Natural Hazards
https://doi.org/10.1007/s11069-021-04806-6

ORIGINAL PAPER

Statistical forecast of the marine surge

Gonzalo Iñaki Quintana[1] · **Pierre Tandeo**[1] · **Lucas Drumetz**[1] · **Laurent Leballeur**[2] · **Marc Pavec**[2]

Received: 27 September 2020 / Accepted: 19 May 2021

Abstract
This paper studies different machine learning methods for solving the regression problem of estimating the marine surge value given meteorological data. The marine surge is defined as the difference between the sea level predicted with the tides equations, and the real measured sea level. Different approaches are explored, from linear regression to multilayer perceptrons and recurrent neural networks. Stochastic networks are also considered, as they enable us to calculate a prediction error. These models are compared with a baseline method, which uses physical equations to calculate the surge. We show that all the statistical models outperform the baseline, being the multilayer perceptron the one that performs the best. (It reaches an R^2 score of 0.68 and an RMSE of 7.3 cm.)

Keywords Machine learning · Data science · Regression · Neural networks · Marine surge · Sea level · Time-series analysis

1 Introduction

When big and violent storms approach, it is desirable to know the maximal and minimal values that the sea level can reach. This is important in order to measure the storm's impact and the flood risk in inhabited areas and, therefore, to take actions and warn the population. The sea level changes periodically due to the gravitational forces exerted by the moon, the sun and Earth's rotation, in a phenomenon we call tides. Fortunately,

✉ Gonzalo Iñaki Quintana
gonzalo-inaki.quintana@imt-atlantique.net

Pierre Tandeo
pierre.tandeo@imt-atlantique.fr

Lucas Drumetz
lucas.drumetz@imt-atlantique.fr

Laurent Leballeur
laurent.leballeur@actimar.fr

Marc Pavec
marc.pavec@actimar.fr

[1] IMT Atlantique, Lab-STICC, UMR CNRS 6285, F-29238 Brest, France

[2] Actimar S.A.S., 36 quai de la douane, 29200 Brest, France

Published online: 31 May 2021

the tides' dynamics are fairly well known, and precise models are available to describe them.

However, to the tides we also have to add all the atmospheric and oceanic processes that sometimes have a nonnegligible effect on the sea level, especially near the coasts, in the bays and estuaries. The instantaneous marine surge is then defined as the difference between the predicted sea level (using the tides equations) and the observed sea level. This difference has mainly a meteorological origin, like anticyclones or storms (strong winds and waves). Nevertheless, the coastline geometry can also influence the value of the surge, as the waves that generate the tidal signal can be amplified or interact with each other in certain bays, ports and river mouths, modifying the total sea level.

We are usually more interested in the high tide surge (defined as an abnormal overshoot of the high tide level) and in the drop in the low tide level (an abnormally low sea level), as they give the total maximal and minimal levels that the sea can reach.

We can denote three principal processes in the surge formation:

- the atmospheric pressure, i.e., when the air is lighter, the water surface rises mechanically, therefore increasing the sea level. A decrease of 1 hPa in pressure implies an increase in the sea level of 1 cm, and vice versa (inverse barometer effect);
- the wind, which generates frictions on the sea surface and movements of water masses. This can sometimes lead to an accumulation of water in the coast, which amplifies the sea level and generates a surge;
- the waves, which contribute by increasing or decreasing the sea level when they hit the coast;

Machine learning models are beginning to be applied to sea-level-related problems. Makarynskyy et al. use multilayer perceptrons (MLPs) to fill the missing values of sea-level time series using measures from neighboring tide gauges and to predict the sea level with its past measurements (Makarynskyy et al. 2004, 2006). A similar approach is presented by Ghorbani et al. (2010) using genetic programming in an island environment. Braakmann-Folgmann et al. (2017) incorporate spatial dependency by combining convolutional neural networks (CNNs) and long short-term memories (LSTMs) to predict a grid of sea-level anomalies (defined as the difference between the observed and the mean sea levels) using its past values. On the other hand, Hieronymus et al. (2019) predict the sea level with oceanic variables (wind and sea level pressure). They use quality-controlled data from 9 different stations in the Baltic sea.

Predicting the oceanic or marine surge is of global importance in a world where climate change will make storms and floods more common. We can see a worldwide growing interest in the oceanic surge: in Europe (Kergadallan et al. 2014; Fernandez-Montblanc et al. 2019), North America (Bernier and Thompson 2014; Soontiens et al. 2016) and Asia (Higaki et al. 2009; Sahoo and Bhaskaran 2019; Chao et al. 2020). The problem of predicting the surge is classically addressed by using numerical methods or ensemble of models, which are complex, computationally expensive and time-consuming (Sahoo and Bhaskaran 2019). Machine learning can be an alternative to these models, as it has shown its practical utility in sea-related problems. In the recent literature, we already find some examples of applications of machine learning to the surge prediction. For instance, Chao et al. (2020) use typhoon parameters (central pressure, maximum typhoon radius, maximum wind speed, typhoon location) and artificial neural networks (ANNs) to predict the storm surge in northeastern Taiwan. Sahoo and Bhaskaran (2019) combine ANNs with wind variables and prevalent tidal and land conditions to

predict the storm surge and the onshore flooding in eastern India. This work intends to continue in that direction.

In this paper, we aim to predict the instantaneous surge with machine learning models that use meteorological variables. As baseline, we use the SEAMER model, a hydrodynamical model used to simulate 2D water levels and currents. SEAMER, developed by Actimar S.A.S as a specific branch of Ifremer's MARS model (Lazure 2008), is a classical primitive-equations model. It solves the shallow water equations (or Saint-Venant equations, shown below), which are the Navier–Stokes equations integrated in the vertical axis, using finite differences.

$$\begin{cases} \frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} - fv = -\frac{1}{\rho_0}\frac{\partial p}{\partial x} - g\frac{\partial \zeta}{\partial x} + v_h\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 v}{\partial y^2}\right) + \frac{\tau_{Sx} - \tau_{Fx}}{\rho_0 h} \\ \frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + fu = -\frac{1}{\rho_0}\frac{\partial p_a}{\partial y} - g\frac{\partial \zeta}{\partial y} + v_h\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 v}{\partial y^2}\right) + \frac{\tau_{Sy} - \tau_{Fy}}{\rho_0 h} \\ \frac{\partial \zeta}{\partial t} + \frac{\partial (hu)}{\partial x} + \frac{\partial (hv)}{\partial y} = 0 \end{cases} \tag{1}$$

In Eq. (1), (x, y) is a pair of Cartesian coordinates, (u, v) the fluid's flow vector, h the instantaneous water height, ζ its deviation from the mean value, f the Coriolis coefficient, v_h the horizontal diffusivity, g the gravitational acceleration at the surface, p the pressure, ρ_0 the fluid's density and τ_S and τ_F the surface and bottom tensions, respectively.

The computational grid of the finite differences solver has a horizontal resolution of 200 m. The model is forced at its offshore boundaries by harmonic tidal components from FES2014 atlas (Lyard et al. 2021). The atmospheric forcing (wind and pressure) comes from the global forecast system (GFS) at a spatial resolution of 0.25° and a temporal resolution of 3 h.

The SEAMER model can predict the surge value three days in the future. However, as its precision deteriorates for times distant from the beginning of the simulation, the model is ran every 6 h, with the updates of the meteorological forcing. The final surge time-series prediction consists of the concatenation of the first 6-h predictions of each simulation.

Nevertheless, the model introduces some errors, due principally to the quantization of the space in a mesh, to the data used (bathymetry, friction on the bottom and meteorological forcing), sub-mesh processes and processes not included in the shallow water equations (tidal and storm waves propagation in bays, resonances, infra-gravity waves, etc). These errors worsen the forecast's performance and are sometimes too difficult and complex to analyze. However, accounting for the ignored processes can be done using a statistical model.

The idea of this work is to implement and compare different statistical and machine learning models to predict the instantaneous surge. These models should be able to forecast the real sea level (by correcting the SHOM[1] tide prediction). It is also desirable that these models provide an estimation of the prediction error, to assess a prediction confidence interval.

[1] Service Hydrographique et Océanographique de la Marine (Hydrographic and Oceanographic Service of the Navy, in French).

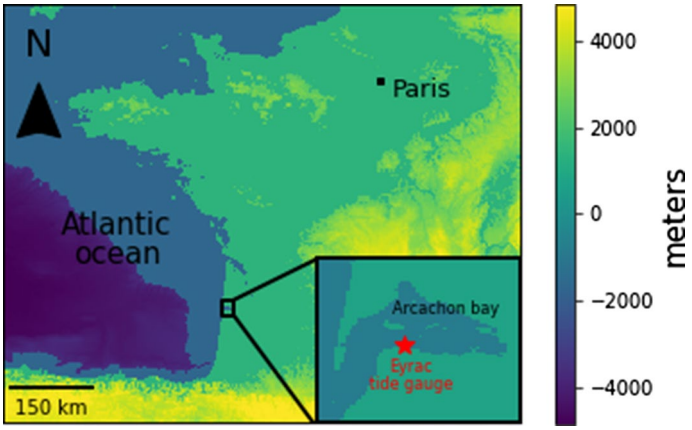


Fig. 1 Location of the Eyrac tide gauge in the Arcachon bay, France. Imagery reproduced from the GEBCO_2019 Grid, GEBCO Compilation Group (2019) GEBCO 2019 Grid (doi:10.5285/836f016a-33be-6ddc-e053-6c86abc0788e)

2 Data preprocessing

We use the data of the Eyrac tide gauge, a meteorological station located in the Arcachon bay in southwest France (Fig. 1). The dataset is composed of 20185 hourly samples of 11 different meteorological and climatological variables (the most relevant being atmospheric pressure, wind speed and significant wave height[2]) together with their associated surge value. This dataset corresponds to the period between July 2017 and October 2019.

Because input variables have different units, the range of values can change significantly between different variables. As this can negatively impact the learning process of the models, the inputs were normalized by subtracting their mean and dividing by their standard deviation.

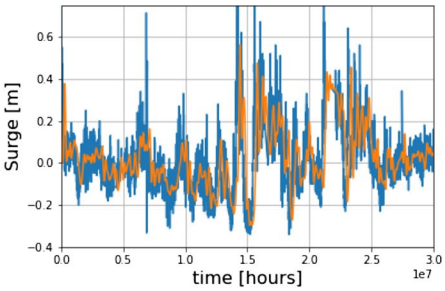
When analyzing the data, three main problems were found in the dataset:

- missing values in the dataset. In the original dataset, 1493 values were missing, that is 7.4% of the dataset;
- presence of outliers;
- too strong variations in subsequent surge values (considered as unrealistic).

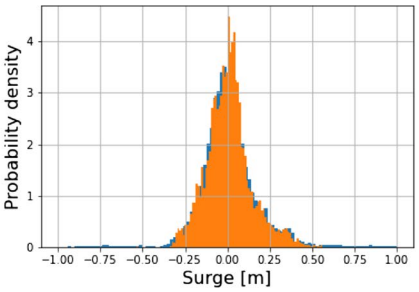
Missing values are mainly due to malfunctions in the sensor. The last two problems are caused by different types of noise in the sensor (electrical, mechanical, optical, etc) or real high-frequency variations in the water level (agitations), which can be considered as noise, since they are not the consequence of natural processes and do not persist in time. The outliers are data that are too isolated and distant from the center of their distribution. In our case, a data record is a vector composed of the surge value and all the other oceanographic and meteorological variables.

The outliers were removed by using the Z-score technique with a threshold of 3, eliminating the data whose distance from the mean of their joint distribution is bigger than three times

[2] The significant wave height is defined as the average height of the 1/3 highest waves, from trough to crest.



(a) Temporal plot of the non filtered (blue) and filtered (orange) data



(b) Surge histograms of the non filtered (blue) and filtered (orange) data

Fig. 2 Filtered and non filtered data (D3 and D2 datasets)

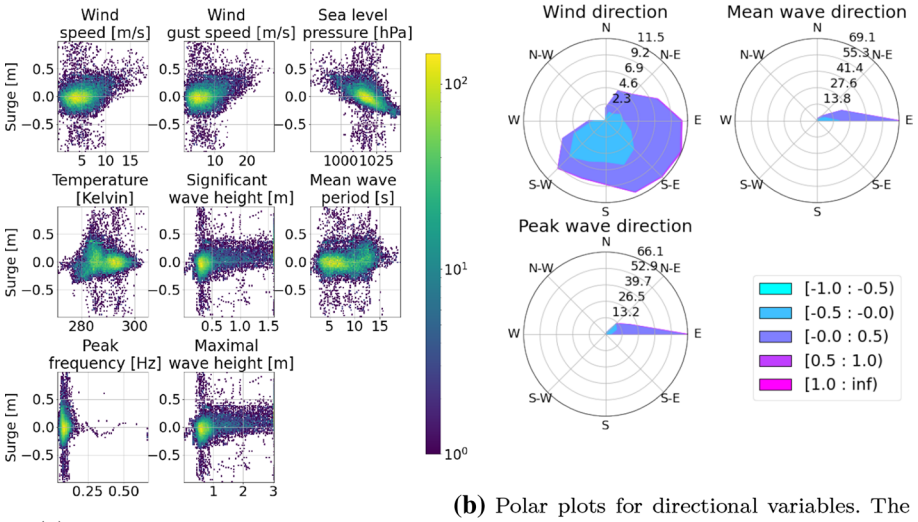
its standard deviation. The data with surges bigger than 1 meter were also considered as outliers and were removed. Most of them had surges of between 4 and 8 meters, which are unrealistic values for the processes in which we are interested. Besides, these values were mostly not preceded or followed by similar surge values, which gives more evidence to think that they were caused by malfunctions of the sensor or by some kind of interference. The missing values were simply ignored. The resulting D1 dataset has 17720 samples.

For recurrent models, a continuity in the time series is needed (i.e., samples for all the sampling instants). We generate the D2 dataset by linearly interpolating all the missing values (i.e., 1493 values, 7.4% of the dataset) as well as the ones removed in the previous quality control stage (i.e., 299 values, 1.4% of the dataset). This results in 1792 interpolated data records, which corresponds to the 8.8% of the dataset. Those values were uniformly distributed along all the time series, and the larger gaps were of 5 consecutive values (i.e., 5 h).

To address the last of the listed problems (too fast surge variations), the dataset D2 was filtered using a 10th-order low-pass Butterworth filter. Its critical frequency was experimentally set to 8.33 μHz, which is associated with a time period of 33 h. In practice, the surge signal should be continuous, without spikes, and high surges should appear progressively, as meteorological phenomena do not act instantly. The idea behind the low-pass filtering is to smooth the temporal series taking this into account: if two contiguous surge values are too far apart within a too short time interval, the filtering will lower their difference to get a more "probable" series of values. In Fig. 2a, we show the time series of the filtered and unfiltered sequences. We can see that the filter smooths the time series and eliminates its unnatural peaks. The resulting dataset is called D3. Figure 2b shows the histograms of the target variable (i.e., the surge) for the D2 and D3 datasets. We can see that the filtering eliminates the heavy tails of the surge distribution, but also makes it less symmetric. We remark that big negative surges are rarer events than big positive ones.

3 Data analysis

The objective of this section is to analyze the relationships between the different meteorological and oceanographic variables and the surge. We have two types of variables: directional (which express directions in the form of angles) and nondirectional (temperature, wind speed, pressure, etc.).



(a) Scalar or non-directional variables

(b) Polar plots for directional variables. The radial axes indicate the occurrence (percentage) and the colors the surge

Fig. 3 2D histograms of the surge and the oceanic and meteorological variables (D1 dataset)

In Fig. 3a, we plot the 2D histograms between each of the nondirectional input variables and the target variable (the surge). These plots are useful to visualize the correlation of each variable with the surge. The areas in yellow and in green are the most dense ones (where there are more samples), while the areas in violet or blue are the least dense. The variables considered here are: the wind speed, the wind gust speed (a brief increase in the speed of the wind, of less than 20 seconds), the atmospheric pressure at the mean sea level, the temperature, the significant wave height, the mean wave period, the peak frequency and the maximum wave height (measured from trough to crest). The peak frequency is defined as the frequency (the reciprocal of the wave period) at which the wave spectrum is at its maximum, i.e., the frequency with the highest energy. For a more detailed description of these parameters, see the Guide to Wave Analysis and Forecasting (World Meteorological Organization 2018).

Firstly, we can see that even though the data were preprocessed, there still are some outliers (points that are far from the centers of the distributions). This analysis highlights that the relationships between the variables in Fig. 3a are mostly linear (even if there are some weak nonlinearities). We can also retrieve some known facts, like the inverse barometer effect (the surge increases approximately linearly when the sea level pressure decreases, by about 1cm for 1hPa). We find a positive correlation of the surge with the wind speed and the wind gust speed, as well as with the significant and maximal wave heights. On the other hand, as we could expect, we can see that the temperature, the mean wave period and the peak frequency do not show a strong and clear correlation with the surge. However, we decided to keep these variables as they can be more significant when combined with other variables.

In Fig. 3b, we plot the normalized histogram of the directional variables (angles) as a function of the surge value, in a polar plot. The radial axes indicate the percentage of occurrence and the color bar the different surge ranges. The directional convention indicates the direction to where winds or waves blow/flow. (We remark that this is not

the usual convention for winds.) There are three directional variables: wind direction, mean wave direction (the mean of all the individual wave directions according to their own energy) and peak wave direction. The peak wave direction is the direction corresponding to the maximum energy of the wave spectrum. This is the most general case, in which the sea contains multiple wave systems (for example, swells and wind waves). However, the sea state often contains a dominant wave system, in which case the mean direction is just the direction of that system, which also corresponds to the peak direction. Firstly, we can see that there are slightly larger surge values when the wind comes from the west, which is reasonable as it accumulates more water in the bay. Oppositely, big negative surges are related to wind blowing out of the bay, in the southwest direction. We can also see that waves come mostly from the west, which also makes sense as the incoming waves come mainly from the Atlantic Ocean, west of the entrance to the Arcachon bay.

4 Data modeling

We are interested in estimating the value of the marine surge for a given time, given all the meteorological and oceanographic variables. That is, we want to estimate a function f that verifies

$$y(t_n) = f(X(t_0), ..., X(t_n)), \tag{2}$$

with $y(t_n)$ being the surge at instant t_n and $X(t_0), ..., X(t_n)$ the set of all the meteorological variables from time t_0 to t_n. For solving this regression problem, several models were considered. All the following models were fitted to data in the training set of the three previously defined datasets (except for the LSTM, which was not applied on the D1). The training set was generated by taking 66% of the samples. The rest of the samples form the test set, which is only used for model evaluation.

4.1 Linear regression

The simplest possible model is to consider that the surge at a given time instant $y(t_n)$ depends only on the input variables at that same time $X(t_n)$, in a linear way. The model is then written as:

$$y(t_n) = \beta X(t_n) + \beta_0, \tag{3}$$

with β the regression coefficients and β_0 the intercept term.

This approach has a particular interest, as we can estimate the uncertainty of the prediction. Once having fitted the linear regressor to the data, by supposing that the error distribution is Gaussian, we calculate the standard deviation between the training data and the predictions. The confidence interval is then centered at each prediction and has a width of $2\Phi^{-1}(1-\alpha/2)\sigma$, where σ is the standard deviation, $1-\alpha$ the confidence level and $\Phi^{-1}(z)$ the inverse of the cumulative normal distribution function. For instance, for a confidence level of $1-\alpha = 0.95$, we have $\Phi^{-1}(1-\alpha/2) = 1.96$.

4.2 Multilayer perceptron

The second possible model in order of complexity is to consider that the function $f(x)$ in Eq. (2) is not necessarily linear. In order to implement this, we use a multilayer perceptron (MLP) of 4 hidden layers and 5885 trainable parameters. We add a batch normalization layer after each fully connected (FC) layer in order to guarantee that the input of each layer is normalized (Bjorck et al. 2013). To prevent overfitting, we include Dropout layers with a drop probability of 0.2 (Srivastava et al. 2014). For training, we use a batch size of 30 and the MSE (mean squared error) as loss function. The loss function is optimized using the Adam optimizer (Kingma and Ba 2014). The same loss function, optimizer and batch size are used for all the following models.

4.3 Recurrent model (LSTM)

We now consider that the surge at each time instant does not only depend on the input variables at that same instant, but also on the k previous ones. That is,

$$y(t_n) = f(X(t_{n-k}), ..., X(t_n)). \tag{4}$$

We implement this regressor using a single-unit LSTM (Hochreiter and Schmidhuber 1997), a type of recurrent neural network (networks incorporating a self-loop that allows them to retain information from the past) that does not suffer from vanishing gradients. LSTMs manage to keep contextual information from past inputs, which means that the output at time t_n also depends on previous inputs. They achieve this by means of a memory vector that is passed from the cell at time t_{n-1} to the one at time t_n. These networks also have a forget gate, which allows them to learn when to forget previous information. As for this model we need to have a complete time series (we cannot have missing or eliminated values between two samples), we can only apply it on the D2 and D3 datasets. We train this network using k=12, i.e., taking into account 12-h delay. On top of the LSTM unit, we connect two FC layers, with a batch normalization at each of their inputs (Bjorck et al. 2013).

4.4 MC Dropout (stochastic FCNN)

Stochastic neural networks are built by introducing random variations into the network. As a consequence, they produce a different output each time the same inputs are forwarded through the network. In this paper, those random variations are introduced by defining a Dropout layer that is also active when the network is used for prediction (and not just during training). That is, at each prediction step, neurons are randomly dropped with a given probability p, which implies that from a same input, different outputs are produced.

We can then generate N predictions for each input vector and calculate the mean and the standard deviation of the predictions (Monte Carlo simulation). The interest of doing this is that we can define an estimation error (set as three times the standard deviation) over the final estimation, which is the mean of the N predictions. This approach is similar to the one used by Gal and Ghahramani (2015), with the difference that here we do not have a regularization term.

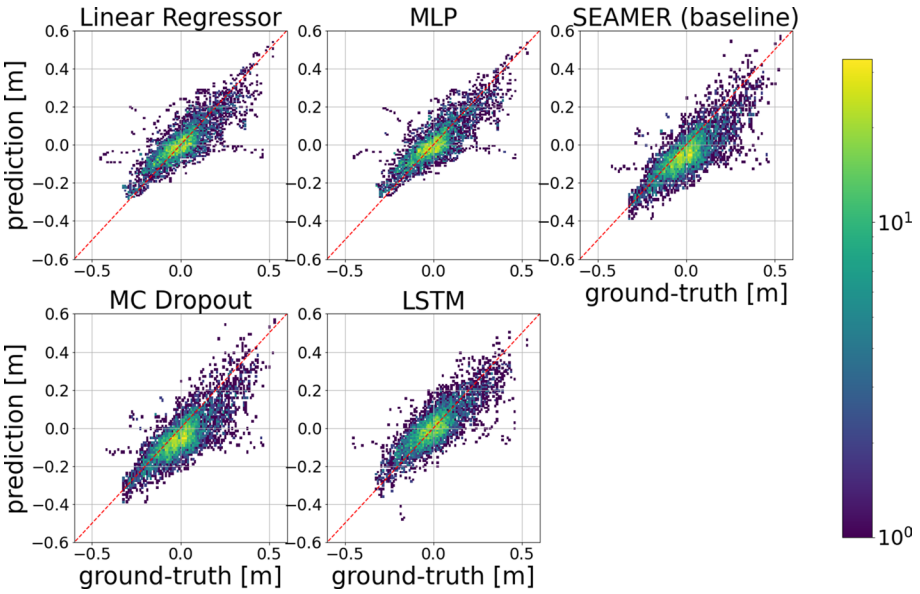


Fig. 4 2D histogram of the predictions and ground truths for the three statistical models and the baseline (D3 dataset)

In order to implement this Monte Carlo Dropout (MC Dropout) model, we add a permanent Dropout layer after each layer of the FCNN defined above. The drop probability is $p = 0.1$.

5 Model evaluations

We apply the models described above and compare their predictions with the available ground truths, coming from the validation set (33% of the dataset).

5.1 2D histograms

We plot the 2D histograms of the model predictions and the ground truth for the test set in the D3 dataset (Fig. 4). If the models' predictions are good, the 2D histogram should resemble a straight line of equation $y = x$ (plotted in red), as this means that the predictions and ground truths are equal. We compare them with the SEAMER model. We observe that all the proposed models show less dispersion than the baseline, which also underestimates the surge (as the distribution in the SEAMER model is denser below the $y = x$ line). We also remark that most errors occur for relatively high values of surge, which the models tend to confuse with relatively small ones.

5.2 Temporal analysis

We plot the different models' predictions together with the corresponding ground truth for the D3 dataset (Fig. 5), for the last 6056 samples. For the linear model, we also

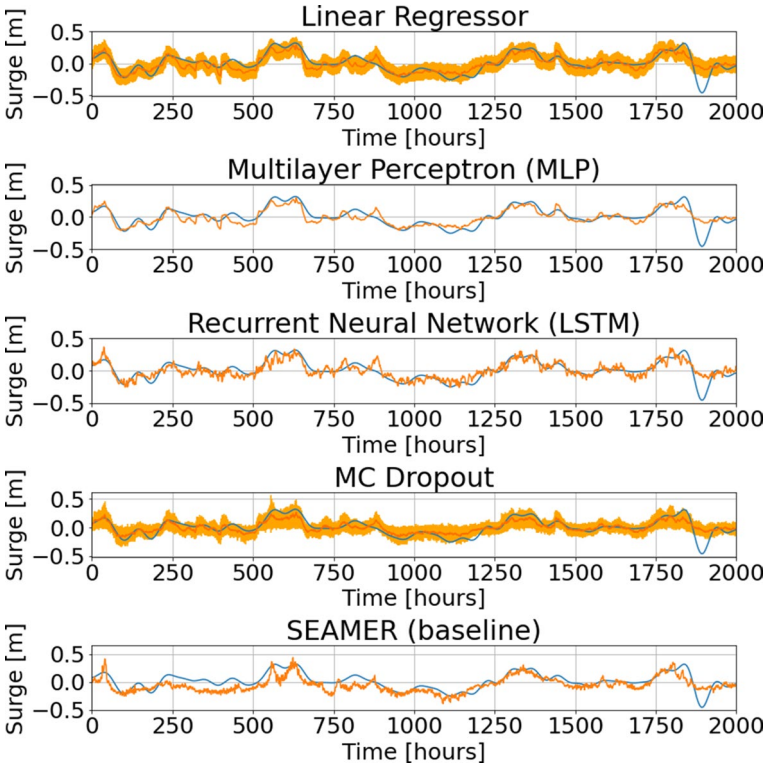


Fig. 5 Temporal plots of the ground truth (blue) and the prediction (orange) for the three statistical models and the baseline (D3 dataset)

plot the 95% confidence interval, and for the MC Dropout we plot the surge prediction (mean) and its associated error bar (three times the standard deviation). The number of Monte Carlo iterations used was 100. (We observed that after 50 iterations the estimations of the mean and standard deviation converge.) We highlight that the SEAMER model underestimates the surge and that none of the models can predict sudden surge changes, like the last event in Fig. 5. This event is comprised neither in the linear model's confidence interval nor in the MC Dropout's error bars.

5.3 Quantitative evaluation

For model evaluation, several metrics are considered: the root mean squared error (RMSE), the bias, the coefficient of determination or R^2 score and the correlation coefficient between the prediction and the ground-truth data (to quantify the information in the 2D histograms of the previous subsection). The correlation coefficient measures how our predictions and the ground truth are correlated and should be 1 for an ideal regressor. The R^2 score is defined as:

Table 1 Evaluation metrics for all studied models in the D1 (up), D2 (middle) and D3 (down) datasets

	RMSE (cm)	Bias (cm)	R^2	Corr.
Linear	**10.6**	– 0.16	0.42	0.65
MLP	10.9	**– 0.15**	**0.46**	**0.68**
MC Dropout	11.6	– 0.23	0.42	0.66
SEAMER (baseline)	13.0	– 5.18	0.28	0.62
Linear	14.3	**0.17**	0.39	0.62
MLP	14.0	– 0.21	0.39	0.62
MC Dropout	15.2	– 0.23	0.34	0.60
LSTM	**13.9**	– 0.34	**0.40**	**0.63**
SEAMER (baseline)	15.4	– 5.31	0.28	0.61
Linear	7.5	0.10	**0.68**	0.82
MLP	**7.3**	**0.08**	**0.68**	0.82
MC Dropout	7.8	0.11	0.65	**0.83**
LSTM	7.5	– 0.32	0.67	0.82
SEAMER (baseline)	9.5	– 5.09	0.44	0.77

$$R^2 = 1 - \frac{\sum(y_{\text{true}} - y_{\text{pred}})^2}{\sum(y_{\text{true}} - \bar{y}_{\text{true}})^2} \tag{5}$$

where y_{pred} is the prediction, y_{true} the ground truth and $\bar{y}_{\text{true}}$ the mean of all the ground truths. This coefficient measures the proportion of the variance of y_{true} that is predicted by our model, and it is smaller than 1.

We compute these metrics for all the evaluated models and show them in Table 1, where the best scores (i.e., minimal values for the RMSE and bias, and maximal values for the R2 score and correlation coefficient) are represented in bold. We first highlight that all the machine learning models outperform the baseline (i.e., the SEAMER model) in all the evaluated metrics.

In the D1 dataset, the RMSE is lower than 12cm for all the proposed statistical models (e.g., for the linear regressor and the MLP) and the bias is lowered in a 96%, with respect to the corresponding metrics of the baseline. The coefficient of determination increases more than 60%, and the correlation coefficient is slightly increased. The best-performing model of this dataset is the multilayer perceptron (MLP).

The performance of all the models worsens in the D2 dataset, as missing values are interpolated instead of being simply ignored. This necessarily adds some error, but enables to incorporate the recurrent model (LSTM). Actually, in this dataset, the best performances are achieved by the LSTM, which has the biggest correlation and determination coefficients and the lowest RMSE. The decrease of the 96% in the bias of dataset D1 is maintained, but the increase in the correlation and determination coefficients is less than in the previous dataset.

On the other hand, all the models perform better in the D3 dataset as data were filtered, therefore eliminating strong variations in the surge which are more difficult to predict. The RMSE is half of that of the D2 dataset for almost all the models; the R^2 and correlation coefficients are increased in 50% and 30%, respectively. MLP is the best-performing model.

In a general way, the baseline was outperformed by all the statistical models. The MLP and the LSTM attain the best performance, but they fail to capture large nonlinearities and extreme values, as shown in Fig. 5.

6 Conclusions

Multiple machine learning models were tested and evaluated for solving the problem of forecasting the marine surge using meteorological variables. All these methods were compared with a baseline (the SEAMER model), a finite differences model that solves the shallow water equations, and most of them were proven to have equivalent or better performances. Simple models, such as linear regression, achieved good results and were not significantly outperformed by more complex models, such as multilayer perceptrons or recurrent neural networks. This makes sense, as the relationships between the input variables and the target variable are, as shown in the Data Analysis section, mostly linear.

The linear regressor also provides a prediction confidence interval, which can be calculated by supposing that the error distribution is Gaussian and by estimating its standard deviation with the training data.

Finally, we highlight that the database was quite noisy and had lots of aberrant values. More experiments should be done with a less noisy database, which could help to increase performances (as the results in the D3 dataset show). Other preprocessing approaches should also be explored, like a gradient criterion for detecting nonphysical variations (too strong surge variations in a small period of time) and more sophisticated techniques for filling the gaps in the surge series (e.g., neural network-based techniques, see Makarynskyy et al. 2006). With respect to the filtering, sensibility tests should be conducted in order to determine the most appropriate cutoff frequency. It would also be interesting to include spatial information, by using data from different tide gauges, or to explore boosting techniques (algorithms that combine different estimators), such as AdaBoost (Freund and Schapire 1996).

Acknowledgements The authors wish to kindly thank the SIBA[3] for letting us use the surge forecasts and tide gauge measurements, and Actimar S.A.S. for providing the datasets.

Author Contributions PT, LD, LL and MP conceived the general idea and problematic of the project and GIQ carried out all the implementations and experiments. PT and LD acted as technical advisors in machine learning and statistics, while LL and MP were the technical advisors in Oceanography and Meteorology. All authors discussed the results and contributed to the final manuscript.

Funding This research was financially supported by Actimar S.A.S. and IMT Atlantique.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Availability of data and material The data used are no publicly available.

Code availability The code is no publicly available.

References

Bernier N, Thompson KR (2014) Deterministic and ensemble storm surge prediction for Atlantic Canada with lead times of hours to ten days. Elsevier, London

[3] Syndicat Intercommunal du Bassin d'Arcachon (Intercommunal Syndicate of the Aracachon Basin).

Bjorck J, Gomes C, Selman B, Weinberger KQ (2013) Understanding batch normalization. ArXiv: 1806:02375

Braakmann-Folgmann A, Roscher R, Wenzel S, Uebbing B, Kusche J (2017) Sea level anomaly prediction using recurrent neural networks. ArXiv. abs/1710.07099

Chao WT, Young CC, Hsu TW, Liu WC, Liu CY (2020) Long-lead-time prediction of storm surge using artificial neural networks and effective typho on parameters: Revisit and Deeper Insight. MPI

Fernández-Montblanc T, Breyiannis G, Vousdoukas MI, Feyen L, Salamon P, Ciavola P, Voukouvalas E, Mentaschi L (2019) Towards robust pan-European storm surge forecasting. Elsevier, New York

Freund Y, Schapire RE (1996) Experiments with a new boosting algorithm. In: CML

Gal Y, Ghahramani Z (2015) Dropout as a Bayesian approximation: representing model uncertainty in deep learning. ArXiv. 1506:02142

Ghorbani MA, Makarynskyy O, Shiri J, Makarynskyy D (2010) genetic programming for sea level predictions in an island environment. Int J Ocean Clim Syst. https://doi.org/10.1260/1759-3131.1.1.27

Hieronymus M, Hieronymus J, Hieronymus F (2019) On the application of machine learning techniques to regression problems in sea level studies. J Atmos Ocean Technol. https://doi.org/10.1175/JTECH-D-19-0033.1

Higaki M, Hayashibara H, Nozaki F (2009) Outline of the storm surge prediction model at the Japan Meteorological Agency Office of Marine Prediction, Japan Meteorological Agency

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9:1735–80. https://doi.org/10.1162/neco.1997.9.8.1735

Kergadallan X, Bernardara P, Benoit M, Daubord C (2014) Improving the estimation of extreme sea levels by a characterization of the dependence of skew surges on high tidal levels. In: Coastal engineering proceedings

Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. Arxiv: abs/1412.6980

Lazure D (2008) An external-internal mode coupling for a 3D hydrodynamical model for applications at regional scale (MARS). Adv Water Resour 31(2):233–250

Lyard FH, Allain DJ, Cancet M, Carrère L, Picot N (2021) FES 2014 global ocean tide atlas: design and performance. Ocean Sci 17:615–649. https://doi.org/10.5194/os-17-615-2021

Makarynskyy O, Makarynska D, Kuhn M, Featherstone W (2004) Predicting sea level variations with artificial neural networks at Hillarys Boat Harbour, Western Australia. Estuar Coast Shelf Sci 61:351–360. https://doi.org/10.1016/j.ecss.2004.06.004

Makarynskyy O, Makarynska D, RusU E, Gavrilov A (2006) Filling gaps in wave records with artificial neural networks. Maritime Transportation and Exploitation of Ocean and Coastal Resources—Guedes Soares, Garbatov & Fonseca (eds). https://doi.org/10.1201/9781439833728.ch131

Sahoo B, Bhaskaran PK (2019) Prediction of storm surge and coastal inundation using artificial neural network—a case study for 1999 Odisha Super Cyclone. Elsevier, New York

Soontiens N, Allen SE, Latornell D, Le Souëf K, Machuca I, Paquin JP, Lu Y, Thompson K, Korabel V (2016) Storm surges in the strait of georgia simulated with a regional model. Atmos Ocean 54(1):1–21. https://doi.org/10.1080/07055900.2015.1108899

Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 15:1929–1958

World Meteorological Organization (2018) Guide to wave analysis and forecasting