

LETTER • OPEN ACCESS

Estimation of global coastal sea level extremes using neural networks

To cite this article: Nicolas Bruneau *et al* 2020 *Environ. Res. Lett.* **15** 074030

View the [article online](#) for updates and enhancements.

Environmental Research Letters



LETTER

Estimation of global coastal sea level extremes using neural networks

OPEN ACCESS

RECEIVED
22 January 2020

REVISED
19 March 2020

ACCEPTED FOR PUBLICATION
16 April 2020

PUBLISHED
2 July 2020

Original Content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.



Nicolas Bruneau^{1,2} , Jeff Polton¹ , Joanne Williams¹  and Jason Holt¹ 

¹ National Oceanography Centre, Joseph Proudman Building, Liverpool L3 5DA, United Kingdom

² Reask UK Ltd, 49 Greek Street, London, W1D 4EG, United Kingdom

E-mail: bruneau.n@gmail.com

Keywords: sea water anomaly, extremes, storm surges, GESLA database, machine learning

Supplementary material for this article is available [online](#)

Abstract

Accurately predicting total sea-level including tides and storm surges is key to protecting and managing our coastal environment. However, dynamically forecasting sea level extremes is computationally expensive. Here a novel alternative based on ensembles of artificial neural networks independently trained at over 600 tide gauges around the world, is used to predict the total sea-level based on tidal harmonics and atmospheric conditions at each site. The results show globally-consistent high skill of the neural networks (NNs) to capture the sea variability at gauges around the globe. While the main atmosphere-driven dynamics can be captured with multivariate linear regressions, atmospheric-driven intensification, tide-surge and tide-tide non-linearities in complex coastal environments are only predicted with the NNs. In addition, the non-linear NN approach provides a simple and consistent framework to assess the uncertainty through a probabilistic forecast. These new and cheap methods are relatively easy to setup and could be a valuable tool combined with more expensive dynamical model in order to improve local resilience.

1. Introduction

Predicting accurately the sea water level variability from short to large time scales is of great importance for coastal communities. The range of impacts and challenges is broad, ranging from harbour management (where minimum water level is required to allow ships to enter the harbour) to life-threatening natural disasters or long-term sea level rise leading to loss of land availability and fertility for agriculture. Coastal flooding due to storm surges is considered as one of the biggest sources of casualties during tropical cyclones; storm surges have large social, economic and environmental impacts [1–4]. Therefore, timely and accurate prediction of sea-level variability and extremes is crucial for global coastal resilience.

Deterministic numerical models have proven to be powerful tools for predicting sea variability. In particular they are effective for simulating storm surge propagation and impacts, and facilitate understanding of the complex physical processes associated with the storms [5–11]. However, they are relatively expensive and complex to set up and run

operationally, with associated additional computation costs if ensemble forecasts are required for analysis of risk or variability.

More generally, machine learning approaches and particularly deep learning have shown huge potential in pattern recognition for a wide range of applications. Recently, these techniques have emerged in climate, meteorological and oceanographic fields with convincing results. For example, convolutional neural networks have been trained to predict variations in the El Niño/Southern Oscillation (ENSO) with skill superior to state-of-the-art dynamical forecast systems [12]. Machine learning algorithms have also been used to aggregate ‘best-estimate’ forecasts from an ensemble for the predictions of ocean waves [13]. Neural networks have also successfully been used to bias-correct measurements leading to more homogeneous climate data records [14].

In sea level and tide processes, regressions have been used to infer meteorological impacts on sea water level and storm surges [15, 16]. Regression models have also been successfully driven by offshore gauge data in New York [17] and statistical

models have been applied to estimate extreme storm surges and associated return periods [18], or as bias-correction to water level predictions along US East Coast [19, 20]. The latter have shown similar performance compared to deterministic hydrodynamic models in capturing extremes in some cases. More recently, storm surges hindcasts in estuarine ports of the UK have been possible using artificial neural networks leading to accurate forecasting coastal flooding [21]. Neural networks have also been used for tide predictions at Mangalore, India [22], and along the Swedish coast to analyse long sea level records [23] where higher performance was obtained when local sea level forcing was prescribed (compared to linear models).

The aim of this study is to describe a general non-tuned machine learning framework, based on neural networks, and apply this around the globe with demonstrable skill in predicting non-tidal sea level residuals and extremes. The manuscript is structured as follows. Section 2 presents the GESLA tide-gauge data and associated pre-processing, the neural network ensemble, and the split between training and test sets as well as the scoring probabilistic measure. The first part of section 3 shows the key results of the study based on performance statistics for over 600 gauges around the world while the second part focuses on two particular regions with contrasting behaviours. Finally, section 4 discusses the results, the benefits and limitations of the approach, and the future steps.

2. Methods

2.1. Global extreme sea level analysis dataset—GESLA

The Global Extreme Sea Level Analysis database (GESLA version 2 [24]) provides unified high-frequency (15 min to 1 hour temporal resolution) quasi-global coastal sea level water information. Only public data (around 1070 gauges) are used in the present study. While data have been standardised, a simple but strict methodology was applied to pre-process each gauge in a systematic and reproducible manner. The key aspect of this stage was the elimination or reduction of potential issues arising from spurious data (e.g. temporal or reference height shifts) as well as removing long-term trends. An example of the pre-processing stage is illustrated in Supplementary figure 1 (stacks.iop.org/ERL/15/074030/mmedia). The following steps were sequentially applied to each gauge:

- Data from 1980 to 2015 are included and overlaps with the atmospheric reanalysis (see 2.2),
- For simplicity, only gauges with scheduled hourly, data are processed (as this only excluded 21 gauges). However, if the time steps are not constant and the percentage of time steps equal to the

statistical mode is less than 95%, the gauge is rejected;

- Sections are defined as time-series records where breaks are less than 7 days long;
- A yearly moving average is computed on each section;
- Given a record that is N years long, tidal harmonic analysis is computed yearly (going backwards) from the most recent one-year period, using the open-source *pytides* python module [25]. The analysis outputs 37 complex constituents per year (Sa, Ssa, Mm, MSF, Mf, 2Q1, Q1, rho1, O1, M1, P1, S1, K1, J1, OO1, 2N2, mu2, N2, nu2, M2, lambda2, L2, T2, S2, R2, K2, 2SM2, 2MK3, M3, MK3, M4, MS4, S4, MN4, M6, N6, M8);
- Spurious years in harmonic analysis are identified by separately inspecting just the M2 and K1 constituents. A gauge year can be rejected by either constituent if it exceeds a threshold separation from the mean (see e.g. Fig S1b). The procedure is as follows: an average complex value is calculated from N yearly values (red square). The average separation from this mean is calculated ($\bar{\epsilon}$) over N years. The complex difference (ϵ_{year_i}) for each year from the mean is independently assessed and the gauge year is rejected if

$$\epsilon_{year_i} > \max(3 \text{ cm}, \min(5 \text{ cm}, 10\% \bar{\epsilon}))$$

(e.g. green diamonds). This procedure ensures the rejection is based on the relative size of the separation from the N -year mean whilst preventing rejection for very small amplitudes.

- The total water signal is re-interpolated over a constant 1 h time vector based on the original temporal resolution excluding rejected periods of data;
- Finally, only gauges with over 3 (not necessarily contiguous) years of data are kept, with at least 2 years for training and one year for testing the model.

At the end of this process, 621 gauges remain and are used in this study. They provide an extensive coverage of the coastlines worldwide. The non-tidal residual is computed as the difference between the observations and the harmonic tide prediction (computed from all remaining sections). The aim was to implement a reasonably simple, robust and consistent pre-processing methodology to objectively deal with the large amount of data available. However, one could define different thresholds or apply different type of pre-processing to the gauge; exploratory analysis suggests that this would not impact the key results of this study.

2.2. High resolution atmospheric and ocean wave reanalysis - ERA5

To assess the impact of atmospheric and ocean wave processes on the non-tidal residual, an ensemble

of hourly physical predictors are extracted from the high-resolution atmospheric reanalysis ERA5 of ECMWF[26]. These are pre-processed over three length scales:

- Local** –10 m wind components and mean sea level pressure at 0.25° resolution as well as significant wave heights (including wind waves and swell) and peak periods (at 0.5° resolution) at the closest grid point from the gauge;
- Neighbourhood** - spatially accumulated precipitation in a 3.5° box centred on the gauge;
- Regional** - maximum and minimum wind speed components, maximum wave heights and minimum mean sea level pressure in a 5° box centred on the gauge.

In addition, proceeding 3 h gradients of all the atmospheric predictors are computed to capture late intensification / de-intensification (for example a low pressure system developing rapidly) as well as for the harmonic tidal level.

2.3. Machine learning

Each gauge is modelled independently using artificial Neural Networks (NNs). Each NN is composed of 3 hidden layers of 48 neurones. The input layer has 33 nodes (one for each environmental predictor described in previous sections combined with 7 hourly time steps of harmonic tide), and the outer layer has a single node providing the non-tidal residual target. While a sigmoid activation function is used for the last layer, the hidden layers consists of Leaky ReLU activation functions [27] combined with batch normalization layer to normalise the activations [28]. The NN had just under 7000 trainable parameters and its schematic view is provided in Supplementary figure 2. Finally, an Adam solver [29] is used to minimise the root mean square error between non-tidal residual predictions and observations; the NN is fitted for 150 iterations or less if the errors is not reduced within 10 consecutive iterations. Due to the large number of gauges available, this configuration has been lightly tuned on three random gauges (namely, a few combinations of the number of neurones, number of hidden layers and type of activation) and then applied to the full set without further adjustment.

For each gauge, the test set consists of the most recent year of recorded data (8784 time steps) while the rest is part of the training set. Therefore depending on the gauge, the training set extends from 2 years to 32 years permitting an analysis of the impact of the training size on the performance. Figure 1(a) shows the number of gauges as a function of the length of the training data.

An ensemble of 20 Neural Networks (NNs) is trained at each gauge location to generate a probabilistic forecast. Each NN is fitted using 50%

of the training set, randomly sampled. While a larger ensemble would have improved our probabilistic forecast, 20 members were chosen as a pragmatic balance between computational cost (over 12 000 NNs have been fitted in this study) and variability in the predictions.

All data (features and targets) have been standardised and normalised. The Neural Networks (NNs) are built with the *Keras* Python module [30] interfacing with *Tensorflow*[31] while processing was mainly done with the *Scikit-Learn* packages [32]. The neural networks have the traditional structure, where each node is connected to every node of the next layer. The temporal evolution of sea water level and non-tidal residual is continuous. Recurrent layers (such as Long Short-Term Memory, LSTM layers [33]) can be used to capture the dynamics of temporal processes. An LSTM neural network structure was implemented and tested for a few gauges but it did not lead to significant improvements of the predictions, and therefore a more simple and traditional structure was kept in this study. Finally, as a baseline, an ensemble of multi-variate linear regressions are fitted and used for predicting sea water level in the same manner as the neural networks for comparison; again for the linear regression no time series model was used.

Note that the neural network described above did not converge for 11 randomly-located gauges. Given the global coverage and the large number of gauges, these 11 gauges have been removed and no further investigation were carried out on these particular gauges.

2.4. Continuous ranked probability score

To assess the skills of the probabilistic predictions, a Continuous Ranked Probability Score (CRPS) is computed, with units cm. In weather forecasting, this is a common qualitative measure of performance for probabilistic forecasts comparing a distribution with observations [34–36].

The CRPS is defined as a quadratic measure of the difference between predicted, $H^p(\eta_r; t)$, and observed, $H^o(\eta_r; t)$, cumulative density functions (CDF). The quadratic measure is integrated over all possible residuals, z , and then averaged over time t to give a CRPS for each gauge

$$CRPS = \frac{1}{N} \sum_{t=1}^N \left(\int_{\eta_r > -\infty}^{\eta_r < \infty} [H^p(\eta_r; t) - H^o(\eta_r; t)]^2 dz \right),$$

where for each gauge $H^p(z; t)$ denotes the probability of an anomaly less or equal to η_r being *predicted* at time t ; and $H^o(\eta_r; t)$ is step-function, denoting the probability of an anomaly less or equal to η_r being *observed* at time t .

Intuitively, an ensemble producing a wide range of outcomes or an ensemble with a mean significantly

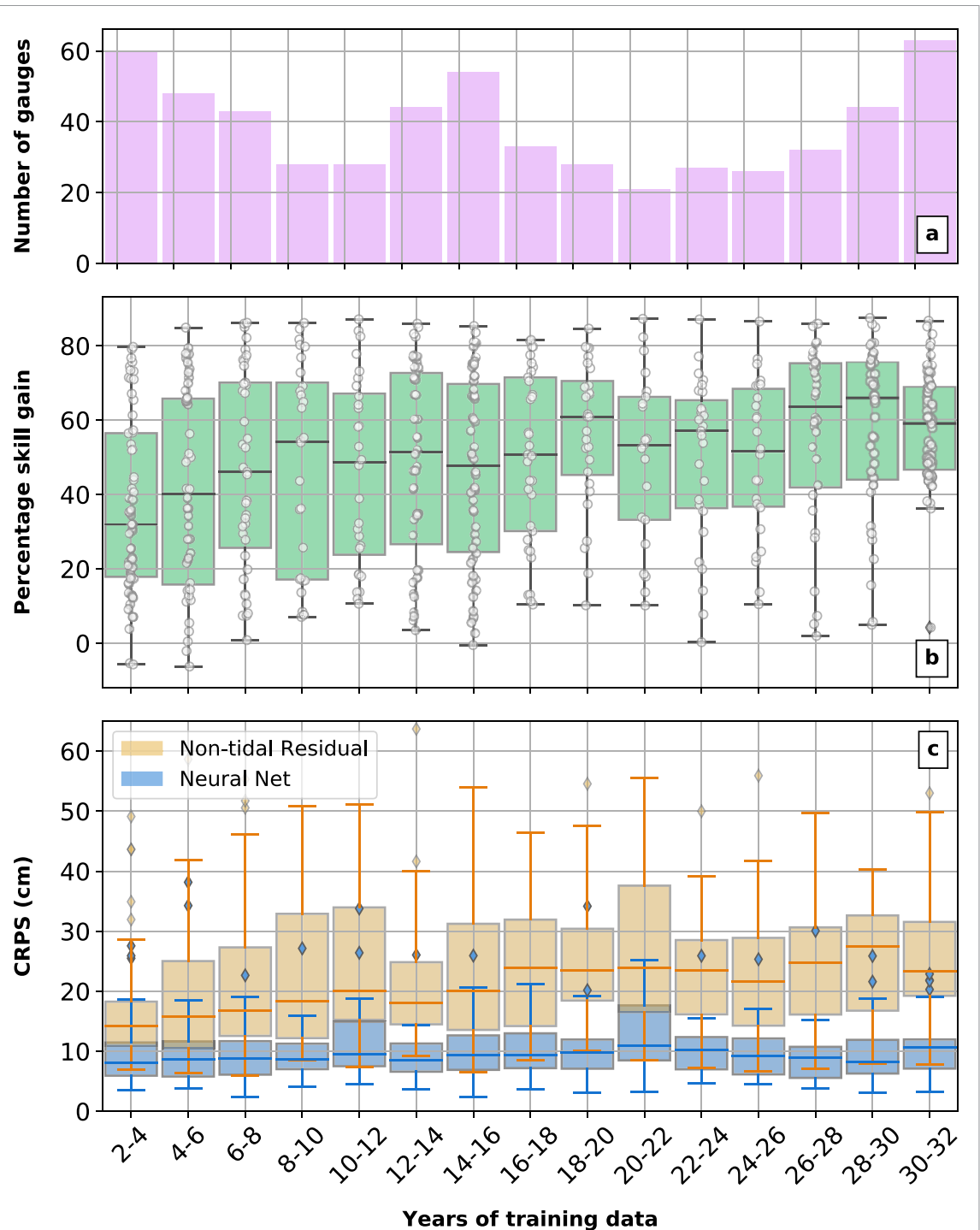


Figure 1. Statistics per number of training years for the 95th percentile. (a) Distribution of the number of GESLA gauges, (b) percentage skill gain from the non-tidal residual to the neural network predictions and (c) the continuous ranked probability score (equivalent to a mean absolute error) for both non-tidal residuals and neural network predictions. The box plots shows the mean, the quartiles and the extend of the distributions excluding outliers (marked as diamonds). Systematic gain of knowledge with the neural network and significant for the full range of training periods.

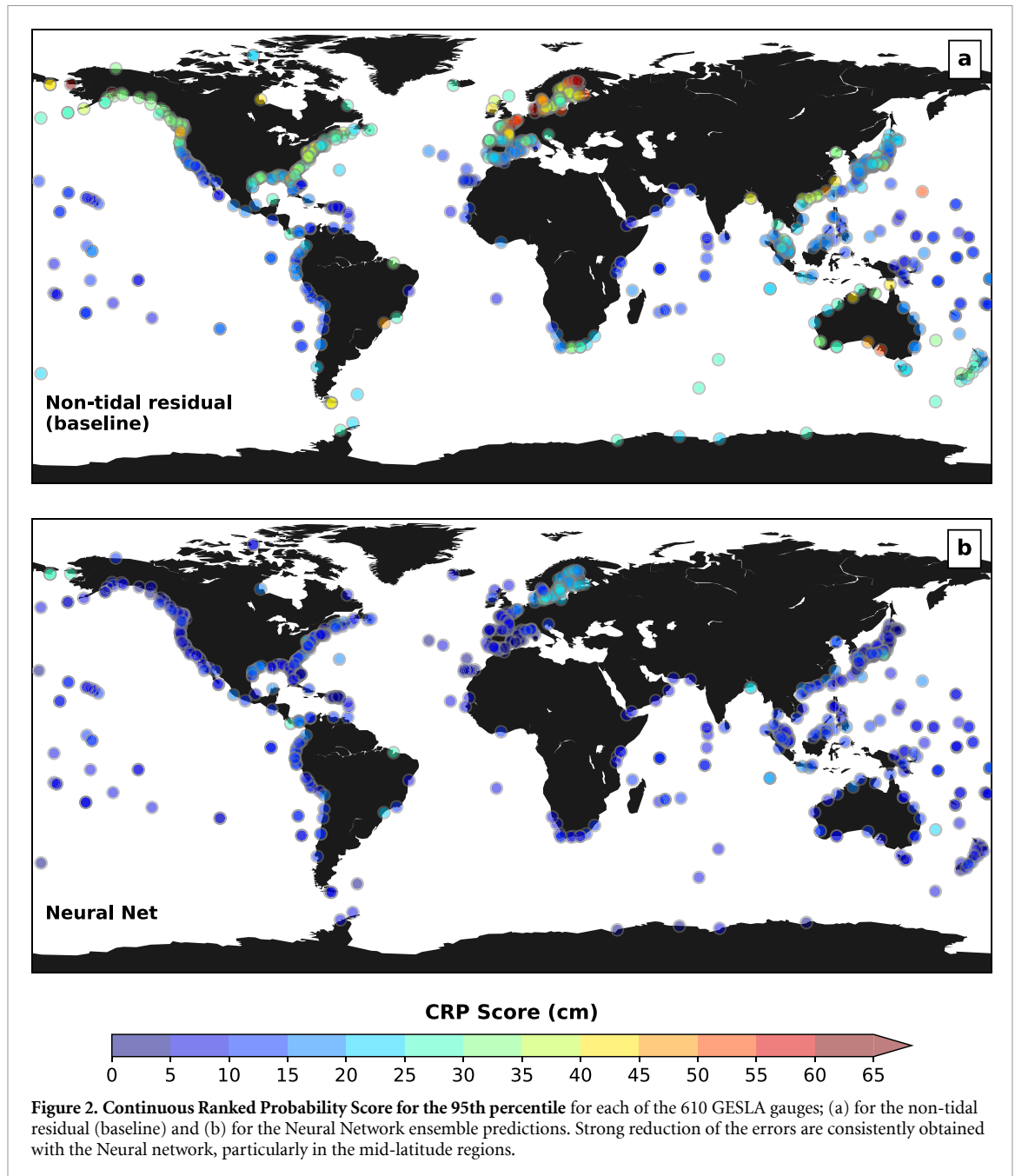
different from the observed values would be heavily penalised while a narrow ensemble centred on the observations would lead to a better score. The CRPS is computed for the one year test period as well as for 95th percentile extreme values (surge). While a 20-member ensemble is not extensive, using a CRPS metric is a better validation approach compared to using the mean or median where the information

contained in the ensemble is mainly lost. The CRPS are computed using the *proprscoring* Python library.

3. Results

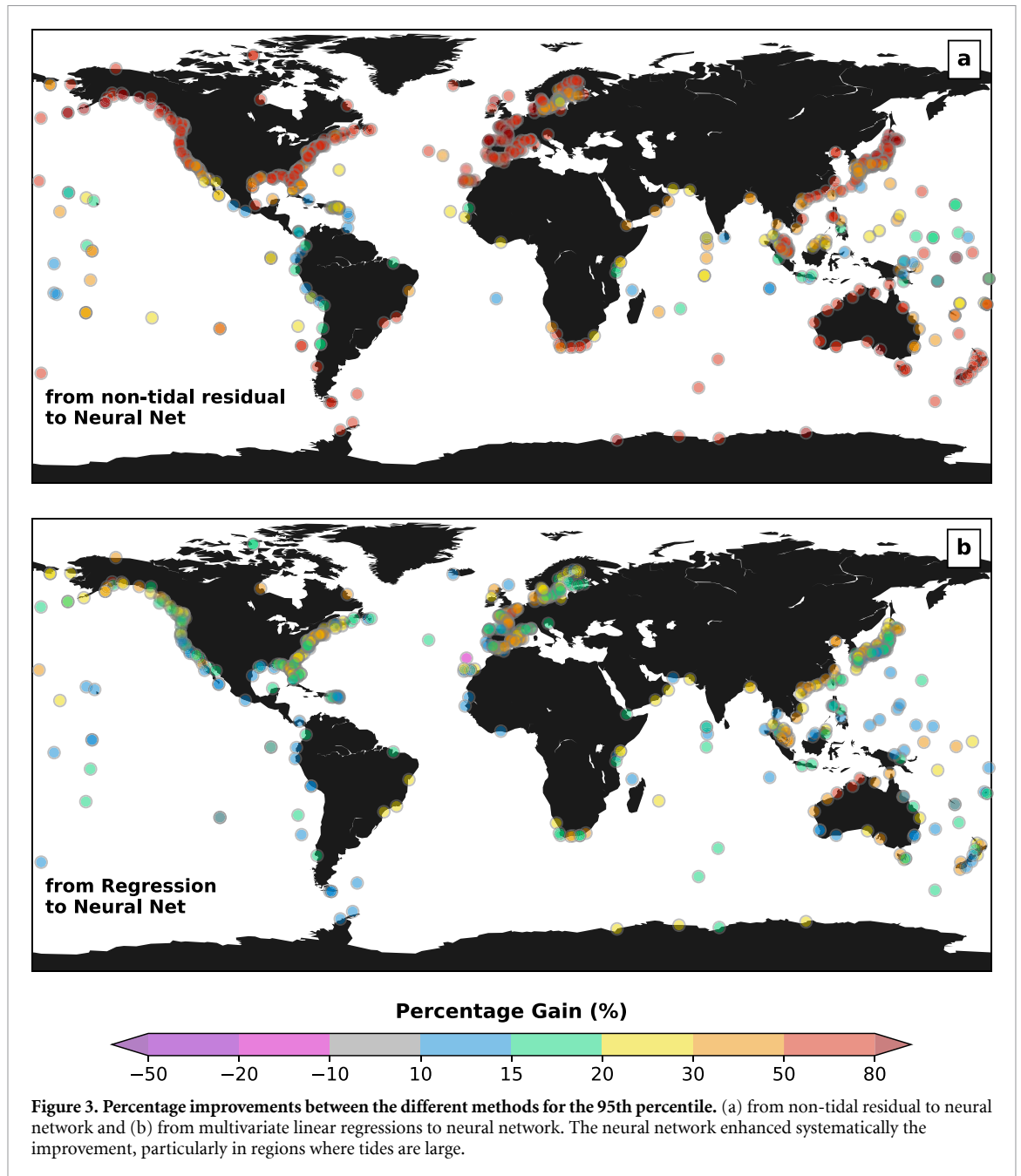
3.1. Global skills of the NN

The CRPS is computed for the observed non-tidal residual to provide a baseline metric for the signal not



captured by the astronomical harmonic analysis. The harmonic analysis does not aim (and has not been designed) to capture this kind of variability; the non-tidal residual simply provides a first-order baseline for comparison based on a 37 constituents harmonic analysis and it is expected that any method should capture parts of the non-tidal signal. The boxplot summarising this baseline skills per number of training years as well as their global distribution for the extreme values (over the 95th percentile anomaly) are presented in figure 1(c) (yellow box) and figure 2(a), respectively. The length of the time series has a weak impact on the CRPS, which ranges from 15 to 25 cm on average. Figure 2(a) illustrates the spatial variability of the CRPS with larger values in mid-latitudes due to consistent winter storms and larger tides compared to tropical regions.

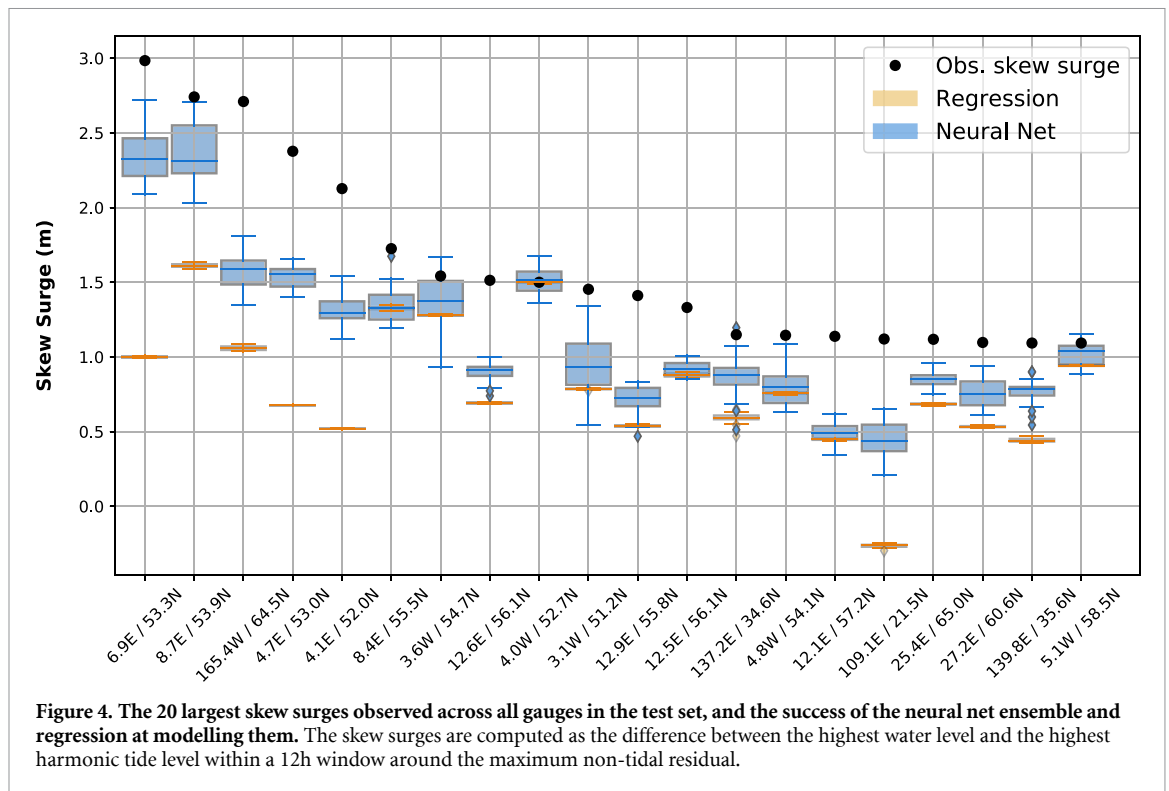
The NNs consistently capture the non-tidal residual due to the effect of atmospheric forcing as well as tide-tide interactions and tide-surge interactions with a mean CRPS of around 10 cm (figure 1(c) - blue box). The CRPS for outlier gauges with large non-tidal residual can be improved from over 50 cm to around 25 cm. Figure 1(b) shows the percentage of non-tidal residual (baseline) captured in the NN predictions ranging from 30 to 60% on average. While longer training period improves the skills, it appears that after 6 years of training data, the performance remains fairly stable. While for any gauge, the NN captures the non-tidal residual (figure 2(b)), the skill varies spatially (figure 3(a)). It is mainly due to the ease of improving a bad skill compared to reducing already good skills (lower than 10 cm).



While the NN approach leads to high skill in reconstructing extremes of non-tidal residual, it is worth considering how a multivariate linear regression would perform in comparison. Figure 3(b) shows the percentage improvement between the two methods. While tropical regions show the lowest improvements using a NN (10–20%), the skills at higher latitude improves by up to 50% with clear regions of the globe emerging as Europe, West coast of North America, Alaska, Chinese Coast, North Australia and the Northern coastline of Japan (facing the Sea of Japan). Except one point in the Canary Islands, the NN outperformed the regression anywhere else; this might be due to a fitting issue at this particular site (not investigated).

Supplementary figures 3, 4 and 5 highlight similar results for the whole 1 year test time series. The skill improvements is not as high as for the extremes but is still significant and systematic. The Baltic sea regions can be pointed here as a region of lower skill improvement from the regression to the NN. This is potentially due to the long time scale sea-level variability that is not included in the predictors used, due to seasonally integrated winds and salinity changes [37]. Similar performance are also obtained for the lowest levels (5th percentile, lowest level being of importance for harbour management) and the 99th percentile (not shown) of the non-tidal residual.

Predicting the full range of non-tidal residuals is key for a broad range of applications. However



assessing the skill of the models in stressed conditions is also of relevant importance. While the usual extreme statistics cannot be applied to this study (the test sets being only one year at each gauge), looking at the most extreme skew surges within over 600 gauges highlights the capability of the models. Figure 4 shows the 20 largest skew surges in the test set. Predictions of these large skew surges are almost always under-estimated compared to observations, but the neural network ensemble shows some skill in capturing them (over 2/3 of the signal) and systematically outperforms the multivariate linear regression. Note that the present neural network and training set have been designed to predict the complete time series and not only the extreme storm surges; therefore the training set is highly unbalanced such that extremes are seen as outliers which penalises the model predictions (more details on the impact of the the training set are provided in discussion).

3.2. Time series at two particular locations

The previous section focus on time-averaged skills in capturing the non-tidal residual. However, it is difficult to assess the highly-complex time variability of this residual. Therefore, two gauges have been selected for a more detailed investigation for their very different characteristics:

- *Anchorage* (149.89 W / 61.24 N - around 14 years of training data - Supp. Figure 6(a), Alaska, USA, located at the end of the Cook inlet and protected from the open ocean. Due to its location, Anchorage is not exposed to extreme surges (less than

1 m in the test year) but the time series exhibits its significant tide-related variability not captured by the harmonic analysis with this constituent set (figure 5(a)),

- *Dunkirk* (2.37E / 51.05 N - also around 14 years of training data - Supp. Figure 6(b)). This gauge was used in the light tuning, mentioned in the method section), North France, located in the English channel, on the North Sea side. For Dunkirk gauge, the test year includes the winter 2013–2014 when severe winterstorm Xaver (Dec. 2013) crossed Northern part of the North Sea and led to significant surges all along the North Sea coast [38]. This was also the highest sea water level anomaly in our 15-year period at Dunkirk (around 2.5 m while the highest peak in the 14 training years was 2.2 m). Finally, the storm occurred far away from Dunkirk where pressure and wind speed did not show any exceptional values but the surge wave travelled around the North Sea, making an interesting and challenging case for the NN (figure 5(b)).

Figure 5 shows a few weeks of non-tidal residual at each selected gauge. The multivariate regression captures fairly well the long-term smoothed variability at Anchorage (figure 5(a)) but cannot capture the high-temporal variability induced by complex tides in the Cook inlet that were not computed in the tidal harmonic analysis; the ensemble variability is also almost non-existent. On the other hand, the NN ensemble captures efficiently the variability (with some spread) leading to a good CRPS (9 cm versus 21 cm for the

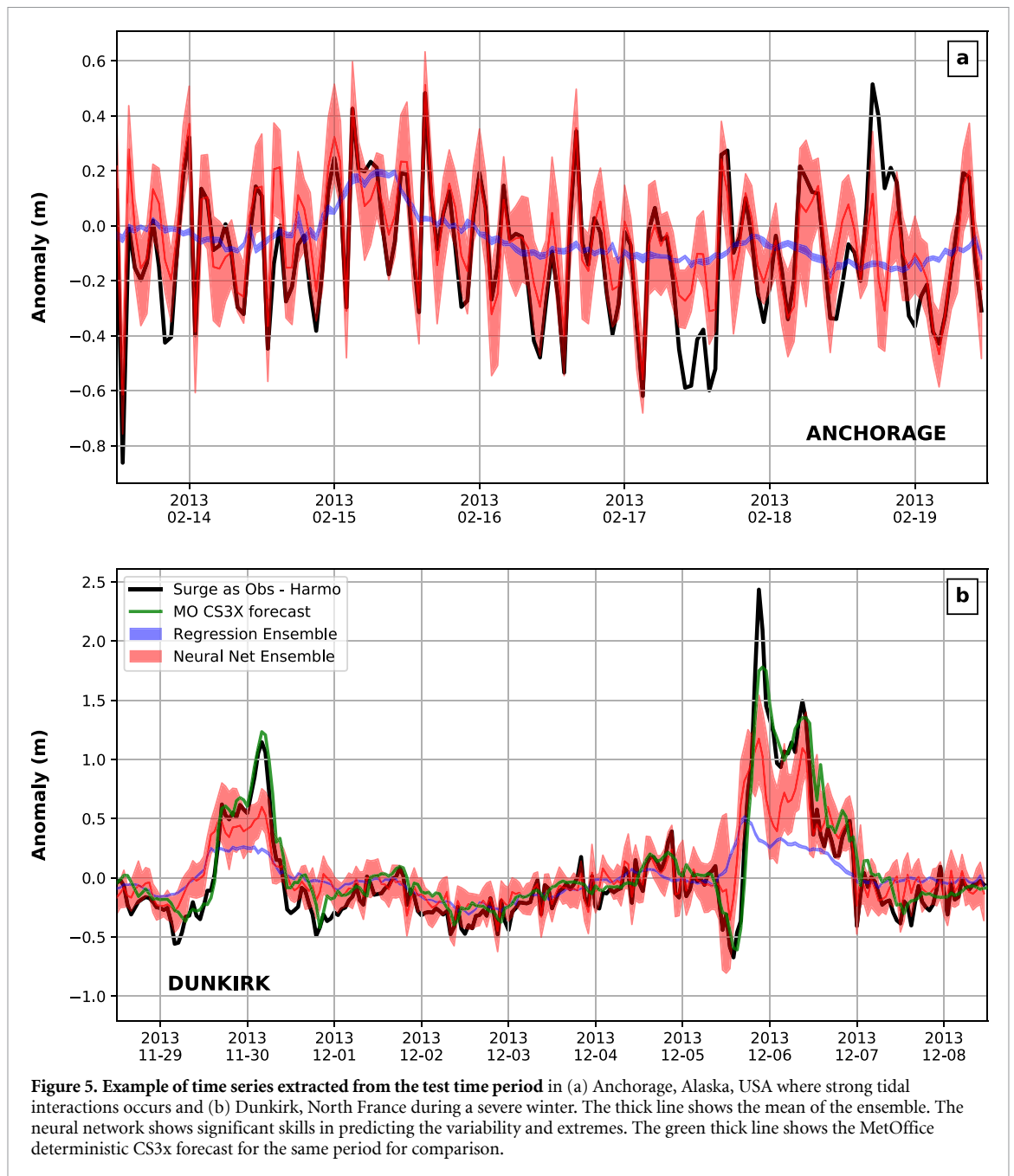
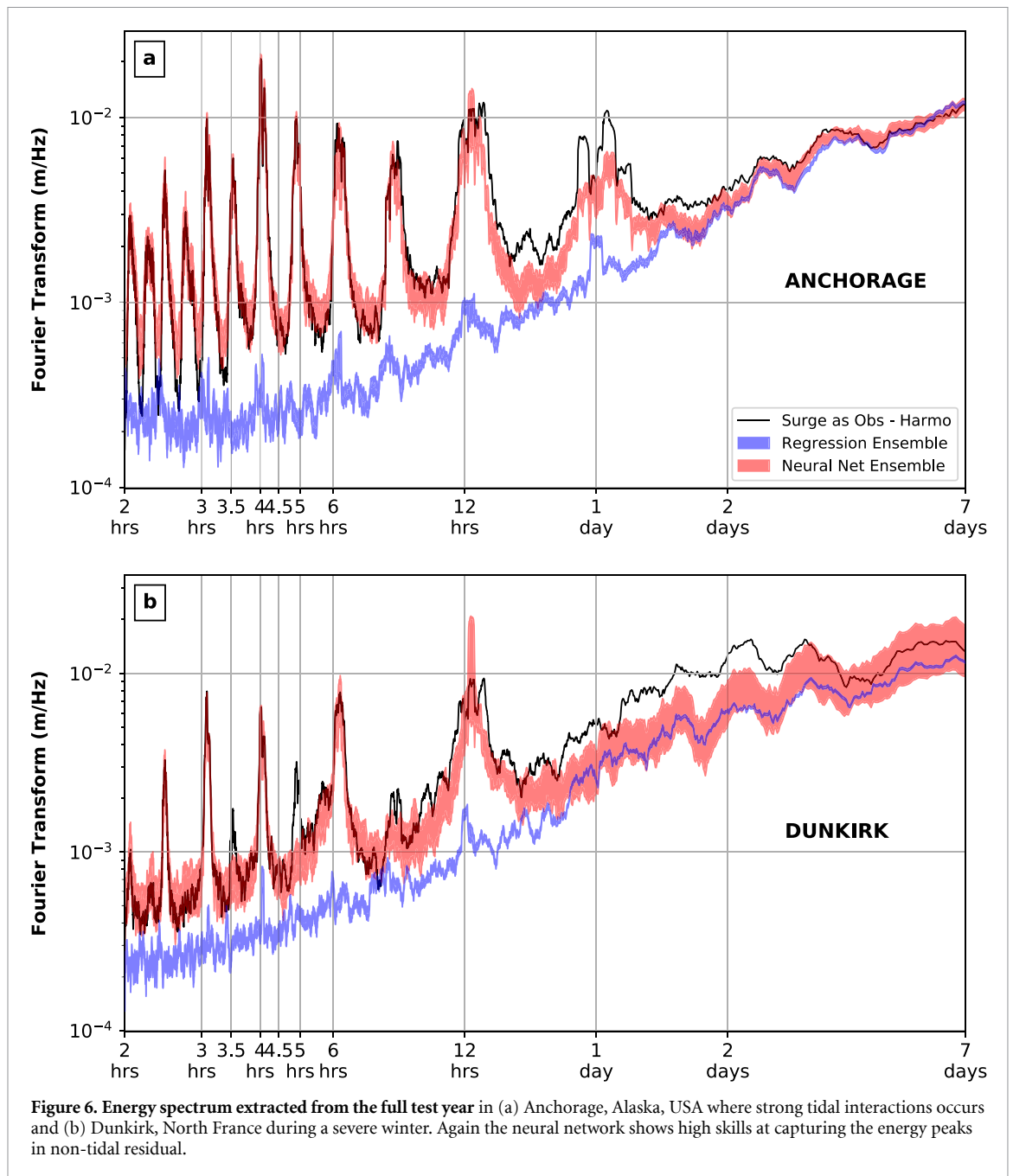


Figure 5. Example of time series extracted from the test time period in (a) Anchorage, Alaska, USA where strong tidal interactions occurs and (b) Dunkirk, North France during a severe winter. The thick line shows the mean of the ensemble. The neural network shows significant skills in predicting the variability and extremes. The green thick line shows the MetOffice deterministic CS3x forecast for the same period for comparison.

regression over the one year test window). A Fourier transform is applied to the one-year signal (figure 6(a)), highlighting the compelling skill of the NN to capture the energy of the system at all time scales while the regression underestimates by an order of magnitude the energy for time scales lower than a day. This shows the capacity of a non-linear NN to predict tide-tide interactions or tide components not included in the harmonic analysis.

Similar conclusions are obtained at Dunkirk. While the extreme storm surges induced by storm Xaver (around 6th December 2013) are underestimated (and so is the previous peak in late November), the prediction is more accurate than the one

predicted with a regression. For comparison, the Met Office CS3x deterministic forecast [39] also underestimates the peak by around 75 cm (figure 5(b)). Over the test year, the NN CRP scores 8 cm and 18 cm for the mean and 95th percentile when the regression gets 13 cm and 32 cm. As for the Anchorage, the energy is well captured by the NN at this gauge except the two smaller peaks for periods of around 3 h 40 min and 4 h 50 min. For periods longer than 1 day, the energy is slightly under-estimated by both the regression and the NN. This highlights skills into predicting sea water anomaly and particularly extreme events using a simple NN forced by a small range of atmospheric and wave data.



4. Discussion and conclusion

An ensemble of NNs have been built for over 600 tide gauges spread around the world in order to predict the non-tidal residual (total sea water level minus an harmonic analysis based on 37 constituents), in term of general behaviours as well as extremes events. The results presented in this study have highlighted the global skill of NNs in capturing non-tidal residual variability and extremes, systematically outperforming predictions based on multivariate linear regressions (in term of CRPS but also in term of correlations). Due to the large amount of available data, the same simple pre-processing and neural network structure were applied to each gauge. A higher level of data quality control or gauge-by-gauge NN tuning

could have been applied, and better performances would then be expected. However, analysis and pre-processing requiring localised intervention was not the aim of the study.

While it was expected that the non-linearity of the NN would play a key role in predicting extreme events through environmental forcing, the results have shown an even better performance of the NNs in their ability to represent tide-tide non-harmonic interactions, treat noise, and express uncertainty. Similar advantages are also reported in the application of Bayesian approaches to the study of tidal currents [40]. Traditional harmonic and response methods [41, 42] have successfully been used for decades to predict tidal amplitudes across the world; however the advent of easily accessible meteorological data

combined with novel applications of methods (for example neural networks, as in this study), could offer a new avenue for improving predictions by capturing non-linear processes.

The model has also shown significant skill in reconstructing extreme surges but still lacks accuracy in the strongest events, in capturing the peak elevation (figures 4 and 5(b) for example). This is partially due to the training data. Extremes can be seen as outliers and are only a fraction of the training set. The machine learning technique minimises a cost function (here, root mean square errors) which generalises common behaviours, and is not well designed for outliers. This leads to bias in the performance toward the average dynamics and not towards the extreme anomalies (positive or negative). Therefore the capability at predicting extremes could be improved by using a differently balanced training set [43]. As a simplistic example, one can draw a similar amount of training data in regular bins covering the range of outcomes (using sampling with replacement technique for bins with a very small amount of data); this leads to a more balanced training set. Supplementary figure 6(a) illustrates the impact of the training set on the model skill at Dunkirk (during storm Xaver in 2013). The NN now captures the amplitude of the peak on the 5th December as well as the deterministic CS3x model, and the peak on the 6th December almost perfectly. The mean of the NN ensemble with a balanced training set is 50 cm higher than the unbalanced result. As seen in Supplementary figure 6(b), in term of energy, the balanced training set is in much better agreement with data for periods longer than 12 hours but it penalises the weaker period where the energy in-between peaks is over-estimated. In terms of CRPS, the mean score decreases by less than 1 cm while the extremes (95th percentile) score improves by 6 cm.

This type of model can be a great tool alongside a deterministic numerical model to improve coastal resilience and potentially set-up warnings in the future as they can also be used to solve classification problems instead of regression ones (as done in the present work) enabling an outcome such as *low risk*, *high risk* and *extreme risk* for example. It was shown here that only a couple of years of training data were enough to get reasonable skills, and there is not significant skill improvement in 30 years training data compared to 6–7 years. In addition, though not shown, even old data collected in the past could be used for present forecasts as long as reference levels have been corrected.

So far the present work has not been extended to locations with no data and the next step would be to build a globally connected tool to predict non-tidal residuals spatially. In addition a better representation of the regional / global atmospheric forcing might help to improve skill. This could be achieved via dimensionality reduction of environmental information

based on unsupervised learning such as principal component analysis or auto-encoder. Finally, investigating more in depth the impact of using a more complex neural network structure adapted to time series (Long- Short Term memory for example) could also be of interest in the future. Setting up high-resolution full physics numerical models in complex inshore regimes is time and computationally expensive and requires physical expertise. These new types of machine learning approaches are appealing for informing stakeholders where there is no capacity for implementing such deterministic weather - surge forecasting systems.

Acknowledgments

Authors NB and JP acknowledge the support from ‘Rapid Tidal Flow Forecasting for Marine Energy Resource Assessment’, National Environmental Research Council (NERC) Innovation Pathfinder award (NE/S005811/1). JW was supported by the NERC under National Capability Official Development Assistance (NC-ODA), ACCORD programme. The GESLA raw data that support the findings of this study are openly available at <https://gesla.org>. Pre-processed data are available from the corresponding author upon reasonable request.

ORCID iDs

Nicolas Bruneau  <https://orcid.org/0000-0002-9017-1000>

Jeff Polton  <https://orcid.org/0000-0003-0131-5250>

Joanne Williams  <https://orcid.org/0000-0002-8421-4481>

Jason Holt  <https://orcid.org/0000-0002-3298-8477>

References

- [1] Blake E, Rappaport E and Landsea C 2007 The deadliest, costliest, and most intense United States tropical cyclones from 1851 to 2006 (and other frequently requested hurricane facts) *Technical report* NHC Miami (<https://www.nhc.noaa.gov/pdf/NWS-TPC-5.pdf>)
- [2] André C, Monfort D, Bouzit M and Vinchon C 2013 Contribution of insurance data to cost assessment of coastal flood damage to residential buildings: insights gained from Johanna (2008) and xynthia (2010) storm events *Nat. Hazards Earth Syst. Sci.* **13** 2003–12
- [3] Needham H F, Keim B D and Sathiaraj D 2015 A review of tropical cyclone-generated storm surges: Global data sources, observations and impacts *Rev. Geophys.* **53** 545–91
- [4] Marsooli R, Lin N, Emanuel K and Feng K 2019 Climate change exacerbates hurricane flood hazards along us atlantic and gulf coasts in spatially varying pattern. *Nat Commun.* **10** 3785
- [5] Dietrich J C *et al* 2010 A high-resolution coupled riverine flow, tide, wind, wind wave and storm surge model for southern Louisiana and Mississippi. part ii: Synoptic description and analysis of hurricanes Katrina and Rita *Mon. Weather Rev.* **138** 378–404

- [6] Rego J L and Li C 2010 Nonlinear terms in storm surge predictions: Effect of tide and shelf geometry with case study from Hurricane Rita *J. Geophys. Res. Oceans* **115** C06020
- [7] Bertin X, Bruneau N, Breilh J-F, Fortunato A B and Karpytchev M 2012 Importance of wave age and resonance in storm surges: The case Xynthia, Bay of Biscay *Ocean Model.* **42** 16–30
- [8] Fortunato A B et al 2017 Operational forecast framework applied to extreme sea levels at regional and local scales *J. Oper. Oceanogr.* **10** 1–15
- [9] Krien Y et al 2017 Towards improved storm surge models in the northern bay of bengal *Cont. Shelf Res.* **135** 58–73
- [10] Suh S-W and Lee H-Y 2018 Forerunner storm surge under macro-tidal environmental conditions in shallow coastal zones of the Yellow Sea *Cont. Shelf Res.* **169** 1–16
- [11] Fernandez-Montblanc T, Vousdoukas M I, Ciavola P, Voukouvalas E, Mentaschi L, Breyiannis G, Feyen L and Salamon P 2019 Towards robust pan-European storm surge forecasting *Ocean Model.* **133** 129–44
- [12] Ham Y-G, Kim J-H and Luo J-J 2019 Deep learning for multi-year ENSO forecasts *Nature* **573** 568–72
- [13] O'Donncha F, Zhang Y, Chen B and James S C 2019 Ensemble model aggregation using a computationally lightweight machine-learning model to forecast ocean waves *J. Mar. Syst.* **199** 103206
- [14] Leahy T P, Llopis F P, Palmer M D and Robinson N H 2018 Using neural networks to correct historical climate observations *J. Atmos. Ocean. Technol.* **35** 2053–9
- [15] Rossiter J R 1959 Research on methods of forecasting storm surges on the east and south coasts of Great Britain *Q. J. R. Meteorol. Soc.* **85** 262–77
- [16] Amin M 01 1982 On analysis and forecasting of surges on the west coast of Great Britain *Geophys. J. Int.* **68** 79–94
- [17] Salmun H, Molod A, Wisniewska K and Buonaiuto F S 2011 Statistical prediction of the storm surge associated with cool-weather storms at the Battery, New York *J. Appl. Meteorol. Climatol.* **50** 273–82
- [18] Lopeman G 2015 M. and Deodatis and G Franco. Extreme storm surge hazard estimation in lower manhattan *Nat. Hazards* **78** 355–91
- [19] Salmun H and Molod A 2015 The use of a statistical model of storm surge as a bias correction for dynamical surge models and its applicability along the U.S. East Coast *J. Mar. Sci. Eng.* **3** 73–86
- [20] Roberts K J, Colle B A, Georgas N and Munch S B 2015 A regression-based approach for cool-season storm surge predictions along the New York-New Jersey coast *J. Appl. Meteorol. Climatol.* **54** 1773–91
- [21] French J, Mawdsley R, Fujiyama T and Achuthan K 2017 Combining machine learning with computational hydrodynamics for prediction of tidal surge inundation at estuarine ports *Procedia IUTAM* **25** 28–35 Symp. on Storm Surge Modelling and Forecasting
- [22] Salim A M, Dwarakish G S, Liju K V, Thomas J, Devi G and Rajeech R 2015 Weekly prediction of tides using neural networks *Procedia Eng.* **116** 678–82 *8th Int. Conf. on Asian and Pacific Coasts (APAC 2015)*
- [23] Hieronymus M, Hieronymus J and Hieronymus F 2019 On the application of machine learning techniques to regression problems in sea level studies *J. Atmos. Ocean. Technol.* (<https://journals.ametsoc.org/doi/full/10.1175/JTECH-D-19-0033.1?mobileUi=0>)
- [24] Woodworth P L, Hunter J R, Marcos M, Caldwell P, Menéndez M and Haigh I 2016 Towards a global higher-frequency sea level dataset *Geosci. Data J.* **3** 50–9
- [25] Sam C 2017 Pytides 0.0.4, python package for the analysis and prediction of tides. *Technical report* (<https://pypi.org/project/pytides/>)
- [26] Copernicus Climate Change Service 2017 ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate *Technical report* Copernicus Climate Change Service Climate Data Store (CDS) (<https://doi.org/10.24381/cds.adbb2d47>)
- [27] Maas A L, Hannun A L and Ng A Y 2013 Rectifier nonlinearities improve neural network acoustic models *30th Int. Conf. on Machine Learning* Atlanta, USA (ICML 2013) (https://ai.stanford.edu/~amaas/papers/relu_hybrid_icml2013_final.pdf)
- [28] Ioffe S and Szegedy C 2015 Batch normalization: accelerating deep network training by reducing internal covariate shift *32th Int. Conf. on Machine Learning* Lille, France (ICML 2015) (<http://proceedings.mlr.press/v37/loff15.pdf>)
- [29] Kingma D P and Ba J L 2015 Adam: A method for stochastic optimization *3rd Int. Conf. on Learning Representations* San Diego, USA (ICLR 2018) (<https://arxiv.org/pdf/1412.6980.pdf>)
- [30] Chollet F 2015 Keras (<https://keras.io>)
- [31] Abadi M et al 2016 Tensorflow: A system for large-scale machine learning *12th Symp. on Operating Systems Design and Implementation (OSDI 16)* pp 265–83 (<https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf>)
- [32] Pedregosa F et al 2011 Scikit-learn: machine learning in Python *J. Mach. Learn. Res.* **12** 2825–30 (<http://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>)
- [33] Hochreiter S and Schmidhuber J 1997 Long short-term memory *Neural Comput.* **9** 1735–80
- [34] Matheson J E and Winkler R L 1976 Scoring rules for continuous probability distributions *Manage. Sci.* **22** 1087–96
- [35] Hersbach H 2000 Decomposition of the continuous ranked probability score for ensemble prediction systems *Weather Forecast.* **15** 559–70
- [36] Zamo M and Naveau P 2018 Estimation of the continuous ranked probability score with limited information and applications to ensemble weather forecasts *Math. Geosci.* **50** 209–34
- [37] Andersson H C 2002 Influence of long-term regional and large-scale atmospheric circulation on the baltic sea level *Tellus A: Dyn. Meteorol. Oceanogr.* **54** 76–88
- [38] Wei X, Brown J M, Williams J, Thorne P D, Williams M E and Amoudry L O 2019 Impact of storm propagation speed on coastal flood hazard induced by offshore storms in the North Sea *Ocean Model.* **143** 101472
- [39] Flowerdew J, Horsburgh K, Wilson C and Mylne K 2010 Development and evaluation of an ensemble forecasting system for coastal storm surges *Q. J. R. Meteorol. Soc.* **136** 1444–56
- [40] Sarkar D, Osborne M A and Adcock T A A 2018 Prediction of tidal currents using Bayesian machine learning *Ocean Eng.* **158** 221–31
- [41] Zetler B, Cartwright D and Berkman S 1979 Some comparisons of response and harmonic tide predictions *Int. Hydrographic Review, Monaco, LVI (2), July 1979* 105–15
- [42] Smith A J E, Ambrosius B A C, Wakker K F, Woodworth P L and Vassie J M 1997 Comparison between the harmonic and response methods of tidal analysis using TOPEX/Poseidon altimetry *J. Geod.* **71** 695–703
- [43] Lemaitre G, Nogueira F and Aridas C K 2017 Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning *J. Mach. Learn. Res.* **18** 1–5 (<http://www.jmlr.org/papers/volume18/16-365/16-365.pdf>)