most cases (but the increments can reach 50 cm in some cases). Figure 6 and Table 4 show the thresholds of the categories for each municipality and the distribution of the scenarios of the learning database (box plots) and validation database (points) inside the categories. It enables to say that:

- Even if the municipalities comprising wetlands (notably Biganos, Audenge, Ares) are more exposed to marine flooding, their urban areas and critical stakes (are impacted for higher TWLmax than the other municipalities.
- The learning scenarios (box plots) are rather well distributed across the categories.
- Historical events do not exceed CAT 2 (with Xynthia) but pseudo-historical events complete the validation scenario relatively well to represent all the categories.

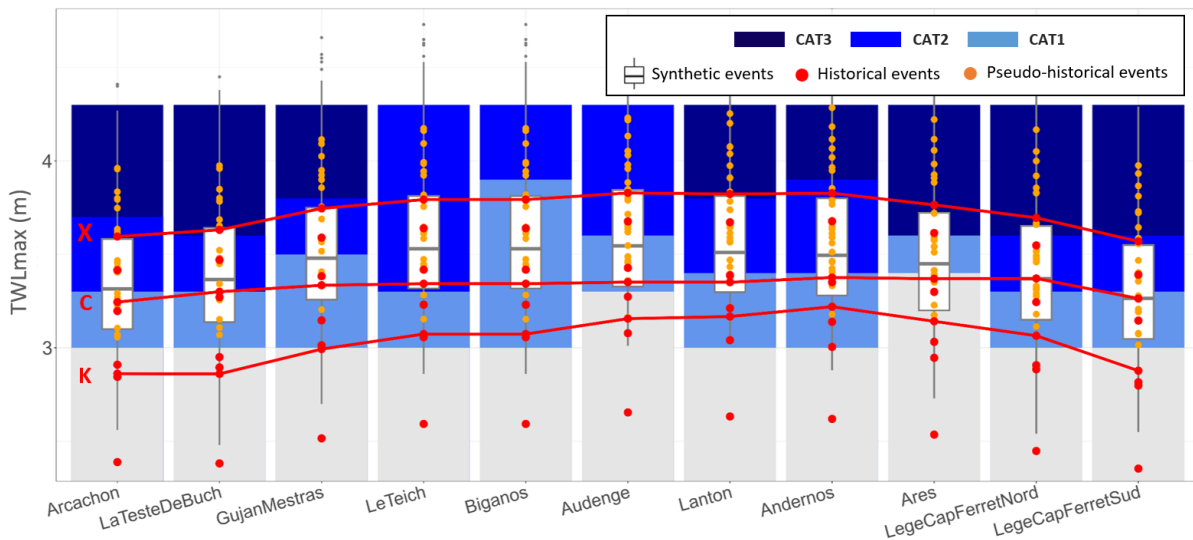


Figure 6: Categories of impact (CAT) related to TWLmax and repartition of the scenarios of the datasets: synthetic (box plots), historical (red points) and pseudo-historical (orange points). Red lines illustrate the variations of TWLmax for Klaus (K), Céline (C) and Xynthia (X).

	Dataset	Arcachon	LaTeste	Gujan	LeTeich	Biganos	Audenge	Lanton	Andernos	Ares	Lege-N	Lege-S
CAT 0	Learn.	36	26	5	2	2	49	0	6	107	28	45
	Histo.	3	3	1	1	1	3	0	1	5	3	3
	Pseudo.	0	0	0	0	0	1	0	0	6	0	0
CAT 1	Learn.	76	70	117	55	187	84	87	91	48	69	79
	Histo.	3	3	5	2	7	3	4	4	1	1	3
	Pseudo.	8	7	7	2	17	6	3	2	7	4	8
CAT 2	Learn.	83	76	60	179	47	103	87	96	0	70	61
	Histo.	2	1	2	5	0	2	2	3	0	3	2
	Pseudo.	10	7	10	22	7	17	12	15	0	9	9
CAT 3	Learn.	41	64	54	0	0	0	62	43	81	69	51
	Histo.	0	1	0	0	0	0	1	0	2	1	0
	Pseudo.	6	10	7	0	0	0	9	7	11	11	7

2 **Table 4: Number of events per category (CAT) in the learning database (Learn.), in the historical**
3 **dataset (Histo.) and the pseudo-historical dataset (Pseudo.) used for validation.**

4

5 This preliminary work highlighted the need to forecast water levels (TWLmax) and the
6 associated categories of impact (CAT) for each municipality first, so that crisis managers can
7 rely on predefined action plan typologies. Forecasting marine flood maps (Hmax) should then
8 be undertaken as a second step, in order to complement and refine the characterization of the
9 event within each impact category.. In the following, we thus propose different methods
10 enabling to articulate the prediction of both TWLmax and Hmax with different levels of
11 accuracy.

12 3 (Meta)modeling methods and performance assessment

13 In this section, we describe the formal framework for setting up the ML-based models
14 (named metamodels). Let us consider the maximum value of water height on land induced by
15 flooding at the N spatial locations denoted $H_{\max}^i$ defined by their geographical coordinates
16 $\mathbf{s}=(s_1; s_2)$. We define $\mathbf{H}_{\max}^i = (H_{\max}^i(\mathbf{s}_1), H_{\max}^i(\mathbf{s}_2), \dots, H_{\max}^i(\mathbf{s}_N))$ as the vector of Hmax
17 related to the i^{th} model run defined by the vector of d offshore conditions $\mathbf{x}^i = (x_1^i, x_2^i, \dots, x_d^i)$,
18 and the $n \times N$ matrix $\mathbf{H}_{\max} = [\mathbf{H}_{\max}^1, \mathbf{H}_{\max}^2, \dots, \mathbf{H}_{\max}^n]'$ built from the n numerical model's
19 outputs. The objective is to predict $\mathbf{H}_{\max}$ given $\mathbf{x}$ based on a statistical predictive model (named
20 metamodel) that is trained with the learning dataset.

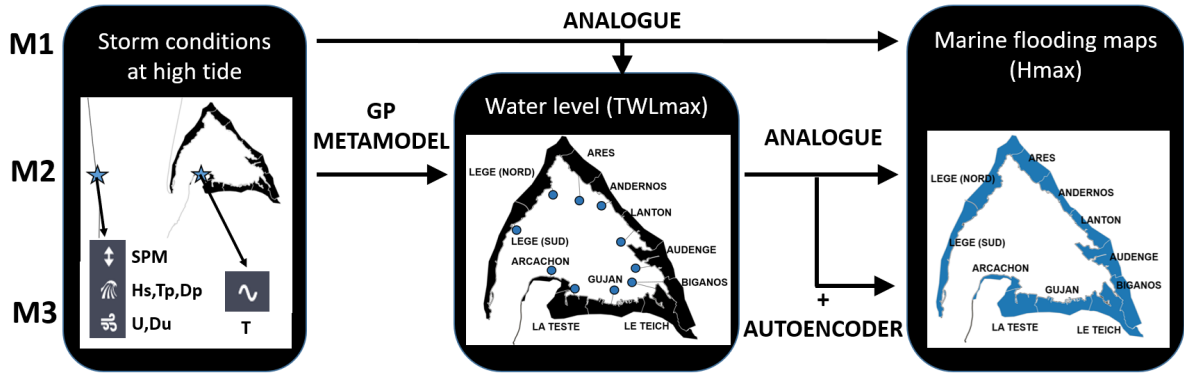


Figure 7: Diagram of the 3 methods named M1, M2 and M3 and main processes involved in the methods.

3.1 Description

3.1.1 M1: Analog-based approach

The first one, named “Analog”, is based on a look-up table approach. For given metocean forcing conditions, it consists in querying the learning database to identify the scenarios whose conditions that are the most similar, and retrieve the corresponding TWLmax and Hmax maps. The result is named “analog” in reference to the approach commonly used in weather forecasting (e.g. Van den Dool, 2007). To measure the similarity, we use the distance named “EC” proposed by Camus et al. (2011) to both handle continuous and circular data (wind and wave direction). In this study, we only query one single analog, because our preliminary tests using multiple analogs (for instance the three analogs whose forcing conditions are the first three closest to the targeted one) have shown minor increase of the predictive capability of this approach.

3.1.2 M2: Combined Gp metamodel and analog

In this approach, we first predict the total water levels TWLmax from the metocean conditions at the reference points (outlined in white dots in Fig. 5) with a Gaussian process (denoted Gp) regression model (Williams and Rasmussen, 2006). On this basis, we identify, for each municipality, the analog flood map in the learning database whose TWLmax values are closest to those predicted by the GP-based model.. The choice of the Gp model is guided by the high performance of this method as shown by multiple real case studies in coastal flooding (see e.g., Rohmer and Idier, 2012; Jia and Taflanidis, 2013; Perrin et al., 2020; Lopez-Lopera et al., 2020). In the following, we assume a Gp model with a linear trend, a Matérn 5/2

kernel model and a nugget effect. The hyper-parameters (parameters of the Gp model) are all evaluated through maximum likelihood estimation (e.g. Roustant et al., 2012). See further technical details in Appendix A.

3.1.3 M3: Combined Gp metamodel, analog and dimension reduction autoencoder

The third approach improves the second approach by predicting $H_{\max}$ from the forcing condition and $TWL_{\max}$ with a prediction based on the combined dimension reduction (denoted DR) – Gp metamodel approach that has been proposed in the literature (see for instance Jia et al. (2013) for hurricane-induced waves, Ma et al. (2022) for hurricane-induced surges, Li et al. (2020) in an estuary context, Perrin et al. (2020) for storm induced coastal flooding). To overcome the high dimensionality of the flood map (related to the number of pixels typically of several 10,000s), the DR technique aims to extract a much smaller number (typically less than 10) of new variables (named latent) to represent the very high-dimensional output. Formally this consists in transforming $\mathbf{H}_{\max}$ into a finite number of latent output variables $\mathbf{Z} = [\mathbf{z}^1, \mathbf{z}^2, \dots, \mathbf{z}^n]'$ with $\mathbf{z}^i = (z_1^i, z_2^i, \dots, z_T^i)$ so that $T \ll N$.

A popular approach is based on principal component analysis, denoted PCA (Jolliffe 2002). However, PCA is a linear technique, and it has been shown in our case to have limits (Rohmer et al., 2024a) due to the complexity of the flood maps, which have discontinuous heterogeneous patterns. Therefore, we preferably rely on a non-linear DR, namely a neural-network-based autoencoder, denoted AE (Baldi, 2012), in place of the popular PCA. For each of the latent variables, a Gp model is trained to link the latent variables to $TWL_{\max}$ forcing conditions. Recently Rohmer et al. (2023) have shown the added value of accounting for the dependence between the latent variables. To do so, we rely on the multi-output Gp model proposed by Gu and Berger (2016) (named the parallel partial kriging model). The AE architecture (number of nodes, number of layers, activation function) is selected so that the validation error is minimized as for Rohmer et al. (2024a) (see Appendix B for the comparison of architectures of autoencoders and Section 3.2 for the description of the validation procedure). Note that, if we recalculate $H_{\max}$ in the flooded area obtained with M2 with the combined dimension reduction (denoted DR) – Gp metamodel, we keep the flood extent from M2 to compensate the tendency of the chosen statistical approach to overestimate the flood extent (see Rohmer et al., 2024a). Thus, M3 can be considered a refinement of M2, as it uses the same methods to compute

TWLmax and the flood extent, while improving the reconstruction of flood depth (Hmax) through the use of an autoencoder.

3.2 Procedure for performance assessment

The objective is to measure to which extent the different approaches are respectively able to correctly predict TWLmax on the reference points and whole maps of Hmax given “yet-unseen” offshore metocean conditions. Two approaches are used:

- The first approach relies on the available learning scenarios using a 10-fold cross-validation procedure (e.g., Hastie et al. 2009). This holds as follows: (1) the initial training dataset of input forcing conditions is randomly split into 10 equal sub-sets; (2) a sub-set is removed from the initial set, and a new Random Forest model is constructed using the remaining set; (3) the sub-set removed from the initial set constitutes the verification set and the differences between the true and the estimated value can then be estimated. In this procedure, the initial random split at step (1) may have some influence on the results. This can be minimized by repeating the procedure given number of times (typically 10 times).

- The second validation approach uses the set of independent samples that have not been used for the training. Here, we use the validation scenarios composed of historical and pseudo-historical events, represented with red and orange points on Fig. 6, that are in the range of the learning database (we exclude storm Domingos that occurred with a tidal coefficient below 40 but we keep all the Pseudo_Domingos_XX whose tidal coefficients are inside the tidal range of the learning database).

The prediction error is measured with the absolute error defined as at each location $\mathbf{s}_{j=1,\dots,N}$:

$$AE^i(\mathbf{s}_j) = \left| H_{\max}^i(\mathbf{s}_j) - \hat{H}_{\max}^i(\mathbf{s}_j) \right|, \quad (1)$$

with $H_{\max}^i(\mathbf{s}_j)$ the true Hmax at the j^{th} location $\mathbf{s}$ ($j=1,\dots,N$) given the i^{th} input offshore conditions ($i=1,\dots,n$), and $\hat{y}^i(\mathbf{s}_j)$ is the reconstructed Hmax map using the metamodel approach.

A spatially averaged error indicator for the i^{th} case is then defined as:

$$MAE^i = 1/N \sum_{j=1}^N AE^i(\mathbf{s}_j), \quad (2)$$

4 Results

4.1 Performance results from the cross-validation procedure

Figure 8 presents the results of the 10-fold cross validation on TWLmax and Hmax for the three methods presented in Section 3. For TWLmax, the error corresponds to the difference between numerical simulations and metamodel-based predictions for each event on the reference points of each municipality (see Fig. 5). For Hmax, the error corresponds to the Mean Absolute Error (or MAE, see Section 3.2 and Eq. (2)) calculated for each municipality for the flooded nodes where $H_{\max} > 0$ in the numerical simulation (results of MAE for $H_{\max}=0$ in the simulation are also provided in appendix C to compare the performances of prediction of the flood extent).

Concerning TWLmax, the analog-based approach (M1) enables to estimate the simulation results with a MAE of 10 cm in average, but for some cases it can reach up to 30 cm. On the other hand, using metamodels (in M2 and thus M3) enables to get closer to the target value within 5 cm most of the time and within 10 cm in all cases. On average, the uncertainty of M2/M3 predictions provided by the Gp model (via the kriging variance see Appendix A) is less than 4 cm which remains far below the ranges between categories of flooding presented in Section 2.4 (around 30 cm). When looking at Hmax, we notice that the MAE decreases by introducing successively metamodel-based approaches to estimate TWLmax and Hmax (from M1 to M3). With M3, the median of the MAE is below 20 cm whereas its ranges between 15 cm and 40 cm depending on the municipality for M1 and M2.

This first performance analysis based on cross validation shows that the two metamodels are correctly trained and improve the accuracy of TWLmax and Hmax predictions. However, we can observe differences between the municipalities, with potentially higher MAE (up to 60 cm) on whose comprising wetlands (Le Teich, Biganos, Audenge, Ares). A first explanation is the deeper Hmax expected on wetlands. To better understand the sources of errors, a deeper analysis is carried out in the next section with the validation dataset.

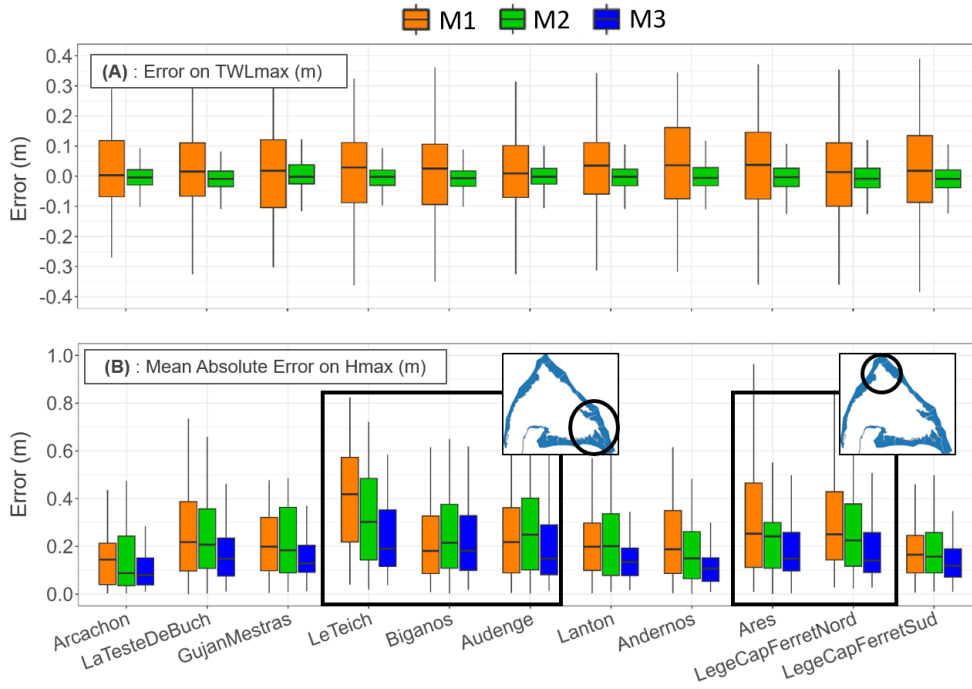


Figure 8: Error on TWLmax (A) and Mean Absolute Errors for Hmax by restricting the analysis to the mesh nodes that are flooded, i.e. for $H_{max} > 0$, in the numerical simulation (B) calculated with the 10-fold cross validation (Section 3.2). Box plots represent the 25th and 75th percentiles.

4.2 Predictive performance on pseudo-historical cases

The analysis of the predictive performance with validation cases presented on Fig. 9 reveals the same ranges of errors than the cross validation with a global improvement of the prediction of TWLmax and Hmax from M1 to M3. However, some differences appear as predictions of TWLmax tend to be slightly overestimated for all the municipalities and M3 seems to perform better than in cross validation for municipalities with wetlands. These differences, due to the limited validation sample available, do not change the conclusions of Section 4.1.

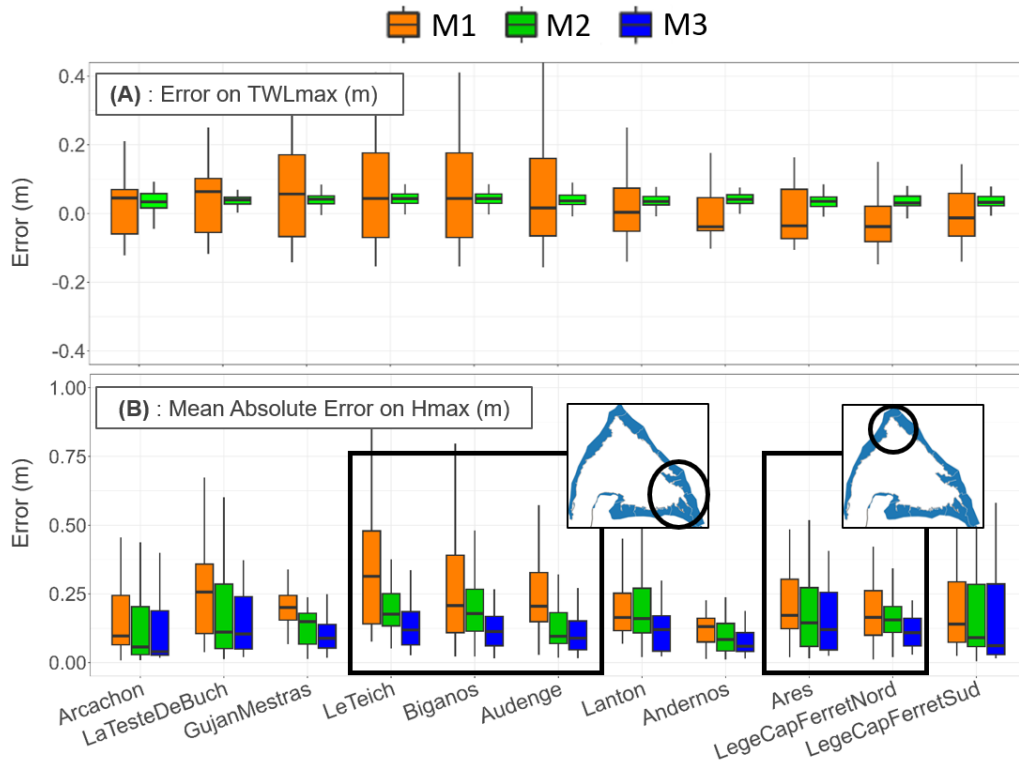


Figure 9: Errors on TWLmax (A) and Mean Absolute Errors on Hmax by restricting the analysis to the mesh nodes that are flooded, i.e. for $H_{max} > 0$, in the numerical simulation (B) calculated with the validation cases (Section 3.2). Box plots represent the 25th and 75th percentiles.

If the MAE on Hmax enables to compare the prediction accuracy between the methods, it also aggregates different sources of errors depending on (1) the accuracy of TWLmax prediction (2) the marine flooding category (3) the exposure of the municipalities. To deepen this analysis, Figure 10 provides the correlation between errors on TWLmax and Hmax for each method (one point corresponds to the MAE of one municipality for one event). These plots show that:

- Higher errors on TWLmax logically lead to higher errors on Hmax but some scenarios among the most important errors on Hmax also present small errors on TWLmax.
- The maximum MAE on Hmax do not always correspond to major flooding events (of CAT 2 or 3 outlined in light blue and cyan).

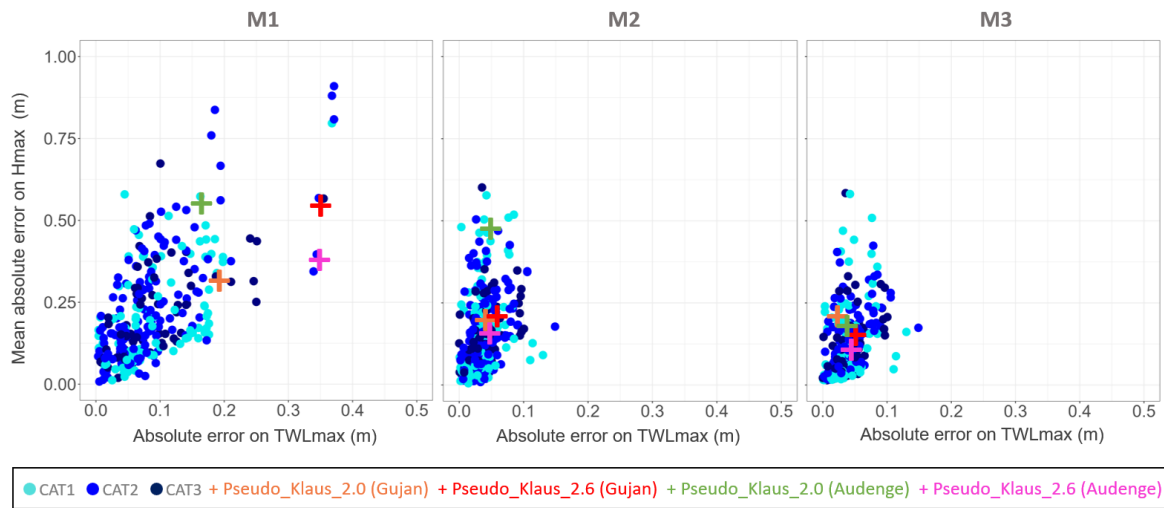
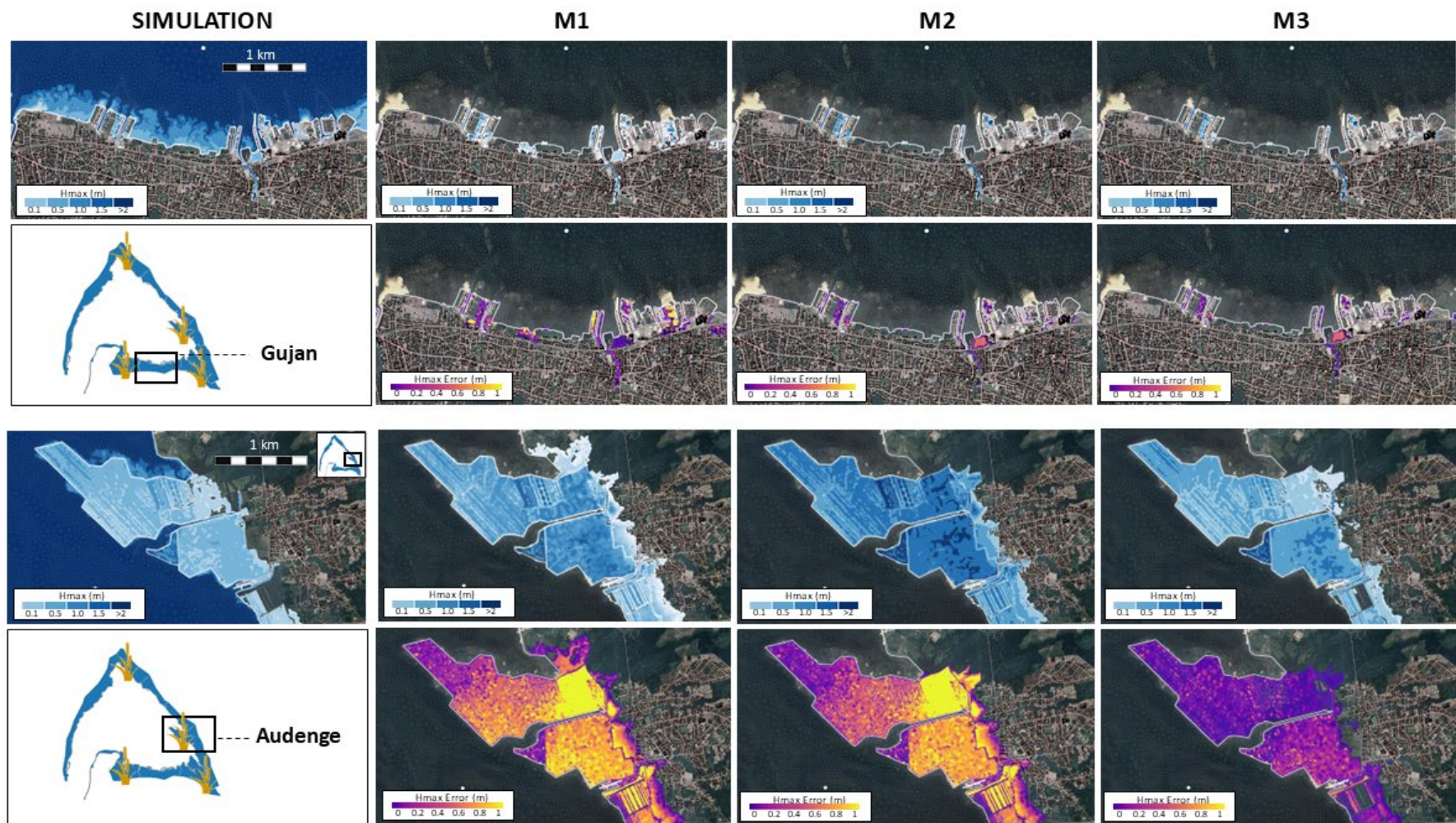


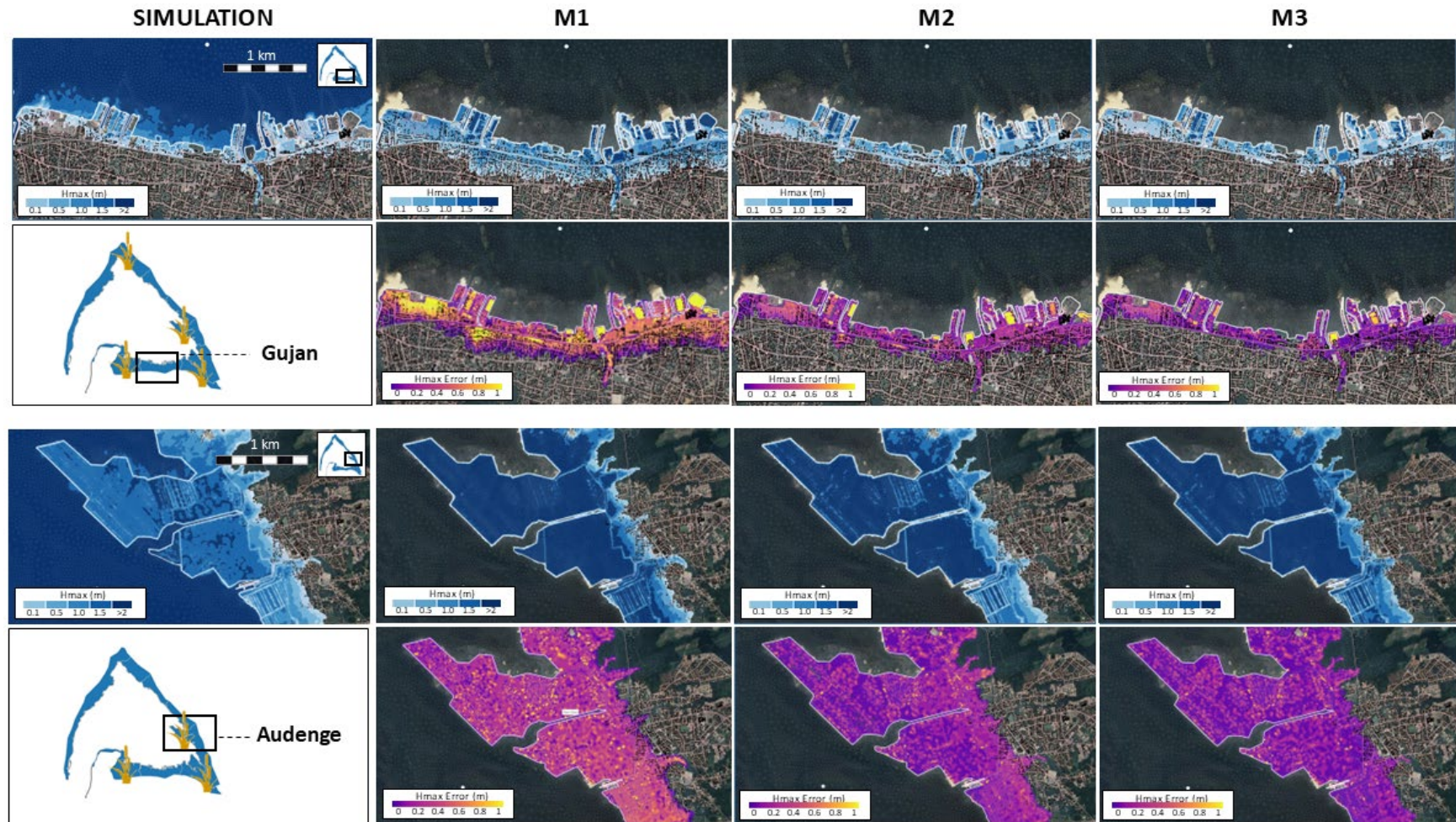
Figure 10: Relation between TWLmax and Hmax errors for each method M1-M3. Each point represents the MAE calculated for one municipality and for one pseudo-historical event.

To investigate further, we focus the analysis on two examples that illustrate different situations in Gujan Mestras and Audenge, namely Pseudo_Klaus_2.0, that is characterised by a low-level category CAT 1, and Pseudo_Klaus_2.6, that is characterised by a high-level category CAT2-3. For these events, MAE on Hmax for M2 and M3 remain quite high (between 20 cm and 50 cm) although TWLmax errors do not exceed 5 cm. The flooding and error maps plotted in Fig. 11 (for Pseudo_Klaus_2.0) and Fig. 12 (for Pseudo_Klaus_2.6) illustrate that, whatever the category of events, the maximum Hmax errors are localized in the seafront quarters (like in Gujan-Mestras) or inside wetlands (like in Audenge). These sectors both present (1) a higher sensitivity to threshold effects on TWLmax which has a direct effect on seafront flooding and controls the quantity of water discharged into low lying areas and (2) greater water heights on land and then potential Hmax prediction errors. Figures 10-12 also illustrate the improvement of marine flooding maps prediction from M1 to M3 both in terms of flooding extent and water height on land. Even for these cases with MAE of Hmax above the median, the flooding maps reproduced with M2 and more particularly M3 are similar to the flooding maps obtained with the simulation of the numerical model. The scatter plots available in appendix D (Figures D1 and D2) confirm these results notably at the scale of the Arcachon lagoon with a significant improvement from M1 to M2 and similar patterns centred on the diagonal between M2 and M3.



Map data: © Google, Maxar Technologies

Figure 11: Hmax and Hmax Absolute Errors (AE) for Pseudo_Klaus_2.0 at Gujan-Mestras (top) and Audenge (bottom) for the three methods.
Map data: © Google, Maxar Technologies



Map data: © Google, Maxar Technologies

Figure 12: Hmax and Hmax Absolute Errors (AE) for Pseudo_Klaus_2.6 on Gujan-Mestras (top) and Audenge (bottom) for the three methods.

Map data: © Google, Maxar Technologies

To complete the comparison of the performance between the three methods, we analyse the confusion matrix on TWLmax and Hmax (Fig. 13) to check how often (all municipalities, mesh nodes and validation scenarios combined) the right category of impact (through TWLmax) and the right Hmax range (with classes of 50 cm commonly used for flood mapping) are predicted. This shows that the categories of impact (CAT) are quite well predicted by M1 (more than 70% of the test cases) but globally better predicted by M2 and M3 (more than 90% of the tests cases). This result agrees with the errors on TWLmax presented in Fig. 9 (potentially up to 20-30 cm for M1 and 10 cm for M2 and M3). The representation of the flood extent improves from M1 to M3 as evidenced by F1 score (that can be interpreted as a harmonic mean of the precision and recall, reaching its best value at 1 and worst value at 0) which reaches 0.94 for M1, 0.96 for M2 and 0.97 for M3. More precisely, the confusion matrix shows that M1 allows to predict the correct order of magnitude only half of the time while M2 and M3 perform better with satisfactory predictions respectively around 65% and 75% of the time. More generally, M2 and M3 has a tendency of overestimation, but it should be underlined that the incorrect class is almost always the closest, i.e. when misclassifying the ‘0.5-1’ class, the predicted one is often the ‘1-1.5’. From a risk perspective, this means that the approach minimises the false positives but tend to over-predict with more false positives.

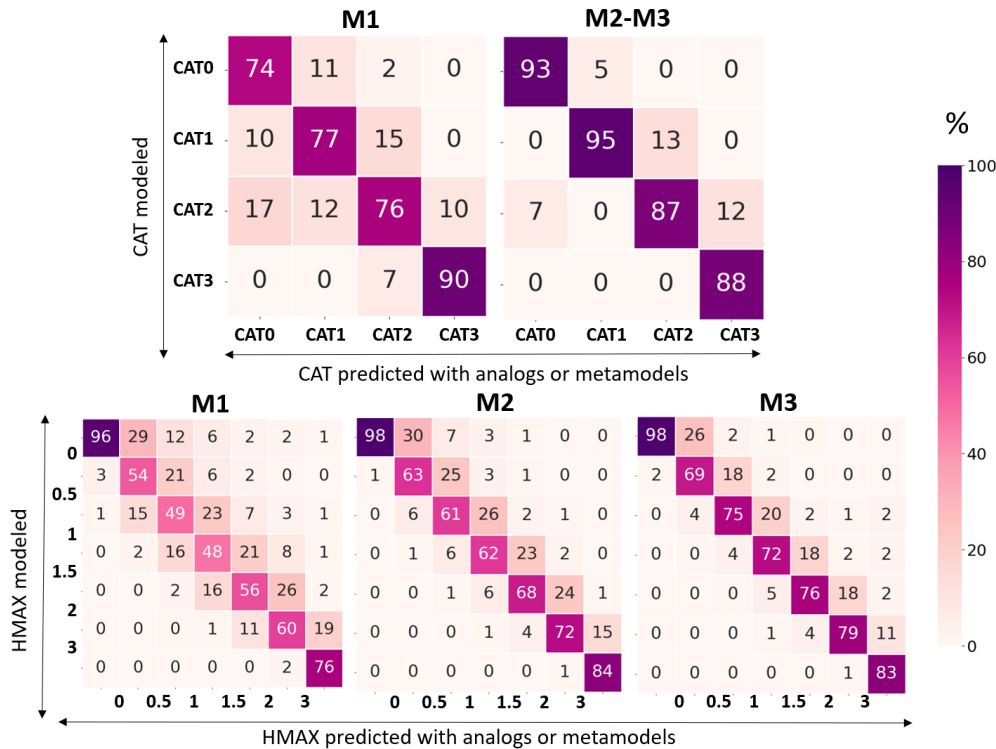


Figure 13: Confusion matrix (or error matrix) for the impact categories denoted CAT (computed based on TWLmax) (top) and Hmax (bottom).

5 Discussion

Section 4 shows that the three methods can all be useful to predict maximum total water levels (TWLmax) as well as marine flooding category (CAT) and water height maps on land (Hmax), each of them bringing some advantages (see Table 5):

- The analog-based approach (M1), although very simple, can provide interesting first order estimates of TWLmax, CAT and Hmax but it can be limited and source of significant errors if the number of scenarios constituting the initial dataset is not enough to find sufficiently close analogs. Besides, the use of 6 input parameters to describe the storm conditions (U, Du, Hs, Tp, Dp, SPM) makes it difficult to evaluate if the selected analog overestimates or underestimates the target scenario, which can be delicate for operational forecast purposes.
- The approach combining Gp metamodels, analog and dimension reduction autoencoder methods (M3) is the most appropriate to jointly estimate with the highest accuracy, TWLmax (with a precision of 5 cm), CAT (90% of the cases) and Hmax (with a precision of 10 cm). It although requires more substantial implementation efforts, higher level of expertise, and uses statistical and machine learning methods that are still in the field of research for such forecasting purposes.
- Finally, the combined Gp metamodel and analog based approach (M2) turns out to be a good compromise between complexity/effort of implementation and accuracy. Through a quite simple statistical approach, it enables to estimate TWLmax and CAT with a good accuracy, which is an important first level of information for the crisis managers. Flooding maps are also quite well estimated, although less accurate than for M3 (with a higher prediction error for Hmax on the order of 20 cm).

To sum-up, the analog-based approach can be seen as a valuable first step to explore the dataset and improve the understanding of flooding phenomena. On the other hand, for an operational forecasting context, the two metamodel-based approaches are more suitable for fast prediction and can be complementary depending on the type of event and the forecast lead time. While M2 may reach sufficient level of accuracy to give first order estimates of categories of impact and marine flooding maps at 72h or 48h lead time (when uncertainties are still important), M3 may be more recommended at 24h or 12h lead time when a detailed vision of the flood map is necessary. As pointed out by the national flood hazard recommendations (MEDDL, 2014) and the municipalities involved in the study, the first 10 cm class of Hmax, is important for emergency services and crisis management as it determines the possibility to circulate easily

for pedestrians and vehicles. Thus, M3 (with a precision of about 10 cm for Hmax) is more suitable to predict Hmax in areas with low Hmax and also minimizes the errors on Hmax classes of 50 cm range.

	Method complexity	Precision				Recommendation of use
		TWL _{max}	CAT	Hmax (h>0)	Flood extent	
M1	+	10 cm	70%	25 cm	+	Prevention and preparedness
M2	++	5 cm	90%	20 cm	++	Forecast (few days lead time)
M3	+++	5 cm	90%	10 cm	+++	Forecast (few hours lead time)

Table 5: Summary of prediction precisions and recommendations of use for the three methods. For TWL_{max} and Hmax (h>0) the precision is an order of magnitude of the mean MAE for all municipalities, and all verification sets of cross validation. For CAT it corresponds to the percentage of correct predictions (all categories mixed up). For the flood extent it is a qualitative appreciation of the quality of the prediction based on Fig. 2B describing the errors for Hmax (h=0).

Analysing the impact of uncertainties and the complementarity of different approaches for operational forecast is worth being further analysed with concrete cases of storm forecasting. This work, initiated in Rohmer et al. (2024b) will be continued and focused on applying these metamodel-based approaches (M2 and M3) with forecasts from Météo-France deterministic and EPS chains on recent events similarly as Dietrich et al. (2013) or Beuzen et al. (2019). These experiments will enable to carry out a complete analysis of the relative sources of uncertainties and to assess to which extent the metamodel error affects the spread of the probabilistic forecast, and whether it can be neglected compared to the variability in the metocean conditions (Rohmer et al., 2026).

Several lines for further improvements of the metamodeling methods are also identified. First, we focused in this study on the cases that led to flooding for the construction of the metamodels. Integrating also the case without flooding deserves to be investigated by testing different approaches either by completing the approach with a classification step (see e.g. Rohmer et al. (2018) for an example using a random forest classification technique) or using variants of Gaussian process metamodels adapted to zero-censored data (Spiller et al., 2023). More generally, this problem is related to the tendency of the metamodels to over-estimate the flood spatial extent, and more advanced approaches should be tested to improve this aspect by

1 relying on more sophisticated deep learning techniques for instance based on generative models
2 (e.g. Ma et al., 2024) combined with appropriate optimisation approaches to select the most
3 optimal values of the ML model's parameters (called hyperparameters, see Bischl et al., 2023).

4 Finally, the global approach based on pre-calculated database is replicable on any type of
5 marine flooding-prone area, and whatever the physical processes at stake, as it manages
6 computation time issues by replacing real time numerical modelling by fast statistical tools.
7 However, the methodological approaches and the associated conclusions regarding solution
8 accuracy may vary depending on the environmental context, notably in wave overtopping prone
9 areas and estuarine environments that are not considered in this study. In any case, the
10 availability of the pre-calculated dataset enables to carry out an interesting preparatory work
11 with potential users from municipalities or state services in preparedness phase to better assess
12 the accurate thresholds, and the type and resolution of information required. This preparatory
13 work is essential to allow the users to determine the criteria that will guide the whole
14 developments and define the right level of complexity required to address operational needs.

16 **6 Conclusion**

17 In this study, we developed and compared three methods to predict marine flooding maps
18 from offshore metocean conditions using successively analog-based approaches, regression
19 type metamodels and deep learning techniques. The comparison, based on both technical
20 accuracy and suitability for operational needs criteria, showed that the three methods all bring
21 some advantages and drawbacks. If the efforts required to develop full metamodel approaches
22 (including deep learning techniques) may be relevant for very urbanized sectors with civil
23 security issues as the pilot site of Arcachon lagoon, the combination of more simple regression
24 type models combined with analogs may be sufficient for applications at large scale or for
25 natural or agricultural areas (where lower precision may be acceptable).

26 In this way, this work offers encouraging perspectives on the use of pre-calculated databases
27 to carry out operational forecasts of marine flooding maps at local scale for different types of
28 sites and geographical coverage. In particular, the capacity to predict quickly marine flooding
29 maps opens up the possibility to address the issue of uncertainties through the production of
30 ensemble forecasts and in-depth sensitivity analyses.

1 Appendix A: Kriging metamodeling

2 For a given $t=1, \dots, T$, each n -dimensional vector $\mathbf{z}_t = (z_t^1, z_t^2, \dots, z_t^n)$ is assumed, in the context
3 of kriging modelling (denoted KM), to be a realisation of a Gaussian process ($Z_t(\mathbf{x})$) with:

- 4 - mean (also named trend) $\mu_t(\mathbf{x}) = \sum_{j=1}^d b_j g_j(\mathbf{x})$ (where g_j are fixed basis functions, and
5 b_j are the regression coefficients of the d input variables);
- 6 - stationary covariance function $k_t(\cdot, \cdot)$ (named kernel) written as $\forall \mathbf{x}, \mathbf{x}', k_t(\mathbf{x}, \mathbf{x}') =$
7 $\text{cov}(Z_t(\mathbf{x}), Z_t(\mathbf{x}'))$.

8 For new offshore forcing conditions $\mathbf{x}^*$, the predictive probability distribution
9 $Z_t(\mathbf{x}^*) | \{Z_t(\mathbf{x}^1) = z_t^1, \dots, Z_t(\mathbf{x}^n) = z_t^n\}$ follows a GP with mean $\hat{z}_t(\mathbf{x}^*)$ and variance $s_t^2(\mathbf{x}^*)$
10 defined using the universal kriging equations (e.g. Roustant et al., 2012) as follows:

$$11 \quad \hat{z}_t(\mathbf{x}^*) = \mathbf{g}(\mathbf{x}^*)' \hat{\mathbf{b}} + \mathbf{c}(\mathbf{x}^*)' \cdot \mathbf{C}^{-1} \cdot (\mathbf{z}_t - \mathbf{G} \hat{\mathbf{b}}), \quad (\text{A1})$$

$$12 \quad s_t^2(\mathbf{x}^*) = V_S + (\mathbf{g}(\mathbf{x}^*)' \hat{\mathbf{b}} - \mathbf{c}(\mathbf{x}^*)' \cdot \mathbf{C}^{-1} \cdot \mathbf{G})' \cdot (\mathbf{G}' \cdot \mathbf{C}^{-1} \cdot \mathbf{G})^{-1} \cdot (\mathbf{g}(\mathbf{x}^*)' \hat{\mathbf{b}} - \mathbf{c}(\mathbf{x}^*)' \cdot \mathbf{C}^{-1} \cdot \mathbf{G}), \quad (\text{A2})$$

13 where $\mathbf{z}_t = (Z_t(\mathbf{x}^1) = z_t^1, \dots, Z_t(\mathbf{x}^n) = z_t^n)$, $\mathbf{C}$ is the $n \times n$ covariance matrix between the points
14 $Z_t(\mathbf{x}^1), \dots, Z_t(\mathbf{x}^n)$ whose element is $C[i, j] = k_t(\mathbf{x}^i, \mathbf{x}^j)$; $\mathbf{c}(\mathbf{x}^*)$ is the n -dimensional vector
15 composed of the covariance between $Z_t(\mathbf{x}^*)$ and the points $Z_t(\mathbf{x}^1), \dots, Z_t(\mathbf{x}^n)$; $\mathbf{g}(\mathbf{x}^*)$ is the d -
16 dimensional vector of trend functions values at $\mathbf{x}^*$, $\mathbf{G} = (\mathbf{g}(\mathbf{x}^1), \dots, \mathbf{g}(\mathbf{x}^n))'$ is the $n \times d$
17 experimental matrix, the best linear estimator $\hat{\mathbf{b}}$ of $\mathbf{b}$ is $(\mathbf{G}' \mathbf{C}^{-1} \mathbf{G})^{-1} \mathbf{G}' \mathbf{C}^{-1} \mathbf{m}_Y$, and $V_S = \sigma^2 -$
18 $\mathbf{c}(\mathbf{x}^*)' \cdot \mathbf{C}^{-1} \cdot \mathbf{c}(\mathbf{x}^*)$ by assuming $k(\cdot, \cdot)$ to be stationary with σ^2 a hyperparameter (named
19 process variance) to be estimated.

20

21

1 Appendix B: Autoencoder

2 This method belongs to the class of deep neural networks whose typical architecture is depicted
3 in Fig. B1. It consists of an input layer (left blue layer in Fig. B1), a given number of hidden
4 layers (light-coloured layers), and an output layer (right green layer). The process of going from
5 the input layer to the hidden layer is called encoding, (i.e. from the original data y to latent
6 variables z), while the process of going from the hidden layer to the output layer is called
7 decoding (i.e. from latent variables z to the variables back-transformed in the physical domain
8 $\hat{y}$). The central layer is named the bottleneck hidden layer and provides the latent variables z .
9 The AE architecture can be parametrised in different manners (Pawar & Attar 2019), i.e. the
10 number of nodes of the hidden layer, the number of nodes of the and bottleneck layer (i.e. the
11 number of latent variables), the type of activation function applied to each node, etc.

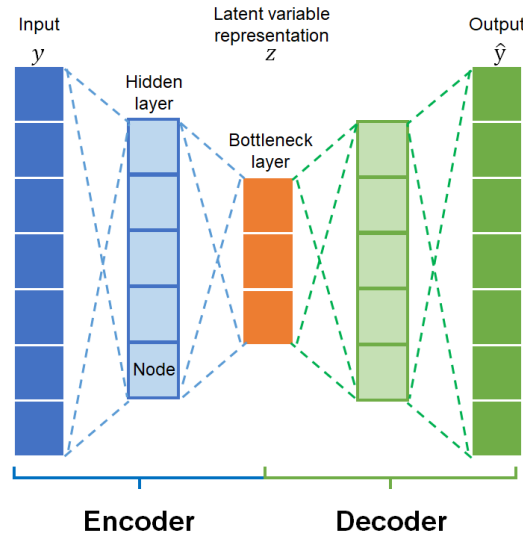


Figure B1: Typical autoencoder architecture.

12 To select the parameters of the AE architecture, we conducted the 10-fold cross validation
13 procedure by assuming different assumptions, namely by varying the number of nodes of the
14 hidden layer from 10 to 50, the number of latent variables from 3 to 5, the type of activation
15 function among *relu* and *sigmoid* for the hidden layer and among *linear*, *tanh* and *selu*. Figure
16 B2 shows the cross-validation prediction error, here measured by the root mean square error
17 denoted RMSE, for two sectors, Gujan-Mestras and Andernos. This indicates that having 10
18 nodes in the hidden layer with *relu* activation, and 3 in the bottleneck with *tanh* allows to
19 achieve the lowest prediction error for mesh nodes with $H_{max} > 0$ and $H_{max} = 0$ in the numerical
20 simulation. This result was also confirmed for the other sectors on the Arcachon lagoon.

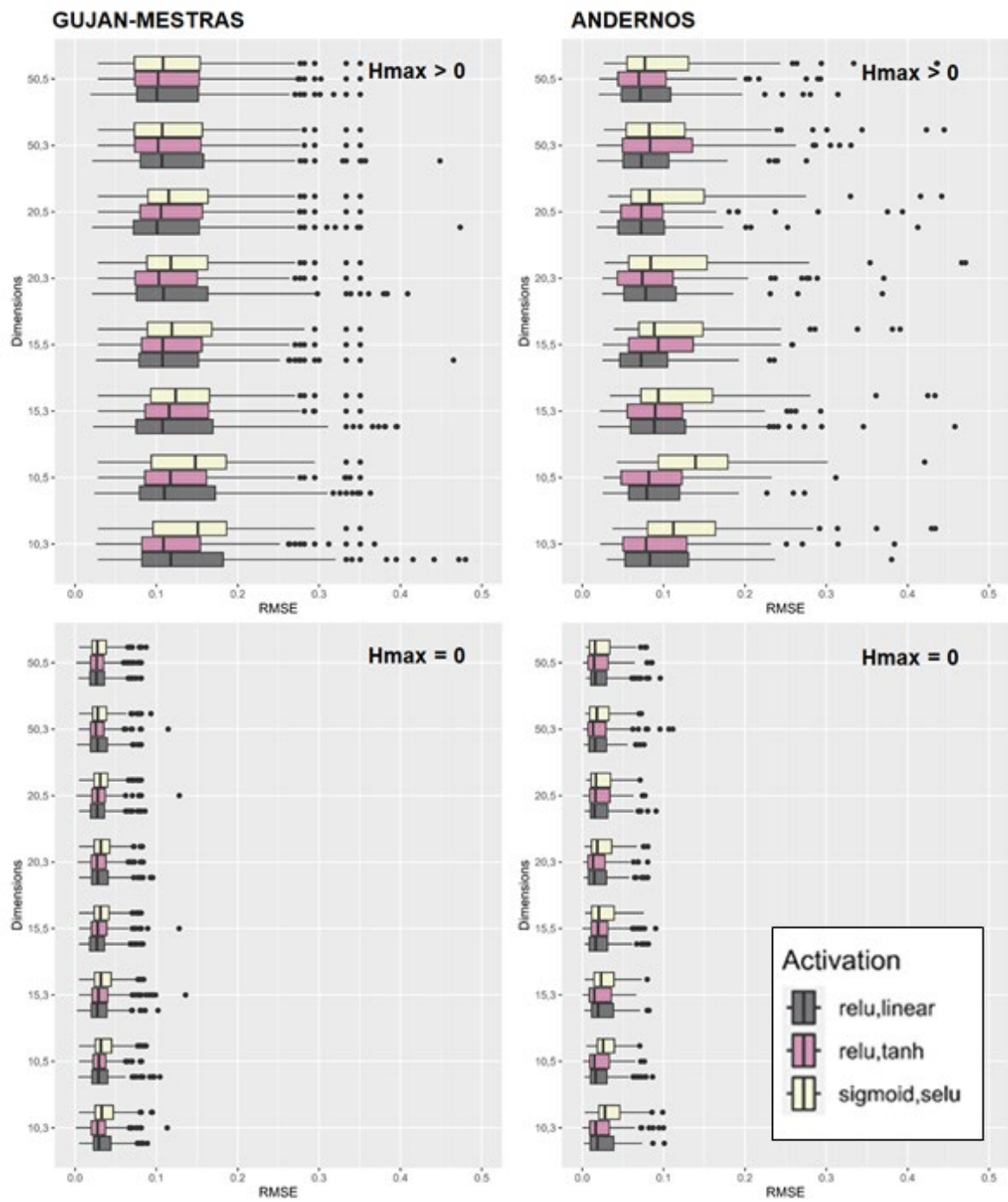


Figure B2: Evolution of the cross-validation prediction error, here measured by the root mean square error denoted RMSE, for two sectors, Gujan-Mestras (left) and Andernos (right) for mesh nodes with $H_{max} > 0$ (top) and $H_{max} = 0$ (bottom) in the numerical simulation.

Appendix C: MAE for hmax=0 in the simulation for the cross validation

The cross-validation-based MAE calculated for non-flooded mesh nodes, i.e. with Hmax=0 in the numerical simulation, shows a clear decrease from M1 to M3. This indicates that M3 achieves the most accurate prediction of the flood spatial extent, i.e. here corresponding to the total number of mesh nodes predicted with Hmax=0 while being non-flooded in the numerical simulation. This suggests also that M3 enables to decrease the rate of false alarms, though we notice that MAE remains non-zero (of the order of less than 1 cm), hence reflecting a remaining tendency of M3 to slightly over-estimate the flood extent.

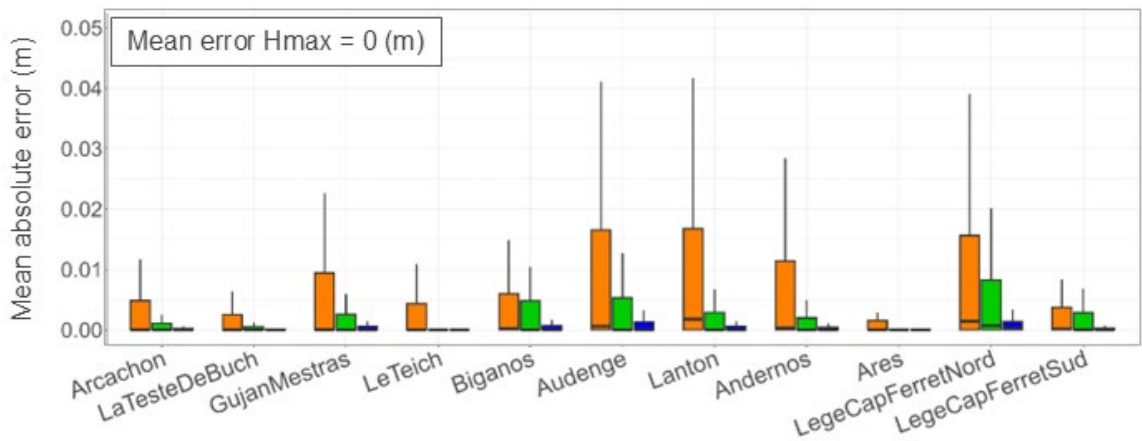


Figure C1: Mean absolute error (MAE) calculated over the non-flooded mesh nodes (with Hmax=0 in the numerical simulations) of each municipality from the cross-validation procedure.

1 **Appendix D: Scatter plots**

2

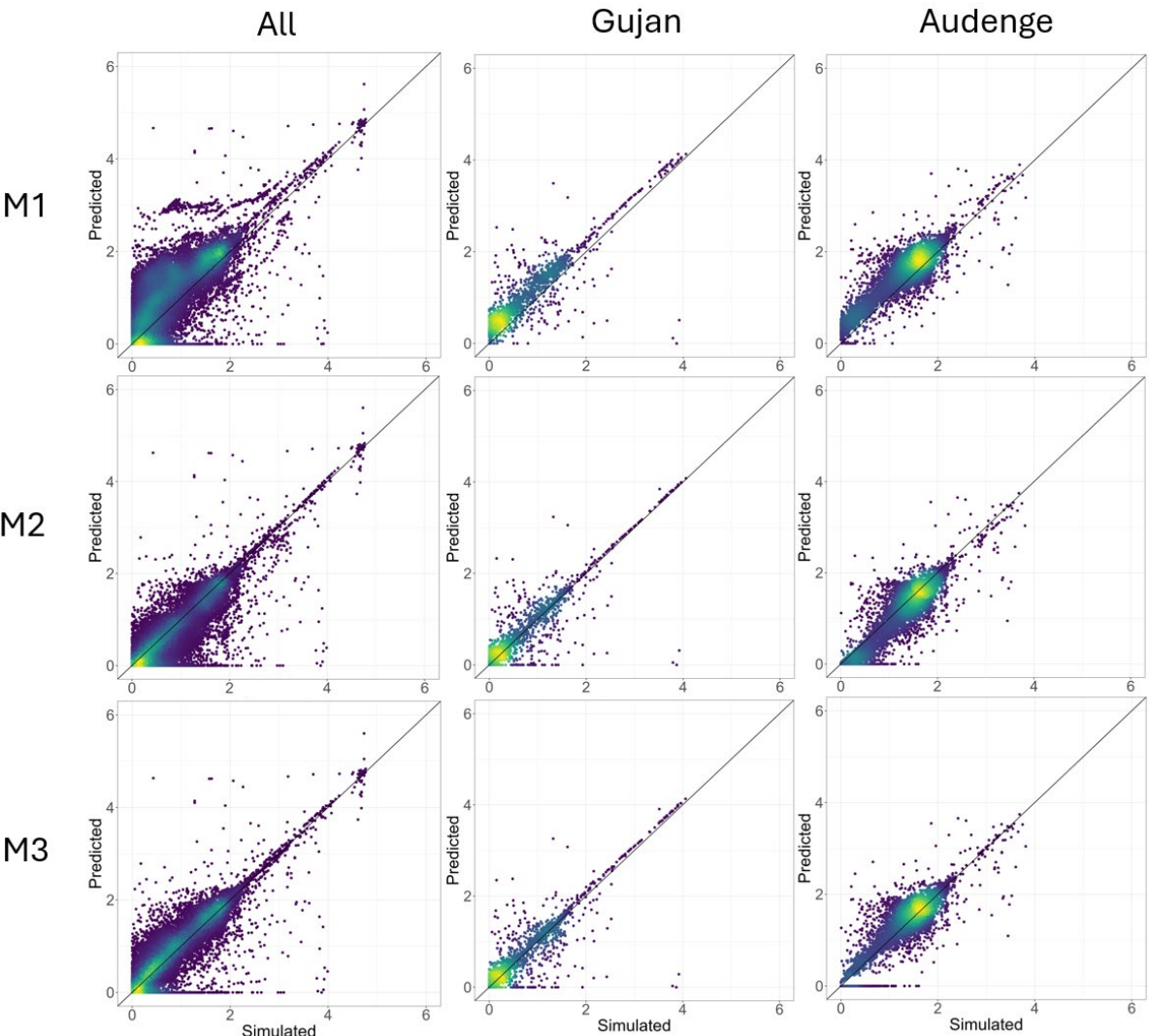


Figure D1: Scatter plot of predicted vs. simulated Hmax for Klaus_2.0.

3
4
5

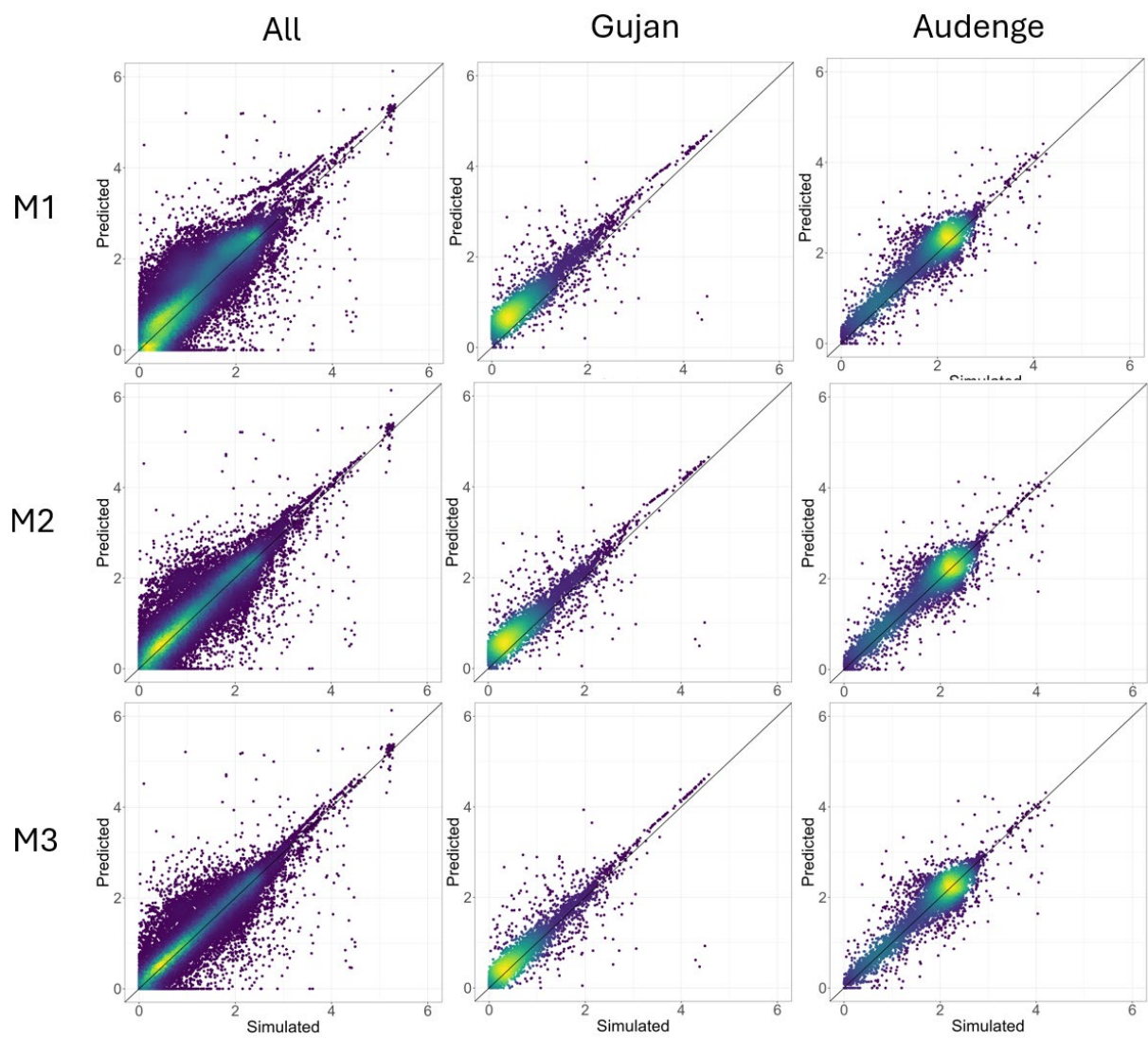


Figure D2: Scatter plot of predicted vs. simulated Hmax for Klaus_2.0.

1 **Data availability**

2 All the data produced as part of the work of this work are available on request

3 **Author contribution**

4 JR and DI guided the development and application of the statistical methodologies. DP, AD
5 and DA provided the meteoceanic conditions of historical storms. AF, RP and SL completed
6 the marine flooding simulations. SG and SL conducted the analysis with local users on marine
7 flooding categories. EM performed the cross and historical validation. SL prepared the
8 manuscript with contributions from all co-authors.

9 **Competing interests**

10 The authors declare that they have no conflict of interest

11 **Acknowledgements**

12 This work is supported by the French National Research Agency within the ORACLES
13 project (ANR – 21 – CE04-0012-01).

14

References

- Amante, C., and Eakins, B.W.: ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS NGDC-24, 19 pp, 2009.
- Biolchi, L. G., Unguendoli, S., Bressan, L., Giambastiani, M.S.B. and Valentini, A.: Ensemble technique application to an XBeach-based coastal Early Warning System for the northwest Adriatic Sea (Emilia-Romagna region, Italy). *Coastal Eng.*, 173, 104081, <https://doi.org/10.1016/j.coastaleng.2022.104081>, 2022.
- Beuzen, T., Goldstein, E.B., and Splinter K.D. : Ensemble models from machine learning: an example of wave runup and coastal dune erosion. *NHESS*, 19, 2295-2309, 2019.
- Bischl, B., Binder, M., Lang, M., Pielok, T., Richter, J., and Coors, S.: Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 13(2), e1484, 2023.
- Bouchet, J.M., Deltreil, J.P., Manaud, F., Maurer, D., and Trut, G.: Etude intégrée du Bassin d’Arcachon Tome 1. Rapport IFREMER R.INT.DEL/97.09/ARCACHON, 1997.
- Camus, P., Mendez, F.J., Medina, R., and Cofiño, A.S.: Analysis of clustering and selection algorithms for the study of multivariate wave climate. *Coastal Engineering*, 58(6), 453-462, 2011.
- Castelle, B., Bonneton, P., Dupuis, H., and Senechal, N.: Double bar beach dynamics on the high energy meso-macrotidal French Aquitanian coast: a review. *Marine Geology*, 245, 141-159, 2007.
- CETE SO: Submersions marines sur le Bassin d’Arcachon. Etude historique. Rapport de novembre 2012, version 2, 47p, 2012.
- Coles, S., Bawa, J., Trenner, L., and Dorazio, P.: An introduction to statistical modeling of extreme values, Vol. 208. London: Springer, 2001.
- Davis, J., Paramygin, V., Forrest, D. and Sheng, Y.P.: Toward the Probabilistic Simulation of Storm Surge and Inundation in a Limited-Resource Environment. *Monthly Weather Review - MON WEATHER REV.* 138. 2953-2974. 10.1175/2010MWR3136.1, 2010.
- Demeritt, D., Stephens, E.M., Créton-Cazanave, L., Lutoff, C., Ruin, I., and Nobert, S.: Communicating and Using Ensemble Flood Forecasts in Flood Incident Management: Lessons from Social Science. In: Duan, Q., Pappenberger, F., Thielen, J., Wood, A., Cloke, H., Schaake, J. (eds) *Handbook of Hydrometeorological Ensemble Forecasting*. Springer, Berlin, Heidelberg, 2016.
- Descamps, L., Labadie, C., and Joly, A.: PEARP Météo-France short range ensemble prediction system. *Q.J.R. Met. Soc.*, 141: 1671-1685, 2015.
- Dietrich, J.C., Dawson, C., Proft, J., and Howard, M.T.: Real-Time Forecasting and Visualization of Hurricane Waves and Storm Surge Using SWAN+ADCIRC and FigureGen. *Computational Challenges in the Geosciences*, Volume 156, p 49-70, 2013.

1 Dupuis, H., Michel, D., and Sottolichio, A.: Wave climate evolution in the Bay of Biscay over
2 two decades, *Journal of Marine Systems*, 63 (3-4), 105-114, 2006.

3 Filippini, A.G., Arpaia, L., Perrier, V., Pedreros, R., Bonneton, P., Lannes, D., and Ricchiuto,
4 M.: An operational discontinuous Galerkin shallow water model for coastal flood
5 assessment. *Ocean Modelling*, 102447, 2004.

6 Fraehr, N., Wang, Q. J., Wu, W., & Nathan, R. (2024). Assessment of surrogate models for
7 flood inundation: The physics-guided LSG model vs. state-of-the-art machine learning models.
8 *Water Research*, 252, 121202.

9 Gu, M., and Berger, J.O.: Parallel partial Gaussian process emulation for computer models with
10 massive output. *The Annals of Applied Statistics* 10(3):1317-1347, 2016.

11 Hastie, T., Tibshirani, R., and Friedman, J.: *The Elements of Statistical Learning: Data Mining,*
12 *Inference, and Prediction*, Springer: Berlin/Heidelberg, Germany, 2009.

13 Hauer, M.E., Hardy, D., Kulp, S.A. Mueller, V., Wrathall, D.J. and Clark P.U. : Assessing
14 population exposure to coastal flooding due to sea level rise. *Nat Commun* 12, 6900.
15 <https://doi.org/10.1038/s41467-021-27260-1>, 2021.
16

17 Heffernan, J. E. and Tawn, J. A.: A conditional approach for multivariate extreme values (with
18 discussion), *J. Roy. Stat. Soc. Ser. B*, 66, 497–546, 2004.
19

20 Irazoqui Apecechea, M., Melet, A., and Armaroli, C.: Towards a pan-European coastal flood
21 awareness system: Skill of extreme sea-level forecasts from the Copernicus Marine Service.
22 *Front. Mar. Sci.* 9:1091844. doi: 10.3389/fmars.2022.1091844, 2023.

23 Irwin, C., Jung, W., Taflanidis, A. A., Nadal-Caraballo, N. C., Aucoin, L. A., & Yawn, M. C.
24 (2025). Emulation of peak storm surge across extended spatial domains using separable
25 Gaussian Process techniques. *Coastal Engineering*, 104945.

26 Jia, G., and Taflanidis, A.A.: Kriging metamodeling for approximation of high-dimensional
27 wave and surge responses in real-time storm/hurricane risk assessment. *Computer Methods in*
28 *Applied Mechanics and Engineering* 261:24-38, 2013.

29 Jolliffe, I. T.: *Principal Component Analysis*. second ed., New York: Springer, 2002.

30 Jung W, Taflanidis AA, Nadal-Caraballo NC, Aucoin LA, Yawn MC (2025) Advances in
31 spatiotemporal storm surge emulation: database imputation and multi-mode latent space
32 projection. *Coastal Engineering*, 104762

33 Le Gal, M., Fernández-Montblanc, T., Duo, E., Montes Perez, J., Cabrita, P., Souto Ceccon, P.,
34 Gastal, V., Ciavola, P., and Armaroli, C. : A new European coastal flood database for low–
35 medium intensity events, *Nat. Hazards Earth Syst. Sci.*, 23, 3585–3602, 2023.

36 Le Gal, M., Fernández-Montblanc, T., Montes Perez, J., Duo, E., Souto Ceccon, P., Ciavola,
37 and P., Armaroli, C.: Influence of model configuration for coastal flooding across Europe,
38 *Coastal Engineering*, Volume 192, 104541, ISSN 0378-3839, 2024.

1 Lecacheux, S., Bonnardot, F., Paris, F., Rousseau, M., and Pedreros, R.: Probabilistic Forecast
2 of Coastal Waves for Flood Warning Applications at Reunion Island. *Journ. of Coastal*
3 *Research: Proceedings of the 15th International Coastal Symposium*: pp. 776 – 780, 2018.

4 Lecacheux, S., Pedreros, R., Filippini, A.G., Arpaia, L., Rohmer, J., Louisor, J., Nicolae Lerma,
5 A., Quique, R., and Pragout, J.: Appui à la mission RDI de Gironde : Cartographie des ZIP «
6 submersion marine » pour le Bassin d’Arcachon. Rapport final. BRGM/RP-73272-FR, 56 p,
7 2023.

8 Li, M., Wang, R.Q., and Jia, G.: Efficient dimension reduction and surrogate-based sensitivity
9 analysis for expensive models with high-dimensional outputs. *Reliability Engineering &*
10 *System Safety* 195:106725, 2020.

11 López-Lopera, A. F., Idier, D., Rohmer, J. and Bachoc, F.: Multi-Output Gaussian Processes
12 with Functional Data: A Study on Coastal Flood Hazard Assessment. *arXiv preprint*
13 *arXiv:2007.14052*, 2020.

14 Lorente, P., García-Sotillo, M., Amo-Baladrón, A., Aznar, R., Levier, B., Sánchez-Garrido, J.
15 C., Sammartino, S., de Pascual-Collar, Á., Reffray, G., Toledano, C., and Álvarez-Fanjul, E.:
16 Skill assessment of global, regional, and coastal circulation forecast models: evaluating the
17 benefits of dynamical downscaling in IBI (Iberia-Biscay-Ireland) surface waters. *Ocean Sci.*,
18 15, 967–996, 2019.

19 Luesink, M., Wolbers, J., Duin, M., and Kuipers, S.: Scenario planning to enable foresight in
20 crisis management, 2024.

21 Ma, P., Karagiannis, G., Konomi, B.A., Asher, T.G., Toro, G.R., and Cox, A.T.: Multifidelity
22 computer model emulation with high-dimensional output: An application to storm surge.
23 *Journal of the Royal Statistical Society Series C Applied Statistics*, doi:10.1111/rssc.12558,
24 2022.

25 Ma, Z., Mei, G., and Xu, N.: Generative deep learning for data generation in natural hazard
26 analysis: motivations, advances, challenges, and opportunities. *Artificial Intelligence Review*,
27 57(6), 160, 2024.

28 Macdonald, E., Tubaldi, E., & Patelli, E. (2025). Robust storm surge forecasts for early warning
29 system: a machine learning approach using Monte Carlo Bayesian model selection
30 algorithm. *Stochastic Environmental Research and Risk Assessment*, 39(7), 2789-2816.

31 MEDDTL. Guide Méthodologique Plans de Prévention des Risques Littoraux, 2014. Available
32 online: [https://www.ecologie.gouv.fr/sites/default/files/Guide%20PPRL%20-](https://www.ecologie.gouv.fr/sites/default/files/Guide%20PPRL%20-%20version%20finale%20mai%202014.pdf)
33 [%20version%20finale%20mai%202014.pdf](https://www.ecologie.gouv.fr/sites/default/files/Guide%20PPRL%20-%20version%20finale%20mai%202014.pdf)

34 Mc-Granahan, G., Balk, D., and Anderson, B.: The rising tide: assessing the risks of climate
35 change and human settlement in low elevation coastal zones, *Environ. Urban.* 19:17-37, 2007.

36 Mugica, J., Bulteau, T., Paris, F., and Pedreros, R.: Caractérisation de l’aléa submersion marine
37 dans le cadre des PPRL du Bassin d’Arcachon, (Gironde), Détermination de l’évènement
38 naturel de référence. Rapport intermédiaire BRGM/RP-61408-FR, 73 p., 38 fig., 11 Table, 2
39 ann, 2014.

1 Pawar, K., and Attar, V.Z.: Assessment of autoencoder architectures for data representation. In
2 Deep Learning: Concepts and Architectures (pp. 101-132). Cham: Springer International
3 Publishing, 2019.

4 Perrin, T.V.E., Roustant, O., Rohmer, J., Alata, O., Naulin, J.P., Idier, D., Pedreros, R.,
5 Moncoulon, D. and Tinard, P.: Functional principal component analysis for global sensitivity
6 analysis of model with spatial output. arXiv preprint arXiv:2005.10285, 2020.

7 Raoult, C., Joly, A., Andreevsky, M., and Joly-Laugel, A.: ANEMOC-3 : Amélioration de la
8 base de données d'états de mer ANEMOC-2 par prise en compte des effets de la marée. Acte
9 de conférence des 16èmes Journées de l'Hydrodynamique - Marseille, 2018, 2018.

10 Saha, S., et al. : The NCEP Climate Forecast System Reanalysis. Bulletin of American
11 Meteorological Society, 91, 1015-1057, 2010.

12 Spiller, E.T., Wolpert, R.L., Tierz, P., and Asher, T.G.: The zero problem: Gaussian process
13 emulators for range-constrained computer models. SIAM/ASA Journal on Uncertainty
14 Quantification, 11(2), 540-566, 2023.

15 Stansby, P., Chini, N., Apsley, D., Borthwick, A., Bricheno, L., Horrillo-Caraballo, J., McCabe,
16 M., Reeve, D., Rogers, B.D., Saulter, A., Scott, A., Wilson, C., Wolf, J. and Yan, K.: Integrated
17 model system for coastal flood prediction. J. Flood Risk Manage, 6: 229-252.
18 <https://doi.org/10.1111/jfr3.12001>, 2013.
19

20 Tolman: User manual and system documentation of WAVEWATCH- III version 4.18.
21 NOAA/NWS/NCEP/MMABTech, 2014.

22 Ragu Ramalingam, N., Johnson, K., Pagani, M., & Martina, M. L. (2025). Advancing nearshore
23 and onshore tsunami hazard approximation with machine learning surrogates. Natural Hazards
24 and Earth System Sciences, 25(5), 1655-1679.

25 Rohmer, J., and Idier, D.: A meta-modelling strategy to identify the critical offshore conditions
26 for coastal flooding. Natural Hazards and Earth System Sciences, 12(9), 2943-2955, 2012.

27 Rohmer, J., Sire, C., Lecacheux, S., Idier, D., an Pedreros, R.: Improved metamodells for
28 predicting high-dimensional outputs by accounting for the dependence structure of the latent
29 variables: application to marine flooding. Stochastic Environmental Research and Risk
30 Assessment, 37(8), 2919-2941, 2021.

31 Rohmer, J., Lecacheux, S., Idier, D., Filippini, A.G., and Pedreros, R.: Fast Prediction of Flood
32 Maps Based on Machine Learning Techniques: Application to Marine Flooding at Arcachon
33 Lagoon (Gironde, France). In: Gourbesville, P., Caignaert, G. (eds) Advances in
34 Hydroinformatics—SimHydro 2023 Volume 1. SimHydro 2023. Springer Water. Springer,
35 Singapore. https://doi.org/10.1007/978-981-97-4072-7_10, 2024a.

36 Rohmer, J., Membrado, E., Lecacheux, S., Idier, D., Filippini, A., Pedreros, R., Dalphiné, A.,
37 Paradis, D., and Ayache, D.: Ensemble forecasts of marine flood maps assisted by probabilistic
38 machine learning techniques: Application at Arcachon Lagoon (France), EGU General
39 Assembly 2024, Vienna, Austria, 14–19 Apr 2024, EGU24-14945,
40 <https://doi.org/10.5194/egusphere-egu24-14945>, 2024b.

1 Rohmer J., Lecacheux S., Idier D. et al. Uncertainty analysis of a machine learning-aided
2 ensemble forecast for coastal flood maps, 11 February 2026, PREPRINT (Version 1) available
3 at Research Square [<https://doi.org/10.21203/rs.3.rs-8758359/v1>]

4 Roustant, O, Ginsbourger, D., and Deville, Y.: DiceKriging, DiceOptim: Two R packages for
5 the analysis of computer experiments by kriging-based metamodeling and optimization,
6 Journal of Statistical Software 51(1):1–55, 2012.

7 SIBA : SLGRI du Bassin d’Arcachon, 2015. Available online : [https://www.siba-bassin-](https://www.siba-bassin-arcachon.fr/sites/default/files/2024-03/slgri-document-final_oct-2017.pdf)
8 [arcachon.fr/sites/default/files/2024-03/slgri-document-final_oct-2017.pdf](https://www.siba-bassin-arcachon.fr/sites/default/files/2024-03/slgri-document-final_oct-2017.pdf)

9 Van den Dool, H.: Empirical methods in short-term climate prediction. Oxford:Oxford
10 University Press, 2007.

11 SHOM: Estimation des valeurs extrêmes de niveau d’eau sur le littoral métropolitain. Rapport
12 d’étude N°22-RE-0188, 2022.

13 Toledano, C., Ghanous, M., Lorente, P., Dalphiné, A., Aouf, L., and Sotillo, M.G.: Impacts of
14 an Altimetric Wave Data Assimilation Scheme and Currents-Wave Coupling in an Operational
15 Wave System: The New Copernicus Marine IBI Wave Forecast Service. J. Mar. Sci. Eng. 2022,
16 10, 457, 2022

17 Turner, I.L., Leaman, C.K., Harley, M.D., Thran, M.C., David, D.R., Splinter, K.D., and Lowe,
18 R.J. : A framework for national-scale coastal storm hazards early warning. Coastal
19 Engineering, 192, 104571, 2024.

20 Williams, C.K., and Rasmussen, C.E.: Gaussian processes for machine learning. Cambridge,
21 MA: MIT press, 2006.

22 Wang, Z., Leung, M., Mukhopadhyay, S., Sunkara, S. V., Steinschneider, S., Herman, J., ... &
23 Ruggiero, P. (2025). Compound coastal flooding in San Francisco Bay under climate
24 change. npj Natural Hazards, 2(1), 3. <https://www.nature.com/articles/s44304-024-00057-0>

25 Wang, W., Wang, Q. J., & Nathan, R. (2025). Gaussian process regression on multiple drivers
26 and attributes for rapid prediction of maximum flood inundation extent and depth. Journal of
27 Hydrology, 649, 132476.

28 Wu, W., Emerton, R., Duan, Q., and Wood, A.W.: Ensemble flood forecasting: Current status
29 and future opportunities. Wiley Interdisciplinary Reviews: Water. 7. e1432.
30 10.1002/wat2.1432, 2020.