

ACCEPTED MANUSCRIPT • OPEN ACCESS

Estimation of global coastal sea level extremes using neural networks

To cite this article before publication: Nicolas Bruneau *et al* 2020 *Environ. Res. Lett.* in press <https://doi.org/10.1088/1748-9326/ab89d6>

Manuscript version: Accepted Manuscript

Accepted Manuscript is “the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an ‘Accepted Manuscript’ watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors”

This Accepted Manuscript is © 2020 The Author(s). Published by IOP Publishing Ltd.

As the Version of Record of this article is going to be / has been published on a gold open access basis under a CC BY 3.0 licence, this Accepted Manuscript is available for reuse under a CC BY 3.0 licence immediately.

Everyone is permitted to use all or part of the original content in this article, provided that they adhere to all the terms of the licence <https://creativecommons.org/licenses/by/3.0>

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions may be required. All third party content is fully copyright protected and is not published on a gold open access basis under a CC BY licence, unless that is specifically stated in the figure caption in the Version of Record.

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

Estimation of Global Coastal Sea level extremes using Neural Networks

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

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

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

E-mail: bruneau.n@gmail.com

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.

PACS numbers: 07.05.Mh, 92.10.Sx, 92.10.hp

Keywords: Sea water anomaly, Extremes, Storm surges, GESLA database, Machine Learning

1
2
3 *Estimation of Global Coastal Sea level extremes using Neural Networks*
4

5 Submitted to: *Environ. Res. Lett.*
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Accepted Manuscript

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, 2, 3, 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, 6, 7, 8, 9, 10, 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

Estimation of Global Coastal Sea level extremes using Neural Networks

4

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

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

2. Methods

2.1. Global Extreme Sea Level Analysis dataset - GESLA

49 The Global Extreme Sea Level Analysis database (GESLA version 2 [24]) provides
50 unified high-frequency (15 min to 1 hour temporal resolution) quasi-global coastal sea
51 level water information. Only public data (around 1070 gauges) are used in the present
52 study. While data have been standardised, a simple but strict methodology was applied
53 to pre-process each gauge in a systematic and reproducible manner. The key aspect
54

5

Estimation of Global Coastal Sea level extremes using Neural Networks

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. 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

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

(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

Estimation of Global Coastal Sea level extremes using Neural Networks

6

small amplitudes.

- The total water signal is re-interpolated over a constant 1h 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 objective 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 3h gradients of all the atmospheric predictors are computed

Estimation of Global Coastal Sea level extremes using Neural Networks 7

100 to capture late intensification / de-intensification (for example a low pressure system
101 developing rapidly) as well as for the harmonic tidal level.

2.3. Machine Learning

103 Each gauge is modelled independently using artificial Neural Networks (NNs). Each
104 NN is composed of 3 hidden layers of 48 neurones. The input layer has 33 nodes
105 (one for each environmental predictor described in previous sections combined with 7
106 hourly time steps of harmonic tide), and the outer layer has a single node providing
107 the non-tidal residual target. While a sigmoid activation function is used for the last
108 layer, the hidden layers consists of Leaky ReLU activation functions[27] combined with
109 batch normalization layer to normalise the activations[28]. The NN had just under
110 7000 trainable parameters and its schematic view is provided in Supplementary Figure
111 2. Finally, an Adam solver[29] is used to minimise the root mean square error between
112 non-tidal residual predictions and observations; the NN is fitted for 150 iterations or less
113 if the errors is not reduced within 10 consecutive iterations. Due to the large number
114 of gauges available, this configuration has been lightly tuned on three random gauges
115 (namely, a few combinations of the number of neurones, number of hidden layers and
116 type of activation) and then applied to the full set without further adjustment.

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

122 An ensemble of 20 Neural Networks (NNs) is trained at each gauge location to
123 generate a probabilistic forecast. Each NN is fitted using 50% of the training set,
124 randomly sampled. While a larger ensemble would have improved our probabilistic
125 forecast, 20 members were chosen as a pragmatic balance between computational cost
126 (over 12000 NNs have been fitted in this study) and variability in the predictions.

Estimation of Global Coastal Sea level extremes using Neural Networks 8

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, 35, 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 gauged location:

1
2
3 *Estimation of Global Coastal Sea level extremes using Neural Networks*
4
5

$$6 \quad CRPS = \frac{1}{N} \sum_{t=1}^N \left(\int_{\eta_r > -\infty}^{\eta_r < \infty} [H^p(\eta_r; t) - H^o(\eta_r; t)]^2 dz \right),$$

7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

156 Intuitively, an ensemble producing a wide range of outcomes or an ensemble with a
157 mean significantly different from the observed values would be heavily penalised while
158 a narrow ensemble centred on the observations would lead to a better score. The
159 CRPS is computed for the one year test period as well as for 95th percentile extreme
160 values (surge). While a 20-member ensemble is not extensive, using a CRPS metric
161 is a better validation approach compared to using the mean or median where the
162 information contained in the ensemble is mainly lost. The CRPS are computed using
163 the *proprscoring* Python library.

164 3. Results

165 3.1. Global skills of the NN

166 The CRPS is computed for the observed non-tidal residual to provide a baseline metric
167 for the signal not captured by the astronomical harmonic analysis. The harmonic
168 analysis does not aim (and has not been designed) to capture this kind of variability;
169 the non-tidal residual simply provides a first-order baseline for comparison based on a
170 37 constituents harmonic analysis and it is expected than any method should capture
171 parts of the non-tidal signal. The boxplot summarising this baseline skills per number
172 of training years as well as their global distribution for the extreme values (over the 95th
173 percentile anomaly) are presented in Figure 1c (yellow box) and Figure 2a, respectively.
174 The length of the time series has a weak impact on the CRPS, which ranges from 15 to
175 25 cm on average. Figure 2a illustrates the spatial variability of the CRPS with larger
176 value in mid latitudes due to consistent winter storms and larger tides compared to

177 tropical regions.

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

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

197 Supplementary Figures 3, 4 and 5 highlight similar results for the whole 1 year
198 test time series. The skill improvements is not as high as for the extremes but is still
199 significant and systematic. The Baltic sea regions can be pointed here as a region of
200 lower skill improvement from the regression to the NN. This is potentially due to the
201 long time scale sea-level variability that is not included in the predictors used, due
202 to seasonally integrated winds and salinity changes[37]. Similar performance are also
203 obtained for the lowest levels (5th percentile, lowest level being of importance for harbour
204 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* (149.89W / 61.24N - around 14 years of training data - Supp. Fig. 6a), 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 significant tide-related variability not captured by the harmonic analysis with this constituent set (Fig. 5a),

- *Dunkirk* (2.37E / 51.05N - also around 14 years of training data - Supp. Fig. 6b). 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.

Estimation of Global Coastal Sea level extremes using Neural Networks 12

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 2.5m 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 (Fig. 5b).

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 (Fig. 5a) 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 regression over the one year test window). A Fourier transform is applied to the one-year signal (Fig. 6a), 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 under-estimated (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 under-estimates the peak by around 75 cm (Fig. 5b). 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 (Fig 4 and Fig 5b

Estimation of Global Coastal Sea level extremes using Neural Networks 14

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 7a 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 7b, in term of energy, the balanced training set is in much better agreement with data for a period 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 be extended to locations with no data and the

315 next step would be to built a globally connected tool to predict non-tidal residuals
316 spatially. In addition a better representation of the regional / global atmospheric forcing
317 might help to improve skill. This could be achieved via dimensionality reduction of
318 environmental information based on unsupervised learning such as principal component
319 analysis or auto-encoder. Finally, investigating more in depth the impact of using a more
320 complex neural network structure adapted to time series (Long- Short Term memory for
321 example) could also be of interest in the future. Setting up high-resolution full physics
322 numerical models in complex inshore regimes is time and computationally expensive
323 and requires physical expertise. These new types of machine learning approaches are
324 appealing for informing stakeholders where there is no capacity for implementing such
325 deterministic weather - surge forecasting systems.

326 5. Acknowledgements

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

334 References

- 335 [1] E. Blake, E. Rappaport, and C. Landsea. The deadliest, costliest, and most intense United States
336 tropical cyclones from 1851 to 2006 (and other frequently requested hurricane facts). Technical
337 report, NHC Miami, 2007.
- 338 [2] C. André, D. Monfort, M. Bouzit, and C. Vinchon. Contribution of insurance data to cost
339 assessment of coastal flood damage to residential buildings: insights gained from Johanna (2008)
340 and xynthia (2010) storm events. *Natural Hazards and Earth System Sciences*, 13(8):2003–2012,
341 2013.

Estimation of Global Coastal Sea level extremes using Neural Networks 16

- [3] H. F. Needham, B. D. Keim, and D. Sathiaraj. A review of tropical cyclone-generated storm surges: Global data sources, observations, and impacts. *Reviews of Geophysics*, 53(2):545–591, 2015.
- [4] R. Marsooli, N. Lin, K. Emanuel, and K. Feng. Climate change exacerbates hurricane flood hazards along us atlantic and gulf coasts in spatially varying patterns. *Nature Communications*, 10(3785), 2019.
- [5] J. C. Dietrich, S. Bunya, J. J. Westerink, B. A. Ebersole, J. M. Smith, J. H. Atkinson, R. Jensen, D. T. Resio, R. A. Luettich, C. Dawson, V. J. Cardone, A. T. Cox, M. D. Powell, H. J. Westerink, and H. J. Roberts. 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. *Monthly Weather Review*, 138(2):378–404, 2010.
- [6] Joo L. Rego and Chunyan Li. Nonlinear terms in storm surge predictions: Effect of tide and shelf geometry with case study from Hurricane Rita. *Journal of Geophysical Research: Oceans*, 115(C6), 2010.
- [7] Xavier Bertin, Nicolas Bruneau, Jean-Francois Breilh, Andr B. Fortunato, and Mikhail Karpytchev. Importance of wave age and resonance in storm surges: The case Xynthia, Bay of Biscay. *Ocean Modelling*, 42:16 – 30, 2012.
- [8] Andr B. Fortunato, Anabela Oliveira, Joo Rogeiro, Ricardo Tavares da Costa, Joo L. Gomes, Kai Li, Gonalo de Jesus, Paula Freire, Ana Rilo, Ana Mendes, Marta Rodrigues, and Alberto Azevedo. Operational forecast framework applied to extreme sea levels at regional and local scales. *Journal of Operational Oceanography*, 10(1):1–15, 2017.
- [9] Y. Krien, L. Testut, A.K.M.S. Islam, X. Bertin, F. Durand, C. Mayet, A.R. Tazkia, M. Becker, S. Calmant, F. Papa, V. Ballu, C.K. Shum, and Z.H. Khan. Towards improved storm surge models in the northern bay of bengal. *Continental Shelf Research*, 135:58 – 73, 2017.
- [10] S.-W. Suh and H.-Y. Lee. Forerunner storm surge under macro-tidal environmental conditions in shallow coastal zones of the Yellow Sea. *Continental Shelf Research*, 169:1 – 16, 2018.
- [11] T. Fernandez-Montblanc, M. I. Vousdoukas, P. Ciavola, E. Voukouvalas, L. Mentaschi, G. Breyiannis, L. Feyen, and P. Salamon. Towards robust pan-European storm surge forecasting. *Ocean Modelling*, 133:129 – 144, 2019.
- [12] Y.-G. Ham, J.-H. Kim, and J.-J. Luo. Deep learning for multi-year ENSO forecasts. *Nature*, 573:568–572, 2019.
- [13] F. O’Donncha, Y. Zhang, B. Chen, and S. C. James. Ensemble model aggregation using a computationally lightweight machine-learning model to forecast ocean waves. *Journal of Marine Systems*, 199:103206, 2019.

Estimation of Global Coastal Sea level extremes using Neural Networks 17

- [14] T. P. Leahy, F. P. Llopis, M. D. Palmer, and N. H. Robinson. Using neural networks to correct historical climate observations. *Journal of Atmospheric and Oceanic Technology*, 35(10):2053–2059, 2018.
- [15] J. R. Rossiter. Research on methods of forecasting storm surges on the east and south coasts of Great Britain. *Quarterly Journal of the Royal Meteorological Society*, 85(365):262–277, 1959.
- [16] M. Amin. On analysis and forecasting of surges on the west coast of Great Britain. *Geophysical Journal International*, 68(1):79–94, 01 1982.
- [17] H. Salmun, A. Molod, K. Wisniewska, and F. S. Buonaiuto. Statistical prediction of the storm surge associated with cool-weather storms at the Battery, New York. *Journal of Applied Meteorology and Climatology*, 50(2):273–282, 2011.
- [18] G. Lopeman, M. and Deodatis and G Franco. Extreme storm surge hazard estimation in lower manhattan. *Nat Hazards*, 78:355–391, 2015.
- [19] H. Salmun and A. Molod. 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. *Journal of Marine Science and Engineering*, 3(1):73–86, 2015.
- [20] Keith J. Roberts, Brian A. Colle, Nickitas Georgas, and Stephan B. Munch. A regression-based approach for cool-season storm surge predictions along the New York-New Jersey coast. *Journal of Applied Meteorology and Climatology*, 54(8):1773–1791, 2015.
- [21] Jon French, Robert Mawdsley, Taku Fujiyama, and Kamal Achuthan. Combining machine learning with computational hydrodynamics for prediction of tidal surge inundation at estuarine ports. *Procedia IUTAM*, 25:28 – 35, 2017. IUTAM Symposium on Storm Surge Modelling and Forecasting.
- [22] Akhil Muhammad Salim, G.S. Dwarakish, Liju K.V., Justin Thomas, Gayathri Devi, and Rajeeesh R. Weekly prediction of tides using neural networks. *Procedia Engineering*, 116:678 – 682, 2015. 8th International Conference on Asian and Pacific Coasts (APAC 2015).
- [23] M. Hieronymus, J. Hieronymus, and F. Hieronymus. On the application of machine learning techniques to regression problems in sea level studies. *Journal of Atmospheric and Oceanic Technology*, 0(0):null, 0.
- [24] P. L. Woodworth, J. R. Hunter, M. Marcos, P. Caldwell, M. Menendez, and I. Haigh. Towards a global higher-frequency sea level dataset. *Geoscience Data Journal*, 3(2):50–59, 2016.
- [25] Sam Cox. Pytides 0.0.4, python package for the analysis and prediction of tides. Technical report, 2017.
- [26] Copernicus Climate Change Service. ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Technical report, Copernicus Climate Change Service Climate Data Store

Estimation of Global Coastal Sea level extremes using Neural Networks 18

- (CDS), 2017.
- [27] A. L. Maas, A. L. Hannun, and A. Y. Ng. Rectifier nonlinearities improve neural network acoustic models. In *30th International Conference on Machine Learning, Atlanta, USA (ICML 2013)*, 2013.
- [28] S. Ioffe and Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In *32th International Conference on Machine Learning, Lille, France (ICML 2015)*, 2015.
- [29] D. P. Kingma and J. L. Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, San Diego, USA (ICLR 2018)*, 2015.
- [30] Francois Chollet. Keras, 2015.
- [31] Martin Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G. Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. Tensorflow: A system for large-scale machine learning. In *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, pages 265–283, 2016.
- [32] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [33] Sepp Hochreiter and Jrgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.
- [34] J. E. Matheson and R. L. Winkler. Scoring rules for continuous probability distributions. *Management Science*, 22(10):1087–1096, 1976.
- [35] H. Hersbach. Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 15(5):559–570, 2000.
- [36] M. Zamo and P. Naveau. Estimation of the continuous ranked probability score with limited information and applications to ensemble weather forecasts. *Mathematical Geosciences*, 50(2):209–234, Feb 2018.
- [37] Helm C. Andersson. Influence of long-term regional and large-scale atmospheric circulation on the baltic sea level. *Tellus A: Dynamic Meteorology and Oceanography*, 54(1):76–88, 2002.
- [38] X. Wei, J. M. Brown, J. Williams, P. D. Thorne, M. E. Williams, and L. O. Amoudry. Impact of storm propagation speed on coastal flood hazard induced by offshore storms in the North Sea. *Ocean Modelling*, page 101472, 2019.

Estimation of Global Coastal Sea level extremes using Neural Networks 19

- 444 [39] J. Flowerdew, K. Horsburgh, C. Wilson, and K. Mylne. Development and evaluation of an ensemble
445 forecasting system for coastal storm surges. *Quarterly Journal of the Royal Meteorological*
446 *Society*, 136(651):1444–1456, 2010.
- 447 [40] D. Sarkar, M. A. Osborne, and T. A. A. Adcock. Prediction of tidal currents using Bayesian
448 machine learning. *Ocean Engineering*, 158:221 – 231, 2018.
- 449 [41] B. Zetler, D. Cartwright, and S. Berkman. Some comparisons of response and harmonic tide
450 predictions. In *International Hydrographic Review, Monaco, LVI (2), July 1979*, pages 105–
451 115, 1979.
- 452 [42] A. J. E. Smith, B. A. C. Ambrosius, K. F. Wakker, P. L. Woodworth, and J. M. Vassie. Comparison
453 between the harmonic and response methods of tidal analysis using TOPEX/Poseidon altimetry.
454 *Journal of Geodesy*, 71(11):695–703, Oct 1997.
- 455 [43] Guillaume Lemaître, Fernando Nogueira, and Christos K. Aridas. Imbalanced-learn: A python
456 toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine*
457 *Learning Research*, 18(17):1–5, 2017.

Estimation of Global Coastal Sea level extremes using Neural Networks

20

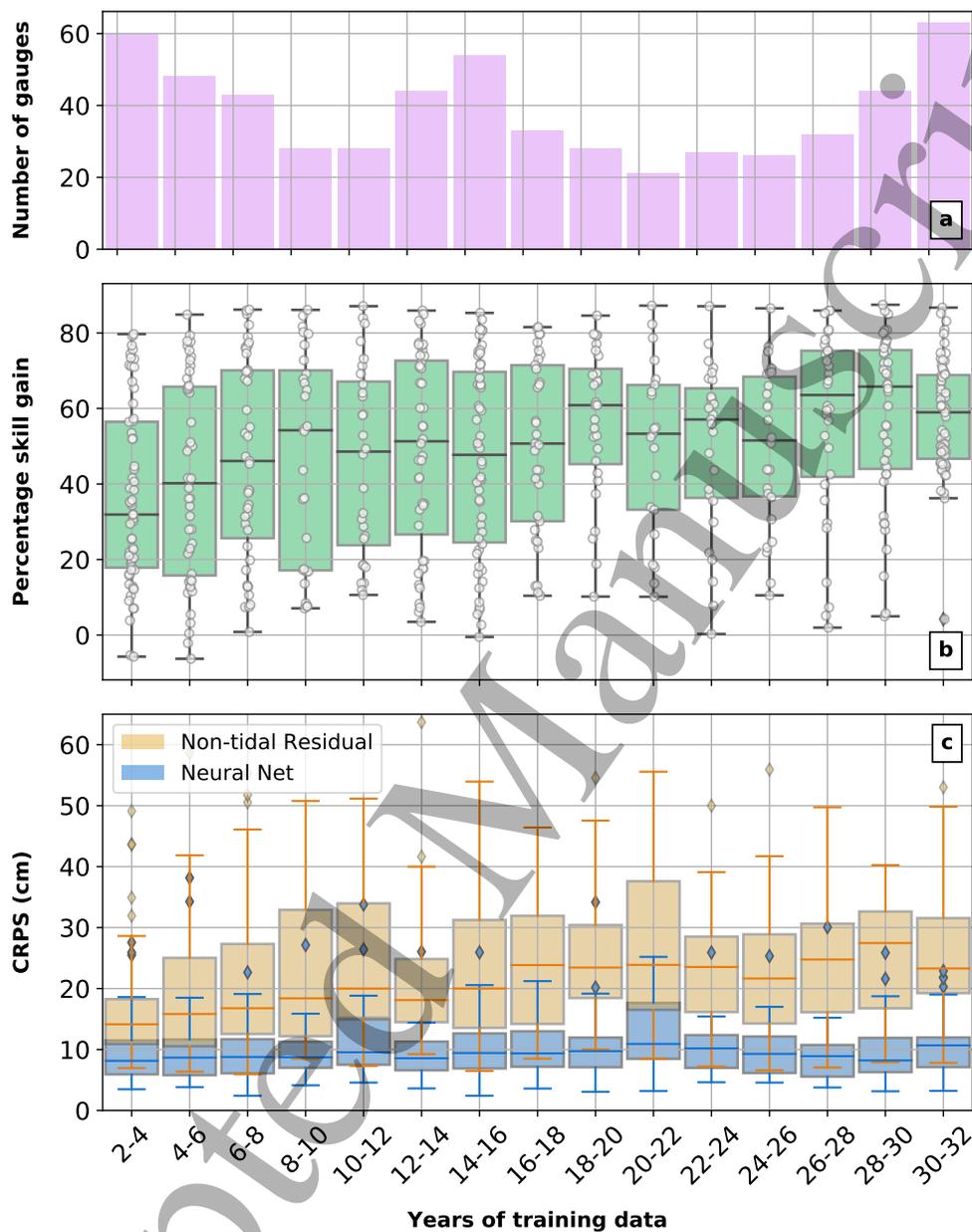


Figure 1. Statistics per number of training years for the 95th percentile.

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.

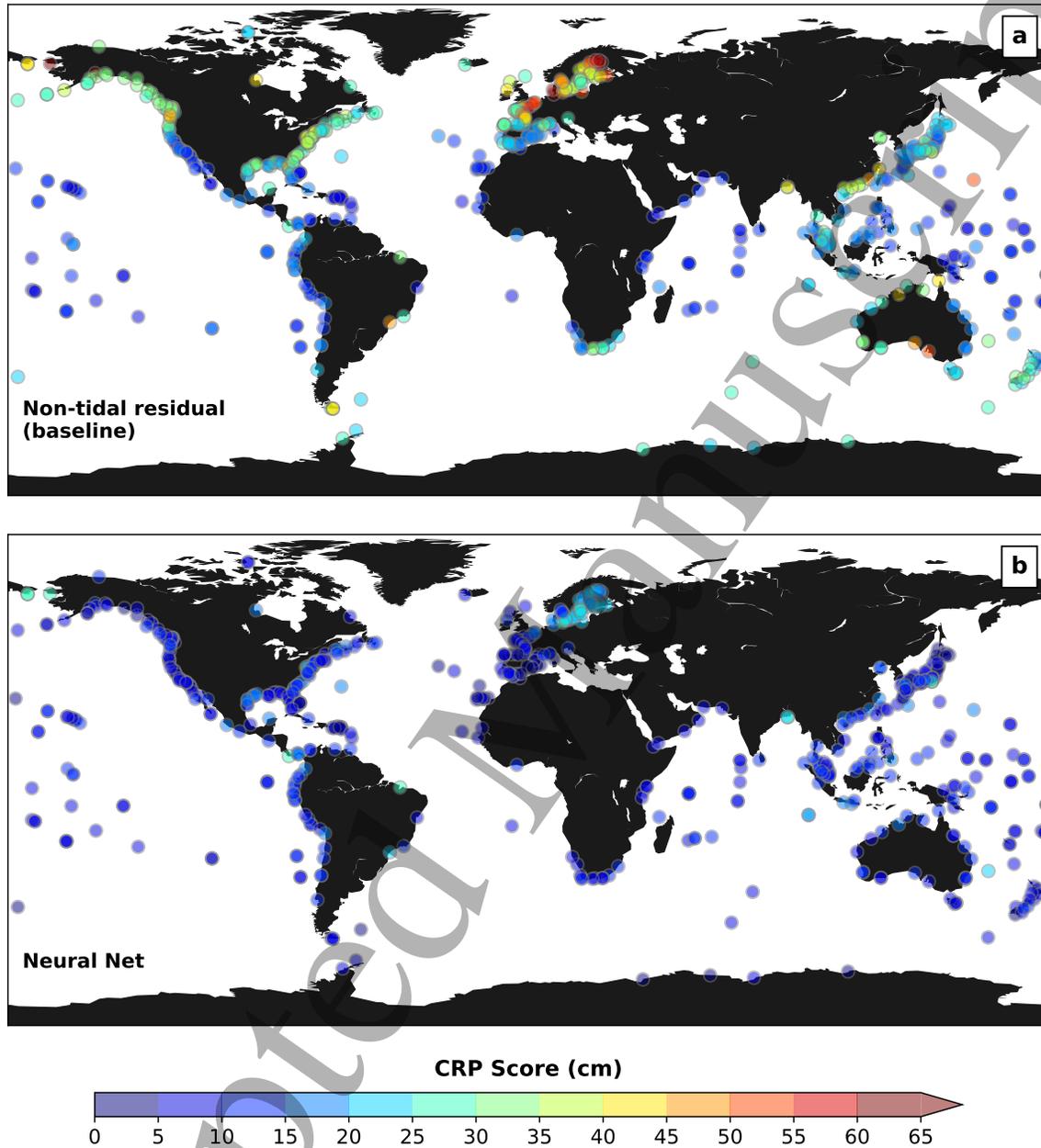


Figure 2. Continuous Ranked Probability Score for the 95th percentile for each of the 610 GESLA gauges; a) for the non-tidal residual (baseline) and b) for the Neural Network ensemble predictions. Strong reduction of the errors are consistently obtained with the Neural network, particularly in the mid-latitude regions.

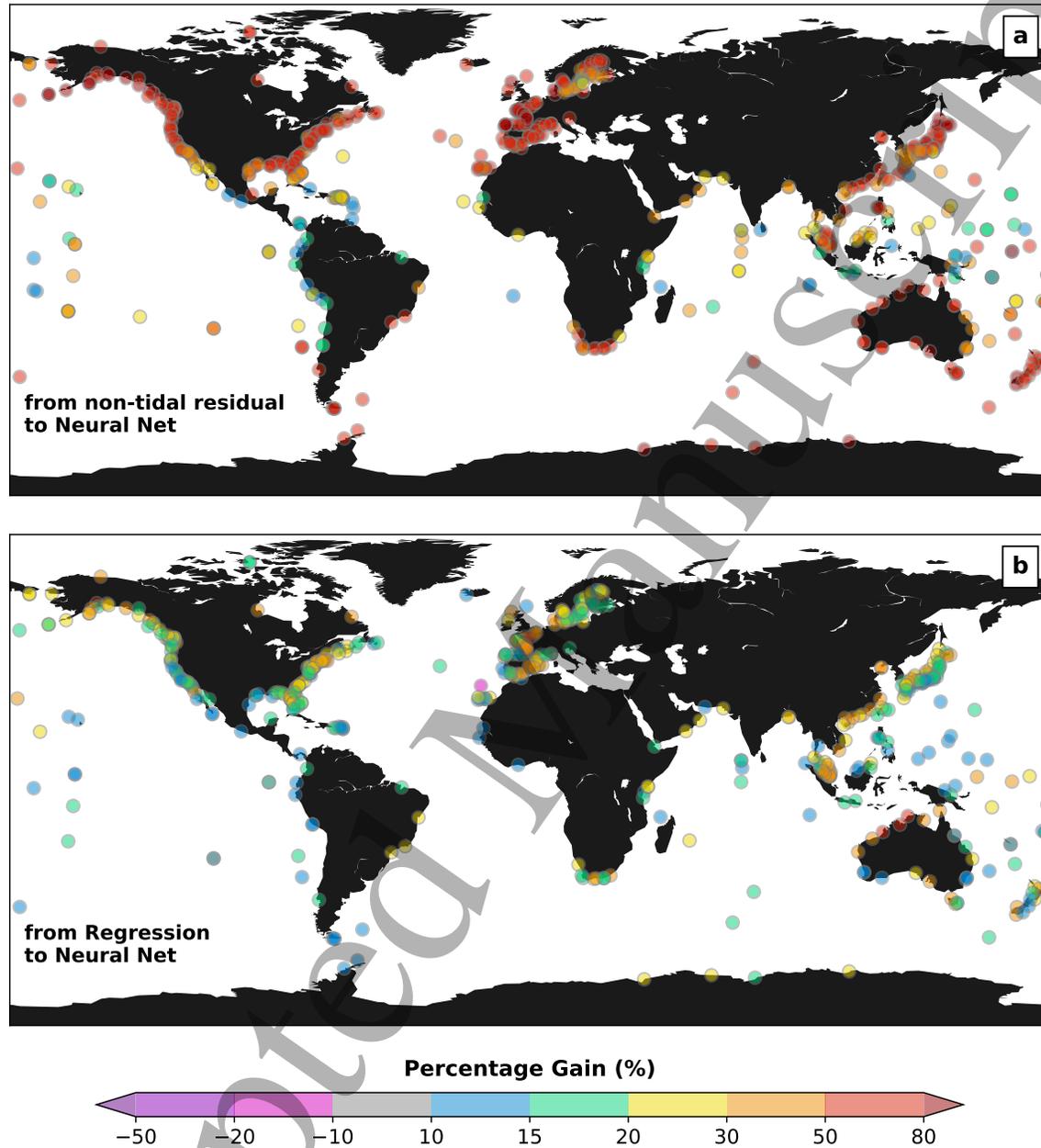


Figure 3. Percentage improvements between the different methods for the 95th percentile. a) from non-tidal residual to neural network and b) from multivariate linear regressions to neural network. The neural network enhanced systematically the improvement, particularly in regions where tides are large.

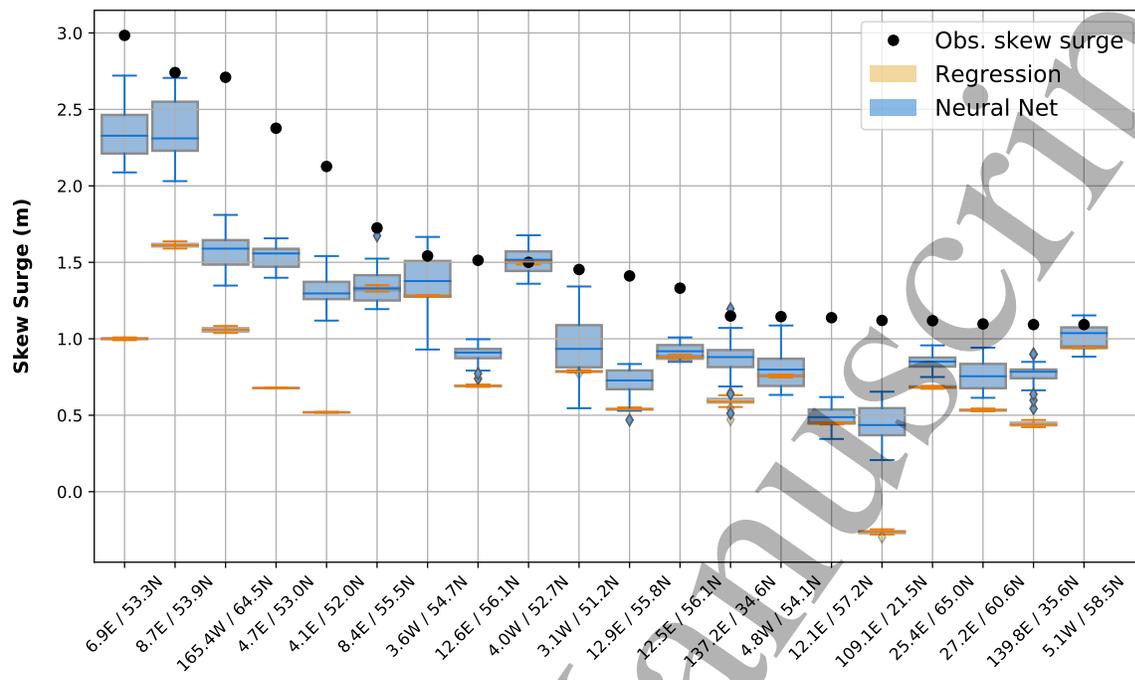


Figure 4. The 20 largest skew surges observed across all gauges in the test set, and the success of the neural net ensemble and regression at modelling them. The skew surges are computed as the difference between the highest water level and the highest harmonic tide level within a 12h window around the maximum non-tidal residual.

Estimation of Global Coastal Sea level extremes using Neural Networks

24

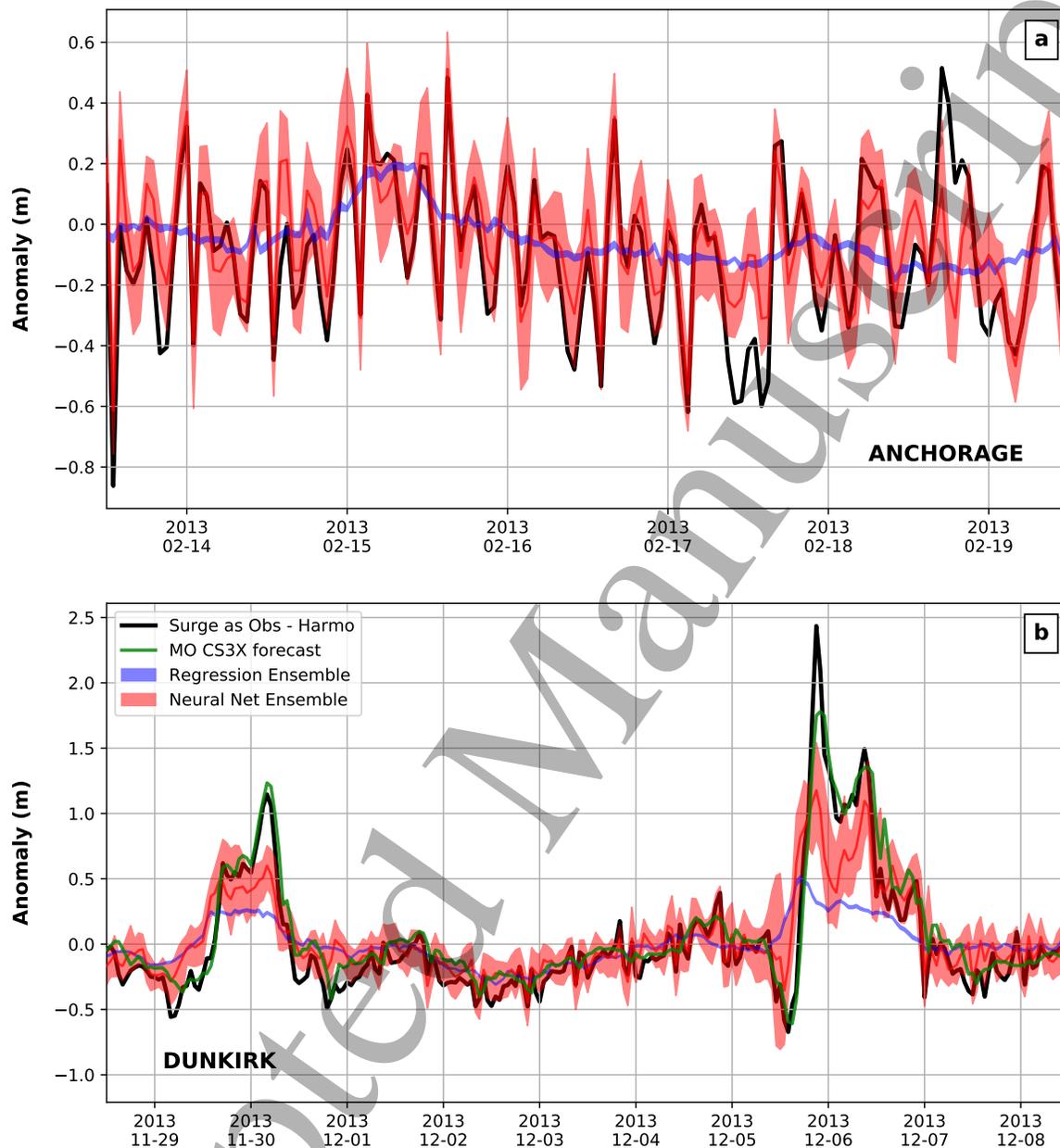


Figure 5. Example of time series extracted from the test time period in a) Anchorage, Alaska, USA where strong tidal interactions occurs and b) Dunkirk, North France during a severe winter. The thick line shows the mean of the ensemble. The neural network shows significant skills in predicting the variability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

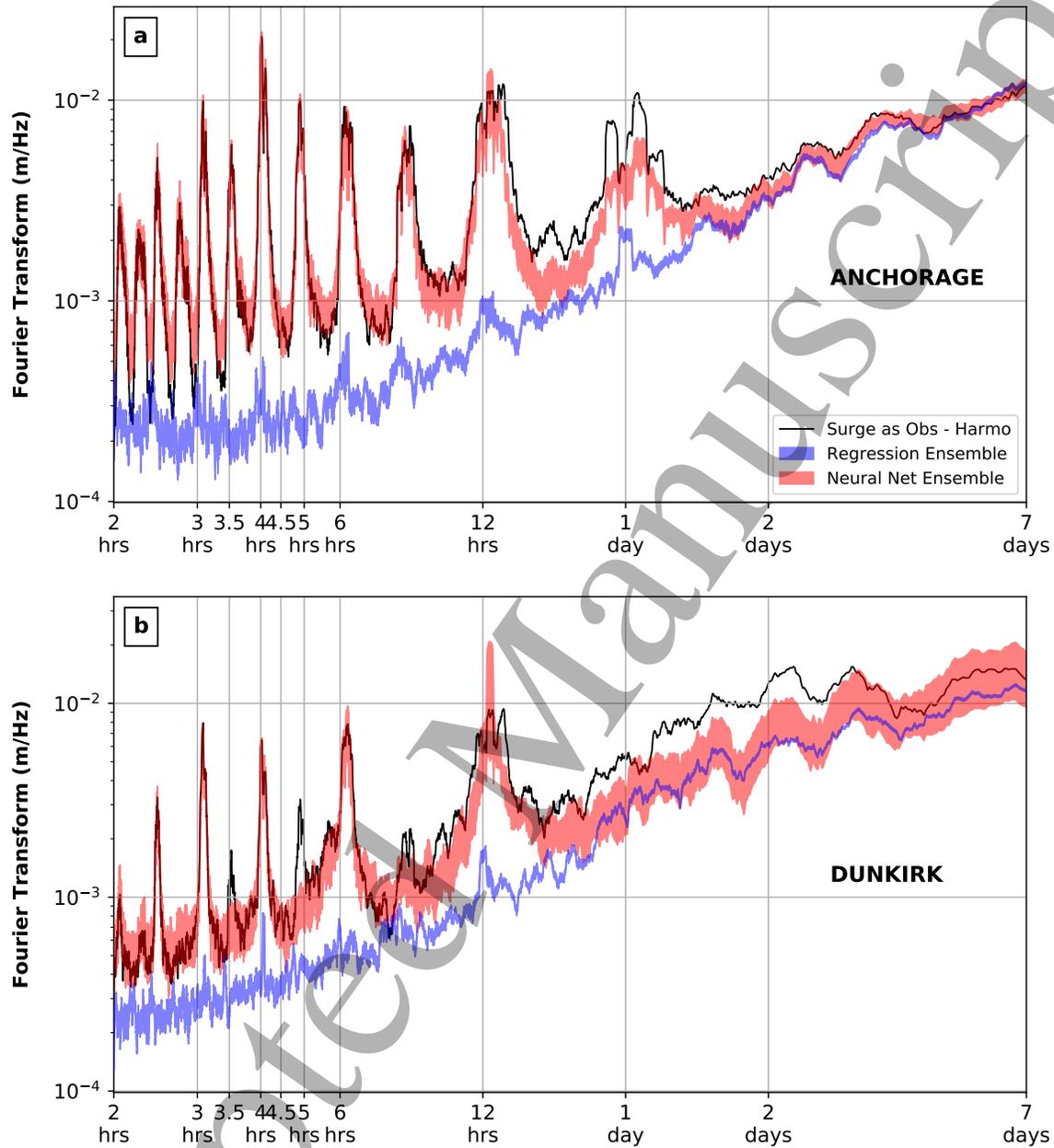


Figure 6. Energy spectrum extracted from the full test year in a) Anchorage, Alaska, USA where strong tidal interactions occurs and b) Dunkirk, North France during a severe winter. Again the neural network shows high skills at capturing the energy peaks in non-tidal residual.