JGR Oceans

RESEARCH ARTICLE
10.1029/2022JC018467

Automated Approaches for Capturing Localized Tsunami Response—Application to the French Coastlines

Daniel Giles[1], Audrey Gailler[2], and Frédéric Dias[1,3]

[1]School of Mathematics and Statistics, University College Dublin, Dublin, Ireland, [2]CEA, DAM, DIF, Arpajon, France, [3]Earth Institute, University College Dublin, Dublin, Ireland

Key Points:

- Automated approaches for capturing local variability in tsunami hazard with minimal runtime requirements are developed
- The independent approaches use either a transfer function (extended Green's Law) or machine learning techniques
- Both methods incorporate local responses and computationally cheap regional maximum wave heights to predict the local variability

Correspondence to:
F. Dias,
frederic.dias@ucd.ie

Citation:
Giles, D., Gailler, A., & Dias, F. (2022). Automated approaches for capturing localized tsunami response—Application to the French coastlines. *Journal of Geophysical Research: Oceans*, *127*, e2022JC018467. https://doi.org/10.1029/2022JC018467

Received 26 JAN 2021
Accepted 19 MAY 2022

Abstract Local bathymetry and onshore features can have a substantial effect on the spatial variability of the hazard from an incoming tsunami. In a warning context, being able to provide localized tsunami forecasts at strategic locations would therefore help mitigate the damage. Despite the recent advancements in computing powers and the development of highly efficient tsunami codes, capturing this local variability can oftentimes be infeasible in a warning setting. Traditional high-resolution simulations which can capture these localized effects are often too costly to run "on-the-fly." Alternative approaches that capture the localized response to an incoming tsunami, which are based upon using the maximum wave heights from a computationally cheap regional forecast, are developed here. These alternative approaches are envisaged to aid in a warning center's ability at providing extremely rapid localized forecasts. The approaches focus upon two different methods: transfer functions and machine learning techniques. The transfer function is based upon a recent extension to the established Green's Law. The extended version introduces local amplification parameters, with the aim of capturing the neglected localized effects. An automated approach which optimizes for these local amplification parameters is outlined and the performance of the transfer function is explored. A machine learning model is also trained and used to predict the localized tsunami hazard. Its performance is compared to the extended Green's Law approach for a site along the French coast. These developed methods showcase promising techniques that a tsunami warning center could use to provide high-resolution warnings.

Plain Language Summary Tsunamis are high-impact, long-duration disasters that in most cases allow for only minutes of warning before impact. Even though there have been significant advancements in warning methodology, pre-disaster preparedness and basic understanding of related phenomena, the vulnerability of bays, harbors, beaches and maritime assets to tsunamis remains poorly understood. The tsunami hazard on such local sites is greatly affected by the offshore and onshore topography. High-resolution simulations which resolve site-specific features can capture this local variability but are computationally too expensive to run. As tsunami early warning centers operate under severe time constraints, it is often not feasible to run these high-resolution simulations in a warning setting. A simulation which is completed after a tsunami has arrived on the coastline of interest, as precise as it might be, is worthless to a warning center. Alternative approaches that are based upon capturing the response of a site to an incoming tsunami ahead of time are developed here. Such approaches could be used by a warning center to quickly forecast the local tsunami hazard in the immediate aftermath of a tsunamigenic event.

1. Introduction

For tsunami warning centers, obtaining real-time coastal inundation forecast maps in a near-field or regional context is a major challenge. Currently, the tsunami warning systems are mainly focused on the first parameters of the earthquake, with the emphasis on earthquake monitoring and seismic data processing, while the tsunami models give only basin scale information of the corresponding risk. Precise coastal water heights and inundations can be computed, but contrary to what happens in the deep ocean, the simulation of tsunami hazard at the coast requires a non-linear model with fine resolution bathymetry and topography to cope with the non-linearity and the complex topography of the coasts. Despite the parallelization of tsunami codes and the power of computers used (Behrens & Dias, 2015; Løvholt et al., 2019), the computing time necessary to carry out a number of high-resolution inundation forecasts in real-time is still non-negligible, particularly in relation to the arrival times for a near-field event. Therefore, in a real-time warning setting, the computational runtime for high-resolution inundation forecasts places a restriction on the extent of the forecast area.

In a near-field and regional context, the rapid determination of water heights at the coast, or even the estimation of run-ups, are part of the objectives of the French tsunami warning center (CENALT) to improve the level of expertise and the information that can be provided to the alert center's correspondents. However, to capture the local variability of water heights at the coast, high-resolution simulations are traditionally required, with a minimum mesh element size $\sim \mathcal{O}(10^1)$ m. Unfortunately, as stated, there exists a balance between mesh resolution and runtime that one must consider when the time constraints of a warning setting are involved (Giles et al., 2021). Therefore, the integration of rapid modeling tools including wave heights at coast and inundation are a priority.

To alleviate the problem of this race against time for the determination of near-field and regional water heights at the coast, the warning centers are developing tools for the rapid prediction of the amplitude of coastal tsunamis, calculated from empirical laws (Gailler et al., 2017; Jamelot & Reymond, 2015; Lalli et al., 2019). This type of linear approximation rapidly provides the tsunami heights along the coastline, with an associated error on the order of a factor 2 at best (Gailler et al., 2017; Jamelot & Reymond, 2015; Reymond et al., 2012). Such approaches are adapted to the near-field context, but not all local effects are effectively resolved and estimates of run-up and horizontal inundation are lacking.

To provide a rapid estimate of the run-up, Wronna et al. (2020) proposed a tsunami runup predictor, relating the accelerating phase of the wave to the length of the beach slope over which the wave is traveling. Other recent studies develop methods involving the use of pre-calculated databases. Gusman et al. (2014) compare the tsunami waveforms simulated at different virtual observation points with the waveforms contained in a database. Each waveform in the database is associated with a high-resolution map of the inundation, pre-calculated with the non-linear model. This allows, once the best scenario of the database is found, to use the corresponding map to quickly estimate the flood. Such pre-calculated databases can be used in deep-learning approaches. Indeed, Mulia et al. (2018) improved the method of Gusman et al. (2014) by using low-resolution maps containing the maximum tsunami heights calculated by a linear model instead of tsunami waveforms. The dimensions of the matrices are reduced using principal component analysis, generally used in computer vision. Recently, Fauzi and Mizutani (2020) and Mulia et al. (2020) applied deep learning algorithms to a tsunami inundation database, and Lee et al. (2020) proposed a tsunami run-up response function including a method to reduce the number of simulations to build this function. The tsunami run-up distribution is therefore decomposed into source run-up and topographic run-up. The source run-up can be modeled by earthquake fault parameters, and the normalized topographic run-up is associated with local topographic characteristics.

Alternative approaches for capturing localized maximum wave heights are introduced here, with one approach also being capable of capturing the inundation. These automated approaches are based upon transfer functions and machine learning techniques. In practice, they are extremely quick to run ($\mathcal{O}$(s)) and are computationally cheap once the local response to an incoming tsunami is captured ahead of time. In a warning context, all that is required are the regional (coarse grid) maximum wave heights, which could be sourced from pre-computed databases (Gailler et al., 2013; Reymond et al., 2012) or obtained from "on-the-fly" regional simulations (Jamelot & Reymond, 2015). The local variability in the tsunami hazard can be rapidly deduced by taking the maximum wave height from a regional scale and transferring it to the high-resolution local area by incorporating the localized response. Due to the lack of available data for the areas considered here, in this work, the goal is to approximate computationally expensive high-resolution localized simulations. It should be stressed that the high-resolution simulations are also not perfect predictors of real tsunamis, leading to an additional source of predictive error that is not treated in this study.

The transfer function method is derived from the established Green's Law, the form of which is outlined in Section 2.1 and the method for using this approach is highlighted in Section 2.1.1. The extended version of Green's Law introduces site specific amplification factors. An automated approach for optimizing these local parameters is given in Section 2.1.1. Details of the machine learning approach are provided in Section 2.2. A robustness study of the optimized site specific amplification parameters is presented in Section 3.1, followed by a comparison against the performance of the machine learning technique in Section 3.2. Due to the large data constraints on the machine learning approach, results for the other sites studied here are presented with the transfer function only and are included in Appendix A. This paper is wrapped up with concluding remarks and proposed future work (Section 4).

2. Methods

2.1. Transfer Functions—Extended Green's Law

The evolution of sea-surface gravity waves over varying bathymetry has captivated scientists for centuries and has been the subject of an extensive body of literature. The classical work of Green dealt with the evolution of shallow water waves in channels with varying depth and width (Green, 1838). His key finding is now referred to as Green's Law, which states that when the bathymetry is slowly varying with respect to the wavelength, the wave height at a shallower point (η_2) is dependent on the wave height at deeper point (η_1) and the ratio of the depths (h_1/h_2). In a 1D channel of constant width the exact form of this functional relationship is given in Equation 1:

$$\eta_2 = \eta_1 \left(\frac{h_1}{h_2} \right)^{\frac{1}{4}}. \tag{1}$$

Green's Law effectively captures the shoaling phenomenon in a 1D case with the wave approaching normal to the shoreline. The theory neglects various 2D physical phenomena, such as reflection and refraction which play an extremely important role in the amplification of shallow water waves. Further, in reality the tsunami wave isn't necessarily approaching normal to the shoreline and instead propagates as a complex wave train over varying bathymetry in the long-shore direction (George et al., 2020). Recently, extended versions of Green's Law have been developed (Lalli et al., 2019; Reymond et al., 2012), which aim to account for the localized effects on the tsunami amplification. These extended versions of Green's Law can capture local variability in tsunami wave height through the incorporation of a local amplification parameter.

Lalli et al. (2019) derived the following extended version of Green's Law

$$\eta_2 = \alpha \eta_1 \left(\frac{h_1}{h_2} \right)^{\frac{1}{4}}, \tag{2}$$

where η_2, η_1, h_2, and h_1 are the same as in Equation 1, and α is defined as the local amplification parameter.

The extended version of Reymond et al. (2012) (not shown here) has been used to forecast maximum tsunami wave heights along the French Polynesia (Jamelot & Reymond, 2015) and Mediterranean coastlines (Gailler et al., 2017). In these works, uniform or coarsely regionalized values of the local amplification parameters were obtained for a whole site (harbor/bay area) via a "trial and error" approach.

The work presented here focuses on the α formulation (Equation 2). However, it can be shown that the α formulation is mathematically equivalent to the empirically derived extended version of Green's Law given in Reymond et al. (2012). In comparison to the formulation given in Reymond et al. (2012) the α formulation possesses additional benefits, namely the absence of an added free parameter and the stronger mathematical underpinning provided by Lalli et al. (2019).

Differing from the "trial and error" approach undertaken in Jamelot and Reymond (2015) and Gailler et al. (2017) an optimization methodology which optimizes for the local amplification parameters finely (i.e., highlights detailed variability within a site) has been developed here (Section 2.1.1). The high-resolution sites studied are located along the French Atlantic and Mediterranean coastline (Figure 1). Due to the lack of tsunami data for the areas of interest a collection of scaled historical events and hypothetical scenarios are simulated to capture a site's response to an incoming tsunami. As these site specific parameters play a vital role in the accuracy of the localized forecasts, a robustness study is carried out (Section 3.1).

2.1.1. Forecasting the Local Maximum Wave Height

A localized forecast of the maximum wave height $\eta_f(\alpha)$ at each point in the fine grid is calculated using Equation 2, which is reproduced here with the dependence on α clearly shown,

$$\eta_f(\alpha) = \alpha \eta_p \left(\frac{h_p}{h_f} \right)^{\frac{1}{4}}, \tag{3}$$

with η_p defined as the wave height at a predictor (deeper) point, h_p/h_f is the ratio of depths (as outlined in Section 2.1) and the value for α is obtained by an optimization procedure introduced in Section 2.1.2. The predictor point at

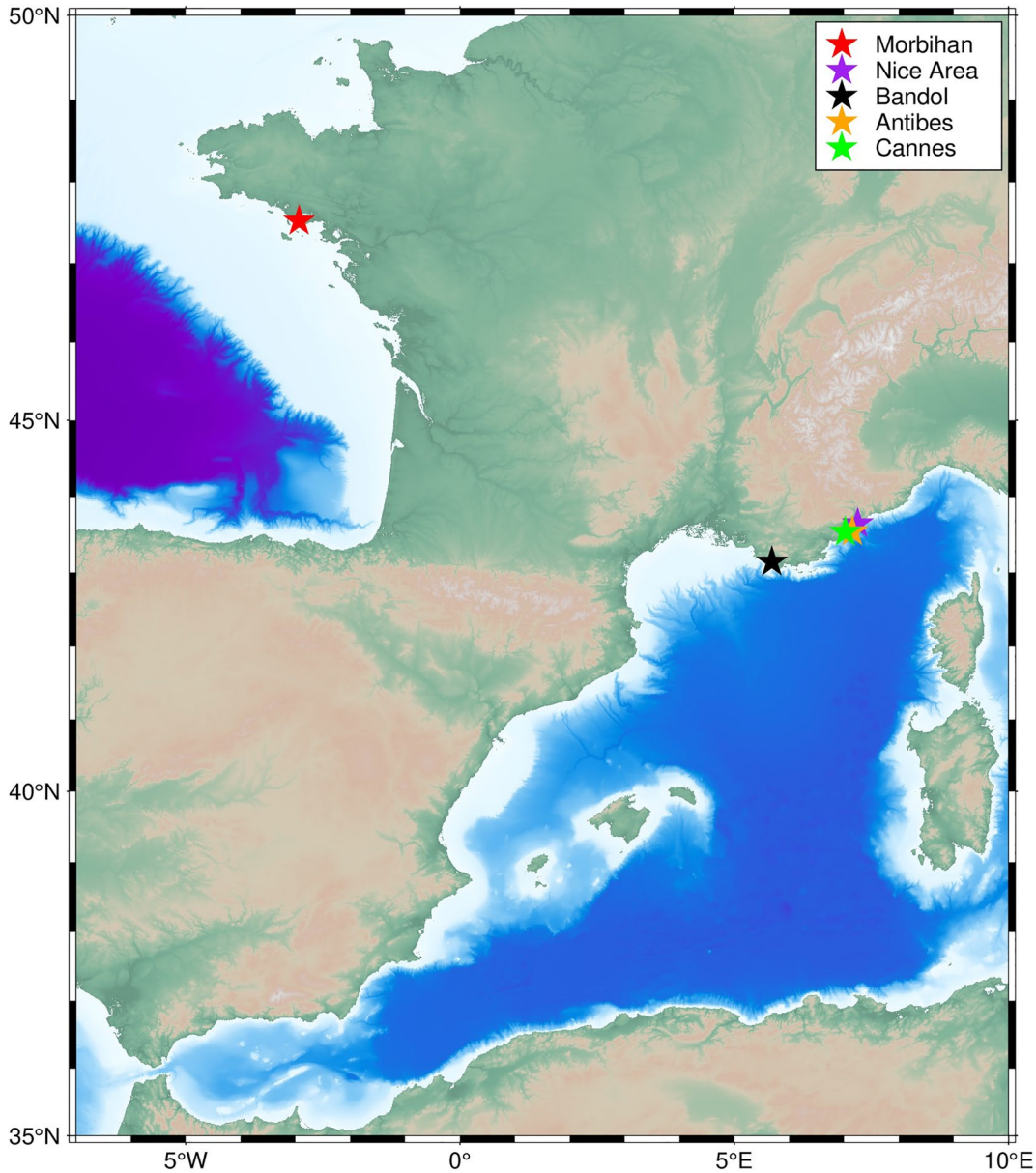

Figure 1. Sites of interest along the French Atlantic and western Mediterranean coastline. Results for Morbihan, Nice Area, Bandol, and Antibes are presented in Appendix A.

each step is chosen to be the nearest neighbor in geographical space which is deeper than the point of interest. The motivation for this choice is given in Section 2.1.3. By using the nearest deeper neighbor as the predictor point the workflow marches serially through the fine grid from the deepest point to the shoreline, predicting the maximum wave height at each location. A search algorithm for the nearest deeper neighbor is implemented once for each site of interest, with the geographical location of the associated neighbor stored and quickly accessed when carrying out the forecasting and optimization. For the deepest point in the fine grid the coarse grid results are used for the predictor point in the optimization of the α value and forecasted wave height.

In summary, the workflow sequentially moves through the fine grid from deepest point (h_1) to the shoreline with the choice of predictor point determined as follows (where h is the value of bathymetry at the point of interest):

1. If $h = h_1$: extract the maximum wave heights and bathymetry values from the nearest coarse grid points in geographical space. An interpolation of these values is defined as the predictor point.

2. If $h_1 < h \leq 1$ m: the predictor point is chosen to be the nearest deeper neighbor in geographical space.
3. If 1 m $< h <$ 0 m: the wave heights are extrapolated to the shoreline using the values obtained at the nearest deeper neighbor with $h < 1$ m.

The code which carries out this workflow is predominantly written in Python with calls to functions written in C to carry out the optimization algorithm (Section 2.1.2). Despite the code operating serially, due to reasons outlined above, the hybrid language (Python and C) implementation allows one to rapidly map the amplification parameters for a site (runtime $\mathcal{O}$(mins)). The Python and C code can be accessed at the following repository: Giles (2022).

2.1.2. Optimization of the Local Amplification Parameters α

Values of the local amplification parameter, α, are calculated at each point in the fine grid by using an automated optimization procedure via a gradient descent algorithm. The localized forecast which is dependent on α (Equation 3) is compared to a computationally expensive high-resolution simulation (ground truth) of the same event. An ordinary least squares optimization of the site specific values has also been carried out (not shown here) and it produces almost identical results to the gradient descent algorithm introduced. The optimized value of α is obtained by minimizing a calculated cost function

$$C(\alpha) = \frac{1}{N}\sum_{i=1}^{N}(\eta_s - \eta_f(\alpha))^2, \tag{4}$$

where η_s is the fine resolution simulated wave height (*ground truth*), $\eta_f(\alpha)$ is the forecasted wave height from Equation 3 and N is the total number of tsunami scenarios simulated.

When selecting a cost function, one must decide whether the approach should be optimized to handle larger wave heights or perform satisfactorily across the range of wave heights present in the training data set. The mean squared error cost function is chosen here, as it will place a greater emphasis on the larger tsunami wave heights. From a warning perspective, tsunami waves of less than 10 cm are not considered to be impacting and therefore the cost function is chosen to reflect this. The choice of evaluation metrics in Section 3 will also focus on larger tsunami wave heights.

Once the local amplification parameters have been pre-calculated, a tsunami early warning center could utilize this same method outlined in Section 2.1.1 in forecasting mode (i.e., without the optimization steps) to obtain localized wave forecasts from "unseen" regional maximum wave heights with minimal runtime required ($\mathcal{O}$(s)).

2.1.3. Motivation and Implications of the Marching Scheme

Marching through the fine grid from the deepest point to the shoreline involves the repeated composition of the extended Green's Law at each prediction point. This strategy results in the optimized α values having a multiplicative effect on the predicted wave heights. To highlight the multiplicative effect of the α values we take a test point with a bathymetry value of h_n. We assume that this test point is close to the shoreline. Therefore $h_n < h_1$, where h_1 is the bathymetry depth of the deepest point in the fine grid. For the extended Green's Law implementation, the predicted maximum wave height at h_n is given as

$$\eta_n = \alpha_n \eta_{n-1}\left(\frac{h_{n-1}}{h_n}\right)^{\frac{1}{4}}, \tag{5}$$

where h_{n-1} is the bathymetry value of the nearest deeper neighbor, η_{n-1} is the predicted wave height at that neighboring point and α_n is the optimized amplification parameter for the test point. As this is just the repeated composition of the transfer function, one can rewrite η_{n-1} as

$$\eta_{n-1} = \alpha_{n-1}\eta_{n-2}\left(\frac{h_{n-2}}{h_{n-1}}\right)^{\frac{1}{4}}. \tag{6}$$

Subbing expression Equation 6 into Equation 5 and repeating until the deeper point corresponds to h_1 (i.e., working your way back to the deepest point in the domain) one gets,

$$\eta_n = (\alpha_1 \times \alpha_2 \times \cdots \times \alpha_n)\,\eta_1 \left(\frac{h_1}{h_n}\right)^{\frac{1}{4}}. \tag{7}$$

The multiplicative effect of the α values on the predicted wave height can now be seen. The α values in the product depend on the individual selection of the nearest deeper neighbor at each predictive step. By repeating this analysis for the "pure" Green's Law implementation (null model) one can show that the predicted wave height at the same test point will be independent of intermediary deeper points and will solely depend on the value of the wave height at the deepest point in the fine grid and the ratio of the bathymetry depths. This implies that the null model will predict uniform wave heights along isobaths of the fine grid domain.

An alternative approach to the marching scheme and resultant reliance on the product of deeper neighbors (Equation 7) would be to use the deepest point in the fine grid as the predictor point for each point in the fine grid. The product of α values in Equation 7 would therefore be replaced by individual optimized values (β) for each point,

$$\eta_n = \beta_n \eta_1 \left(\frac{h_1}{h_n}\right)^{\frac{1}{4}}. \tag{8}$$

Before comparing this alternative approach to the marching scheme, it is worth reiterating that the marching scheme involves the repeated optimization of separate cost functions with different depth ratios at each step. Despite the mathematical reasoning for Equation 7, the current implementation involves using a gradient descent algorithm which is unaware of the product of α values and is solely optimizing for α_n in Equation 5. Further, as forecasted wave heights at the predictor point are used in the marching scheme the resultant optimization set up is different when compared to the alternative approach. Therefore, it can be shown that $(\alpha_1 \times \alpha_2 \times \cdots \times \alpha_n) \neq \beta_n$. However, when the two approaches are used to predict events the resultant wave heights are found to be identical. The key benefit therefore for the marching scheme is the computational efficiency, as the optimization generally requires fewer iterations to converge to the optimal value (α_n is generally closer to 1).

2.2. Multi-Layer Perceptron

A Multi-Layer Perceptron (MLP) is trained and tested for predicting maximum wave heights and inundation in the Cannes region. The authors were restricted to the Cannes region as training a neural network requires a substantial number of high-resolution model simulations as input data, which are computationally expensive to produce. As seen in Appendix A only a limited number of scenarios were available for the other sites studied. The MLP consists of an input layer, two hidden layers and an output layer. A cropped array of the maximum wave heights from the coarse regional results (mother grid Figure 2) is flattened and fed into the input layer. The output layer corresponds to a flattened array of the maximum wave heights in a subset of the Cannes region (daughter grid Figure 2). The MLP is then trained on the mother/daughter grid pairs, with standard gradient descent and back-propagation techniques used to optimize for the weights and bias terms in the network.

A python code which utilizes the Keras and TensorFlow packages (Abadi et al., 2015) was developed to construct and train the MLP. The mother grid has dimensions 150×150, therefore the input to the MLP is a flattened array of dimensions $22{,}500 \times 1$. It is a fully connected network with the two hidden layers having 256 neurons each. The output dimension is $120{,}000 \times 1$, as the daughter grid has dimensions 300×400. The activation function in each layer is the popular Relu function. L2 regularization and neuron dropout are implemented in each of the hidden layers to avoid overfitting. Identical to the extended Green's law approach, the loss function is chosen to be the mean squared error. As stated this choice of loss function will place a greater emphasis on the larger wave height events. The evaluation metrics used to compare the performance of the MLP and extended Green's Law approach (Section 3) reflect this emphasis on larger tsunami wave heights. To generalize well the MLP model is trained with the full training set (Section 2.3) for 200 epochs. Further information on the activation function, layer features and loss function can be found on the Keras API website (https://keras.io/api/). The MLP implementation developed here can also be located at the referenced repository (Giles, 2022). Finally, to plot the predictions of the MLP or to carry out analysis on them, the outputs are reshaped back into two dimensional arrays (300×400).

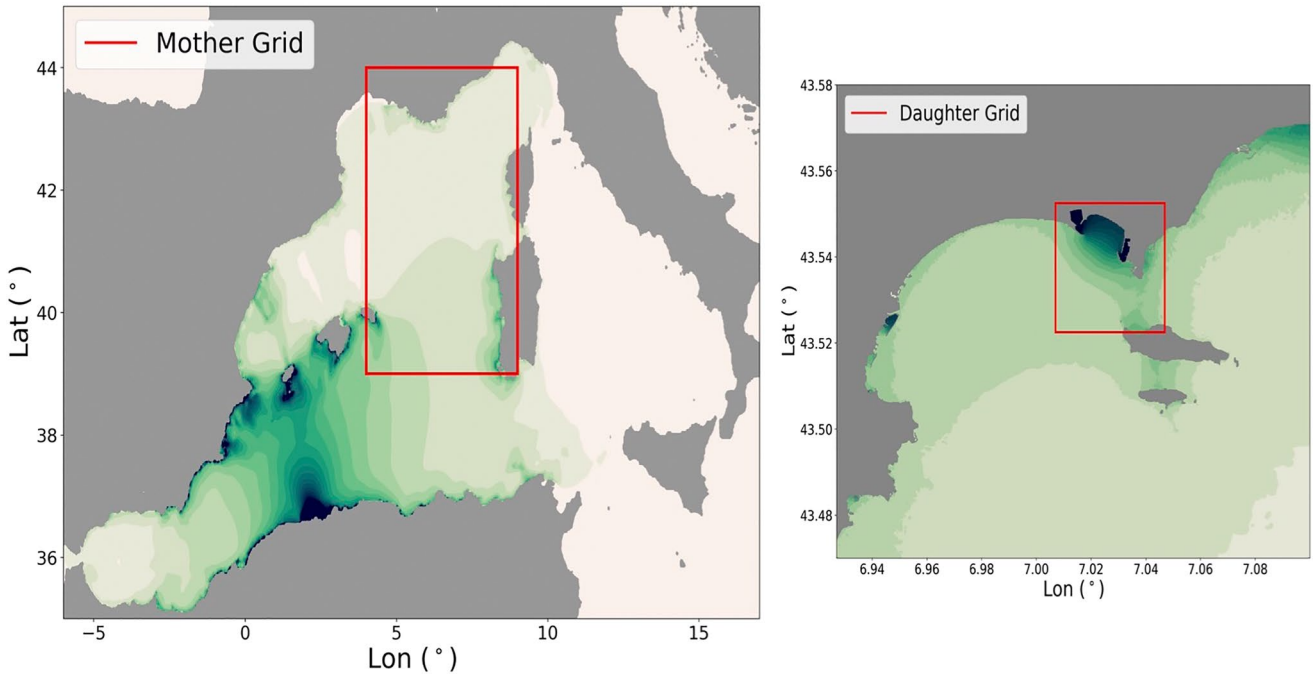

Figure 2. Left: cropped region of the regional maximum wave heights which are fed in as the input grids to the Multi-Layer Perceptron (MLP; mother grid). Right: cropped region of the localized maximum wave heights which are the targets of the MLP (daughter grid).

2.3. Data Set

A large number of high-resolution simulations for the Cannes region performed with the Taitoko code (Heinrich et al., 2021 down to 10 m resolution, which have been carried out under a different framework (Souty and Gailler (2020) within the NARSIS project, Foerster et al. (2020)), are used for this study. The total number of tsunami scenarios for the Cannes region is 120 simulation pairs (consisting of both coarse and fine grid simulation results), drawn from the seismogenic zones of the north Algerian margin and the Ligurian Sea (see Figure 3), the latter being the most tsunamigenic for the target area. This data set is built from tsunamigenic earthquakes following the western Mediterranean basin active tectonic structures (French tsunami warning center fault database, Gailler et al. (2013)) with magnitudes ranging from 6.2 M_w to 7.5 M_w. The data set is split into 10 subsets (9 training subsets and 1 test subset), with each data subset containing 12 scenarios and representing a stratified sample, that is, each subset contains a proportional amount of scenarios from each seismogenic zone.

3. Results

3.1. Robustness of the Local Amplification Parameters

The robustness of the local amplification parameters is investigated by increasing the amount of data being used in the optimization method. If the values are stable they can be considered to be well constrained by the input high-resolution model results. To test the robustness of the optimized amplification parameters, the values are optimized with increasing amounts of "seen" data (Figure 5), one data subset (12 scenarios) at a time.

The optimized α values obtained with one training data subset (12 tsunami scenarios) are plotted in Figure 4. One can see that the optimized values are close to 1. However, due to the multiplicative effect of the α values in the marching scheme (see Section 2.1.3) the percentage differences from 1 have a bearing on the predicted wave heights.

Changes in the calculated α values with increasing amounts of "seen" data are shown in the subplots of Figure 5, where the incremental relative differences

$$\Delta_{\text{rel}}\alpha = \frac{|\alpha^n - \alpha^{(n-1)}|}{\alpha^n} \tag{9}$$

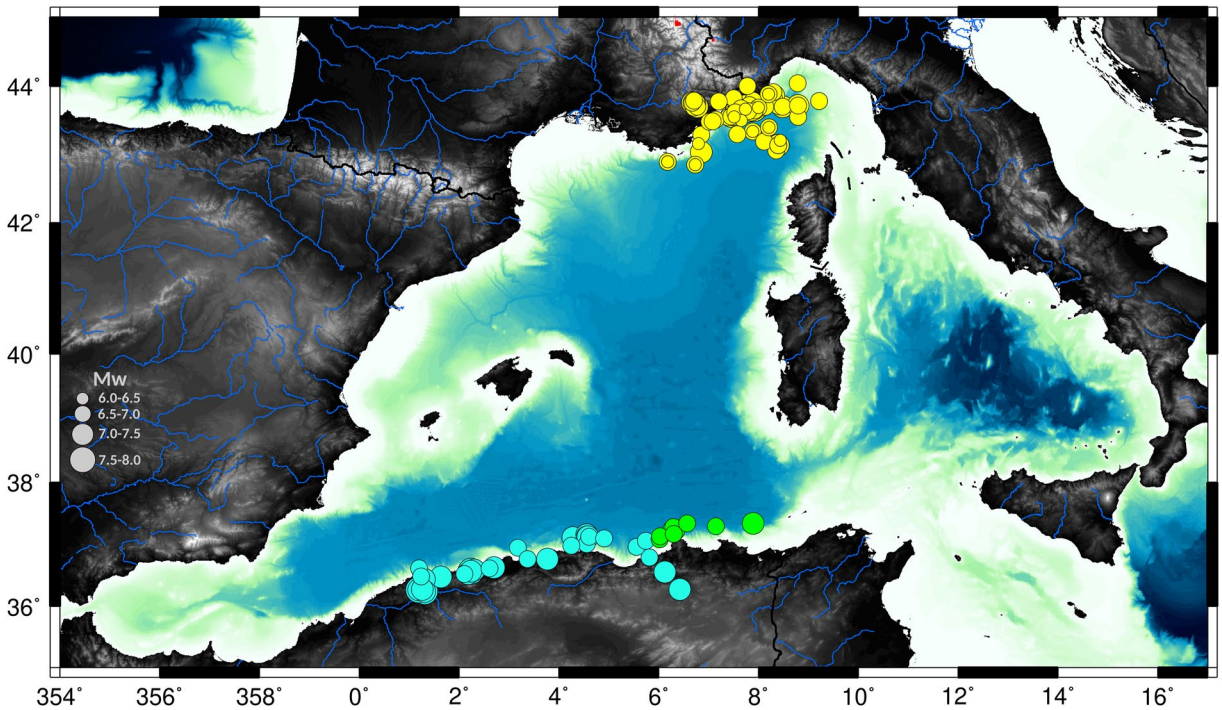

Figure 3. Map of the location of the seismogenic scenarios used in this study. The seismogenic zones are derived from Sørensen et al. (2012) and the scenarios come from Souty and Gailler (2020). The tsunamigenic potential for the Cannes region mainly comes from the North Algerian margin (turquoise and green) and from the Ligurian Sea (yellow).

are calculated at each point in the fine grid, with the superscripts $n-1$ and n referring to the number of subsets used. Each subset contains 12 scenarios. The value of n increases as one moves from the top left to bottom right subplot (Figure 5). The convergence of the optimized parameters with increasing data can be observed in each subplot of Figure 5. This is a promising result as it highlights the robustness of the optimized α values and minimal amounts of scenarios (i.e., costly high-resolution simulations) are required to accurately capture the localized response.

3.2. Comparison Between Extended Green's and MLP Model

The MLP is trained on the same Cannes region data set as the one that was introduced above (Section 2.3). The performance of the MLP model is compared to the extended Green's law approach on the same testing subset. The α values obtained in Section 3.1, where 12 and 108 tsunami scenarios are used in the optimization, are chosen for the comparison. The two sets of local amplification parameters are selected for the comparison to explore the minimal data requirements of the extended Green's approach.

The performance of the approaches at forecasting the tsunami events is demonstrated by comparing the forecasted wave heights (η_f) to the fine grid simulation results (*ground truth*) (η_s) at virtual wave gauges. The locations of the virtual wave gauges are selected to be at a depth of 2 m and are marked as red stars in the bathymetry subplot, see top figure in Figure 6. Log-log plots of the forecasted versus *ground truth* wave heights can be seen in the middle row of Figure 6. Minimal differences can be observed between the extended Green's Law model which was trained using $N = 12$ and 108 scenarios. The extended Green's approach generally overpredicts the maximum wave heights of the smaller wave heights and more accurately forecasts for the larger values. As the mean squared error function (Equation 4) was chosen for the loss function, thus placing a greater emphasis on larger wave heights, this behavior was to be expected. The MLP model performs better across the range of maximum wave heights, despite the same choice of loss function. As stated in Section 2.1.3, Green's Law (null model) predicts uniform wave heights along iso-baths of the fine grid domain. This is clearly seen as the virtual gauges are all located at a depth of 2 m.

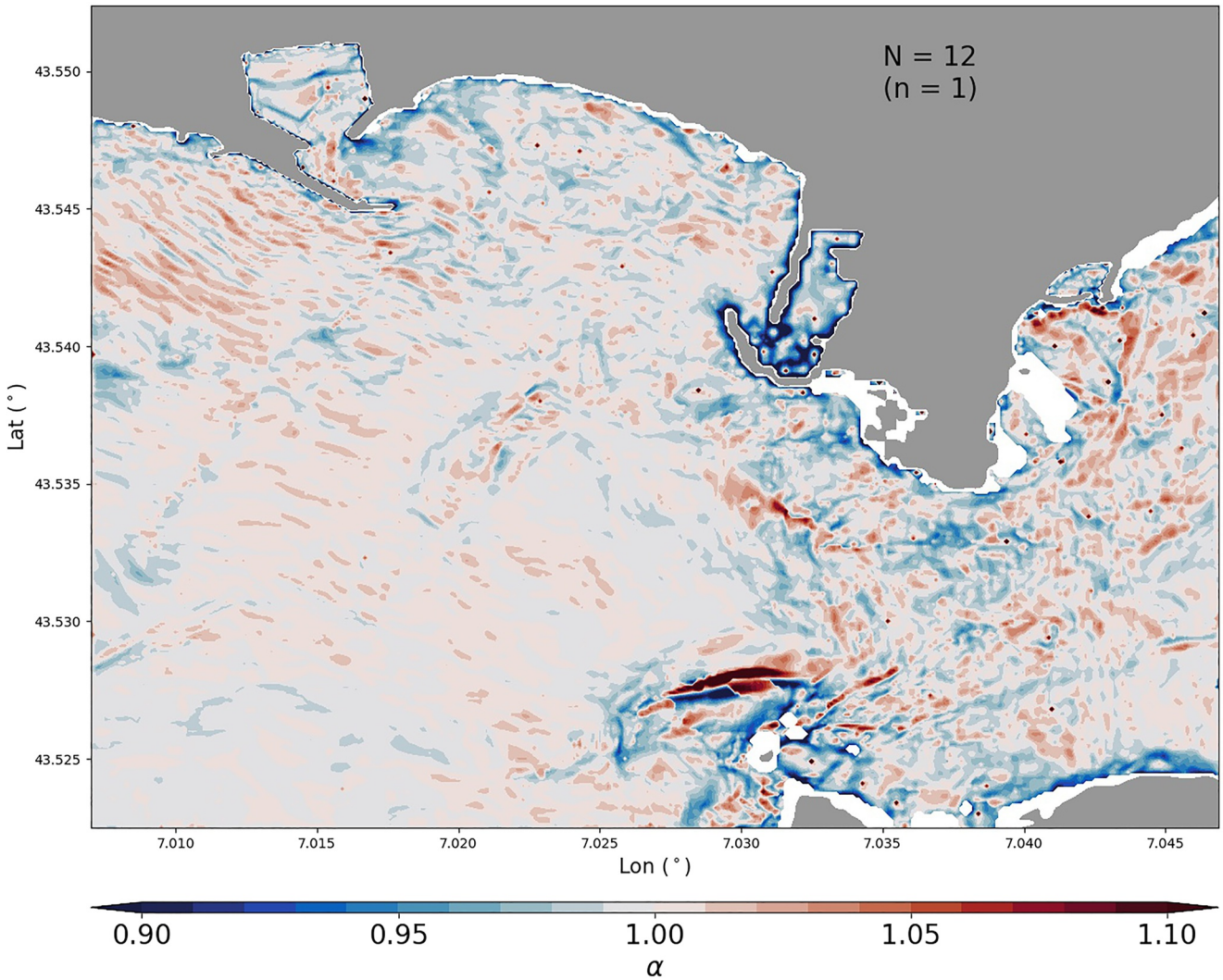

Figure 4. The optimized local amplification parameters (α in Equation 5) for the Cannes region, where the number of tsunami scenarios (N) used in the optimization is 12, which corresponds to one training data subset ($n = 1$).

To emphasize the performance of the extended Green's Law and MLP approaches on larger wave heights, subplots of the percentage relative error versus *ground truth* wave height are included in the bottom row of Figure 6. The current implementation of the automated approaches offer an improvement when compared to the null model results, with the percentage relative error on the majority of waves greater than 0.5 m constrained to be within ±50%.

To further highlight the performance of the approaches on larger wave heights, a tailored percentage mean relative error (Equation 10), which only includes scenarios with a maximum tsunami wave height greater than 10 cm, is calculated at each point in the daughter grid and plotted in Figure 7:

$$\text{Mean Relative Error } \% = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\eta_f - \eta_s}{\eta_s}\right) \times 100, \tag{10}$$

where η_f is the forecasted maximum wave height coming from either the extended Green's Law or MLP model, η_s is the simulated wave height and N is the number of scenarios. From a warning perspective, tsunami waves of less than 10 cm are not considered to be hazardous and therefore the cost function and choice of evaluation metric is chosen to reflect this.

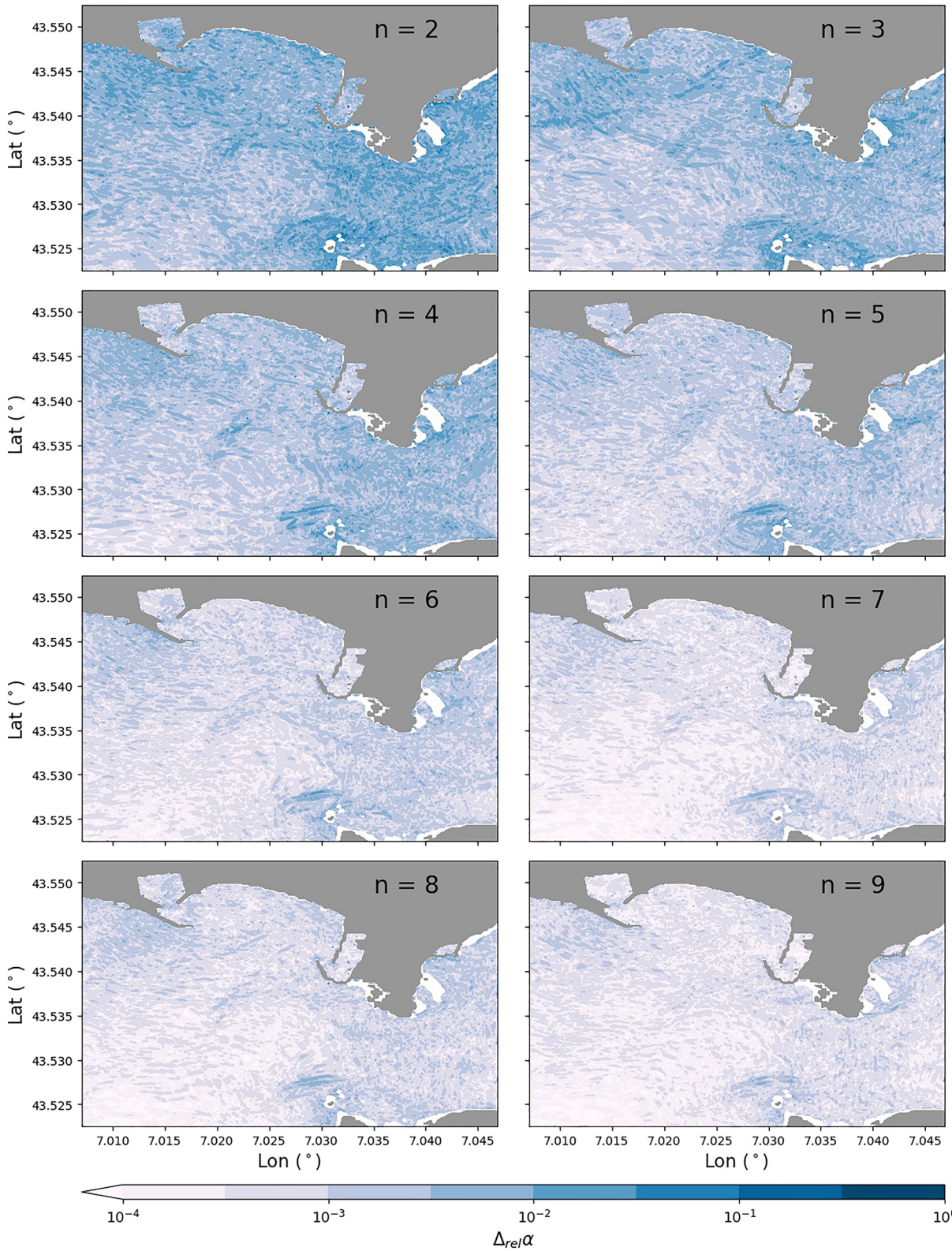

Figure 5. The effect of increasing the amount of tsunami scenarios in the optimization of the local amplification parameters (α in Equation 5) in the Cannes region. From top left to bottom right, the number of training subsets (n) and therefore tsunami scenarios (N) increases in each subplot. To showcase the convergence of the optimized parameters, the relative difference $\Delta_{rel}\alpha$ (Equation 9) at each point is plotted.

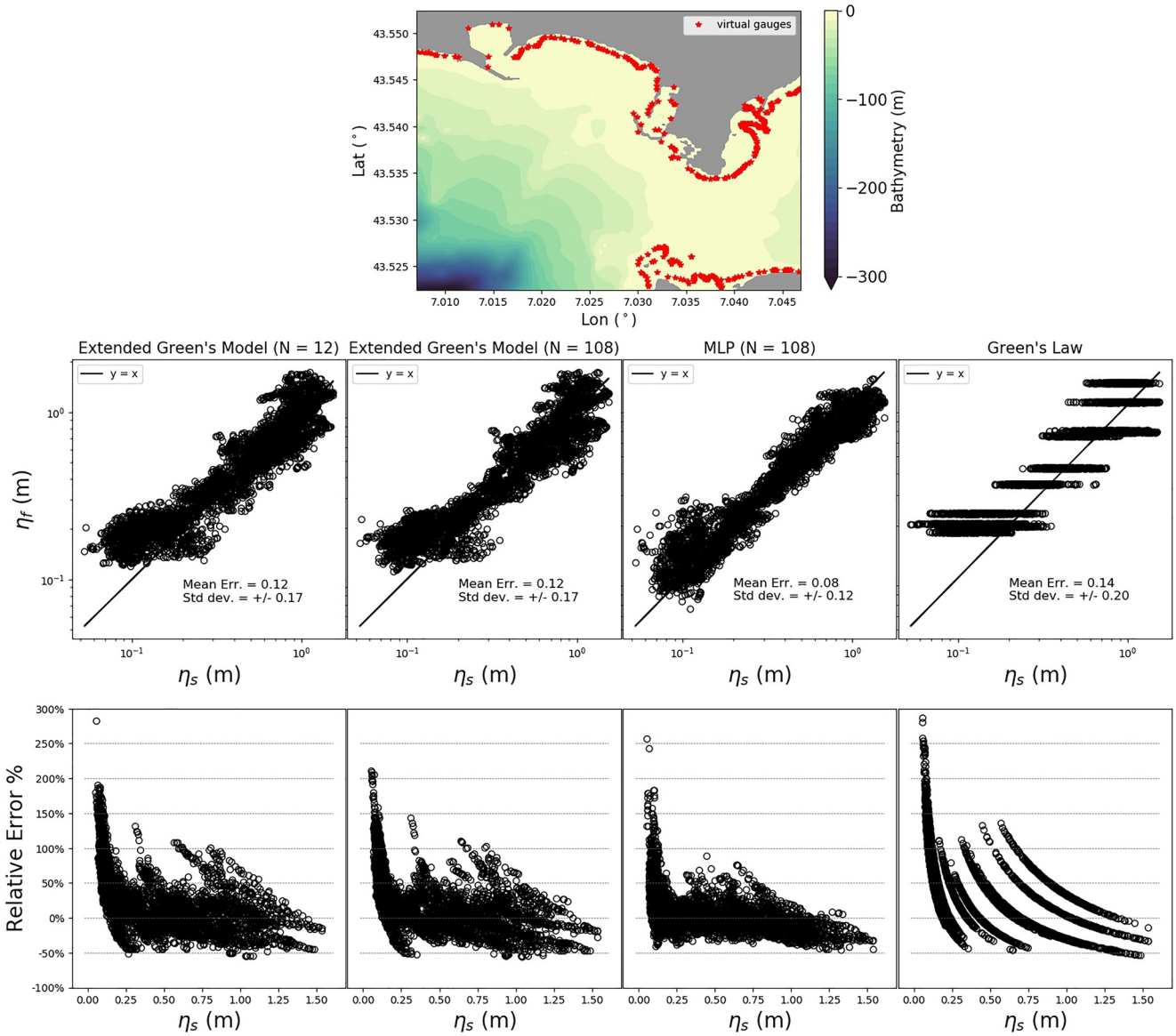

Figure 6. Top row: the fine grid bathymetry of Cannes with the location of the virtual wave gauges marked with red stars. Middle row: log-log plots of the simulated (*ground truth*) and forecasted maximum wave heights at the virtual wave gauge points for the various automated approaches and the null model. The black line is a plot of the $\eta_f = \eta_s$ line showing perfect agreement. Bottom row: percentage relative error between forecasted and simulated wave height versus *ground truth* at virtual wave gauge points produced using the various approaches.

Focusing on the performance of the extended Green's model, one can see that in Figure 7 (top row) there are slight differences between the performance of the extended Green's model optimized with 12 and 108 scenarios with the maximum mean relative error being less for the latter set up. Despite this the extended Green's model trained with only 12 scenarios does offer an improvement when compared to the null model. This behavior highlights that with minimal data requirements the extended Green's approach can be used to improve localized forecasts. Notable differences between the MLP and extended Green's Law model trained with 108 scenarios can be observed within the harbors. The null model showcases the largest mean relative errors, particularly in the eastern region of the domain. In Figure 7 a key benefit of the MLP approach is also highlighted: the ability to predict inundation. The MLP predicts a mean relative error of <50% for the inundated points. The inundated regions are masked in white in the extended Green's Law model and null model subplots. Owing to the breakdown of the Green's Law formulation in the immediate nearshore and onshore region, additional methods would have to be employed to take the offshore values and predict inundation (Smart et al., 2016).

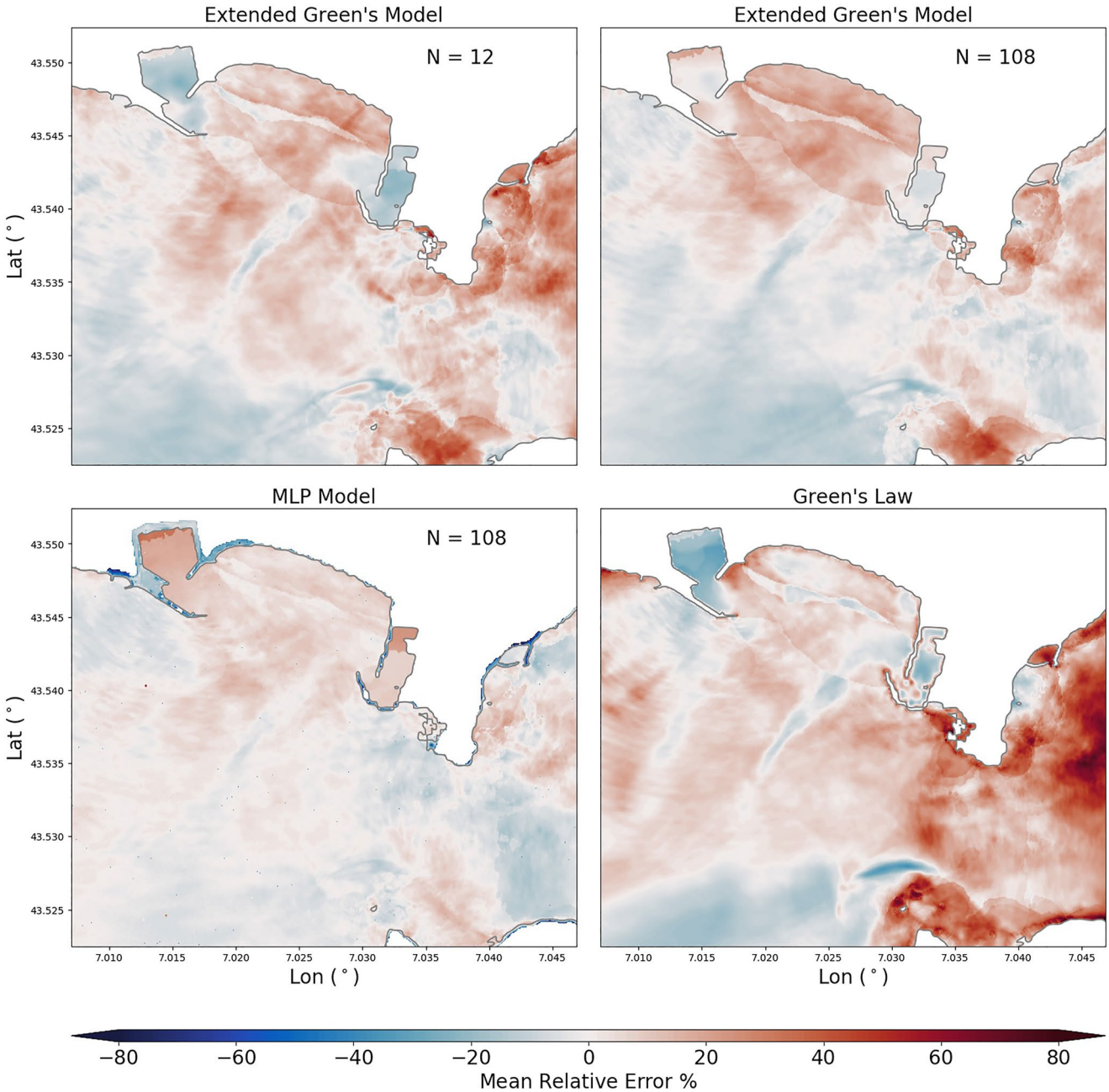

Figure 7. Tailored percentage mean relative error (Equation 10) for larger tsunami wave heights calculated at each point in the grid across the 12 "unseen" scenarios. Top row: extended Green's Law model with 12 scenarios (left) and 108 scenarios (right) used in the optimization. Bottom row: Multi-Layer Perceptron model trained with 108 scenarios (left) and the null model (right).

To provide a concrete example of the expected outputs for a warning center, Figure 8 has been included to demonstrate the performance of the approaches on a particular scenario drawn from the testing set (12 scenarios). In Figure 8, one observes the *ground truth* maximum wave heights coming from the high-resolution simulation and the corresponding predictions made by the automated approaches and the null model. The errors between the forecasted (η_f) and simulated (η_s) maximum wave heights can also be seen. Differences between the errors of the extended Green's models trained with 12 and 108 scenarios respectively can be seen, in particular for the western harbor region. The MLP model accurately predicts the extent of inundation but shows some errors in the predicted flow depth. The null model again features the largest errors when compared to the automated approaches. Overall, the areas of enhanced amplification are generally well captured by both automated approaches. If computed

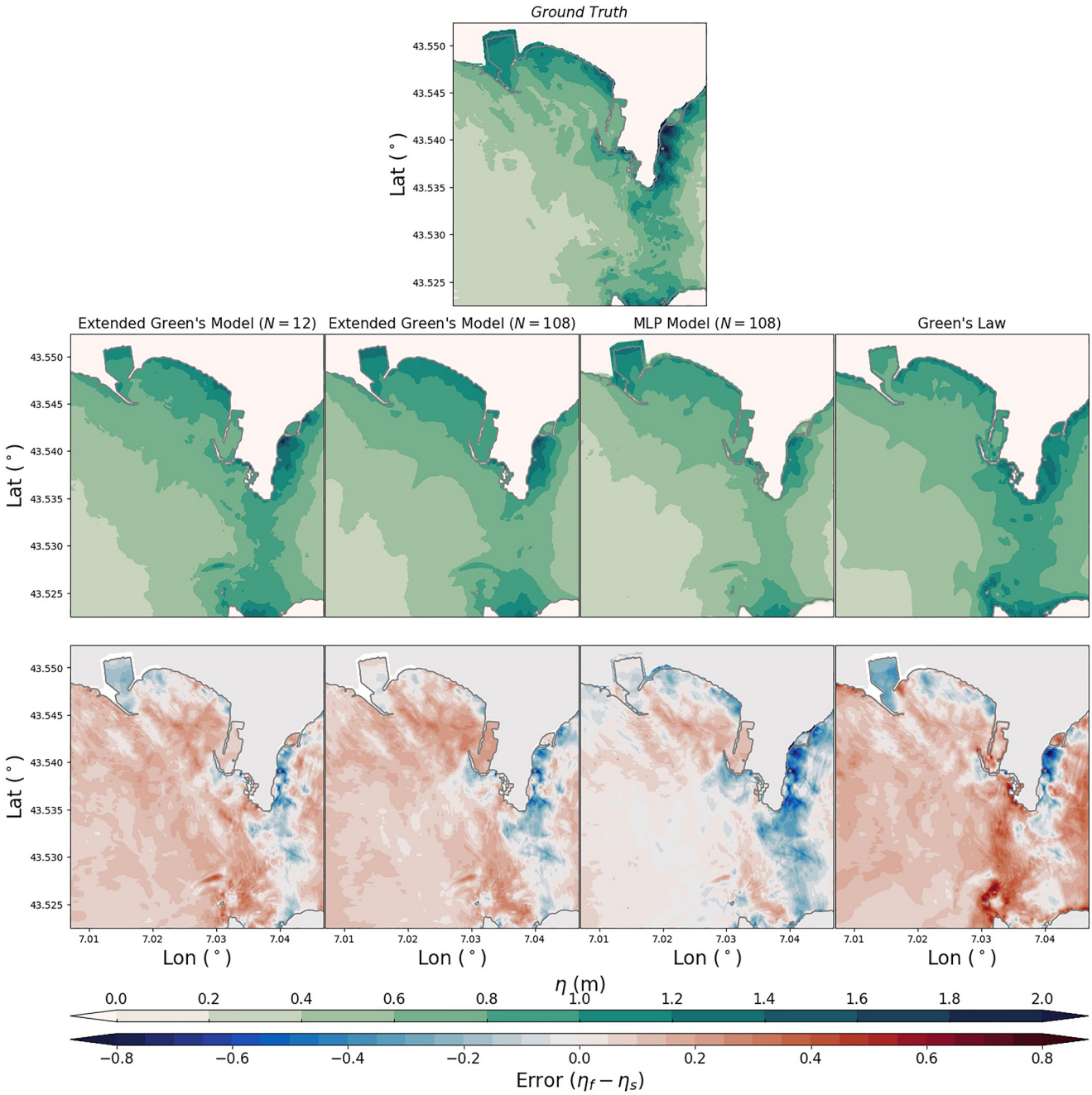

Figure 8. Top row: *Ground truth* maximum wave heights for a particular scenario drawn from the testing set of 12 scenarios. Middle row: Maximum wave heights for the same scenario as forecasted by the extended Green's Law (left and left center), Multi-Layer Perceptron (MLP; right center) and the null model (right). N in the extended Green's and MLP model results refers to the number of scenarios utilized in the training. Bottom row: The respective errors between the forecasted and *ground truth* maximum wave heights. The particular *ground truth* scenario is drawn from seismogenic zone 5 (Figure 3) and corresponds to a 7.3 M_w event.

ahead of time, these approaches could be leveraged by a tsunami warning center to capture the local variability in tsunami amplitude with minimal added runtime ($\mathcal{O}$(s)).

4. Concluding Remarks and Future Work

In a warning setting, and especially in a near field context, capturing the local response is often too computationally expensive, despite local bathymetry greatly influencing the variability in the tsunami hazard at a site. Two approaches, which can rapidly predict the local tsunami hazard when calculated ahead of time, are inves-

tigated. Sites located along the French Atlantic and Mediterranean coastlines are studied. The local response is captured by incorporating regional and computationally expensive local tsunami simulation results. Scaled historical and hypothetical earthquake sources form the initial conditions of these simulations. An automated method for calculating the site specific parameter in the formulation of the extended Green's Law proposed by Lalli et al. (2019) has been developed. This automated method differs from the "trail-and-error" approach undertaken in previous works (Gailler et al., 2017; Jamelot & Reymond, 2015). The method relies upon optimization techniques.

The robustness of the optimized local amplification factors is explored by utilizing a large database (120 scenarios) of pre-computed pairs of regional and local maximum tsunami wave heights for the Cannes region (Section 3.1). Incrementally increasing the amount of data (scenarios) in the optimization procedure results in the convergence of the local amplification parameters (α). This allows one to conclude that the value of the parameter is robust and that the extended Green's Law approach can be utilized in a setting where minimal pre-computed high-resolution data is available.

An "off-the-shelf" MLP network is then trained and tested on the same 120 scenarios for the Cannes region. The network is considered "off-the-shelf" as no hyperparameters are optimized for and the architecture of the network is a standard form. The performance of the network at predicting maximum wave heights is compared to the predictions made by the optimized transfer function method for the same 12 unseen scenarios. The relative data requirements for each approach should be highlighted, with the MLP model requiring a training set of 108 scenarios to produce satisfactory results, whilst the α values optimized using only 12 scenarios perform comparable to those optimized with 108 scenarios. However, a key benefit of the MLP model compared to the extended Green's Law approach can be seen in its ability to predict inundation.

Due to the limited number of high-resolution results for the other French sites studied here (Appendix A), only the extended Green's approach is carried out. Results for Morbihan, Bandol, Nice, and Antibes are presented in Figures A1–A4, respectively. As is the case with the Cannes results (Figure 6), in its current implementation the extended Green's Law method is shown to be able to rapidly predict the maximum wave heights, with a good agreement between the simulated (*ground truth*) and forecasted wave heights for larger events. These figures highlight the performance of the approach at capturing "seen" data, with percentage relative errors between $\pm 25\%$ for the larger values at virtual wave gauges located along the coastline. Comparing against "seen" data does not provide an accurate representation of the generalization of the approach, that is, how will the approach perform for an "unseen" event. Therefore, it is noted that a greater number of high-resolution simulations for the sites studied would allow for a more in depth testing of the optimized amplification parameters.

In the current form of the extended Green's Law (Equation 2), improved results could be obtained by optimizing for the choice of prediction point. The choice of the loss function can also play a role in the performance of the approach. A mean squared error cost function was chosen here to place a greater emphasis on larger wave heights. Further, an alternative method similar to the transfer function approach would be generalized additive models (Wood, 2017). These statistical models capture functional responses between variables and offer an alternative approach to calculating local values from off-shore wave heights.

Future work involving machine learning methods could focus on developing customized machine learning models, where convolutional neural networks could be used. The use of deep neural networks for tsunami forecasting tasks is an active area of research, with various approaches being developed (Fauzi & Mizutani, 2020; Makinoshima et al., 2021; Mulia et al., 2020). To build on these existing works, a customized convolutional neural network could be built which leverages stacked layers of the bathymetry, wave heights and flow velocities. These variables could form the input for a traditional three channel RGB (Red, Green, and Blue) image classification network. This approach of stacking the bathymetry, maximum wave heights and flow velocities into three layers would have the extra benefit of predicting the maximum flow velocities in conjunction to the maximum wave heights and inundated areas. Transfer learning techniques could also be investigated. One could incorporate well established convolutional neural networks, which have been used for similar image classification tasks, to minimize the training data requirements.

Finally, an extension of the present work could be the coupling of pre-computed localized response models with a computationally cheap statistical emulator. This would greatly improve a warning center's capabilities at providing extremely localized tsunami hazards with the associated uncertainties in a warning setting. This coupling would allow for the uncertainties on the rapid source characterization to be propagated through to the localized

tsunami wave heights. The regional maximum wave heights needed for forecasting the localized response would be computationally cheap to obtain. Then the emulator could be trained on the output of the localized response approach (transfer function or machine learning model) and then used to predict the uncertainty on the tsunami hazard. This would differ from the results presented in Giles et al. (2021) as the underlying mesh resolution could be refined by a factor more than 10, but with a reduction/minimal addition in the runtime.

At present tsunami warning center's provide rapid estimates of regional tsunami hazard maps along with predictions at various forecast points. These forecasts could be greatly augmented by incorporating the methods developed here. The maximum wave heights predicted at a regional scale form the input and the resultant output will be forecasted maximum wave heights on a localized level (harbor/port scale)—with minimal additional runtime required. The localized information will greatly benefit any resultant warnings issued.

It should be noted, that the approaches developed here do not lend themselves to rigorous uncertainty quantification. The main source of uncertainty will center on the use of computationally expensive high-resolution simulations as the "ground truth." This choice was due to the lack of historical data available for tsunami events in the regions studied. As stated the numerical errors associated with the computational approach, despite using high-resolution simulations, are not accounted for here and are an added source of prediction error.

To wrap up this work, two separate automated approaches which are capable of capturing the local variability in maximum tsunami wave height have been developed. Both approaches offer an improvement when compared to the null model. For the extended Green's Law approach, improvements can be seen with minimal data needed in training and optimization of the local amplification parameters. The robustness of the amplification parameters is showcased by increasing the amount of training data in the extended Green's Law optimization. However, as the values of the amplification parameters converge so to does its performance. This limiting nature and the additional benefit of the MLP (ability to predict inundation) allows one to conclude that—when sufficient training data is available the MLP model provides more promising results. However, if computed ahead of time both of the developed approaches allow for extremely rapid localized forecasts of tsunami wave heights.

Appendix A: Additional Results

The data set of high-resolution simulation results for the following French coastal sites is limited to 9 tsunami scenarios for Morbihan and 8 scenarios for the Mediterranean sites (Nice Area, Bandol, and Antibes), with the seismic source parameters listed in Table A1. The sources used for the Western Mediterranean echo those defined in Gailler et al. (2017). Due to the limited availability of training data, only the extended Green's Law model is used to capture the local response. This decision is justified by the previous findings on the robustness of the local amplification parameters (Section 3.1).

Nevertheless, the minimal data set of high-resolution simulations restricts our ability to carry out a proper performance evaluation of the optimized site specific parameters. A training and testing data split cannot be afforded with such minimal high-resolution simulation data. Therefore, the performance of the extended Green's Law approach is non-optimally compared against the same scenarios used for optimization. This testing against the training data set is not ideal and therefore work is ongoing to increase the number of high-resolution simulations for the sites studied here. As before the *ground truth* results were obtained as part of a separate study, which involved running computationally expensive high-resolution nested grid simulations with the Taitoko code (Heinrich et al., 2021), down to 20 and 10 m resolution for the Atlantic and Mediterranean sites respectively.

In Figures A1–A4, the high-resolution bathymetry and optimized α parameters are plotted. Areas of enhanced amplification are associated with larger α values. The performance of the obtained amplification factors at forecasting the tsunami events is demonstrated by comparing the forecasted wave height to the fine grid simulation results (*ground truth*) at virtual wave gauges, the locations of which are marked as red stars in the bathymetry subplots. A log-log plot of the forecasted versus *ground truth* (η_s) wave heights can be seen in Figures A1–A4. As in Section 3.2, the effects of the design of the loss function in the optimization procedure can be clearly seen. Across the different sites, the extended Green's Law model underpredicts the smaller wave heights and more accurately forecasts for the larger values. If one's motivation was to predict wave heights across a broad scale of values, a cost function which is based upon a relative error formulation would be more appropriate.

Table A1
Source Parameters for the 9 Atlantic and 8 Mediterranean Reference Scenarios

Scenario	M_w	Lon (°)	Lat (°)	Depth (km)	Slip (m)	Strike (°)	Dip (°)	Rake (°)	Length (km)	Width (km)
GC01_80	8	−12.26	33.67	12.91	3.5	30	35	90	200	45
GC01_85	8.5	−12.26	33.67	16.27	7.8	30	35	90	370	55
GC05_85	8.5	−9.01	35.94	16.27	7.8	270	35	90	370	55
GC07_85	8.5	−10.64	36.38	16.27	7.8	235	35	90	370	55
GF02_85	8.5	−13.49	38.77	27.9	7.8	102	85	170	370	55
AZ01_75	7.5	−26.06	40.25	15.2	1.4	117	60	−90	100	35
AZ01_85	8.5	−26.06	40.25	24.32	7.8	117	60	−90	370	55
GCMT1_80	8	−32.77	52.62	10	10.5	186	80	−4	150	20
GCMT1_85	8	−35.1	35.17	10	10.5	282	87	15	150	20
S_7.1	7.1	3.34–4.01	36.66–37.12	2.4–22.9	128 × 0.007	54	47	90	128 × 4	128 × 4
Y_7.5	7.5	3.56	36.80	12.0	2.5	60	42	84	85	35
J_7.1	7.1	5.47–5.73−6.15	36.95–37.08−37.18	7.0	1.0–1.5−1.5	75−85−75	40	90	25−37−44	3 × 22
J_7.8	7.8	5.47–5.73−6.15	36.95–37.08−37.18	7.5	3 × 6.0	75−85−75	40	90	25−37−44	3 × 20
413_6.5	6.5	6.95	37.29	4.0	0.45	91	50	90	45	10
413_7.0	7.0	6.95	37.29	7.0	1.5	91	50	90	42	18
413_7.5	7.5	6.95	37.29	12.0	2.6	91	50	90	85	29
413_7.5s	7.5	6.95	37.29	9.0	8.0	91	50	90	33	22

Note. The Azores-Gibraltar source parameters (GC0*, GF*, and AZ*) are adapted from Matias et al. (2013) and are taken from the French tsunami warning center (CENALT) fault database Gailler et al. (2013). The GCMT scenarios are located along the Mid-Atlantic Ridge and correspond to real events that occurred in 2014 and 2015, with a magnitude increased to 8.0 in a conservative approach. The Mediterranean sources cover the range of source location, magnitude and azimuth most representative of the main tsunami sources for the French Riviera area, coming form the Algerian margin. The S*, Y*, and J* sources are based on historical events (1856 and 2003), and the 413* sources are derived from the No. 413 fault of the CENALT pre-computed scenario database. See Gailler et al. (2017) for more details.

Subplots of the percentage relative error versus *ground truth* (η_s) wave height for values greater than 10 cm are included in Figures A1–A4. The current implementation of the automated optimization procedures is found to capture the localized response to larger wave heights well, with percentage relative errors between ±25% for the larger values.

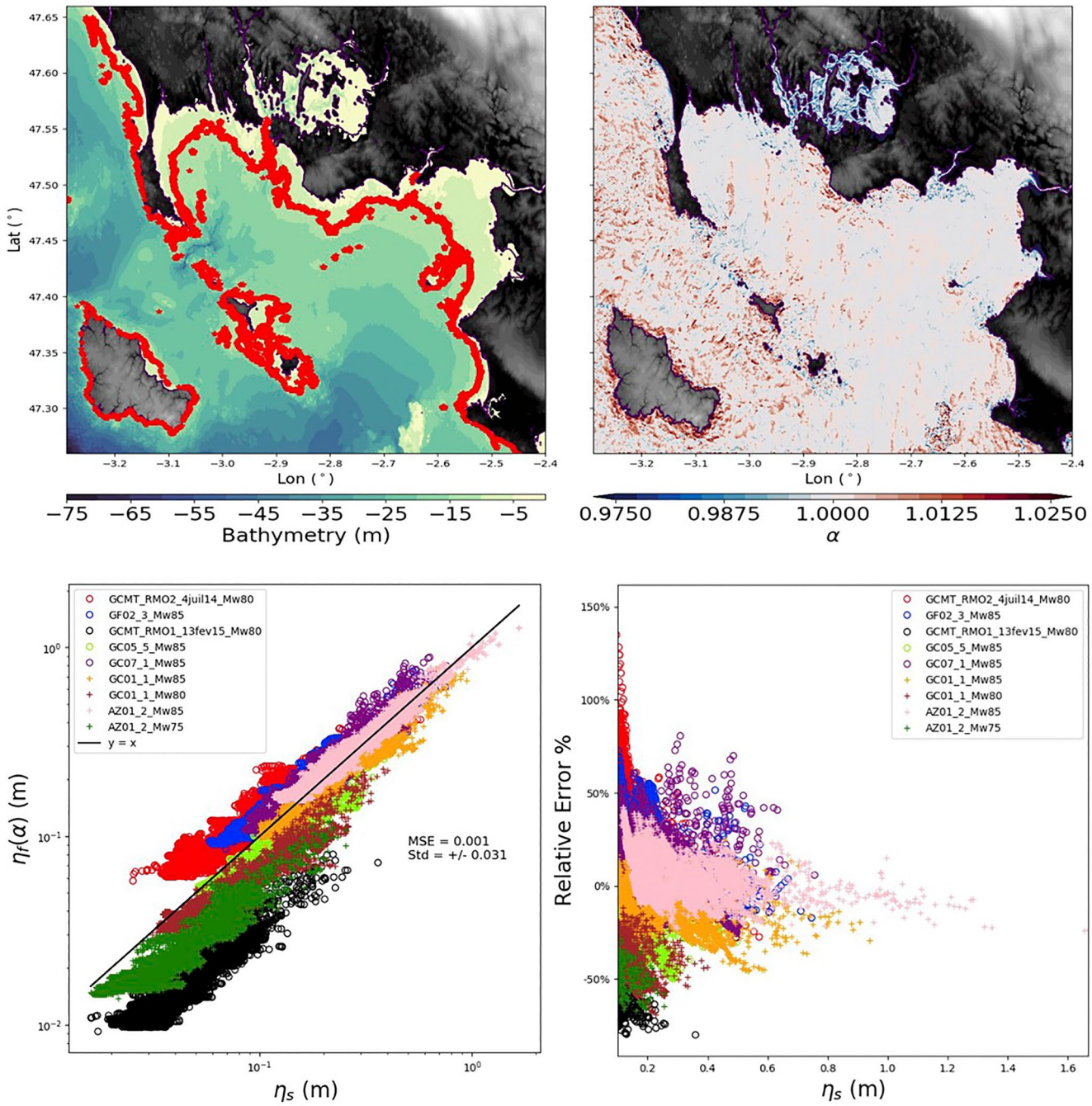

Figure A1. Top left: the fine grid bathymetry of Morbihan with the location of the virtual wave gauges marked with red stars. Top right: the optimized α (Equation 5) values. Bottom left: log-log plot of the simulated (*ground truth*) and forecasted maximum wave heights at the virtual wave gauge points. The black line is a plot of the $\eta_f = \eta_s$ line showing perfect agreement. Bottom right: percentage relative error between forecasted and simulated wave height versus *ground truth* at virtual wave gauge points for wave heights >10 cm. The wave heights from the various earthquake sources have been color-coded.

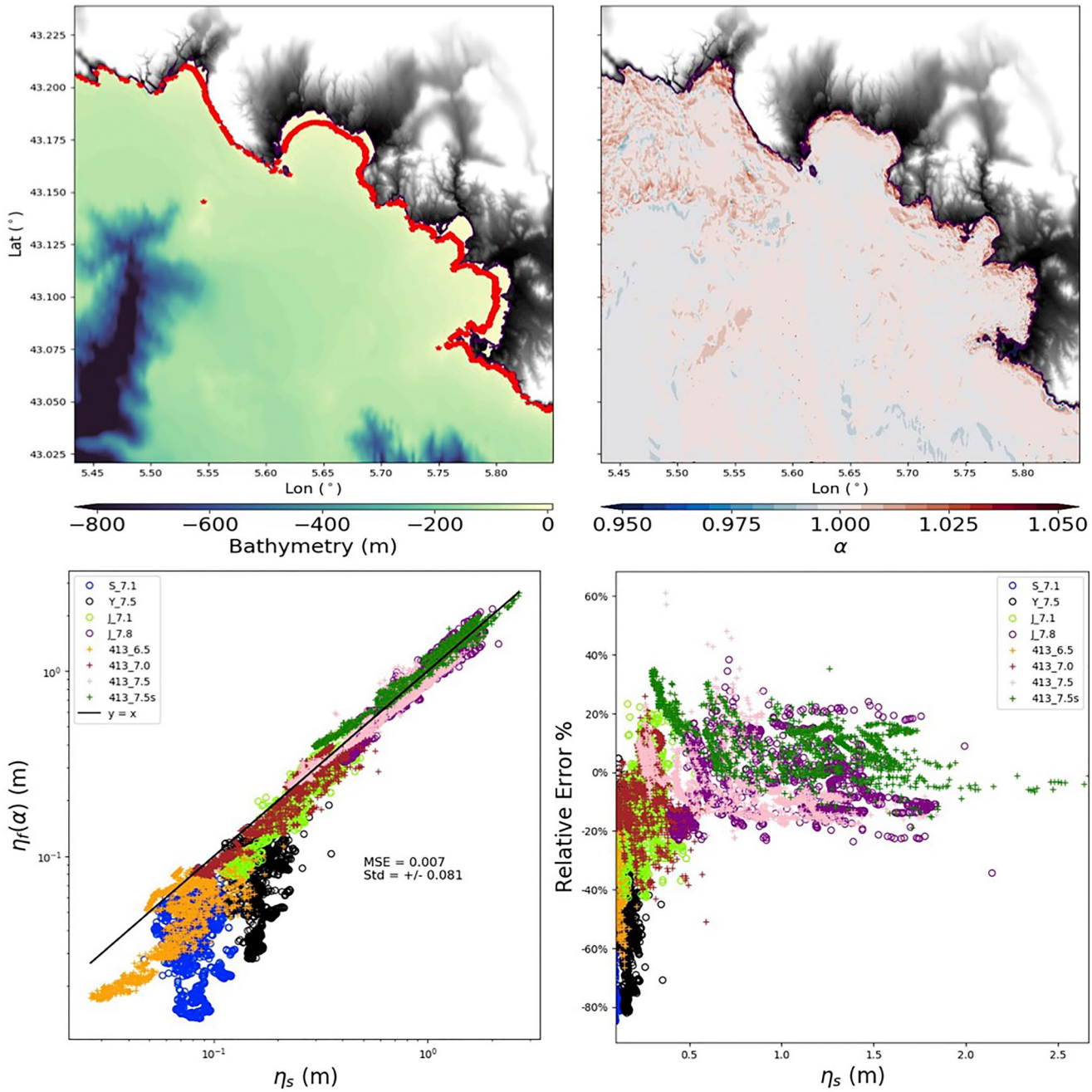

Figure A2. Top left: the fine grid bathymetry of Bandol with the location of the virtual wave gauges marked with red stars. Top tight: the optimized α (Equation 5) values. Bottom left: log-log plot of the simulated (*ground truth*) and forecasted maximum wave heights at the virtual wave gauge points. The black line is a plot of the $\eta_f = \eta_s$ line showing perfect agreement. Bottom right: percentage relative error between forecasted and simulated wave height versus *ground truth* at virtual wave gauge points for wave heights >10 cm. The wave heights from the various earthquake sources have been color-coded.

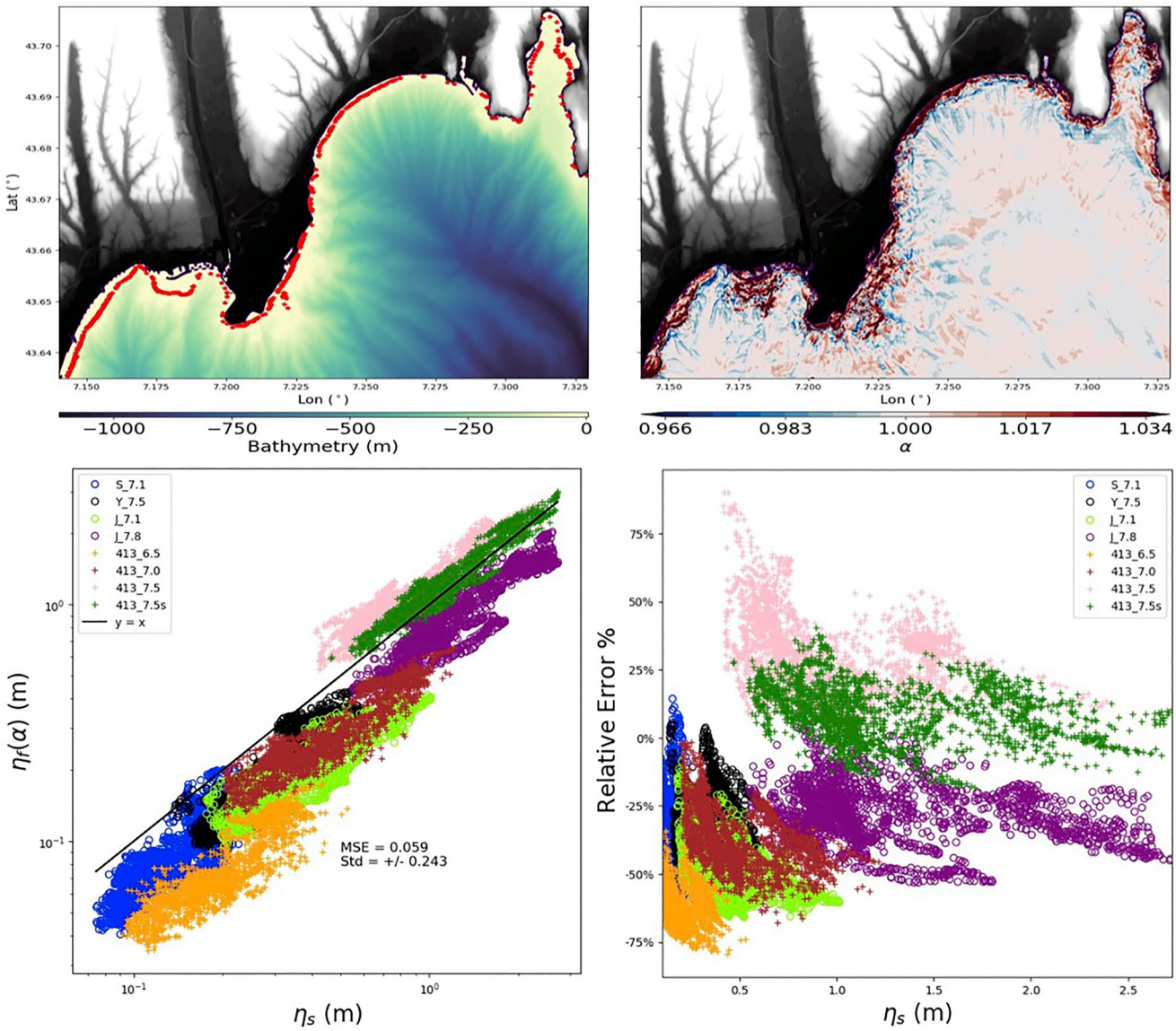

Figure A3. Top left: the fine grid bathymetry of the Nice Area with the location of the virtual wave gauges marked with red stars. Top right: the optimized α (Equation 5) values. Bottom left: log-log plot of the simulated (*ground truth*) and forecasted maximum wave heights at the virtual wave gauge points. The black line is a plot of the $\eta_f = \eta_s$ line showing perfect agreement. Bottom right: percentage relative error between forecasted and simulated wave height versus *ground truth* at virtual wave gauge points for wave heights >10 cm. The wave heights from the various earthquake sources have been color-coded.

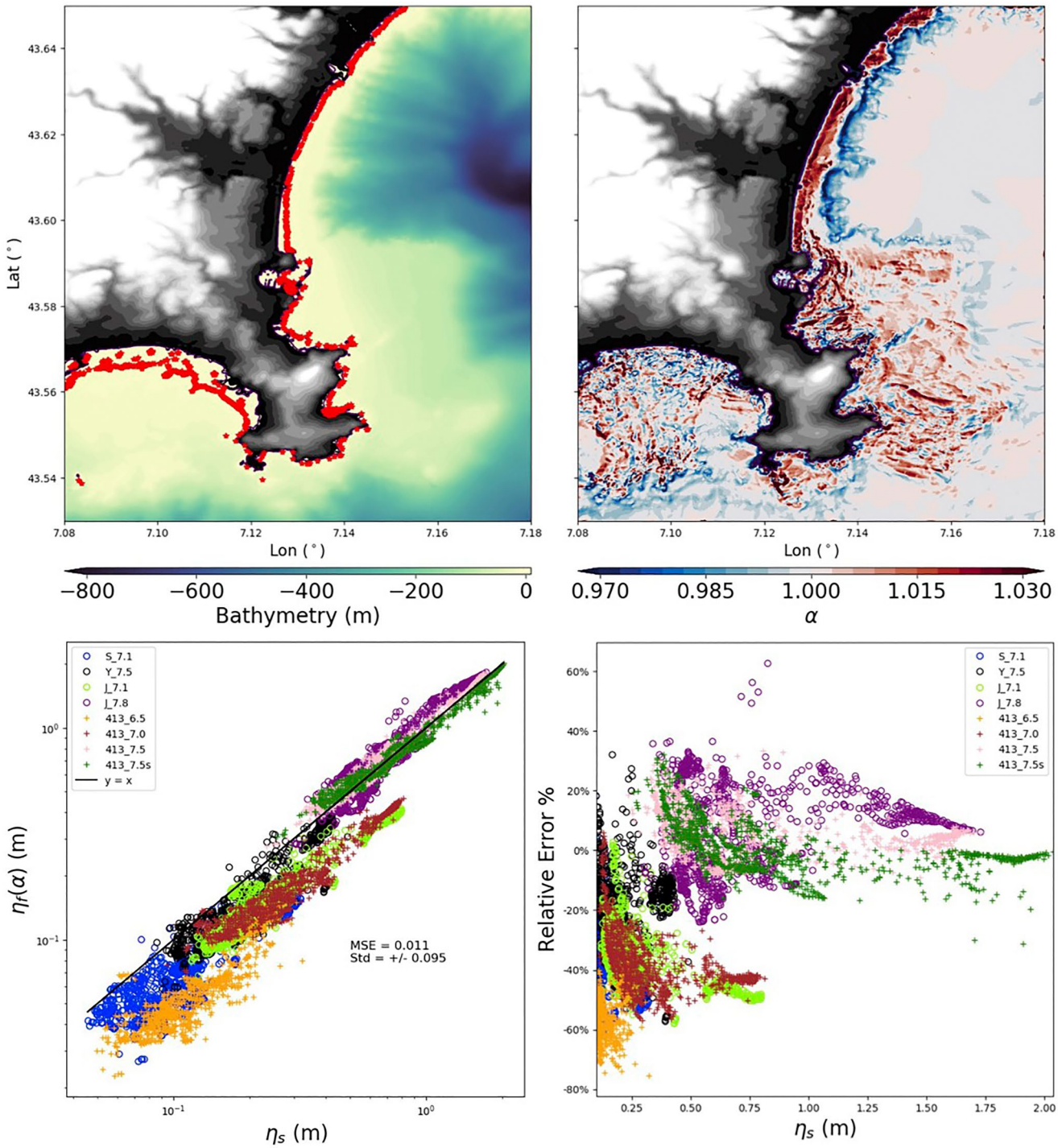

Figure A4. Top left: the fine grid bathymetry of Antibes with the location of the virtual wave gauges marked with red stars. Top right: the optimized α (Equation 5) values. Bottom left: log-log plot of the simulated (*ground truth*) and forecasted maximum wave heights at the virtual wave gauge points. The black line is a plot of the $\eta_f = \eta_s$ line showing perfect agreement. Bottom right: percentage relative error between forecasted and simulated wave height versus *ground truth* at virtual wave gauge points for wave heights >10 cm. The wave heights from the various earthquake sources have been color-coded.

Data Availability Statement

The software (version 1.0.0) developed for this manuscript is licensed under BSD 2-Clause "Simplified" License and published on GitHub (Giles, 2022).

Acknowledgments
D. Giles was supported by a Government of Ireland Postgraduate Scholarship from the Irish Research Council (GOIPG/2017/68). The authors would like to acknowledge Jessie Levillain for her efforts during her internship at University College Dublin. The authors would also like to thank Viviane Souty and Audrey Chouli for their upstream work. The scenario database and simulation results were obtained thanks to the NARSIS (New Approach to Reactor Safety Improvements, Horizon 2020) H2020 project (http://www.narsis.eu/), with the grant agreement No. 755439. The Service Hydrographique et Océanographique de la Marine (SHOM) is gratefully acknowledged for making bathymetry and topography data available free of charge on their portal data.shom.fr.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems*. Retrieved from http://tensorflow.org/

Behrens, J., & Dias, F. (2015). New computational methods in tsunami science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences*, *373*(2053), 20140382. https://doi.org/10.1098/rsta.2014.0382

Fauzi, A., & Mizutani, N. (2020). Machine learning algorithms for real-time tsunami inundation forecasting: A case study in Nankai region. *Pure and Applied Geophysics*, *177*(3), 1437–1450. https://doi.org/10.1007/s00024-019-02364-4

Foerster, E., Raimond, E., & Guigueno, Y. (2020). Probabilistic safety assessment for internal and external events – European projects H2020-NARSIS and FP7-ASAMPSA_E. *EPJ Nuclear Sciences and Technologies*, *6*, 38. https://doi.org/10.1051/epjn/2019012

Gailler, A., Hébert, H., Loevenbruck, A., & Hernandez, B. (2013). Simulation systems for tsunami wave propagation forecasting within the French tsunami warning center. *Natural Hazards and Earth System Sciences*, *13*(10), 2465–2482. https://doi.org/10.5194/nhess-13-2465-2013

Gailler, A., Hébert, H., Schindelé, F., & Reymond, D. (2017). Coastal amplification laws for the French tsunami warning center: Numerical modeling and fast estimate of tsunami wave heights along the French Riviera. *Pure and Applied Geophysics*, *175*(4), 1429–1444. https://doi.org/10.1007/s00024-017-1713-9

George, J., Ketcheson, D. I., & LeVeque, R. J. (2020). Shoaling on steep continental slopes: Relating transmission and reflection coefficients to Green's law. *Pure and Applied Geophysics*, *177*(3), 1659–1674. https://doi.org/10.1007/s00024-019-02316-y

Giles, D. (2022). Localised-tsunami-response [Software]. Zenodo. https://doi.org/10.5281/zenodo.6334280

Giles, D., Gopinathan, D., Guillas, S., & Dias, F. (2021). Faster than real time tsunami warning with associated hazard uncertainties. *Frontiers of Earth Science*, *8*, 597865. https://doi.org/10.3389/feart.2020.597865

Green, G. (1838). On the motion of waves in a variable canal of small depth and width. *Transactions of the Cambridge Philosophical Society*, *6*, 457.

Gusman, A. R., Tanioka, Y., MacInnes, B. T., & Tsushima, H. (2014). A methodology for near-field tsunami inundation forecasting: Application to the 2011 Tohoku tsunami. *Journal of Geophysical Research: Solid Earth*, *11*(119), 45–50. https://doi.org/10.1002/2014JB010958

Heinrich, H., Jamelot, A., Cauquis, A., & Gailler, A. (2021). Taitoko, an advanced code for tsunami propagation, developed at the French Tsunami Warning Centers. *European Journal of Mechanics – B: Fluids*, *88*, 72–88. https://doi.org/10.1016/j.euromechflu.2021.03.001

Jamelot, A., & Reymond, D. (2015). New tsunami forecast tools for the French Polynesia tsunami warning system Part II: Numerical modelling and tsunami height estimation. *Pure and Applied Geophysics*, *172*(3–4), 805–819. https://doi.org/10.1007/s00024-014-0997-2

Lalli, F., Postacchini, M., & Brocchini, M. (2019). Long waves approaching the coast: Green's Law generalization. *Journal of Ocean Engineering and Marine Energy*, *5*(4), 385–402. https://doi.org/10.1007/s40722-019-00152-9

Lee, J., Irish, J. L., & Weiss, R. (2020). Rapid prediction of alongshore run-up distribution from near-field tsunamis. *Natural Hazards*, *104*, 1157–1180. https://doi.org/10.1007/s11069-020-04209-z

Løvholt, F., Lorito, S., Macias, J., Volpe, M., Selva, J., & Gibbons, S. (2019). Urgent tsunami computing. In *IEEE(2019 IEEE/ACM HPC for Urgent Decision Making (UrgentHPC))* (pp. 45–50).

Makinoshima, F., Oishi, Y., Yamazaki, T., Furumura, T., & Imamura, F. (2021). Early forecasting of tsunami inundation from tsunami and geodetic observation data with convolutional neural networks. *Nature Communications*, *12*(1), 2253. https://doi.org/10.1038/s41467-021-22348-0

Matias, L. M., Cunha, T., Annunziato, A., Baptista, M. A., & Carrilho, F. (2013). Tsunamigenic earthquakes in the Gulf of Cadiz: Fault model and recurrence. *Natural Hazards and Earth System Sciences*, *13*(1), 1–13. https://doi.org/10.5194/nhess-13-1-2013

Mulia, I. E., Gusman, A. R., & Satake, K. (2018). Alternative to non-linear model for simulating tsunami inundation in real-time. *Geophysical Journal International*, *3*(214), 2002–2013. https://doi.org/10.1093/gji/ggy238

Mulia, I. E., Gusman, A. R., & Satake, K. (2020). Applying a deep learning algorithm to tsunami inundation database of megathrust earthquakes. *Journal of Geophysical Research: Solid Earth*, *9*(125), e2020JB019690. https://doi.org/10.1029/2020JB019690

Reymond, D., Okal, E. A., Hébert, H., & Bourdet, M. (2012). Rapid forecast of tsunami wave heights from a database of pre-computed simulations, and application during the 2011 Tohoku tsunami in French Polynesia. *Geophysical Research Letters*, *39*(11), 1–6. https://doi.org/10.1029/2012GL051640

Smart, G., Crowley, K., & Lane, E. (2016). Estimating tsunami run-up. *Natural Hazards*, *80*(3), 1933–1947. https://doi.org/10.1007/s11069-015-2052-8

Sørensen, M. B., Spada, M., Babeyko, A., Wiemer, S., & Grünthal, G. (2012). Probabilistic tsunami hazard in the Mediterranean Sea. *Journal of Geophysical Research*, *117*(B1), B01305. https://doi.org/10.1029/2010JB008169

Souty, V., & Gailler, A. (2020). Tsunami hazard associated to earthquakes along the French coasts. A probabilistic approach (PTHA). In *EGU General Assembly 2020*. https://doi.org/10.5194/egusphere-egu2020-5554

Wood, S. (2017). *Generalized additive models: An introduction with R* (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781315370279

Wronna, M., Baptista, M.-A., & Kânoğlu, U. (2020). A new tsunami runup predictor. *Natural Hazards*, *105*(2), 1571–1585. https://doi.org/10.1007/s11069-020-04366-1