A machine learning model for the estimation of hourly non-tidal water levels in the Dutch coastal zone

Based on satellite altimetry observations and pressure and wind fields from ERA5

Sofie Schijvenaars

Delft University of Technology

TU Delft

A machine learning model for the estimation of hourly non-tidal water levels in the Dutch coastal zone

Based on satellite altimetry observations and pressure and wind fields from ERA5

by

Sofie Schijvenaars

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Thursday June, 20, 2024 at 12:00.

Student number:	4659333	
Project Duration:	May, 2023 – June, 2024	
Graduation committee:	Dr. Ir. C. Slobbe	TU Delft, chair
	Dr. M. Schleiss,	TU Delft, supervisor
	Dr. Ir. B. Wouters,	TU Delft, supervisor
	Dr. M. Eleveld,	Deltares, daily supervisor
	Dr. Ir. M. Gawehn,	Deltares, supervisor
In collaboration with:	Deltares, Delft, Netherlands	

Cover:	Courtesy of Bing

Enabling Delta Life

Preface

Throughout the course of this thesis, many people have helped and motivated me. First and foremost, I would like to express what their support and contribution meant to me.

Firstly, I would like to share my greatest gratitude towards my daily supervisor, Marieke. You have been incredible. Your enthusiasm, unrelenting support and willingness to help are deeply appreciated. Secondly, my graduation committee helped me a lot throughout the entire process. Cornelis, thank you for your critical thinking and attention to detail. Your guidance has been appreciated every step of the way. Many thanks to Marc, for pointing me in the right direction to learn as much as I could about machine learning, even though you sometimes scared me with terms I had never heard before. Matthijs and Bert, thank you for asking the right questions at the right time.

Thanks to Deltares, specifically the Data Science and Water Quality department, for hosting me and welcoming me into your team. Additionally, Stendert Laan provided me with DCSM data and Jonathan Nuttall helped me with his insight into machine learning and neural networks. Thank you both, also for your interest in my thesis and the tips.

Deltares has been actively involved in the Earth Observation Advanced Science Tools for Sea Level Extreme Events (EOatSEE) project, which was funded by ESA, with the reference AO/1-10758/21/I-DT. Thanks to the EOatSEE consortium for welcoming me to some of their technical meetings. Thanks to Olivier Lauret from CLS, and to Florence Birol and Fabien Léger from LEGOS-CTOH for providing satellite altimetry data with the latest X-TRACK reprocessing.

Finally, I would like to thank my friends and family who have been standing by my side and who I know will always have my back. Thank you to my biggest supporter, who has watched his daughter grow into the person she is now. Dad, your limitless knowledge about data processing, Python skills, machine learning and so many other things have saved me multiple times. Thank you for pushing me to achieve my goals and teaching me to never give up despite the countless memory errors. Thanks to my mom, who always believed in me and supported me unconditionally. I hope you realise the heroes you represent for me in my life.

A special thanks to Bernice, for sticking with me from the beginning and for all the hearty discussions, brainstorming sessions, coffee breaks and late-night dinners. I strive to be the friend you are to me. Thank you, Esther, for brightening my day every time I see you, and to Nadine, for all the dinners you cooked for me and Bernice when we did not have time. Finally, thanks to Fabian for catching me when I fall, no matter what. I love you all.

Sofie Schijvenaars
Delft, Tuesday 12th June, 2024

Abstract

Globally, coastal communities face increasing risks from hazards such as flooding, shoreline erosion, and salt intrusion due to climate change. These hazards pose threats to both people and their environment, with extreme sea level events increasing these risks. Satellite altimetry allows for global observation of the sea level, reaching remote regions that are not covered by unevenly distributed tide gauges, as these are concentrated in densely populated regions of Western cultures. However, their 10- to 35-day repeat cycles complicate the capture of extreme sea level events. Machine learning offers a promising approach to combine direct satellite altimetry observations with ERA5 pressure and wind speed fields into a data-driven model. As opposed to global and regional numerical models, which require substantial time and expertise to develop, machine learning models are time efficient and require relatively low effort to develop and expand.

This study presents a shallow neural network that effectively estimates hourly non-tidal water levels in the Dutch coastal zone, using X-TRACK retracked and reprocessed satellite altimetry observations and ERA5 hourly pressure and wind speed fields. Reprocessed satellite altimetry observations from 11 missions are used to provide more accurate coastal observations. Tide gauge records are used as ground truth. Both tide gauge and satellite altimetry data are corrected for harmonic tidal signals before training. A 48-hour time window is applied, using all data from 48 hours to 1 hour prior to the estimates as input into the network. The area of interest covers most of the North Sea, from the Strait of Dover to the northern North Sea, excluding the Danish and Norwegian coasts. The neural network is trained and tested at three locations: Scheveningen, Vlissingen, and the Europlatform.

Results show that the neural network can estimate hourly non-tidal water levels with mean squared errors ranging from 0.011 to 0.018 m, mean absolute errors from 0.078 to 0.101 m and standard errors from 0.100 to 0.134 m. K-fold cross-validation with $K = 4$ indicates high robustness, with mean squared errors varying by 0.004 m, mean absolute errors by 0.012 m and standard errors by 0.017 m. The model performs best for hourly and high water levels at the Europlatform and worst for high water levels at Scheveningen. This is partly due to the location of the Scheveningen tide gauge in a harbour with more localised disruptions of the water level compared to the tide gauge at the Europlatform. The ERA5 longitudinal wind speed component contributes most to the estimation of non-tidal water levels, accounting for $\pm18\%$ of all weights corresponding to the input variables. Key regions for the estimation of non-tidal water levels include the Dutch coast and the northern North Sea.

When compared to a local numerical model, the developed neural network does not perform with the same accuracy. However, several upsides of the model are identified, such as high computational efficiency for single locations and easy implementation options for refinement of the model. Recommendations for future research focus mostly on improving the model's performance on high water levels and applicability to different regions.

Nomenclature

Abbreviations

Abbreviation	Definition
Adam	Adaptive Moment Estimation
ALES	Adaptive Leading Edge Subwaveform
AT	Astronomical Tide
CART	Classification And Regression Trees
C3S	Copernicus Climate Change Service
CDF	Cumulative Density Function
CDS	Climate Data Store
CMEMS	Copernicus Marine Environment Monitoring Service
CNN	Convolutional Neural Network
CRPS	Continuous Ranked Probability Score
CTOH	Centre for the Topography of Oceans and the Hydrosphere
DAC	Dynamic Atmosphere Component
DAHITI	Database for Hydrological Time Series over Inland Waters
DCSM-FM	Dutch Continental Shelf Model - Flexible Mesh
DOY	Day Of Year
ECMWF	European Centre for Medium-Range Weather Forecasts
EM	Electromagnetic
ERA5	Fifth generation of ECMWF atmospheric reanalysis of the global climate
ERS	European Remote Sensing Satellite
ESA	European Space Agency
GDR	Geophysical Data Records
GESLA	Global Extreme Sea Level Analysis
GFO	GeoSat Follow-On
GIM	GPS-derived Ionosphere Maps
GNSS	Global Navigation Satellite System
GOCE	Gravity field and steady-state Ocean Circulation Explorer
GPD	Generalised Pareto Distribution
GPS	Global Positioning System
GRACE	Gravity Recovery And Climate Experiment
GTSM	Global Tide and Surge Model
HF	High-frequency
HY-2	Haiyang-2
IB	Inverse Barometer
IGS	International GNSS Service
ISG	International Service for the Geoid
KNMI	Royal Netherlands Meteorological Institute
LEGOS	Laboratory of Space Geophysical and Oceanographic Studies
LF	Low-frequency
MAD	Median Absolute Deviation
MAE	Mean Absolute Error
MARS	Multi-Adaptive Regression Spline
ME	Mean Excess
ML	Machine Learning
MOG-2D	Two-dimensional Gravity Waves Model
MSE	Mean Squared Error

Abbreviation	Definition
MSL	Mean Sea Level
MSS	Mean Sea Surface
MSSH	Mean Sea Surface Height from XTRACK
NASA	National Aeronautics and Space Administration
NAO	North Atlantic Oscillation
NAP	Normal Amsterdam Level
NCEP	United States National Centers for Environmental Prediction
NetCDF	Network Common Data Form
NTR	Non-tidal residual
OCOG	Offset Centre Of Gravity retracker
POT	Peak Over Threshold
PRF	Pulse Repetition Frequency
PT	Pole Tide
QC	Quality-control
RA	Radar Altimeter
ReLU	Rectified Linear Unit
RWS	Netherlands Ministry of Infrastructure and Water Management
S3A	Sentinel-3A
SAR	Synthetic Aperture Radar
SLA	Sea Level Anomaly
SNR	Signal-to-noise ratio
SSA	Singular Spectrum Analysis
SSB	Sea State Bias
SSH	Sea Surface Height
ST	Solid-Earth tide
SWAN	Simulating WAves Nearshore
SWH	Significant Wave Height
TEC	Total Electron Content
TG	Tide Gauge
T/P	TOPEX / Poseidon

Symbols

Symbol	Definition	Unit
α	Learning rate	-
β_1	Exponential decay of the first moment of the gradient of the loss	-
β_2	Exponential decay of the second moment of the gradient of the loss	-
γ	Correction term for the 18.6-year tidal cycle	-
Δh	Correction term for satellite altimetry	[m]
Δh_{dry}	Try tropospheric correction	[m]
Δh_{ib}	Inverse barometer low-frequency correction	[m]
Δh_{iono}	Ionospheric correction	[m]
Δh_{ssb}	Sea surface bias correction	[m]
Δh_{wet}	Wet tropospheric correction	[m]
ϵ	Numerical stability constant for the Adam optimiser	-
η	Scale parameter of the GPD	-
λ	Geographic longitude	[degrees]
μ	Mean of a dataset	-
ξ	Shape parameter of the GPD	-
ρ_w	Water vapour in the atmosphere	[kg/m^3]
σ	Standard deviation of a dataset	-
σ_ε	Standard error	[m]
ϕ	Geodetic latitude	[degrees]
ω_i	Angular velocity of tidal component i	[rad/hr]
A_i	Amplitude of tidal component i	[m]
a_0	Geophysical correction parameter	-
a_1	Geophysical correction parameter	-
a_2	Geophysical correction parameter	-
a_3	Geophysical correction parameter	-
a_4	Geophysical correction parameter	-
b_{nj}, b_y	Neural network bias term	[m]
c	Speed of the radar pulse	[m/s]
d	Haversine distance	[m]
d_E	Longitudinal component of the distance between a TG and a satellite altimetry observation	[m]
d_N	Lateral component of the distance between a TG and a satellite altimetry observation	[m]
dt	Time difference between the ML output and a satellite altimetry observation	[s]
Dwf	Vector that holds the normalised differences between consecutive gates	-
f	Radar frequency of the altimeter	[GHz]
g_i	Phase of tidal component i	[rad]
H	Orbit height of the satellite	[m]
k	Constant within ionospheric correction term	[mGHz^2/TECU]
$L(w,b)$	Loss function	-
m	Number of input features	-
m_1	First moment of the gradient of the loss	-
m_2	Second moment of the gradient of the loss	-
N	Number of tidal constituents	-
n_1, n_j	Neuron $j = 1...32$	-
P_0	Sea surface pressure	[hPa]
P_{ref}	Reference pressure	[hPa]
P_s	Total atmospheric pressure	[hPa]
p	ERA5 sea surface pressure	[Pa]
q	Constant to repeat high water levels during resampling	-

Symbol	Definition	Unit
R	Radius of the Earth	[m]
$R_{corrected}$	Corrected range	[m]
R_{obs}	Observed range	[m]
r	The current batch	-
s	Batch size	-
T	Threshold defined by the POT method	[m]
TECU	Total Electron Content Unit	[10^6 electrons/m^2]
t	Two-way travel time of the mid-point of the leading edge	[s]
U	Wind speed derived from backscatter coefficient	[m/s]
$U10$	ERA5 longitudinal wind speed	[m/s]
u	Correction term for the 18.6-year tidal cycle	-
$V10$	ERA5 lateral wind speed	[m/s]
v_0	Phase of the equilibrium tide at $t = 0$	[rad]
w_{ij}	Neural network weight term for input feature i and neuron j	-
x'	Min-max normalised neural network input feature	-
$\bar{x}$	Sample mean of the performance metrics	-
x_i	Neural network input feature	-
$\hat{y}$	Estimated water level	[m]
y	Ground truth water level	[m]
z_i	Z-score normalised neural network input feature	-
z_s	Height of the surface with respect to the geoid	[m]

List of Figures

1.1 Global map of tide gauge locations (Haigh et al., 2022) 2
1.2 An overview of (a) the GTSM grid points (Muis et al., 2020) and (b) the DCSM grid points (Zijl, Groeneboom, et al., 2022). In (a), red points denote ocean points and blue points denote coastal points. In (b), the colours indicate the grid size (yellow: ±4 nautical mile (nm); light green: ±2 nm; blue: ±1 nm; red: ±0.5 nm, cyan: ±0.25 nm, dark green: ±200 m and orange ±100 m) 3
1.3 A time series of non-tidal residuals (NTR) from a tide gauge in Eastport, USA. The black dots represent monthly maxima from the tide gauge and orange crosses refer to monthly maxima from satellite altimetry (Lobeto et al., 2018). 5

2.1 Forces that cause extreme coastal water levels (Mullan et al., 2005). 7
2.2 Propagation of the M2 tide in the North Sea with co-tidal lines radiating away from the amphidromic points and co-range lines encircling them. The co-tidal lines show that the phase increases counter-clockwise around the amphidromic point (typical of NH amphidromes). The co-range lines show the tidal range increasing away from the node (Bosboom & Stive, 2021). 8
2.3 A timeline of modern radar altimeters from the 1990s to the next decade, including expected or announced mission extensions beyond the nominal satellite lifespan. Past missions are in red (Aviso+, 2024). 10
2.4 The reflected power of the satellite altimeter (a) when approaching the coast and (b) over homogeneous ocean surface. The red line in (a) shows land contamination, meaning that the signal does not conform to the Brown model anymore (COASTALT, 2023). The named parts of the signal shown in (b) correspond to different geophysical parameters associated with the water level (Idris et al., 2014). 11
2.5 Theoretical Brown ocean waveform shape and related ocean parameters (Gommenginger et al., 2011). 13
2.6 A pulse-limited altimeter in a coastal region. The lower panel shows a nadir view of the pulse-limited footprint corresponding to each waveform gate. The variable H corresponds to the range between the satellite and the ocean's surface, B corresponds to the bandwidth of the altimeter, and c is the speed of the radar pulse. The variables τ_0, τ_1 and τ_2 refer to the pulse emission time and received time delays respectively. For this altimeter, the onboard tracker has determined the altimeter analysis window to be between gates 1 and 128 and has identified gate 44 as the tracking gate (Gommenginger et al., 2011). 14
2.7 Example of Jason 2 track 9 in the Mediterranean Sea. Shown are (a) the location of the observations and (b) the standard deviation of the corresponding SLA time series (in metres) before (in red) and after (in blue) the X-TRACK procedure (Birol et al., 2021). 15
2.8 Schematic overview of the design of the shallow neural network developed in this study. The blue boxes show the input variables, design parameters and output values, while the red boxes show the trainable parameters. $[x_1 \quad x_2 \quad \cdots \quad x_m]$ denote the input variables, $[n_1 \quad n_2 \quad \cdots \quad n_j]$ refer to the neurons within the hidden layer, and $\hat{y}$ shows the output value. The matrices and vectors of w and b refer to the weights and biases that are estimated during the training of the model. 16
2.9 Examples of (a) an underfit, (b) a good fit and (c) an overfit. The dots represent the training data and the line represents the machine learning model output (Shah, 2023). 17
2.10 Visualisation of the ReLU activation function. 17

3.1 Extent of the area of interest. The red line denotes the edge of the area of interest. Any data within the IJsselmeer has been discarded, since this is considered inland water. The selected tide gauges on which the model will be trained are shown in black. 19

3.2 Locations of the (a) Vlissingen and (b) Scheveningen tide gauge, taken from Google Earth imagery. . . . 20
3.3 Timeline of satellite altimeters since 1992 which are used in this thesis (updated from Shu et al., 2021) . . . 22
3.4 Numbered tracks of (a) GeoSat Follow-On, (b) Sentinel-3A, (c) ERS-2, Envisat and SARAL, (d) TOPEX/Poseidon and Jason-1/-2/-3, (e) TOPEX/Poseidon and Jason-1/-2 interleaved orbit and (f) Haiyang-2. Taken from Aviso+ (2023). . . . 23
3.5 Illustration of a nominal track (red line and black stars), with their reference cycles (one colour for each cycle). For each shown cell, the nominal track is defined by the mean of the coordinates/MSSH of the reference cycles (Birol & Passaro, 2020). . . . 24
3.6 Representation of the EGG2015 geoid (Denker, 2013) . . . 25

4.1 The tracks available within the area of interest. The outer red shape marks the edge of the area in which all data is gathered, with the exception of the IJselmeer. The black lines visualise all tracks available within the area of interest. Each track has its own file. . . . 27
4.2 Examples of an ERA5 dataset on 2018-01-01 00:00:00, for (a) sea level pressure, (b) longitudinal 10-m wind speed component and (c) lateral 10-m wind speed component. The data is clipped to the area of interest and masked with a land-sea mask that masks all pixels which have a mask larger than 0.5. The outer red shape marks the edge of the area in which all gridded data is gathered, with the exception of the IJsselmeer. . . . 28
4.3 Water level measurements from tide gauges at Vlissingen, Scheveningen and Europlatform from 2017/01/01 to 2017/01/15. The dots represent the hourly values which are used as ground truth data in the machine learning model. . . . 29
4.4 Water level measurements from tide gauges at Vlissingen, Scheveningen and Europlatform from 2017/01/01 to 2017/01/15, after applying the Hatyan method to correct for harmonic tidal signals. The values are hourly. . . . 30
4.5 Step-wise overview of the method used to prepare the input data for developed machine learning model. . . . 30
4.6 Data availability for the full period of interest. The upper graph shows the number of observations within the 48-hour subsets. The lower graph shows the period at which altimetry data from different missions is available. HY-2 refers to Haiyang-2, S3A to Sentinel-3A, GFO to GeoSat Follow-On, TPN to Topex/Poseidon and Jason-1/-2 on an interleaved track, ERS to ERS-1/-2, Envisat and SARAL and TP to Topex/Poseidon and Jason-1/-2/-3. The black dashed lines refer to the start/end of missions and the red dotted lines refer to the four validation datasets for K-fold cross-validation and the testing dataset. . . . 31
4.7 Example of padding satellite altimetry, where stars (*) refer to satellite observations and one row corresponds to one hour. The hour with the largest number of observations is taken (fourth row), after which all other hours are padded with zeros until every hour has the same shape, including hours which have no observations (second row). In reality, this matrix is much larger for each 48-hour subset (48 rows and 514 columns). . . . 31
4.8 Overview of how the input data is split into testing, validation and training data for training and K-fold cross-validation. The folds are shown as orange boxes (validation data) or grey boxes (testing data). . . . 32
4.9 Observed water level with (a) no declustering and (b) declustering with a time window of 48 hours. On the y-axis, the water level observed by the tide gauge is shown, referenced to NAP. The dotted line refers to the threshold and the red dots to the selected peaks that exceed the threshold. . . . 35

5.1 Time series reconstruction for the Scheveningen model. The shaded areas in (a) correspond to the months shown in (b), (c), (d), (e), (f) and (g). The green lines represent the TG observations, the blue lines refer to the ML model output. The shaded areas in the sub-graphs show severe storms catalogued by the KNMI (KNMI, 2024). The y-axis unit is in metres and depicts the non-tidal water level referenced to NAP. . . . 38

5.2 Time series reconstruction for the Vlissingen model. The shaded areas in (a) correspond to the months shown in (b), (c), (d), (e), (f) and (g). The green lines represent the TG observations, the blue lines refer to the ML model output. The shaded areas in the sub-graphs show storms catalogued by the KNMI (KNMI, 2024). The y-axis unit is in metres and depicts the non-tidal water level referenced to NAP. 40
5.3 Time series reconstruction for the Europlatform model. The shaded areas in (a) correspond to the months shown in (b), (c), (d), (e), (f) and (g). The green lines represent the TG observations, the blue lines refer to the ML model output. The shaded areas in the sub-graphs show storms catalogued by the KNMI (KNMI, 2024). The y-axis unit is in metres and depicts the non-tidal water level referenced to NAP. 41
5.4 Histograms of (a) sea level pressure, (b) longitudinal wind speed and (c) lateral wind speed for (left) a good performance sample on 2015-07-25 at 12:00, (middle) a constant output sample on 2016-12-31 at 00:00 and (right) a good performance sample on 2018-01-03 at 17:00. The y-axis corresponds with the density of each bin. 43
5.5 Histograms of (a) non-tidal residual, (b) longitudinal distance and (c) lateral distance for (left) a good performance sample on 2015-07-25 at 12:00, (middle) a constant output sample on 2016-12-31 at 00:00 and (right) a good performance sample on 2018-01-03 at 17:00. The y-axis corresponds with the number of observations in each bin. 44
5.6 Selected high water levels from tide gauge data for (a) Scheveningen, (b) Vlissingen and (c) Europlatform. Thresholds are shown as red dashed lines and high water levels as red dots. A declustering window of 48 hours is used. 46
5.7 Selected high water levels from tide gauge data for (a) validation dataset 1, (b) validation dataset 2, (c) validation dataset 3 and (d) validation dataset 4. The threshold is shown as a red dashed line and high water levels as red dots. A declustering window of 48 hours is used. 47
5.8 Observed high water levels (x-axis) vs their corresponding estimated water levels (y-axis) for (a) the machine learning model and (b) DCSM. The blue dots refer to the Vlissingen model, the orange dots to the Scheveningen model and the green dots to the Europlatform model. . 49
5.9 Second-layer weight $[w_{n1} \quad w_{n2} \quad \cdots \quad w_{n32}]$ and first-layer bias $[b_{n1} \quad b_{n2} \quad \cdots \quad b_{n32}]$ graphs for the (a, b) Scheveningen, (c, d) Vlissingen and (e, f) Europlatform model. Blue dots refer to the weights and orange dots to the biases. The red dots show the ones where the weight is so close to zero that their inputs are nullified. 50
5.10 First-layer weights summed over the 48-hour time window for the Scheveningen model for (a) the non-tidal water level, (b) the longitudinal component of the distance d_E, (c) the lateral component of the distance d_N and (d) the time difference between observation and tide gauge dt. Each coloured line corresponds to a neuron. The black lines show the mean of all 32 lines. 51
5.11 First-layer mean weights of the non-tidal water level observations from satellite altimetry. The first and last hour of the 48-hour time window is shown for (a, d) Scheveningen, (b, e) Vlissingen and (c, f) Europlatform. High positive weights are shown in red, high negative values in blue and weights close to zero in yellow. 52
5.12 First-layer mean weights of the longitudinal distance between the satellite altimetry observations and the tide gauge. Only the first and last hour of the 48-hour time window is shown for (a, d) Scheveningen, (b, e) Vlissingen and (c, f) Europlatform. High positive weights are shown in red, high negative values in blue and weights close to zero in yellow. 52
5.13 First-layer mean weights of the lateral distance between the satellite altimetry observations and the tide gauge. The first and last hour of the 48-hour time window is shown for (a, d) Scheveningen, (b, e) Vlissingen and (c, f) Europlatform. High positive weights are shown in red, high negative values in blue and weights close to zero in yellow. 53
5.14 First-layer mean weights of ERA5 sea surface pressure. The first and last hours of the 48-hour time window is shown for (a, d) Scheveningen, (b, e) Vlissingen and (c, f) Europlatform. High positive weights are shown in red, high negative values in blue and weights close to zero in yellow. 54
5.15 First-layer mean weights of ERA5 longitudinal wind speed. The first and last hours of the 48-hour time window is shown for (a, d) Scheveningen, (b, e) Vlissingen and (c, f) Europlatform. High positive weights are shown in red, high negative values in blue and weights close to zero in yellow. 55

5.16 First-layer mean weights of ERA5 lateral wind speed. The first and last hours of the 48-hour time window is shown for (a, d) Scheveningen, (b, e) Vlissingen and (c, f) Europlatform. High positive weights are shown in red, high negative values in blue and weights close to zero in yellow. 55

A.1 An example of a Jason-1 track crossing the western Mediterranean Sea. Blue dots indicate the footprint of the altimeter and green circles show the extent of the main beam of the radiometer. The figure illustrates where the radiometer observations are contaminated by land (Andersen & Scharroo, 2010). 76

B.1 Flow diagram of ALES retracking algorithm for each waveform (Passaro et al., 2014). 78
B.2 Examples of real recorded waveforms with their ALES retracked waveforms for (left) open ocean with SWH = 0.75 m, (middle) coastal ocean with corrupted trailing edge and SWH = 1.65 m and (right) open ocean with SWH = 9.448 m (Passaro et al., 2014). 79
B.3 Wet tropospheric correction for TOPEX/Poseidon pass 137, cycle 200. Black represents the raw correction from the GDR. Red represents the valid correction points after the second filter. Purple represents the data after interpolation. The x-axis represents the time on 1998-02-22. The y-axis corresponds to the wet tropospheric correction in metres (Roblou et al., 2011). . 80
B.4 Examples of (a) ionospheric correction for Jason-1, track 222, cycle 10 and (b) wet tropospheric correction for cycle 8 in the Mediterranean Sea. The yellow circles indicate the corrections before and the red line indicates the corrections after X-TRACK processing (Birol et al., 2017). 81

D.1 Amplitude and phase of all considered tidal constituents for Scheveningen. Amplitude is given in metres and phase in degrees. 87

E.1 Distribution of the day of the year 88
E.2 Distribution of the non-tidal residuals for (a) Scheveningen, (b) Vlissingen and (c) Europlatform. 89
E.3 Distribution of (a) non-tidal residuals for satellite altimetry observations (b) the time difference between each satellite altimetry observation and the time stamp of interest. Note that each satellite altimetry observation appears 48 times in 48 successive input datasets. 90
E.4 Distribution of the distance between satellite altimetry observations and the tide gauge of interest in longitudinal direction for (a) Scheveningen, (b) Vlissingen and (c) Europlatform. While the distribution is the same, the three histograms are slightly shifted with respect to one another, depending on the longitude of the tide gauge. 91
E.5 Distribution of distance between satellite altimetry observations and the tide gauge of interest in lateral direction for (a) Scheveningen, (b) Vlissingen and (c) Europlatform. While the distribution is the same, the three histograms are slightly shifted with respect to one another, depending on the latitude of the tide gauge. 92
E.6 Distribution of ERA5 sea surface pressure, wind speed in longitudinal direction and wind speed in lateral direction. 93

F.1 Mean excess plots for (a) Scheveningen, (b) Vlissingen and (c) Europlatform. The black lines indicate where the mean excess plot is approximately linear. The shaded area refers to the 95% confidence interval. 95
F.2 Parameter stability plots of the shape ξ and scale η parameters of the GPD for (a) Scheveningen, (b) Vlissingen and (c) Europlatform. The blue shaded area refers to the 95% confidence interval and the grey shaded area to the threshold range defined with the ME plots. The black line refers to the selected thresholds for each location. 96

G.1 Time-series reconstruction for validation set 1 of the Scheveningen model. The shaded areas in (a) correspond to the months shown in (b), (c), (d) and (e). The green lines represent the tide gauge observations, the blue lines refer the machine learning model output. The shaded areas in the sub-graphs show storms catalogued by the KNMI (KNMI, 2024). The unit of the y-axis is in metres and depicts the non-tidal water level referenced to NAP. 97

G.2 Time-series reconstruction for validation set 2 of the Scheveningen model. The shaded areas in (a) correspond to the months shown in (b), (c), (d) and (e). The green lines represent the tide gauge observations, the blue lines refer the machine learning model output. The shaded areas in the sub-graphs show storms catalogued by the KNMI (KNMI, 2024). The unit of the y-axis is in metres and depicts the non-tidal water level referenced to NAP. 98
G.3 Time-series reconstruction for validation set 3 of the Scheveningen model. The shaded areas in (a) correspond to the months shown in (b), (c), (d) and (e). The green lines represent the tide gauge observations, the blue lines refer the machine learning model output. The shaded areas in the sub-graphs show storms catalogued by the KNMI (KNMI, 2024). The unit of the y-axis is in metres and depicts the non-tidal water level referenced to NAP. 99
G.4 Time-series reconstruction for validation set 4 of the Scheveningen model. The shaded areas in (a) correspond to the months shown in (b), (c), (d) and (e). The green lines represent the tide gauge observations, the blue lines refer the machine learning model output. The shaded areas in the sub-graphs show storms catalogued by the KNMI (KNMI, 2024). The unit of the y-axis is in metres and depicts the non-tidal water level referenced to NAP. 100

H.1 First-layer weights summed over the 48-hour time window for the Vlissingen model for (a) the NTR, (b) the longitudinal component of the distance d_E, (c) the lateral component of the distance d_N and (d) the time difference between observation and tide gauge dt. 101
H.2 First-layer weights summed over the 48-hour time window for the Europlatform model for (a) the NTR, (b) the longitudinal component of the distance d_E, (c) the lateral component of the distance d_N and (d) the time difference between observation and tide gauge dt. 102
H.3 Difference between hour -48 and -1 of satellite altimetry first-layer weights corresponding to (a) non-tidal water level, (b) longitudinal distance and (c) lateral distance for Scheveningen. 102
H.4 Difference between hour -48 and -1 of satellite altimetry first-layer weights corresponding to (a) non-tidal water level, (b) longitudinal distance and (c) lateral distance for Vlissingen. . . 102
H.5 Difference between hour -48 and -1 of satellite altimetry first-layer weights corresponding to (a) non-tidal water level, (b) longitudinal distance and (c) lateral distance for Europlatform. 103
H.6 Difference between hour -48 and -1 of ERA5 first-layer weights corresponding to (a) sea surface pressure, (b) longitudinal wind speed and (c) lateral wind speed for Scheveningen. . 103
H.7 Difference between hour -48 and -1 of ERA5 first-layer weights corresponding to (a) sea surface pressure, (b) longitudinal wind speed and (c) lateral wind speed for Vlissingen. . . . 103
H.8 Difference between hour -48 and -1 of ERA5 first-layer weights corresponding to (a) sea surface pressure, (b) longitudinal wind speed and (c) lateral wind speed for Europlatform. . 104

List of Tables

2.1 Principal tidal constituents with equilibrium amplitudes from Apel (1987) 12

3.1 Data sources 20
3.2 Most important metadata variables for the tide gauges of interest. The column on the left denotes the variable names as they are provided in the record files (Haigh et al., 2022). . . . 21
3.3 Overview of satellite altimeters provided (Shu et al., 2021; CEOS and ESA, 2024) 22
3.4 Parameters provided in X-TRACK NetCDF files (Aviso+, 2022) 24
3.5 Parameters provided in ERA5 NetCDF files (Hersbach et al., 2018) 25

4.1 Python packages and their version used in this methodology 26
4.2 Final satellite altimetry variables suited for machine learning. 28
4.3 Final ERA5 variables suited for machine learning. 29
4.4 Input variables needed for machine learning. 30
4.5 Group size per data source. 32
4.6 Validation subsets used for the K-fold cross-validation. 32
4.7 Design parameters for the shallow neural network 34
4.8 Selected thresholds for each location. The thresholds are selected with a POT method to assess the model performance on high water levels. 35

5.1 Severe storms as registered by the Koninklijk Nederlands Meteorologisch Instituut (KNMI), flagged as a storm if one of the KNMI weather stations spread throughout the Netherlands records a wind speed higher than 24.5 m/s (KNMI, 2024). Shaded in grey are the storms that occurred during the ML model testing period. The highest wind speed recorded is averaged over one hour. The highest wind gust is instantaneous. The location refers to where these wind speeds and gusts were recorded. The last column indicates where the storm was coming from. 37
5.2 Error between machine learning estimates and tide gauge non-tidal water levels for severe storms as registered by the KNMI (KNMI, 2024) for the testing period. The last column indicates where the storm was coming from. Errors are given in metres and show the TG observations minus the ML estimates. 39
5.3 Statistics on the constants that each model converges towards for the testing dataset. Given are the constants in metres, the total duration where the model output is equal to the constant in hours, the interval of the TG observations in which 95% of the constant output appears in metres and the percentage of constant outputs within this interval. 42
5.4 Model performance metrics for each location. All metrics are based on data that the model has not used in the training stage. Shown are the mean squared error (MSE), the signal-to-noise ratio (SNR), the mean absolute error (MAE) and the standard error (σ_ε), also known as the standard deviation of the error. Units of MSE, MAE and σ_ε are in metres. The ML model metrics are shown above the DCSM metrics, followed by the K-fold cross-validation metrics. 45
5.5 Sample mean ($\bar{x}$) and range of the performance metrics for the Scheveningen model. 45
5.6 Model performance metrics for each location for high water levels, including the harmonic tidal signals. Shown are the mean squared error (MSE), the signal-to-noise ratio (SNR), the mean absolute error (MAE) and σ_ε. Units of MSE, MAE and σ_ε are in metres. The ML model metrics are shown above the DCSM metrics, followed by the K-fold cross-validation metrics. 48
5.7 Sample mean ($\bar{x}$) and range of the performance metrics for the Scheveningen model for high water levels. 48

5.8 Weight contributions of input variables in percentages. The weights are summed over the 48-hour time window and all active neurons for every model. 56
5.9 Computation time for one year for the ML model developed in this study vs DCSM-FM 100m and DCSM-FM 0.5nm. Computation times are shown in seconds. Computational power is shown in the number of available cores. The time step refers to the time step of the estimates. 56

A.1 The four parameters for the BM4 model for eight different altimeters. The a_0 parameter is added to ensure that $\Delta h_{\mathrm{ssb}} = 0$ when SWH = 0 (Scharroo & Lillibridge, 2005) 77

B.1 X-TRACK thresholds for detecting erroneous data in corrective terms (Roblou et al., 2011). 80

C.1 Technical Details of ERS-1 and ERS-2 Radar Altimeters (RA-1) (ESA, 2023a) 82
C.2 Technical Details of Envisat Radar Altimeter 2 (RA-2) (Resti et al., 1999) 83
C.3 Technical Details of SARAL radar altimeter (AltiKa) (ESA, 2023a) 83
C.4 Technical Details of GeoSat Follow-On radar altimeter (GFO-RA) (Walker et al., 1993) . . . 83
C.5 Technical Details of HY-2 Radar Altimeters (RA/HY-2) in orbit (Dong et al., 2004) 84
C.6 Technical Details of Sentinel-3A Radar Altimeter (SRAL) (ESA, 2023a) 84
C.7 Technical Details of TOPEX/Poseidon Radar Altimeter (NRA) (Fu et al., 1994) 85
C.8 Technical Details of Jason Radar Altimeters (Poseidon-2/-3/-3B) 85

D.1 Constituents used in the tidal analysis. A_0 is not a tidal constituent, but the mean of the time series (Veenstra & Kerkhoven, 2020). 86
D.2 Components with significant amplitudes according to `Hatyan` with their respective period in hours and angular velocity in radians per hour (Veenstra & Kerkhoven, 2020). The components with the largest amplitudes are shown in grey. 87

Contents

Preface i

Abstract ii

Nomenclature iii

List of Figures xi

List of Tables xiii

1 Introduction 1
1.1 Research motivation . . . 1
1.2 Related work . . . 2
1.2.1 Numerical models . . . 2
1.2.2 Machine learning . . . 3
1.2.3 Satellite altimetry . . . 4
1.3 Research objectives . . . 5
1.4 Thesis outline . . . 6

2 Background 7
2.1 Definition of sea surface height . . . 7
2.1.1 Astronomical tide . . . 8
2.1.2 Dynamic atmosphere correction . . . 9
2.1.3 Sea level anomaly . . . 9
2.2 Satellite altimetry . . . 9
2.2.1 Derivation of sea surface height . . . 11
2.3 Tidal and atmospheric components . . . 11
2.4 Retracking methods . . . 13
2.4.1 X-TRACK/ALES post-processing software . . . 14
2.5 Machine learning . . . 15
2.5.1 Neural networks . . . 15
2.5.2 Design parameters . . . 16

3 Study area and data overview 19
3.1 Dutch Coast and North East Atlantic Ocean . . . 19
3.2 Tide gauge records . . . 20
3.3 Satellite altimetry . . . 21
3.4 ERA5 . . . 24
3.5 Geoid . . . 25

4 Method 26
4.1 General approach . . . 26
4.2 Data pre-processing . . . 26
4.2.1 Satellite altimetry . . . 27
4.2.2 ERA5 . . . 28
4.2.3 Tide gauge . . . 29
4.3 Input preparation . . . 30
4.3.1 Clipping time series . . . 31
4.3.2 Padding satellite altimetry . . . 31
4.3.3 Grouping . . . 32
4.3.4 Splitting into training and validation sets . . . 32
4.3.5 Normalising data . . . 32
4.3.6 Resampling . . . 33

4.4 Neural network design . . . 34
4.5 Performance analysis . . . 34
4.5.1 Performance on high water levels . . . 35
4.5.2 Weight analysis . . . 36

5 Results **37**
5.1 Time series reconstruction . . . 37
5.1.1 Scheveningen . . . 38
5.1.2 Vlissingen . . . 40
5.1.3 Europlatform . . . 41
5.1.4 Scenario examples . . . 42
5.2 Model performance . . . 44
5.3 Estimation of high water levels . . . 45
5.4 Weight analysis . . . 49
5.4.1 Satellite altimetry features . . . 50
5.4.2 ERA5 features . . . 54
5.4.3 Weight importance . . . 56
5.5 Computation time . . . 56

6 Discussion **58**
6.1 Potential of machine learning . . . 58
6.2 Assumptions and uncertainties . . . 59
6.3 Estimation of coastal water levels . . . 60
6.4 Estimation of high water levels . . . 60
6.5 K-fold cross-validation results . . . 61
6.6 Weight analysis . . . 61
6.6.1 Satellite altimetry . . . 62
6.6.2 ERA5 . . . 62
6.6.3 Regular events over high water level events . . . 63

7 Conclusion and outlook **64**
7.1 Conclusion . . . 64
7.2 Outlook . . . 65
7.2.1 Estimation of high water levels . . . 66
7.2.2 Applicability to different regions . . . 67

References **68**

A Satellite altimetry corrections **75**

B Adaptive Leading-Edge Subwaveform retracker **78**

C Additional information on satellite altimetry missions **82**

D Hatyan tidal analysis **86**

E Input variables **88**

F POT threshold selection **94**

G K-fold reconstructions **97**

H Weight analysis **101**
H.1 Satellite altimetry . . . 101
H.2 ERA5 . . . 103

1
Introduction

1.1. Research motivation

Coastal communities worldwide face severe challenges from coastal hazards, including flooding, salinisation of groundwater, inundation of low-lying coastal regions and increased erosion, all expected to rise in the coming decades due to climate change (Oppenheimer et al., 2019). As the coastal population grows, the risks of economic damage and casualties increase (Nicholls & Cazenave, 2010). Even in the most optimistic of mitigation scenarios, communities remain increasingly vulnerable to the impacts of coastal hazards (Nicholls & Tol, 2006). Studying these coastal hazards allows us to understand and assess the susceptibility of coastal regions to these hazards and quantify the risks in order to develop optimal strategies to increase the resilience of coastal communities.

Many of the coastal hazards that threaten communities and ecosystems around the world are linked to extreme sea level events (Almar et al., 2021; Dullaart et al., 2021; Hanson and Larson, 2008; Jongman et al., 2012; Needham et al., 2015). These events cause coastal flooding, erosion, damage to infrastructure and forced displacement of populations (Almar et al., 2021; Hallegatte et al., 2011; Nicholls, 2011; Parise et al., 2009). This highlights the fact that good monitoring and prediction of extreme sea level events is important for the safety of coastal populations and environmental habitats. Their importance is further emphasised by the consequences of climate change. Sea level rise and global warming contribute significantly to the intensification and increasing frequency of sea level extremes (Church et al., 2006; Marcos et al., 2015), accelerating the need to have accurate prediction models for the implementation of mitigation techniques to increase protection along vulnerable coasts. Reliable, accurate and precise estimates of extreme sea levels are vital for understanding trends and designing adequate protection strategies.

Tide gauges (TG) accurately record water levels, making them ideal for extreme value analysis due to their suitable measurement frequency and their long observation periods (Ji & Li, 2020). However, tide gauges have an inhomogeneous spatial distribution over the world's coasts (Figure 1.1), limiting their usefulness in sparsely populated coastal regions (Adebisi et al., 2021).

It has been found that large numerical models such as the Global Tide and Surge Model (GTSM) (Muis et al., 2020) or the Dutch Continental Shelf Model (DCSM) (Zijl, Groeneboom, et al., 2022) provide accurate estimates of water levels. Nevertheless, constructing such numerical models demands a substantial investment of time, often spanning several years, and the involvement of numerous experts who have specialised knowledge in Earth system physics and hydrodynamics. Given the global extent of satellite data and the fast-paced development of machine learning (ML), we can explore the potential to estimate these water levels effectively by combining these two technologies. Satellite altimetry has been used abundantly to tackle the limitations of tide gauges as a data source for the estimation of high water levels (Andersen et al., 2015; Izaguirre et al., 2011; Ji et al., 2019; Lobeto et al., 2018). Furthermore, ML methods have proven to be useful in combination with satellite altimetry (Gharineiat and Deng, 2015; Passaro and Juhl, 2023) and numerical models (Den Bieman et al., 2023; Hieronymus and Hieronymus, 2023; Xie et al., 2023). However, ML has yet to be explored for the specific purpose of estimating coastal water levels with satellite altimetry. Especially at locations with complicated coastal processes, such methods could improve large-scale numerical solutions.

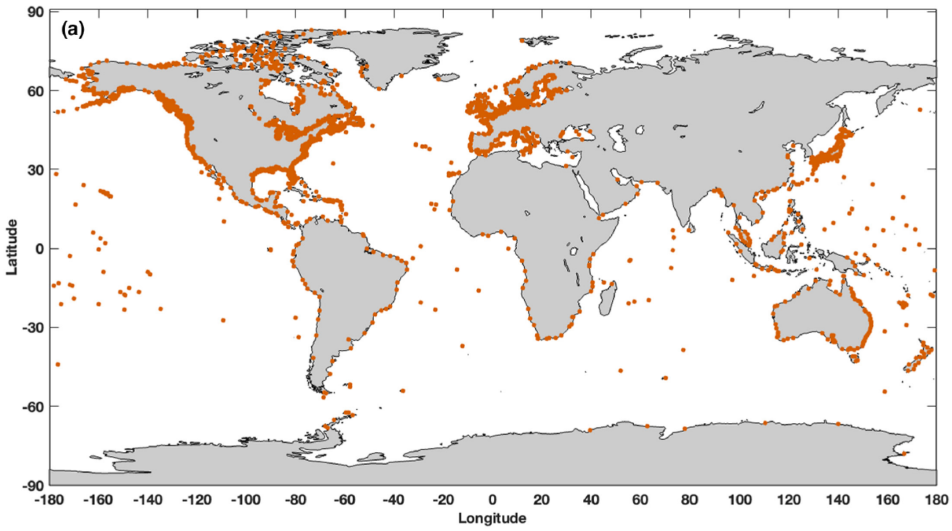


Figure 1.1: Global map of tide gauge locations (Haigh et al., 2022)

1.2. Related work

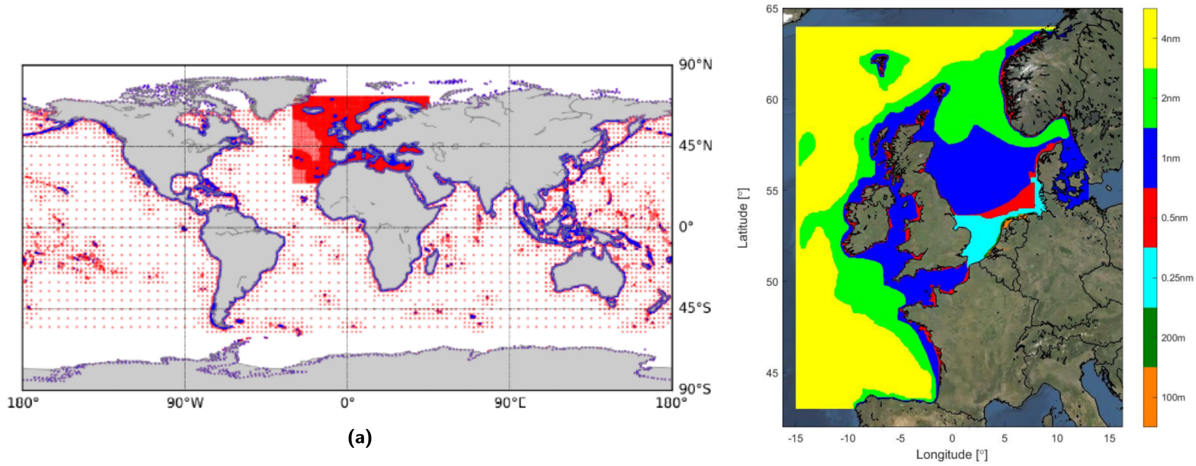
1.2.1. Numerical models

GTSM provides an elaborate hindcast of water levels and a number of forecasts for different climate scenarios (Muis et al., 2020). The numerical model is based on hydrodynamic equations, such as the shallow water equations and the continuity equation, to simulate tides and surges. It has a flexible resolution, ranging from 25 km in open ocean to 2.5 km along coasts, which has been improved to 1.25 km for Europe (Figure 1.2a). Based on 591 tide gauges, the root mean squared error (RMSE) for annual maxima is 0.26 m, with a standard deviation of 0.73 m. However, since the process-based model is partly based on the shallow water equations, the model performs poorly on coasts with a steep near-shore slope, meaning deep waters near the coast. This induces large spatial variability in the performance of the model, as evident in the high standard deviation.

While global models are useful for a great many applications, regional processes are often better represented in regional models. For the North Sea, the Dutch Continental Shelf Model (DCSM) has been developed at the request of Rijkswaterstaat (RWS), the Ministry of Infrastructure and Water Management in the Netherlands. Originally developed by Gerritsen et al. (1995), DCSM has undergone many updates to improve the model, resulting in accurate estimates of tides and surges, as well as useful estimations for water quality and ecology studies, oil spill modelling and boundary conditions for more detailed local models (Zijl, Groeneboom, et al., 2022). Like GTSM, its resolution is flexible, and ranges from 4 nautical miles offshore to 100 metres along the Dutch coast, including the Wadden Sea and the Dutch estuaries (Figure 1.2b).

The most recent versions are DCSM-FM 0.5nm, DCSM-FM 100m and 3D DCSM-FM (Zijl, Zijlker, Laan, and Groeneboom, 2022b; Zijl, Groeneboom, et al., 2022; Zijl, Zijlker, Laan, and Groeneboom, 2022a), where all three models have a flexible mesh with varying spatial resolution up to 0.5 nautical mile for DCSM-FM 0.5nm and 3D DCSM-FM and 100 metres for DCSM-FM 100m. The mean RMSE's are 0.085, 0.081 and 0.078 m for DCSM-FM 0.5nm, 3D DCSM-FM and DCSM-FM 100m respectively. The performance is highest in offshore areas with a RMSE between 0.048 and 0.055 m, and decreases only slightly in coastal areas with a RMSE between 0.063 and 0.068 m. For the Wadden Sea and the southwestern delta, the RMSE is larger, ranging from 0.085 to 0.115 m.

The main limitations of these kinds of numerical models are computation time and complexity. When ensemble forecasts or additional studies are required, the computation time is large and models such as DCSM are difficult to set up, despite their widespread use. Another important note is that most of these models are process-based, meaning they are based on hydrodynamic equations such as the shallow water equations. Validation is then based on tide gauge data. In some cases, tide gauge data is even used as calibration data. These models often exclude satellite altimetry data, which is a promising alternative to the sparse and limited spatial information tide gauges provide. Additionally, these models do not make use of the increasing potential of ML, which offers flexibility, low computational effort once trained, and efficient continual improvement.



(a) (b)

Figure 1.2: An overview of (a) the GTSM grid points (Muis et al., 2020) and (b) the DCSM grid points (Zijl, Groeneboom, et al., 2022). In (a), red points denote ocean points and blue points denote coastal points. In (b), the colours indicate the grid size (yellow: ±4 nautical mile (nm); light green: ±2 nm; blue: ±1 nm; red: ±0.5 nm, cyan: ±0.25 nm, dark green: ±200 m and orange ±100 m)

1.2.2. Machine learning

While numerical models such as GTSM or DCSM rely on hydrodynamic equations, ML models derive their relationships entirely from data. These data-driven models do not contain pre-defined conditions, equations, or assumptions to guide their predictions. Implementations of ML methods to improve sea surface height estimates have been developed for a number of cases, from large-scale global models (Bruneau et al., 2020; Passaro and Juhl, 2023) to regional studies (Den Bieman et al., 2023; Gharineiat and Deng, 2015; Hieronymus and Hieronymus, 2023; Xie et al., 2023).

Global studies

Bruneau et al. (2020) have developed a neural network method to estimate global coastal non-tidal residuals with ERA5 assimilated data (Hersbach et al., 2018), mainly 10-metre wind speed components, mean sea level pressure, significant wave heights, peak periods and precipitation. They have used a continuous ranked probability score (CRPS) to assess the performance of their method, and reach a mean CRPS of ±0.1 m.

Regional studies

Passaro and Juhl (2023) have tested the potential of a Random Forest Regression method to interpolate daily sea level anomalies (SLA) from satellite altimetry to a regional grid, which performs better than the Copernicus Marine Environment Monitoring Service (CMEMS) Level 4 gridded SLAs derived from satellite altimetry. They used altimetry data from Jason-1, Envisat, TOPEX/Poseidon and GeoSat Follow-On for the North Sea in 2004 (CMEMS Level 3 SLAs) to train the Random Forest Regression model, after which it was validated with tide gauge observations. The main results show a RMSE between 0.02 and 0.12 m.

Gharineiat and Deng (2015) have compared a multi-adaptive regression spline (MARS) model with a multivariate regression model to predict sea levels. They have combined tide gauge observations with data from satellite altimetry, creating a method that shows a strong correlation (±99%) between modelled and observed sea levels. In general, the RMSE varies between ±0.03 and ±0.16 m. For sea levels during cyclones, the RMSE varies between 0.04 and 0.21 m.

Hieronymus and Hieronymus (2023) have created a machine-learning-based bias correction to apply to sea levels within a climate model in the Baltic Sea. They have used ERA-Interim data (Dee et al., 2011) to train their neural network and did additional analysis on the high values within their sea level estimations. The RMSE, which has been normalised with the standard deviation of the observed sea level, varies between 0.4 and 0.8.

Xie et al. (2023) have developed a deep learning model based on a training dataset generated by a numerical model for the Pearl River and the East China Sea, which can estimate storm surges effectively. The reported average RMSE is 0.16 m.

Den Bieman et al. (2023) use a ML model built with classification and regression trees (CART) to improve wave field forecasts from a wave forecasting model. SWAN model predictions are fed into the ML model with wind and wave measurement data from 14 locations along the Dutch coast. The results show a reduction of the RMSE by 21.7% for the energy density and 25.3% for the mean wave direction. For the spectral wave height, the RMSE decreased from 0.21 m to 0.14 m (33.3%). For the spectral wave period, the RMSE decreased from 0.67 to 0.41 s (38.8%).

Notably absent from these studies is the inclusion of local estimates for coastal water levels derived directly from detailed satellite altimetry data rather than from assimilated products such as ERA5 or CMEMS. Bruneau et al. (2020) showed promising results using wind speed and wave heights from ERA5, which stacks all available satellite altimetry data and interpolates it onto an hourly 0.25°grid. However, it has been found that ERA5 data tends to underestimate extreme wind speeds (Dullaart et al., 2020; Haakenstad et al., 2021). Using satellite altimetry sea surface heights (SSH) could contribute to the current study focussing on coastal water levels. Many of the mentioned studies incorporate satellite altimetry, which has a couple of significant advantages over in-situ observations and complicated numerical models.

1.2.3. Satellite altimetry

The principle of satellite altimetry is to measure the travel time of emitted radar pulses to the Earth's surface and back. Analysing this travel time allows for accurate estimates of the SSH once corrected for atmospheric disturbances and signal-wave interactions. Due to their large spatial coverage, satellite altimetry missions can capture remote locations where tide gauges are sparse and consequently provide limited information on the coastal water levels. Several studies have used this to study coastal water levels and high water level events either regionally or globally.

Global studies

Extreme wave climates have been studied globally, focussing on monthly maxima of significant wave heights (SWH) (Izaguirre et al., 2011). Several seasonal variability aspects have been found, the most prominent ones being high wave heights for a 20-year return period in the boreal winter in the North Atlantic Ocean (±17 m south of Iceland and southwest of Ireland) and the North Pacific Ocean (17.65 m south of the Aleutians). In the austral winter, the highest wave heights are observed in the Southern Ocean between South Africa and Australia (15.5 m).

It has been found that satellite altimetry underestimates the amplitude of the SWH (Alves and Young, 2003; Timmermans et al., 2020) and extreme water levels (Darko et al., 2023) due to undersampling of the extreme events, although it can also overestimate extreme SSHs due to insufficient spatial interpolation (Figure 1.3). Jiang (2020) has studied altimeter undersampling in global wind and wave estimates. A bias of 0.491 m/s for 10-m sea surface wind speed (U10) observations and 0.126 m for SWH observations was reported for the 90% percentile in a period with a limited number of active satellite altimetry missions (1992). For a period with more active missions (2017), this undersampling decreased to 0.167 m/s and 0.042 m respectively. The 99% percentile metrics report more extreme undersampling, with 1.781 m/s and 0.722 m for U10 and SWH respectively for 1992 and 0.777 m/s and 0.339 m for 2017.

Regional studies

To battle this undersampling, scale factors were formulated by Lobeto et al. (2018) and Bij de Vaate (2023). Along the North American East Coast, Lobeto et al. (2018) achieved a 76% reduction in the mean relative error for return periods of up to 50 years by applying a scale factor to correct the satellite altimetry non-tidal residuals (NTR). Before correction, the satellite altimetry NTR values ranged from 1.5 to 2.5 when normalised against tide gauge derived return periods. After scaling, these values were reduced to between 0.8 and 1.25. Bij de Vaate (2023) applied a similar scaling on a global grid, who found that the tropics need a larger scaling than higher latitudes, but the higher latitudes contain larger errors for these scaling factors.

Satellite altimetry in combination with tide gauges has been used to improve local storm surge forecast models for the Danish Coast and the northeast coast of Australia (Andersen et al., 2015). They illustrated the importance of having multiple active satellite missions for capturing storm surges in the North Sea. Ji et al. (2019) have used satellite altimetry complementary to tide gauge data to monitor storm surges in coastal areas of China, concluding the same for that region.

Additionally, various retracking algorithms such as the Adaptive Leading Edge Subwaveform retracker (ALES) used in coastal processing by X-TRACK, developed by the Centre for the Topography of Oceans and the

Hydrosphere (CTOH) (Birol et al., 2021), aim to improve SSH estimations close to the coast (5-10 km). Schwatke et al. (2015) developed an approach to compute inland water levels from satellite altimetry data, generating the Database for Hydrological Time Series over Inland Waters (DAHITI) using an extended outlier rejection and a Kalman filter for the processing of satellite altimetry data. J. Guo et al. (2019) introduced a singular spectrum analysis retracker (SSA), which combines three existing retracking methods. All of these retracking studies have improved coastal water level observations significantly.

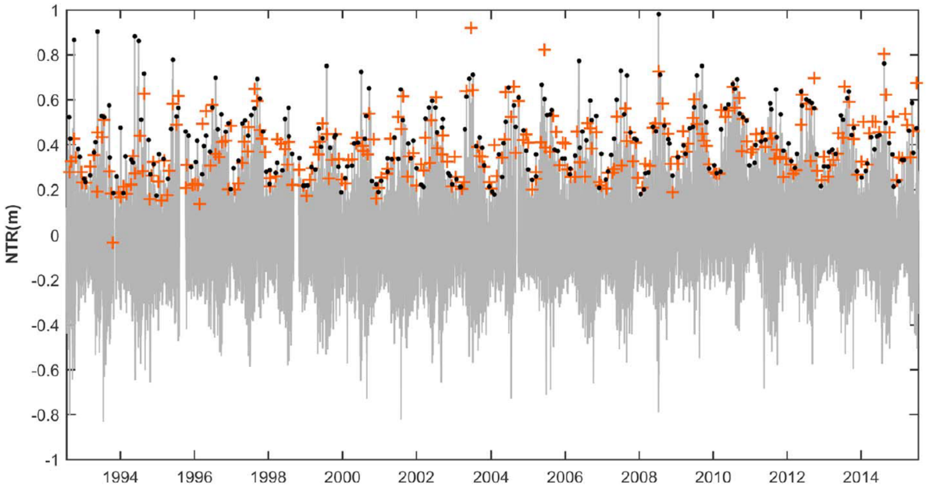


Figure 1.3: A time series of non-tidal residuals (NTR) from a tide gauge in Eastport, USA. The black dots represent monthly maxima from the tide gauge and orange crosses refer to monthly maxima from satellite altimetry (Lobeto et al., 2018).

1.3. Research objectives

Exploring ML in combination with satellite altimetry will help us understand how data-driven models can be used to study water level variations. It is required to assess the potential of ML in this context and to obtain the main drivers that influence the quality of a ML model. This study's main research objective is **to develop a machine learning model that produces hourly non-tidal water level estimates at the location of a tide gauge, based on satellite altimetry observations and ERA5 pressure and wind fields within a 48-hour time window and across an area including the North Sea and the Strait of Dover**. A window of 48 hours is defined beforehand to ensure the model captures the full extent of storms that travel over the North Sea (Tijssen & Diermanse, 2010). The main research objective is broken up into three sub-questions which are defined below.

1. ***How does the performance of the ML model compare to tide gauge observations and a regional numerical model?***

   To assess the performance of the developed ML model, it will be compared with DCSM and TG data with statistical metrics such as the mean squared error, mean absolute error and standard error. Using examples, specific advantages and limitations will be highlighted. Matching the performance to DCSM will be a secondary focus to illuminating the differences and similarities.

2. ***How well does the ML model estimate high water levels?***

   High water levels are defined by a peak over threshold (POT) method in this study, and will be compared with DCSM and TG data to assess the performance of the model related to high water levels. Several severe storms are used as case studies to define potential advantages or limitations of the model.

3. ***Which factors affect the estimation of hourly non-tidal water levels estimated with the ML method?***

   To gain a better understanding of the variables that contribute most to water level variations along the Dutch coast, the weights trained within the ML model are analysed. Additionally, this study will focus on the parts of the ML model with good or bad performance and determine which input variables cause deviations from the overall performance.

1.4. Thesis outline

Chapter 2 provides theoretical background information on coastal water level variations, focussing on its main components and their corresponding physical processes. Satellite altimetry and machine learning are also explained, with additional information given in Appendix A and B. Next, the considered study area and datasets are presented in Chapter 3. More information on the individual satellite missions is given in Appendix C. Chapter 4 contains the method which is used to pre-process the data, set up the ML model, train it and compute the performance. Some pre-processing steps are further elaborated upon in Appendix D, E and F. Next, the output of the ML model and the model performance are presented in Chapter 5. These results consist of time-series reconstruction, performance metrics and a weight analysis. Additional results are presented in Appendix G and H. A discussion of the results is given in Chapter 6, after which the research questions will be answered and several recommendations will be given in Chapter 7.

2

Background

2.1. Definition of sea surface height

At every low-lying coastal location, it is important to understand the processes that cause the sea surface to vary. These processes range from the interaction of the Earth with the Sun and the Moon to the forcings induced by weather conditions within a certain vicinity of the location. Factors such as the orientation of the coast and the adjacent water body also play significant roles. The interactions of these processes and factors result in a complicated variability in sea level, with potential mutual amplification or dampening effects. This section picks apart the largest forcings on sea level variability in the North Sea, along with the response of the sea level to them.

The sea surface height (SSH) can be defined as the sum of the mean sea surface (MSS), the astronomical tide (AT), the dynamic atmosphere correction (DAC) and the sea level anomaly (SLA), as defined in (2.1). High values in the SSH are often the result of a combination of extremes of the AT, DAC and SLA, such as spring tide and large pressure gradients. This is schematised in Figure 2.1, where AT is defined as *expected high tide* and DAC is defined as *storm surge*, which is caused by low pressure and extreme winds. The SLA includes the remaining factors such as wave set-up and wave run-up, though these factors are negligible when going further offshore. Normally, the SSH is referenced to the ellipsoid, but for some applications, the sea level referenced to the geoid is preferred. In that case, the MSS (which is referenced to the ellipsoid) is changed into the mean sea level (MSL), which is referenced to the geoid.

$$\mathrm{SSH} = \mathrm{MSS} + \mathrm{AT} + \mathrm{DAC} + \mathrm{SLA} \tag{2.1}$$

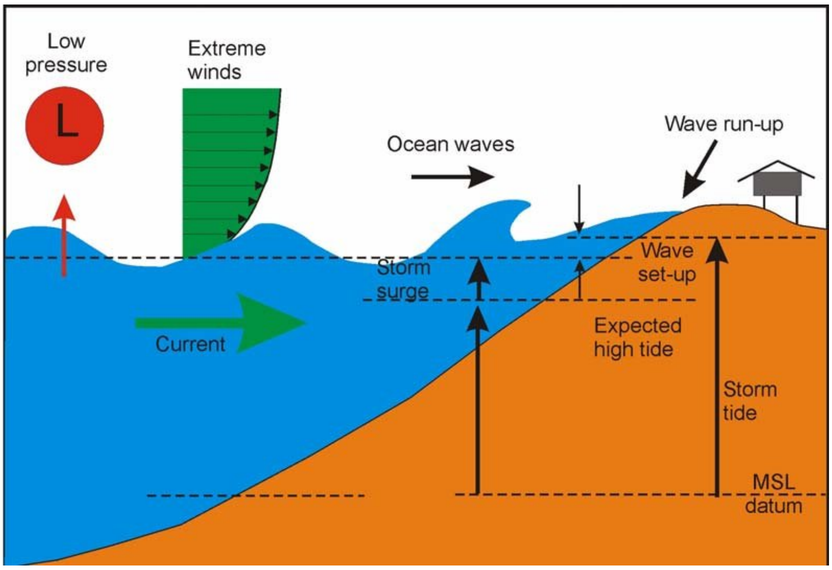


Figure 2.1: Forces that cause extreme coastal water levels (Mullan et al., 2005).

2.1.1. Astronomical tide

The largest forcing which influences the water level in the Dutch coastal zone is the astronomical tide (AT) (Idier et al., 2019). Tidal components can cause the water level to rise up to a few metres in this region, which can, in combination with pressure and wind forcing, cause high water levels (Figure 2.1). The tides are mostly regulated by the orientation of the Moon relative to the Earth and the Earth relative to the Sun. These planetary positions and their orbits cause a variety of different tidal cycles, with the most well-known ones the spring-neap tide and semi-diurnal tides (Eleveld et al., 2014). These tidal components represent stationary oscillations, like the lunar and solar cycles, which are estimated by hydrodynamic, empirical or mixed models (Stammer et al., 2014).

The propagation of these tides is dependent on friction, resonance from ocean basin shapes and depths, and the Coriolis effect. Due to the Coriolis effect and the land masses around the North Sea, the tidal waves propagate counter-clockwise in the Northern Hemisphere around nodes with a tidal amplitude of zero. These nodes are also referred to as amphidromic points. The North Sea contains amphidromic points between the Dutch and English coast, ±2 degrees west of the coast of Denmark and at the southern tip of Norway.

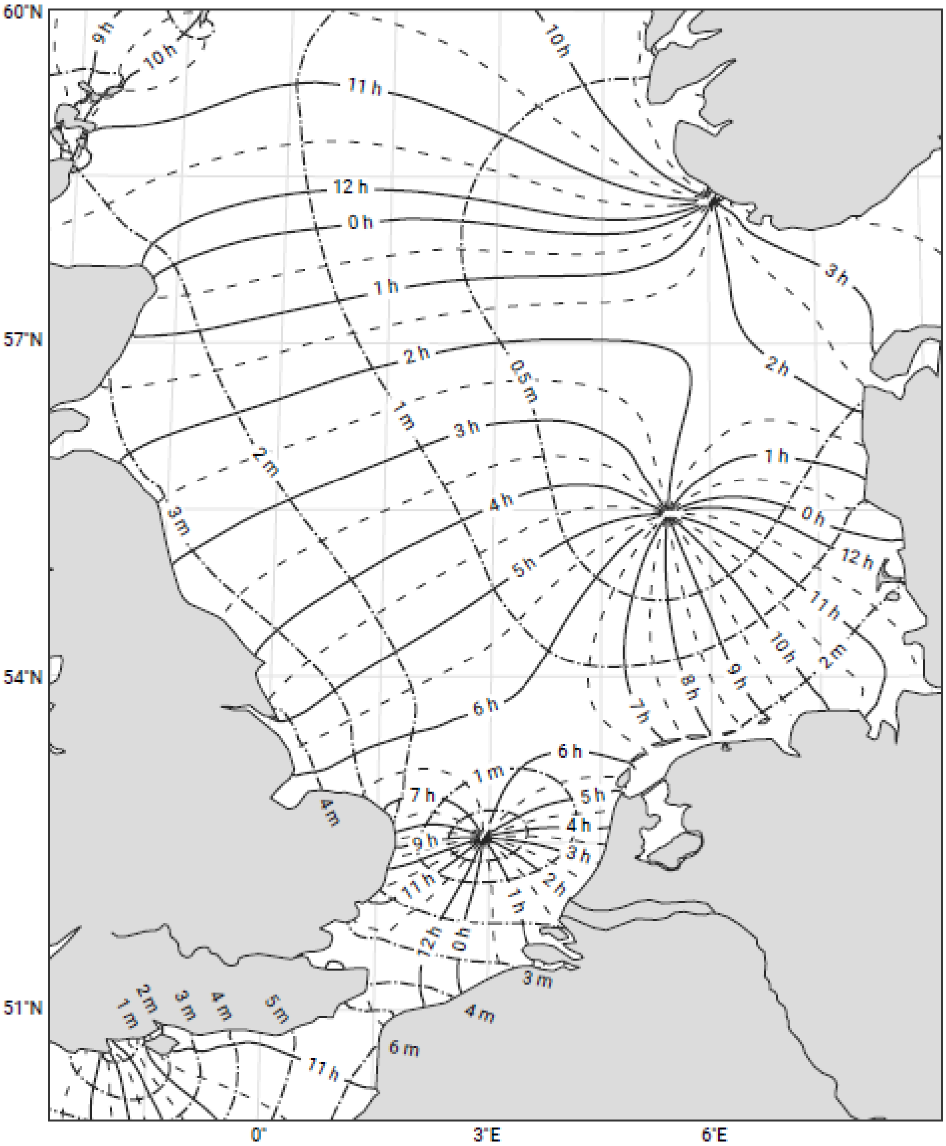


Figure 2.2: Propagation of the M2 tide in the North Sea with co-tidal lines radiating away from the amphidromic points and co-range lines encircling them. The co-tidal lines show that the phase increases counter-clockwise around the amphidromic point (typical of NH amphidromes). The co-range lines show the tidal range increasing away from the node (Bosboom & Stive, 2021).

2.1.2. Dynamic atmosphere correction

Forcings that correspond to atmospheric variables are represented in the dynamic atmosphere correction (DAC). This factor includes a pressure and a wind component, which both contribute to the SSH (Harris, 1963). Firstly, low-pressure areas are meteorological forcings that can increase the coastal water level and are described by the inverse barometer (IB) effect (Wunsch & Stammer, 1997). The difference between the low-pressure centre of a storm and the high-pressure boundary causes a difference in water levels in the open ocean or along the open coast. One hectopascal drop in atmospheric surface pressure causes one centimetre rise in water level. This effect is assumed to only hold in open oceans or along open coasts where the water is not too shallow (Harris, 1963). If this pressure disturbance is moving at a speed comparable to the wave speed near the coast, this water level disturbance can even be greatly amplified by resonance (Gertsenshtein, 1962).

Secondly, the wind fields over the North Sea can be sufficiently strong that they alter currents in the surface layer of the ocean (Harris, 1963). This creates a wind set-up, a phenomenon where the downwind side of the wind field experiences a larger water level than the upwind side. This usually causes high-frequency modulations (<20 days), while the IB effect is considered a low-frequency response (>20 days) (Carrère & Lyard, 2003).

2.1.3. Sea level anomaly

The remaining parameters that influence the SSH are summed in the sea level anomaly (SLA). Among these are salinity variations, temperature variations, freshwater influxes, ocean circulation conditions, local bathymetry and topography, orientation of the coast and depth of the water body. Higher salinity increases water density, while lower salinity results in less dense, buoyant water. Changes in salinity, due to freshwater influxes, precipitation and evaporation, can change the salinity of the water, expanding or compressing the water (Antonov et al., 2002; Cabanes et al., 2001; Ishii et al., 2006). Aside from changing the salinity, the freshwater influxes can also influence the SLA, especially on regional scales (Lombard et al., 2009). Next, thermal expansion is a source of variability within the SLA, its importance only increasing with global sea level rise (Cabanes et al., 2001).

On a larger scale, ocean circulation patterns such as the North Atlantic Oscillation (NAO) cause sea level variability. Positive NAO phases typically strengthen westerly winds and enhance ocean circulation, leading to higher sea levels around a large part of the northwestern European Continental Shelf (Iglesias et al., 2017). In addition, the region's local bathymetry and topography play important roles (Harris, 1963). In the North Sea, the Strait of Dover is a narrow funnel-like channel that separates the English Channel from the North Sea. This causes the North Sea to behave like a semi-enclosed basin, allowing for water accumulation with the right wind direction and strength (Bosboom & Stive, 2021). Furthermore, the orientation of the coastline relative to prevailing winds and currents, as well as the depth of the adjacent water body, can amplify or dampen wind-driven effects on the water level.

Any non-linear interactions between the tides, atmospheric forcings and other processes in the North Sea are not accounted for in the AT or the DAC. However, these interactions induce significant sea level variations and have been a topic of interest for many coastal and river studies over the years (L. Guo et al., 2023; S. Hu et al., 2023; Jones and Davies, 2008; Moftakhari et al., 2024; Prandle and Wolf, 1978). These interactions can cause temporary sea level rises that exceed the sum of their individual effects, and are left in the SLA when decomposing the SSH from satellite altimetry observations or tide gauge measurements.

Finally, some localised effects can cause the SLA to vary, such as land-sea interactions by anthropogenic constructions such as harbours, dams and breakwaters or effects from passing boats or reflected waves. The wave field during a storm can create wave set-up, which is defined as the piling of water near the shore under the direct influence of waves (Harris, 1963). This is caused by near-shore wave breaking and is highest for open coasts with steep near-shore bathymetry slopes, since the waves will dissipate and break closer to the shore. Wave run-up is defined as the increase in coastal water level due to the waves rolling up the slope of the coast. However, tide gauges often do not observe wave set-up or run-up, since they are often located further offshore or in sheltered bays or harbours.

2.2. Satellite altimetry

One way to accurately measure the SSH is from space, specifically through satellite altimetry. The objective of satellite altimetry missions is to measure the height of the Earth's surface, with an increased focus on the world's oceans and ice surfaces. Satellite altimetry observations are therefore a reliable proxy, and consequently useful for a reliable risk assessment of high water level events. Figure 2.3 shows an overview of past, present and future satellite altimetry missions, with decommissioned missions shown in red, active missions in orange, and future missions in yellow.

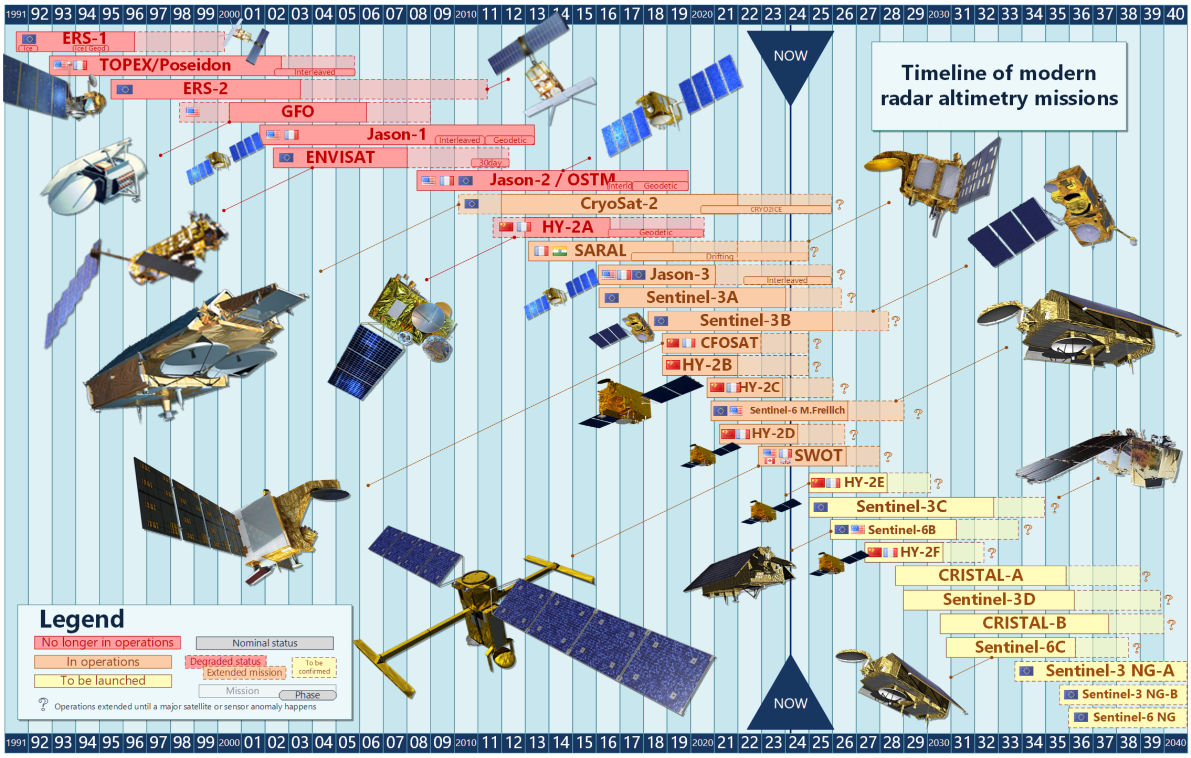
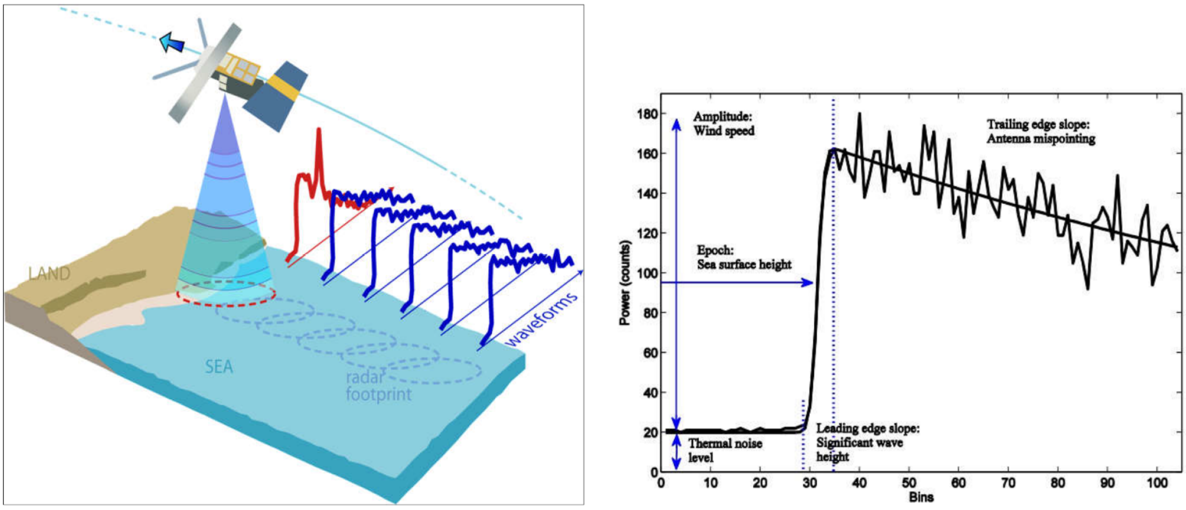


Figure 2.3: A timeline of modern radar altimeters from the 1990s to the next decade, including expected or announced mission extensions beyond the nominal satellite lifespan. Past missions are in red (Aviso+, 2024).

From the missions shown in Figure 2.3, TOPEX/Poseidon, Jason-1, -2 and -3, Sentinel-6 M. Freilich, Sentinel-6B, Sentinel-6C and Sentinel-6 NG have a repeat cycle of 10 days (ESA, 2023a). This means that the satellites revisit the same location on Earth every 10 days. CFOSAT has a revisit time of 13 days, all HY-2 satellites of 14 days and GFO of 17 days. The SWOT mission has a repeat cycle of 21 days, and all Sentinel-3 missions of 27 days. ERS-1, ERS-2, Envisat and SARAL have repeat orbits of 35 days, and the longest repeat orbits are for the CRISTAL-A and -B missions with 367 days and for CryoSat-2 with 369 days.

The hardware used on satellite altimetry missions is a radar altimeter (RA). This sensor emits short pulses of microwave signals towards the Earth's surface, and collects the reflected power of the signal. This gives information on the radar footprint (Figure 2.4a). With the use of this reflected power and accurate information on the orbit of the satellite, the SSH can be estimated.

The observed return signal resembles the received power over time and is called the waveform (Figure 2.4b). The power is received in bins, also often referred to as gates, which represent small windows of 3.125 nanoseconds over the range between the water level and the satellite (Passaro et al., 2014). The waveform in open ocean can be described by the Brown model (Brown, 1977). With this, the SSH can be estimated by taking the mid-point of the leading edge and the SWH by the slope of the leading edge. The mid-point is defined by fitting the Brown model to the received signal and searching for the gate that contains the mid-point of the leading edge. When the satellite altimeter travels closer to the coast (Figure 2.4a), land signals can cause severe distortions in the signal, which is called land contamination. The waveform does not conform to the original Brown model anymore, which means the reliability of the observations decreases (Gommenginger et al., 2011).



(a) (b)

Figure 2.4: The reflected power of the satellite altimeter (a) when approaching the coast and (b) over homogeneous ocean surface. The red line in (a) shows land contamination, meaning that the signal does not conform to the Brown model anymore (COASTALT, 2023). The named parts of the signal shown in (b) correspond to different geophysical parameters associated with the water level (Idris et al., 2014).

2.2.1. Derivation of sea surface height

The SSH is derived from satellite altimetry by computing the distance between the satellite and the water surface with $R_{obs} = c\frac{t}{2}$, where t is the observed two-way travel time of the mid-point of the leading edge and c is the speed of the radar pulse through vacuum (Andersen & Scharroo, 2010). To obtain the SSH, the observed range R_{obs} needs to be corrected for several disturbance factors, as given in (2.2). The height of the satellite with respect to the ellipsoid is given as H.

$$\begin{aligned} \text{SSH} &= H - R_{corrected} \\ &= H - (R_{obs} + \Delta h) \end{aligned} \tag{2.2}$$

The correction term Δh is categorised into atmospheric range corrections and a geophysical correction as shown in (2.3). Atmospheric range corrections (Δh_{dry}, Δh_{wet} and Δh_{iono}) adjust the range to account for any disturbances when the signal travels through the ionosphere and troposphere. The geophysical correction (Δh_{SSB}) focusses on adjusting the range caused by the physical properties of the ocean's surface and how the signal interacts with it, particularly in terms of reflection back to the satellite. This correction consists of three main components, mainly an electromagnetic bias (EM), a skewness bias and a tracker bias. Further information on these corrections is given in Appendix A.

$$\Delta h = \Delta h_{dry} + \Delta h_{wet} + \Delta h_{iono} + \Delta h_{SSB} \tag{2.3}$$

When all correction terms have been applied, the resulting variable equals the sea surface height with respect to the ellipsoid. Decomposing the SSH can help make sense of the different forcings that make up the water level (see also (2.1)). Especially the AT and the DAC have been extensively investigated and represented by numerous models (Carrère and Lyard, 2003; Lyard et al., 2021; Stammer et al., 2014). Datasets often provide the SLA, where the mean sea surface (MSS) and the appropriate corrections have already been removed from the satellite observations. The non-tidal residual (NTR) is used in satellite altimetry studies as well (Lobeto et al., 2018), and corresponds to the SLA plus the DAC.

2.3. Tidal and atmospheric components

Next to the astronomical tide (AT), two additional tidal components influence the SSH. These include the solid-Earth tide (ST) and the pole tide (PT), and are often used to obtain the SLA from the SSH derived from satellite altimetry observations discussed in the previous section.

Astronomical tide

The astronomical tide consists of an ocean tide and a loading tide. The loading tide is the smaller of the two, with an amplitude of only 4-6% of the ocean tide, and is a result of the deformation of the ocean floor due to the weight of the water column on top of it (Andersen & Scharroo, 2010). Since the loading tide is in phase with the ocean tide, they can simply be added together and modelled accurately according to the lunar and solar cycle (Stammer et al., 2014).

The ocean tide is represented as the sum of a finite number of tidal harmonic constituents, of which Table 2.1 shows the most prominent ones. The semi-diurnal components refer to harmonic tidal signals that occur twice a day, which are numbered with a "2" in their names. Diurnal components are numbered with a "1" and refer to tidal signals which occur once a day. Longer modulations of the tidal signal are subscripted with individual letters, such as "f" for fortnightly, "m" for monthly and "sa" for semi-annual signals. Table 2.1 also shows the equilibrium amplitude, which is a measure of the amplitude of the harmonic tidal signal in open ocean according to the equilibrium theory (Bosboom & Stive, 2021).

Table 2.1: Principal tidal constituents with equilibrium amplitudes from Apel (1987)

Tidal constituents	Name	Equilibrium Amplitude [m]	Period [h]
Semi-diurnal			
Principal lunar	M2	0.24	12.42
Principal solar	S2	0.11	12.00
Lunar elliptical	N2	0.046	12.66
Lunar-solar declinational	K2	0.031	11.97
Diurnal			
Lunar-solar declinational	K1	0.14	23.93
Principal lunar	O1	0.10	25.82
Principal solar	P1	0.047	24.07
Lunar elliptical	Q1	0.019	26.87
Long period			
Fortnightly	Mf	0.042	327.9
Monthly	Mm	0.022	661.3
Semi-annual	Ssa	0.019	4383

There are many more constituents, which take into account more than just the solar, lunar and earthly cycles, but also the influence of shallow water on these signals. Shallow-water tides, also known as overtides, arise from nonlinear effects in coastal waters (Bosboom & Stive, 2021), resulting in higher harmonics of the primary tidal constituents such as M2 and S2 shown in Table 2.1. These higher harmonics, such as M4 (with a period half that of M2) and M6 (one-third of M2), are generated by interactions with bottom friction, variations in water depth and the continuity of water flow. Interaction tides such as MS4, which results from M2 and S2 interactions, also contribute to the asymmetry of the tidal elevations. Appendix D gives more information on which constituents are applied in this study.

Solid-Earth tide

The solid-Earth tide originates from the elastic response of the Earth's crust due to the gravitational tide potential, caused by the Sun and the Moon. This has been estimated by Cartwright and Tayler (1971) and Cartwright and Edden (1973). The solid-Earth tide can reach values of ± 0.2 m (Andersen & Scharroo, 2010).

Pole tide

The final tidal correction relates to the seasonal variation of the Earth's axis of rotation. This variation is known as the *Chandler Wobble*, and the pole tide is caused by the centrifugal forces, which change according to the variation of the Earth's axis (Andersen & Scharroo, 2010). The method developed by Desai et al. (2015) and Ries (2017), based on the method by Wahr (1985), is used to estimate the pole tide.

Dynamic atmosphere correction

The DAC consists of two main contributors, categorised into a low-frequency (LF) and a high-frequency (HF) part. The low-frequency contribution has a period of longer than 20 days and is also known as the inverse barometer (IB) effect. This is the result of pressure fields in the atmosphere that press the ocean down in high-pressure areas and let the ocean rise in low-pressure areas. The IB effect can be directly computed from the surface pressure by

$$\Delta h_{ib} \approx -0.99484\,(P_0 - P_{ref}) \tag{2.4}$$

where P_0 is the same sea surface pressure as the one used to compute the dry tropospheric correction Δh_{dry} (see Appendix A) and P_{ref} is a reference pressure value, which can be taken as a stationary value or a variable one. Usually, the reference pressure varies between 1009 and 1013 hPa (Andersen & Scharroo, 2010). The high-frequency contribution of the DAC is caused by wind fields, creating wind set-up at the coast, and is modelled with the MOG2D model (two-dimensional gravity waves model) (Carrère & Lyard, 2003), which uses the shallow water continuity and momentum equations for computation. The MOG2D_IB model combines the high- and low-frequency components and provides atmospheric corrections from approximately -0.1 m to 0.15 m. A new version of the MOG2D model, which uses a higher FES2014 resolution mesh, improved bathymetry fields and includes the dominant tide forcing (M2, N2, S2, K1, O1), is called TUGO (Carrère et al., 2019).

2.4. Retracking methods

As mentioned in previous sections, land contamination causes a disturbance in the waveform of satellite altimetry observations. Generally, the waveforms of a returned signal are captured using onboard trackers (Gommenginger et al., 2011). These trackers predict the next signal's likely position based on the recently returned signals to ensure that the returned signal is kept within the altimeter analysis window. These computations are done on-board the satellite, but since the waveforms are subject to changes in the vicinity of land (Figure 2.6), the identification of the tracking gate, defined as the bin/gate number of the mid-point of the leading edge, may be difficult for these trackers.

To improve the range retrieval for the waveforms, ground-based retracking can be performed, which is defined as a reprocessing of the recorded waveforms that the altimeters send to Earth. These retracking methods can be based on empirically fitted forms or physical models, both proven to increase the accuracy of coastal SSH estimation (Gommenginger et al., 2011). Most retrackers are built to reconstruct waveforms in coastal areas from the recorded noisy, land-contaminated waveforms.

Empirical retrackers are based on years of observations, such as the offset centre of gravity retracker (OCOG) (Wingham et al., 1986), the (improved) threshold retracker (Bao et al., 2009; C. Davis, 1995; C. Davis, 1997; Fenoglio-Marc et al., 2010; Hwang et al., 2006; Lee et al., 2008) and the β-parameter retracker (Deng and Featherstone, 2006; Martin et al., 1983; Zwally and Brenner, 2001). Additionally, the Brown-Hayne Theoretical Ocean Model is a physical model based on the original Brown model (Brown, 1977) (Figure 2.5), and redefined by Hayne (1980).

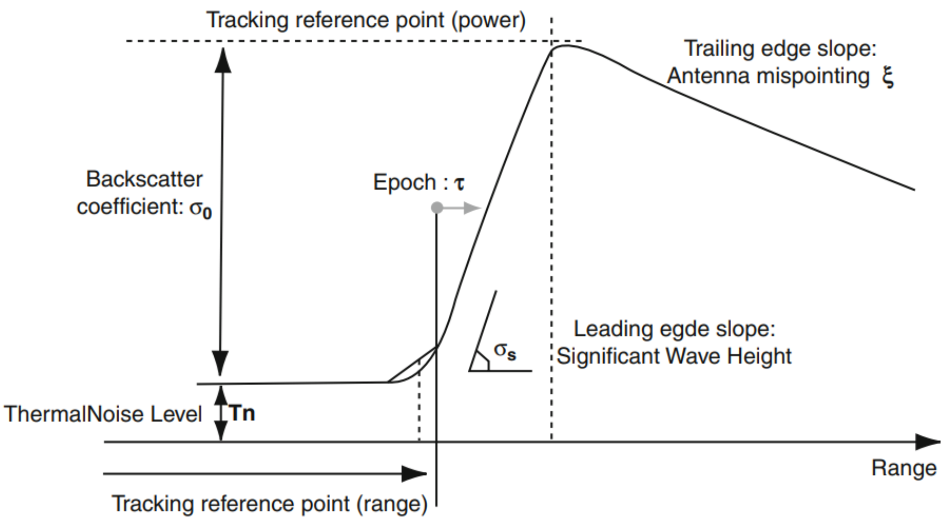


Figure 2.5: Theoretical Brown ocean waveform shape and related ocean parameters (Gommenginger et al., 2011).

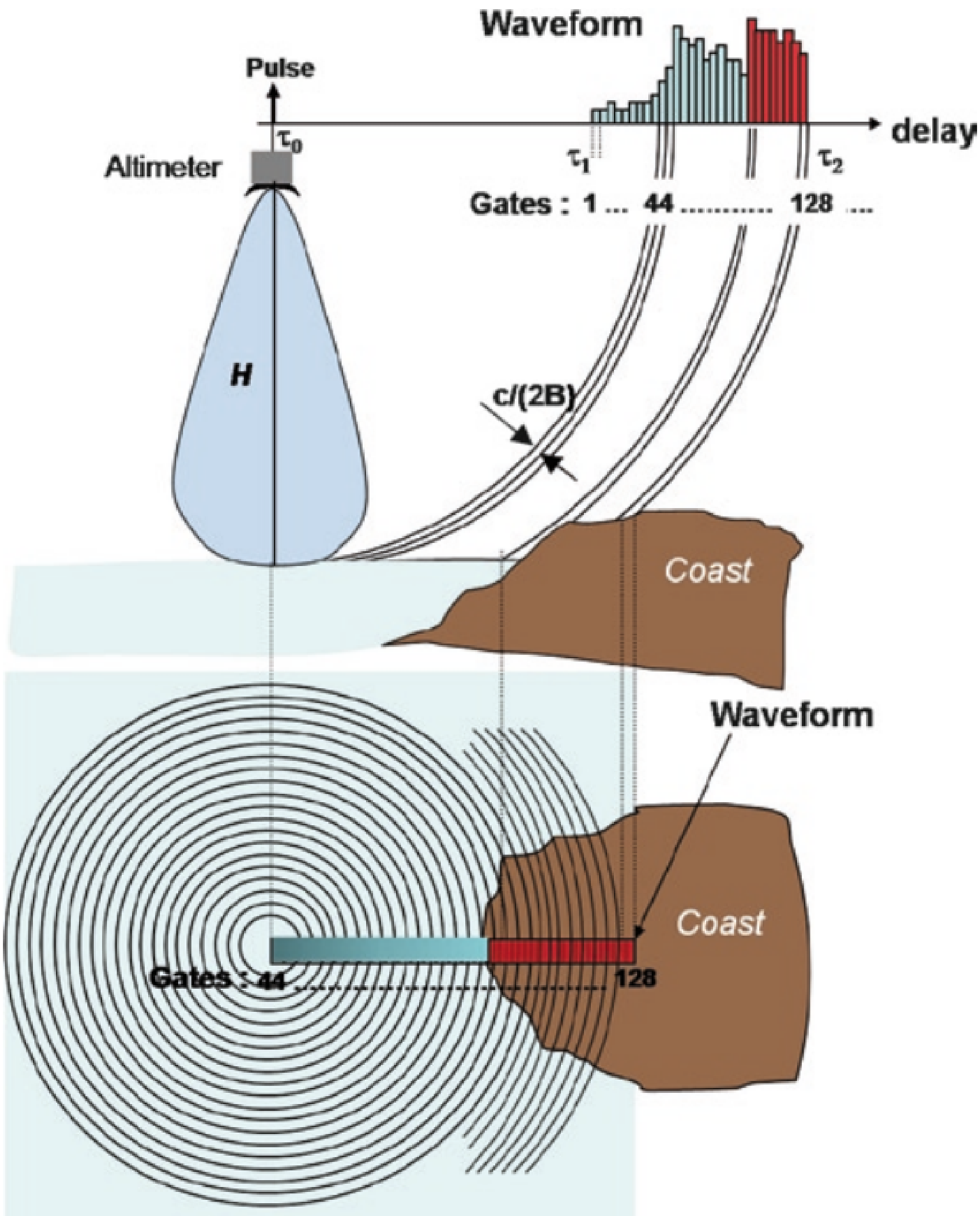


Figure 2.6: A pulse-limited altimeter in a coastal region. The lower panel shows a nadir view of the pulse-limited footprint corresponding to each waveform gate. The variable H corresponds to the range between the satellite and the ocean's surface, B corresponds to the bandwidth of the altimeter, and c is the speed of the radar pulse. The variables τ_0, τ_1 and τ_2 refer to the pulse emission time and received time delays respectively. For this altimeter, the onboard tracker has determined the altimeter analysis window to be between gates 1 and 128 and has identified gate 44 as the tracking gate (Gommenginger et al., 2011).

2.4.1. X-TRACK/ALES post-processing software

This study makes use of data processed by the X-TRACK algorithm, which is developed by the *Center of Topography of the Ocean and Hydrosphere* (CTOH) in Toulouse for the specific purpose of improving coastal water level estimations observed by satellite altimetry (Birol et al., 2017; Roblou et al., 2011). The Adaptive Leading-Edge Subwaveform (ALES) retracker has been applied to the satellite altimetry observations (Passaro et al., 2014), based on the classic Brown model for open ocean. Additionally, the method provides updated propagation and geophysical corrections. The reprocessed product (see Figure 2.7 for an example) is distributed by the AVISO+ operational centre for twenty-seven coastal regions (AVISO+, 2022). More information on the algorithm is given in Appendix B.

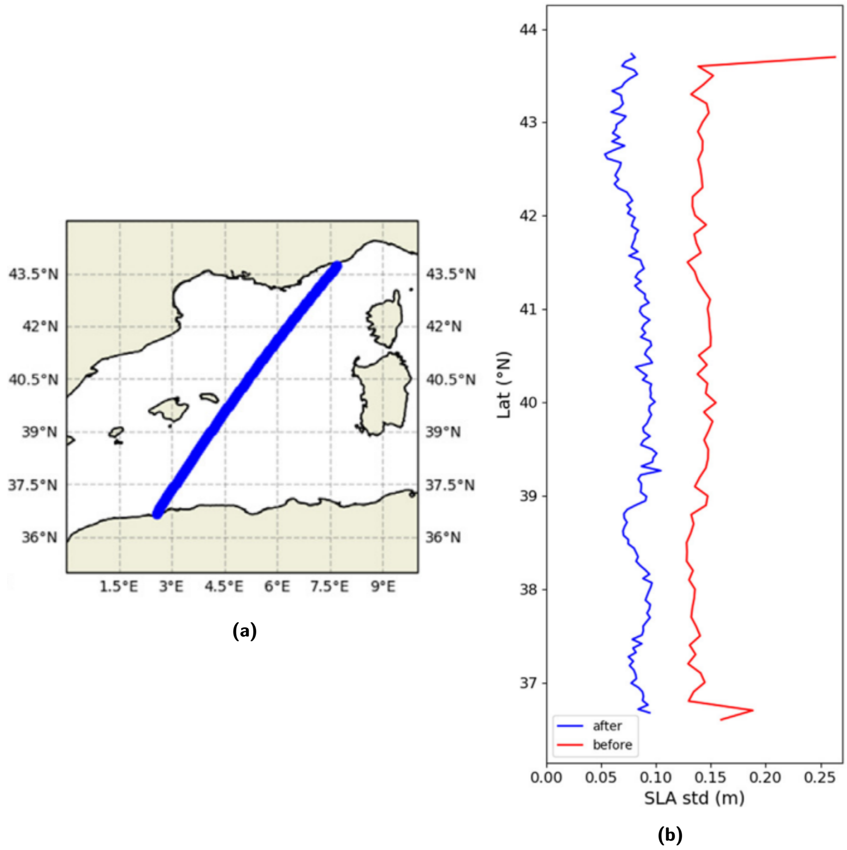


Figure 2.7: Example of Jason 2 track 9 in the Mediterranean Sea. Shown are (a) the location of the observations and (b) the standard deviation of the corresponding SLA time series (in metres) before (in red) and after (in blue) the X-TRACK procedure (Birol et al., 2021).

2.5. Machine learning

Due to its global coverage, satellite altimetry could be a useful resource in addition to tide gauges to study water levels. However, the repeat cycle of the individual missions can cause a problem when trying to capture storm surges. To decrease the risk of missing storm surges, the observations can be stacked, meaning that data from overlapping tracks are combined and interpolated to reference points along the altimetry reference tracks. This consequently can improve the temporal resolution to a couple of days, and can create elaborate global time series (Figure 3.3). A downside to this is that the spatial resolution decreases, as one needs overlapping tracks. This trade-off between temporal and spatial resolution still poses a problem when using the data to estimate high water levels.

Machine learning is a powerful technique that can identify relationships between datasets without being explicitly provided or programmed. It can recognise patterns and insights which are sometimes not even discernible by people. Within the scope of this study, machine learning is applied to solve a regression problem. With tide gauge data being used as ground truth data, this is a supervised learning case, and will be tackled with a neural network as inspired by Bruneau et al. (2020).

2.5.1. Neural networks

Built on the principle of biological neural networks within the body of an organism, artificial neural networks are created with artificial neurons, which represent the biological neurons, and are interconnected, symbolising the synapses of the biological neurons. Like a pulse of energy travels through a biological neural network, information flows through an artificial one. They are able to identify correlated patterns within large datasets with different parameters, such as high water levels during storm events with low-pressure fields and large wind speeds.

Figure 2.8 shows a simple representation of a fully connected, shallow neural network, showing all major components that need to be defined before applying any neural network to any dataset. The left-hand side shows the input dataset $[x_1 \quad x_2 \quad \cdots \quad x_m]$, which travels through the neural network through a hidden layer and gives an output variable $\hat{y}$.

During the training of the neural network, weights and biases are fitted to the data. These trainable parameters define the relationships that the input variables have with the output variables. The main goal of training a neural network is to optimise these trainable parameters, trying to find a calibration that ensures the output of the neural network is as close to the ground truth data as possible.

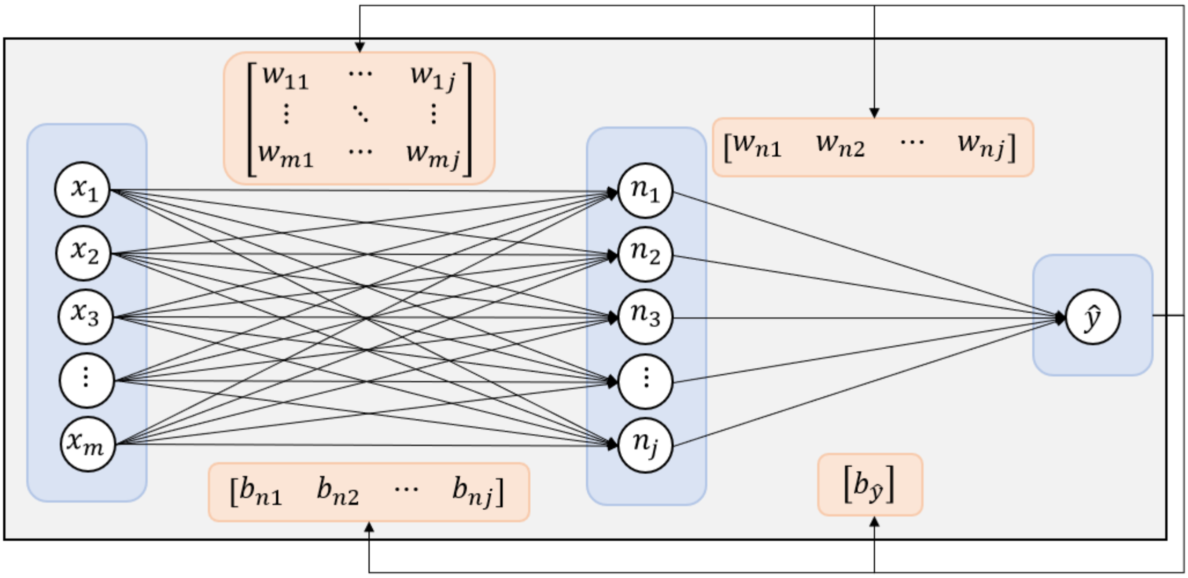


Figure 2.8: Schematic overview of the design of the shallow neural network developed in this study. The blue boxes show the input variables, design parameters and output values, while the red boxes show the trainable parameters. $[x_1 \quad x_2 \quad \cdots \quad x_m]$ denote the input variables, $[n_1 \quad n_2 \quad \cdots \quad n_j]$ refer to the neurons within the hidden layer, and $\hat{y}$ shows the output value. The matrices and vectors of w and b refer to the weights and biases that are estimated during the training of the model.

The developed shallow neural network needs some manually defined variables, from here on referred to as design parameters. These design parameters are defined as variables that are non-trainable by the model itself. These need to be pre-defined and influence for example the complexity of the model, non-linearity, training speed and weight change.

2.5.2. Design parameters

The parameters explained in this section are known as design parameters or hyper parameters. They cannot be trained by the model and have to be manually defined before the training of the model starts. They remain constant throughout the training stage.

Number of neurons

Neurons within a neural network act as nodes connecting the input data with the output. All input data are fed into each neuron, after which a weighted sum is applied according to (2.5). The output of one neuron equals one value, which is used as input for the next layer (in this case the output layer). The same weighted sum, with different weights and a different bias, is applied, resulting in an estimate of the output.

$$\text{Weighted Sum neuron j} = \sum_{i=1}^{m} (x_{ij} * w_{ij}) + b_{nj} \tag{2.5}$$

The number of neurons within the hidden layer has to be chosen with care. Too few neurons and the neural network is not able to find the relevant correlations between the input variables and the output, which is called underfitting (Figure 2.9a). Too many neurons and the model can overfit (Figure 2.9c). Overfitting happens when the model fits so precisely on the training data, that it performs poorly on data it has never seen before, which is - essentially - the goal of such a model.

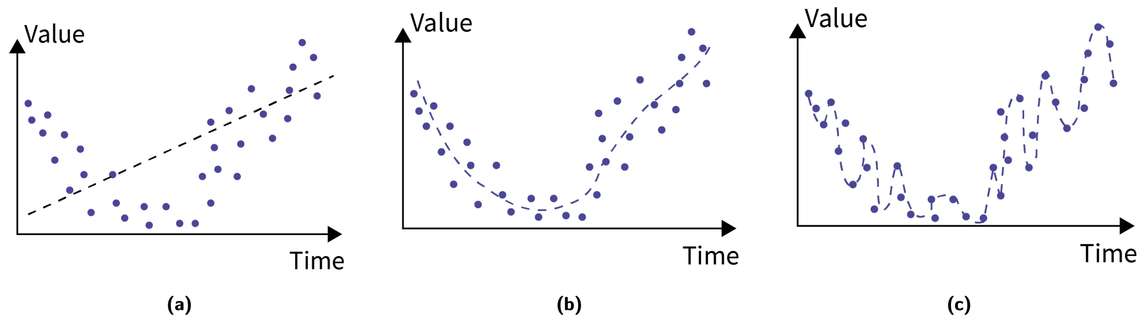


Figure 2.9: Examples of (a) an underfit, (b) a good fit and (c) an overfit. The dots represent the training data and the line represents the machine learning model output (Shah, 2023).

Activation functions

After the weighted sum within a neuron is applied, an additional step is applied before the weighted sum is fed into the output layer. The weighted sum of each neuron is fed through an activation function, which can be linear or non-linear. The rectified linear unit (ReLU) activation function is often applied on the hidden layer (Fukushima, 1969) due to its simplicity and ability to decrease the risk of the vanishing gradient problem (Glorot et al., 2011). The vanishing gradient problem occurs during the training of a neural network, where the updates to the model's weights become too small. This slows down the training and in some cases even stops the model from improving. The weighted sum that is passed through the function (x) will be set to zero if it is < 0 before it is passed to the output layer. Any weighted sum > 0 will be passed to the output layer without any changes according to (2.6) (Figure 2.10). The output layer often has a linear activation function, since any non-linearity has been accounted for within the hidden layer.

$$f(x) = x^+ = \max(0, x) = \frac{x + |x|}{2} = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.6}$$

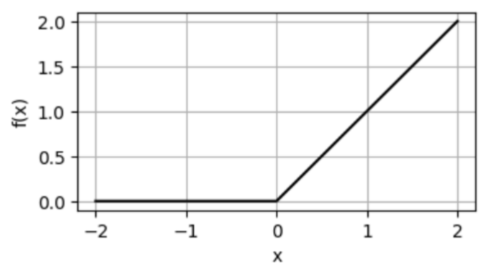


Figure 2.10: Visualisation of the ReLU activation function.

Loss function

The weights and biases of the neural network are changed based on a loss function. This function quantifies how well the neural network has estimated the output. The goal of the model is to minimise this loss function by changing its weights and biases accordingly. Examples of loss functions are the (Root) Mean Squared Error or Mean Absolute Error as shown in (2.7), where t denotes the number of output values, $\hat{y}_i$ is the estimated value and y_i is the ground truth value of the output.

$$\text{MSE} = \frac{1}{t}\sum_{i=1}^{t}(y_i - \hat{y}_i)^2 \qquad \text{MAE} = \frac{1}{t}\sum_{i=1}^{t}|y_i - \hat{y}_i| \tag{2.7}$$

Number of epochs

One of the most important design parameters is the number of epochs. This parameter defines how many training iterations are needed to let the model converge to the final set of weights and biases. This is a trade-off between over- and underfitting (Figure 2.9).

Optimisation algorithm

An optimisation algorithm can be used to update the weights within a neural network to optimise the training stage. One of the most widely used optimisation algorithms is the Adaptive Moment Estimation (Adam), developed by Kingma and Ba (2014). It computes the first and second moment of the gradient of the loss function with respect to the weights and biases for every batch in the training dataset. Given a weight w, a bias b and a loss function $L(w, b)$, these moments can be derived by computing $\partial L/\partial w$ and $\partial L/\partial b$.

$$\frac{\partial \text{MSE}}{\partial w} = \frac{1}{s}\sum_{i=1}^{s} 2\left(\hat{y}_i - y_i\right)\cdot x_i \tag{2.8}$$

$$\frac{\partial \text{MSE}}{\partial b} = \frac{1}{s}\sum_{i=1}^{s} 2\left(\hat{y}_i - y_i\right) \tag{2.9}$$

In (2.8) and (2.9), x_i denotes the input connected to the weight w, y_i refers to the ground truth value of the model output $\hat{y}_i$ and s denotes the batch size. For each w and b, the first m_1 and second m_2 moment is computed with

$$m_{1,r} = \beta_1 \cdot m_{1,r-1} + (1-\beta_1)\cdot \nabla_\theta L \tag{2.10}$$

$$m_{2,r} = \beta_2 \cdot m_{2,r-1} + (1-\beta_2)\cdot \left(\nabla_\theta L\right)^2 \tag{2.11}$$

where $\nabla_\theta L$ equals either $\partial L/\partial w$ or $\partial L/\partial b$, depending on which weight/bias is being updated. The parameters β_1 and β_2 refer to the exponential decay rate of the first moment and second moment respectively, and have to be manually defined before training the model. r denotes the current batch, of which the maximum is the number of samples in the training dataset divided by the batch size (s). The first initialisation usually puts the weights and biases to random values and the moments $m_{1,0}$ and $m_{2,0}$ to zero. Next, with the first and second moments of the gradients, they are corrected for a bias with (2.10) and (2.11) to account for the induced bias of setting the initial moments to zero.

$$\hat{m}_{1,r} = \frac{m_{1,r}}{1-\beta_1^r} \tag{2.12}$$

$$\hat{m}_{2,r} = \frac{m_{2,r}}{1-\beta_2^r} \tag{2.13}$$

(2.12) and (2.13) are used to update all weights and biases according to

$$w_r = w_{r-1} - \alpha\frac{\hat{m}_{1,r}}{\sqrt{\hat{m}_{2,r}}+\epsilon} \tag{2.14}$$

$$b_r = b_{r-1} - \alpha\frac{\hat{m}_{1,r}}{\sqrt{\hat{m}_{2,r}}+\epsilon} \tag{2.15}$$

where α is the learning rate and ϵ is a small constant meant for numerical stability, meaning that the term avoids any division by zero. When the weights are updated, the algorithm moves on to the next batch in the training dataset. Upsides to this algorithm include simplicity, computational efficiency, little memory requirements and suitability for large datasets, non-stationary target variables and noisy gradients. It also requires little to no tuning of its parameters. Its default parameters are $\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1 * 10^{-8}$, as defined by Kingma and Ba (2014).

Learning rate and batch size

During the training process, the input data are divided into chunks, called batches, and the model's trainable parameters (weights and biases) are updated based on the loss computed over each batch, using the optimisation algorithm to do so. The choice of batch size can impact training speed, memory requirements, and the stability of the learning process. Larger batch sizes often lead to faster training but may require more memory, while smaller batch sizes may result in slower training but can offer more accurate parameter updates.

3

Study area and data overview

3.1. Dutch Coast and North East Atlantic Ocean

The study area of this project will encompass the North Sea as shown in Figure 3.1. This part of the world is rich in tide gauges and has an elaborate coastal variability, which makes it suitable for developing this pilot model. The study will mostly focus on the Dutch coastal zone. The coastal areas around Denmark and Norway have been left out, since any information in these areas is not expected to have much effect on the water levels along the Dutch coast. This is due to the locations of the amphidromic points in the North Sea. They cause the tidal propagation to reach the Dutch coast before it reaches the coasts of Denmark and Norway due to the counter-clockwise progression of the phase (Figure 2.2). This suggests that any land-sea interactions induced by the tides and present in the non-tidal water levels around the Danish and Norwegian coasts do not interact with the Dutch coast and therefore, would cause only noise in the data. The large inland water body of the IJsselmeer is also excluded from the area of interest because this water is disconnected from the North Sea by a dam and as such will not provide useful information on water levels along the Dutch coast. Data over land is also not used.

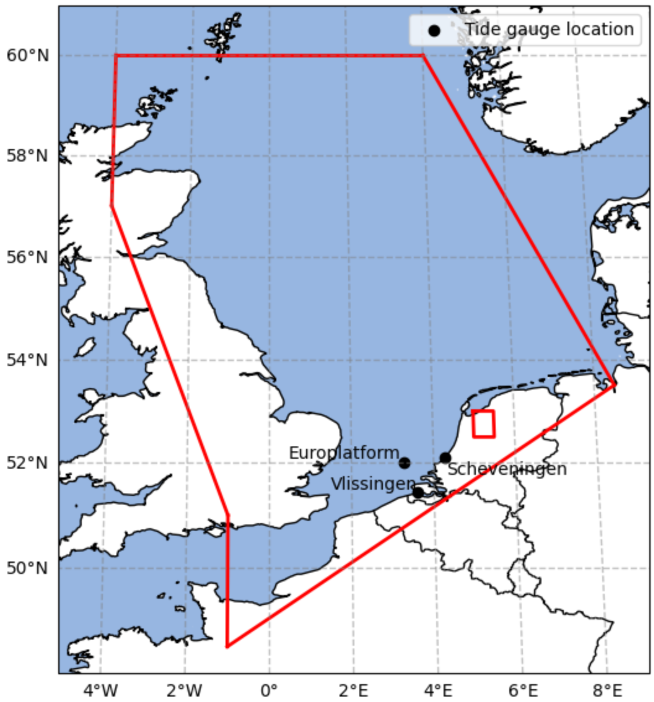


Figure 3.1: Extent of the area of interest. The red line denotes the edge of the area of interest. Any data within the IJsselmeer has been discarded, since this is considered inland water. The selected tide gauges on which the model will be trained are shown in black.

The data needed within this area include satellite altimetry data, ERA5 assimilated data, tide gauge records and information on the geoid. These datasets are acquired from different sources (Table 3.1). The tide gauge records act as ground truth datasets. Retracked and reprocessed satellite altimetry data from X-TRACK is chosen, because the X-TRACK algorithm can conserve more satellite observations in coastal regions than traditionally processed satellite altimetry data. ERA5 data is added to provide information on the sea level pressure fields and wind speeds. Finally, the geoid is used to reference the satellite altimetry observations to.

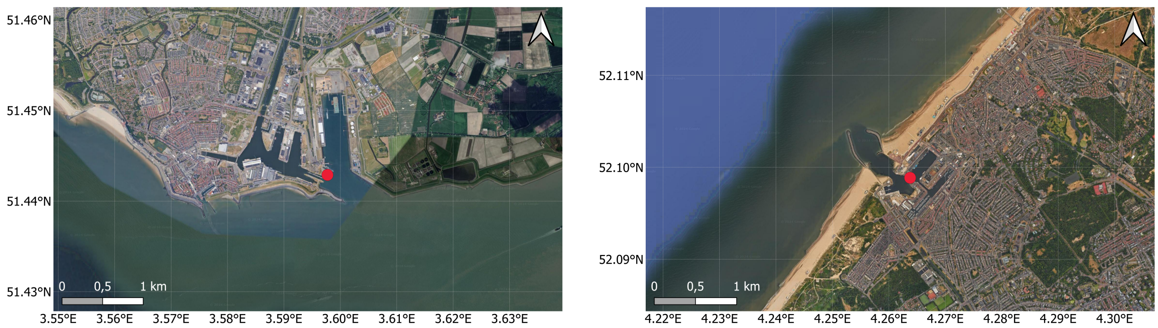
Table 3.1: Data sources

Data	Source	Reference
Tide gauge	GESLA	Haigh et al., 2022
Satellite altimetry	X-TRACK	Birol et al., 2021
ERA5	Climate Data Store (CDS)	Hersbach et al., 2018
Geoid	International Service for the Geoid (ISG)	Denker, 2013

3.2. Tide gauge records

For the development of a machine learning model that can produce reliable estimates of water levels, accurate control data is needed on which the ML model can be trained. This data is taken from the GESLA version 3 database, which is a global database of high-frequency tide gauge time series (Haigh et al., 2022). Three tide gauges in the Dutch coastal area are selected to serve as ground truth stations: Vlissingen, Scheveningen and Europlatform (Figure 3.1). These locations are chosen intentionally for their extensive and continuous data records, and to represent the large variability of the Dutch coastal region while simultaneously account for any time constraints.

Vlissingen lies in a sheltered bay in the estuary of the Westerschelde, experiencing a large tidal variability and more land-sea interactions of the water level than Scheveningen or Europlatform (Figure 3.2a). This can increase flood probabilities during storms and induce larger risks than in other locations. Scheveningen is facing the North Sea more openly, experiencing larger effects of wind set-up and less tidal variability than Vlissingen (Figure 3.2b). The population density is also highest around this region, making it essential for flood risk analysis due to its high possible damage factor. Finally, the Europlatform is surrounded by sea on all sides, experiencing the fewest land-sea interactions and tidal variability. Including this station allows for comparisons between the stations and assess how much the land-sea interactions matter in estimating the non-tidal water levels. The data for the Dutch coast is collected by the Dutch Ministry of Infrastructure and Water Management (Rijkswaterstaat, 2024). After downloading the data from GESLA (2024), some metadata is provided for each file. The most important metadata variables are presented in Table 3.2.



(a) (b)

Figure 3.2: Locations of the (a) Vlissingen and (b) Scheveningen tide gauge, taken from Google Earth imagery.

Table 3.2: Most important metadata variables for the tide gauges of interest. The column on the left denotes the variable names as they are provided in the record files (Haigh et al., 2022).

SITE NAME	Vlissingen	Scheveningen	Euro_platform
LATITUDE	51.44230800	52.09903300	51.99861100
LONGITUDE	3.59605700	4.26356300	3.27638900
START DATE/TIME	1863/09/01 07:40:00	1946/01/01 01:40:00	2001/06/30 23:00:00
END DATE/TIME	2018/09/08 00:00:00	2018/09/07 23:50:00	2018/09/07 23:50:00
DATUM INFORMATION	NAP	NAP	NAP
NULL VALUE	-99.9999	-99.9999	-99.9999

For all three locations, the datum information provided is Normal Amsterdam Level (NAP). This is nearly equivalent to the geoid, reporting water levels in the order of a few metres. The recorded data is provided in five columns:

1. Date yyyy/mm/dd
2. Time hh:mm:ss
3. Observed sea level (m)
4. Observed sea level QC flag

   0 - no quality control
   1 - correct value
   2 - interpolated value
   3 - doubtful value
   4 - isolated spike or wrong value
   5 - missing value

5. Use-in-analysis flag (1 = use, 0 = do not use)

The quality control (QC) flags for column 4 refer to a quality control applied to the data by the provider, in this case, Rijkswaterstaat. For column 5, all instances where the QC flags indicate a 0, 1 or 2 have been flagged with a 1.

3.3. Satellite altimetry

The altimetry data used in this study were developed, validated, and distributed by the CTOH/LEGOS, France (Aviso+, 2023), and have been provided to Deltares for the project Earth Observation Advanced science Tools for Sea Level Extreme Events (ESA, 2023b). The data have already been reprocessed using the X-TRACK retracking method (Birol et al., 2021). This method has an improved derivation for coastal observations of the sea surface height. Since we are interested in coastal water levels, this data is preferable over traditionally processed altimetry data.

Retracked data from the satellite missions TOPEX/Poseidon, ERS-2, GeoSat Follow-On (GFO), Jason-1, Envisat, Jason-2, Haiyang-2 (HY-2), SARAL, Jason-3 and Sentinel-3A (S3A) are provided (Figure 3.3). All satellites carry an active radar sensor that uses ranging techniques to observe the sea surface, as explained in further detail in Section 2.2. Operational times, repeat cycles and instruments differ per mission, consequently also varying in data availability and accuracy (Table 3.3). For more information on the individual missions, see Appendix C.

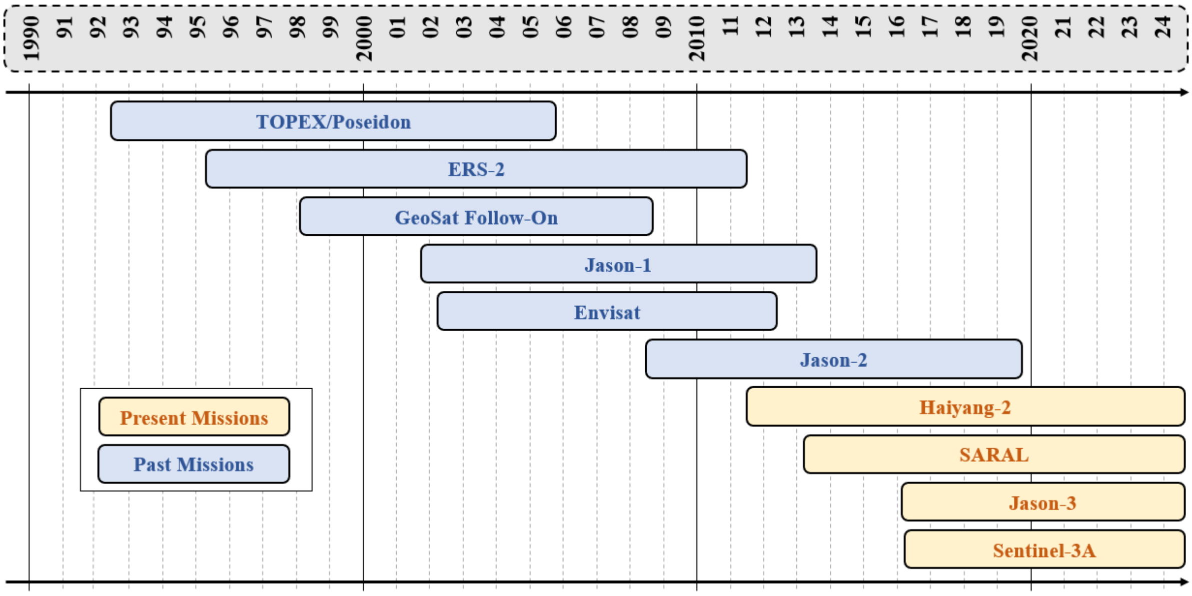


Figure 3.3: Timeline of satellite altimeters since 1992 which are used in this thesis (updated from Shu et al., 2021)

Table 3.3: Overview of satellite altimeters provided (Shu et al., 2021; CEOS and ESA, 2024)

Altimetry mission	Operational time		Repeat cycle	Accuracy (RMSE)	Instrument
	Start	End			
TOPEX/Poseidon	10-08-1992	09-10-2005	10 d	2.4/2.5 cm	NRA/Poseidon-1 (SSALT-1)
ERS-2	21-04-1995	05-07-2011	35 d	10 cm	RA-1
GeoSat Follow-On	10-02-1998	17-09-2008	17 d	3.5 cm	GFO-RA
Jason-1	07-12-2001	21-06-2013	10 d	3.9 cm	Poseidon-2 (SSALT-2)
Envisat	01-03-2002	08-04-2012	35 d	4.5 cm	RA-2
Jason-2	20-06-2008	01-10-2019	9.9 d	3.9 cm	Poseidon-3
Haiyang-2	16-08-2011	-	14 d	4 cm	RA/HY-2
SARAL	25-02-2013	-	35 d	3.4 cm	AltiKa
Jason-3	17-01-2016	-	9.9 d	3.4 cm	Poseidon-3B Altimeter
Sentinel-3A	16-02-2016	-	27 d	3 cm	SRAL

A couple of missions are not used in this study, such as ERS-1, CryoSat, GeoSat and Sentinel-6. Either the operational period of these missions does not coincide with the period of interest, or the data of these missions have not been provided/reprocessed by CTOH/LEGOS.

The retracked altimetry data have been stored in NetCDF files, containing several variables. Each track has its own filename, with mission, zone and track number included (`ctoh.sla.ref.<MISSION>.<ZONE>.<TRACK_NUMBER>.nc`). The track number is based on the geographic location of repeated overpasses as shown in Figure 3.4, where missions with the same orbit are combined to create six graphs that encompass all tracks.

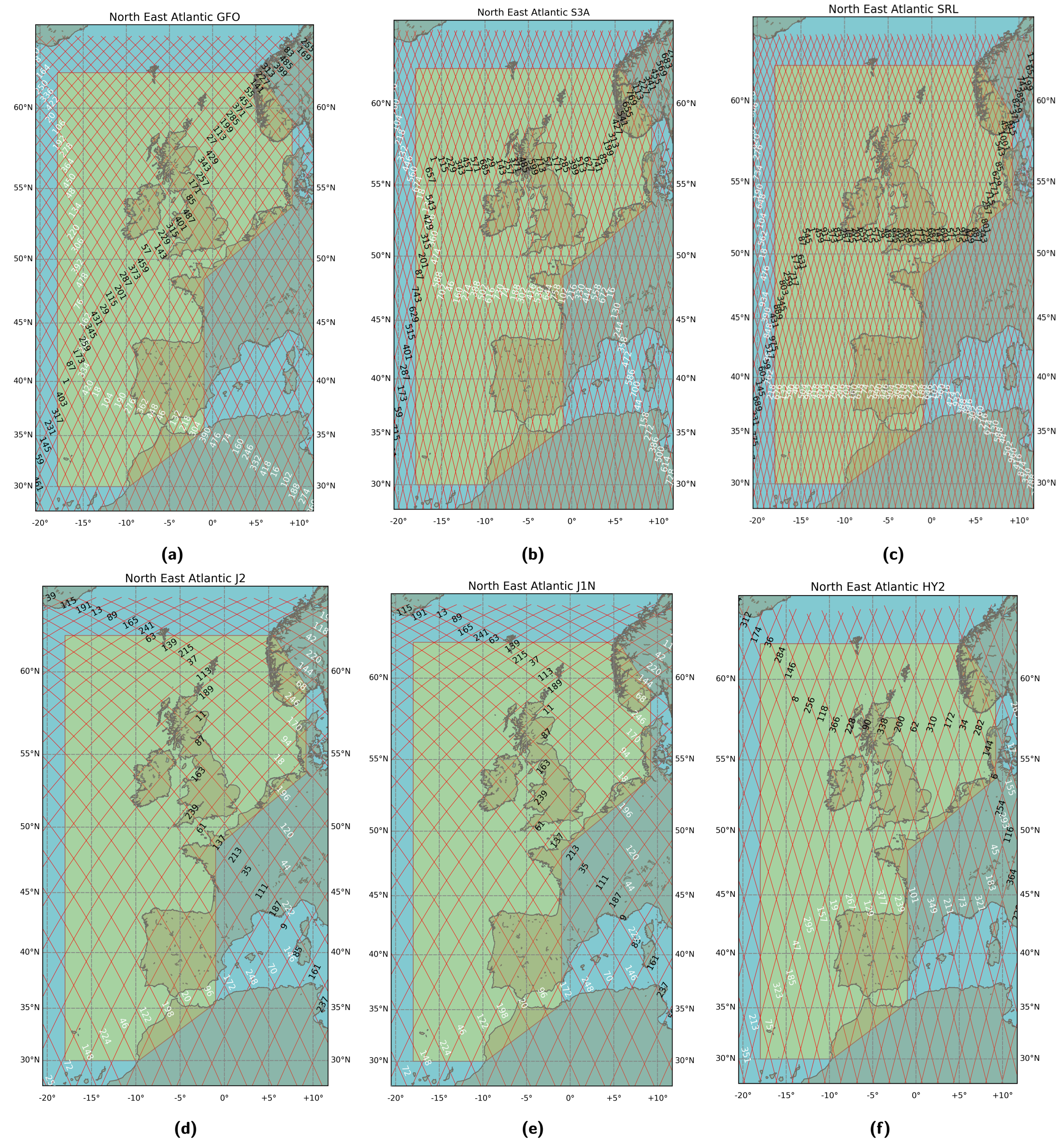


Figure 3.4: Numbered tracks of (a) GeoSat Follow-On, (b) Sentinel-3A, (c) ERS-2, Envisat and SARAL, (d) TOPEX/Poseidon and Jason-1/-2/-3, (e) TOPEX/Poseidon and Jason-1/-2 interleaved orbit and (f) Haiyang-2. Taken from Aviso+ (2023).

The variables taken from the XTRACK reprocessed data are summarised in Table 3.4. The time of each observation is given in days, referenced to 1950-01-01. Both the latitude and the longitude are given in degrees and refer to the nominal track (Figure 3.5). The DAC has been produced by the TUGO HF model forced with ERA5 pressure and wind fields and the IB LF contribution from (2.4) for the altimetry data before 02/2016 (Aviso+, 2022). After 02/2016, the DAC has been estimated with MOG2D HF forced with ECMWF pressure and wind. The SLA has been provided in the data, which has been computed with

$$\begin{aligned} \text{SLA} = & H - R_{obs} - \Delta h_{iono} - \Delta h_{dry} - \Delta h_{wet} - \Delta h_{SSB} \\ & - \text{ST} - \text{AT} - \text{PT} - \text{DAC} - \text{MSSH} - b_{MSSH} \end{aligned} \tag{3.1}$$

where all components have been discussed in Section 2.1 and 2.2 except for the global mean sea level bias (b_{MSSH}). This bias is applied to account for the interpolation that happens when the observations are resampled onto reference points along a nominal track (Figure 3.5). The MSSH is computed as the arithmetic mean of all observations in their respective cell within a reference period (Birol et al., 2017), where the b_{MSSH} accounts for the differences in the MSSH between the actual observation and the closest reference point, essentially correcting for any geoid variation (Birol & Passaro, 2020). The reference periods are defined below Figure 3.5.

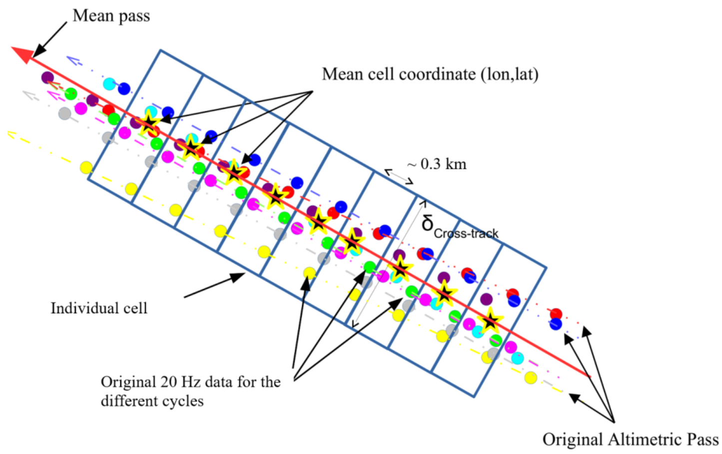


Figure 3.5: Illustration of a nominal track (red line and black stars), with their reference cycles (one colour for each cycle). For each shown cell, the nominal track is defined by the mean of the coordinates/MSSH of the reference cycles (Birol & Passaro, 2020).

- 1993-03-09 - 2021-03-05 for TOPEX/Poseidon and Jason-1/-2/-3
- 1995-06-18 - 2016-03-11 for ERS-2, Envisat and SARAL
- 2000-01-28 - 2008-01-22 for GFO
- 2002-09-25 - 2011-09-25 for TOPEX/Poseidon and Jason-1/-2 interleaved orbit
- 2014-04-19 - 2016-03-19 for HY-2
- 2016-03-12 - 2022-03-08 for S3A

The FES2014b model has been used to compute the astronomical tide (Lyard et al., 2021). The 34 constituents included in this model are K1, M2, N2, O1, P1, Q1, S1, S2, K2, 2N2, EPS2, J1, L2, T2, La2, Mu2, Nu2, R2, M3, M4, M6, M8, MkS2, MN4, MS4, N4, S4, MSF, Mf, Mm, MSqm, Mtm, Sa and Ssa. The remaining corrections and coefficients are computed as explained in Appendix A and B, and in further detail in Aviso+ (2022). The variables that are taken from each file are presented in Table 3.4, along with their units and dimensions.

Table 3.4: Parameters provided in X-TRACK NetCDF files (Aviso+, 2022)

Parameter	Description	Units	Dimensions
Time	Time of observation	Days since 1950-01-01	[Cycle, Lat, Lon]
Lat	Latitude of observation	Degrees North	[Lat]
Lon	Longitude of observation	Degrees East	[Lon]
MSSH	XTRACK MSS	Metres	[Lat, Lon]
SLA	Sea Level Anomaly	Metres	[Cycle, Lat, Lon]
DAC	Dynamic Atmosphere Correction	Metres	[Cycle, Lat, Lon]

3.4. ERA5

Additional information to teach the developed ML model is given as ERA5 hourly grids (Hersbach et al., 2018). From the Copernicus Climate Data Service (C3S), hourly fields for mean sea level pressure, wind speed and a land-sea mask are downloaded in NetCDF format. Within these files, the wind speed is split into a longitudinal component (U10) and a lateral component (V10), and given in metres per second (Table 3.5). The mean sea level pressure (p) is given in Pascal. The land-sea mask is provided as a value between

0 and 1. Pixels with a land-sea mask value of 0.5 and below are defined to be consisting of water (Hersbach et al., 2018). The spatial resolution of this data is ±31 km, and the area downloaded is from 48°N to 60°N and 4°W to 9°E.

Table 3.5: Parameters provided in ERA5 NetCDF files (Hersbach et al., 2018)

Parameter	Description	Units	Dimensions
Time	Time of data field	Hours since 1900-01-01	[Time]
Lat	Latitude	Degrees North	[Lat]
Lon	Longitude	Degrees East	[Lon]
p	Mean sea level pressure	Pascal	[Time, Lat, Lon]
U10	Longitudinal wind speed component at 10 metres above the Earth's surface	Metres/second	[Time, Lat, Lon]
V10	Lateral wind speed component at 10 metres above the Earth's surface	Metres/second	[Time, Lat, Lon]
lsm	Land-sea mask	-	[Lat, Lon]

3.5. Geoid

Since the datum for the tide gauge records is given in NAP, it is preferred to work with satellite altimetry data referenced to the geoid. Referencing satellite altimetry data to the geoid adds high-frequency spectral content due to the complex variations of the geoid surface. This can introduce additional properties to the signals that would be missed by the ML model if the SSH referenced to the ellipsoid was used. The geoid used in this study is the EGG2015 gravimetric model developed by Denker (2013). The model represents geoid heights referenced to the GRS80 ellipsoid in metres (Figure 3.6). It spans Europe from 25°N to 85°N and from 50°W to 70°E. The grid spacing in lateral and longitudinal direction is 0.1667° and 0.25° respectively.

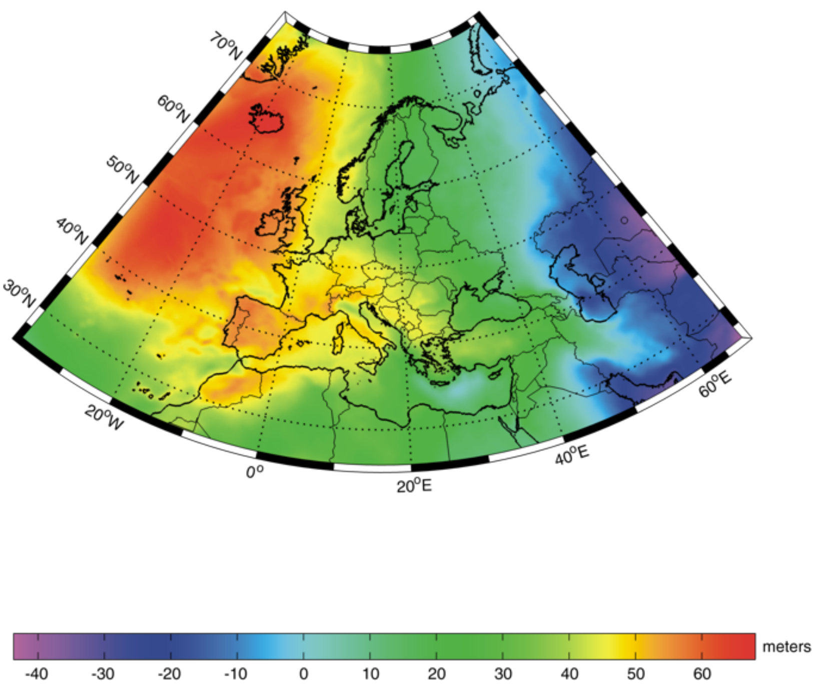


Figure 3.6: Representation of the EGG2015 geoid (Denker, 2013)

4 Method

To explore the potential of machine learning within the boundaries of this study, a simple neural network has been created that searches for the relationship between the coastal water level and the forcings that influence this water level. This chapter elaborates on the developed neural network with a bottom-up approach, focussing on the processing of the inputs and outputs, the design of the neural network, and a performance analysis to assess the robustness and efficiency of the final model. The chapter starts with a general outline of the method used in this study, after which the individual steps are further elaborated upon.

4.1. General approach

The first step is to pre-process the available datasets that are used to represent the forcings that influence the variability of the water level, namely satellite altimetry observations and ERA5 assimilated hourly datasets. Tide gauge records are considered as the ground truth data within this study, which also need pre-processing before they can be used to train the neural network. When all datasets are pre-processed, they are shaped into features to fit the requirements for the input into the ML model. Next, the neural network is designed to be simple yet complicated enough that it can identify different (non-linear) correlations between the input features and the ground truth data. The robustness of the model is assessed with a K-fold cross-validation, and the performance with tide gauge data. Finally, an additional analysis involving the trained parameters of the model is presented.

This method has been developed in Python 3.11.7, with the following packages and their versions as shown in Table 4.1.

Table 4.1: Python packages and their version used in this methodology

Package name	Version	Package name	Version
`cartopy`	0.22.0	`pandas`	2.2.0
`hatyan`	2.7.0	`pyextremes`	2.3.2
`keras`	2.15.0	`scikit-learn`	1.4.1
`matplotlib`	3.8.3	`scipy`	1.12.0
`netCDF4`	1.6.5	`tensorflow`	2.15.0
`numpy`	1.26.4		

4.2. Data pre-processing

The satellite altimetry observations are stored in NetCDF files grouped by mission and track. The ERA5 hourly gridded datasets are stored in NetCDF files grouped by year. The tide gauge records are stored in a file format including five columns, representing the date, time, the observed water level and two quality flags. For the data to be easier to work with, these files are pre-processed into one file for each data source. The preprocessing is done with the Python packages `netCDF`, `numpy`, `pandas` and `hatyan` (Veenstra & Kerkhoven, 2020).

4.2.1. Satellite altimetry

Within each NetCDF file of each mission, geographic coordinates of its nominal track are given. The tracks only span the North-East Atlantic coastal zone (Figure 3.4). The pre-processing steps of this dataset include selecting and masking the tracks so all observations fall within the area of interest and referencing the observations to include or exclude the desired signals within the observations.

Step 1: selection and masking

The first step is to select the files of which the track crosses the area of interest. The geographic locations of each track are retrieved, after which only the ones that cross the area of interest are selected (Figure 4.1). These files, along with a mask that defines which indices within the array of track locations fall within the area of interest, are saved for further processing. Note that the IJsselmeer is marked as being *outside* of the area of interest, since it is assumed that this inland water does not add important information to indicate the water level at the coast. It is expected to only cause noise since the lake is cut off from the sea.

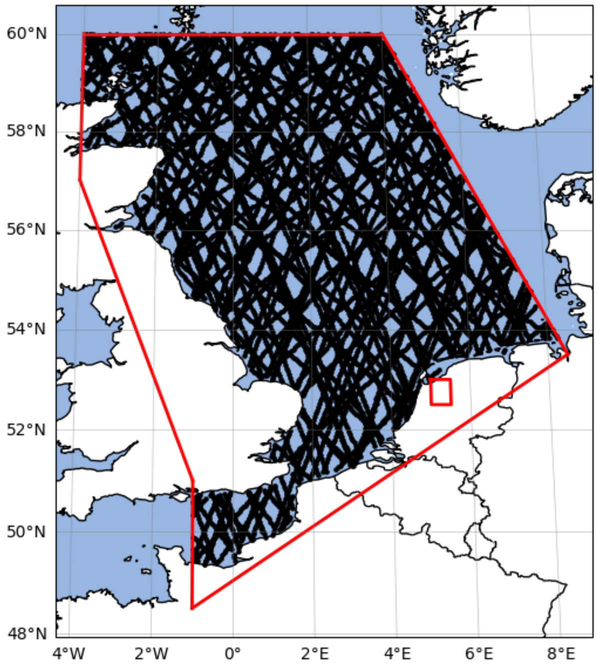


Figure 4.1: The tracks available within the area of interest. The outer red shape marks the edge of the area in which all data is gathered, with the exception of the IJselmeer. The black lines visualise all tracks available within the area of interest. Each track has its own file.

Step 2: referencing

From an altimetry file, the timestamp, mean sea surface (MSS), sea level anomaly (SLA), dynamic atmosphere correction (DAC), latitude and longitude of each observation are retrieved (Table 3.4). The timestamp is a decimal number that returns the days since 1950-01-01, with an accuracy of less than one second. This is converted to the correct date and stored with the remaining five variables.

The hourly non-tidal water level that is stated in the objective is best represented by the water level referenced to the geoid and not corrected for any atmospheric effects, defined as the non-tidal residual (NTR). Since the SLA retrieved from the satellite altimetry files includes both the MSSH and the DAC, and not the geoid, the retrieved variables are processed to represent the NTR with (4.1), using the provided data and the geoid given by Denker (2013).

$$\text{NTR} = \text{SLA} + \text{MSS} - \text{Geoid} + \text{DAC} \tag{4.1}$$

For each altimetry observation, the closest geoid pixel is taken to reference the observations to the geoid. The closest geoid pixel is found by using the Haversine distance method, where d denotes the distance between the altimetry observation and the centre of a geoid pixel in metres.

$$d = 2R \arcsin\left(\sqrt{\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \cos\varphi_1 \cdot \cos\varphi_2 \cdot \sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right) \tag{4.2}$$

In (4.2), R refers to the radius of the Earth, which is assumed to be constant at $6.371 * 10^3$ metres. The variables φ and λ refer to the latitude and longitude of the satellite altimetry observation, expressed in radians. The distance d is then computed in metres. Next to the NTR, two more variables are added in this step, namely the distance between each observation to the tide gauge of interest in lateral and longitudinal direction, computed by (4.3) and (4.4) respectively. Like the computation of the Haversine distance, it is assumed that the Earth is a perfect sphere, and both directional distances are computed with the assumed constant radius of the Earth of $6.371 * 10^3$ metres. It is assumed that this approach is sufficiently accurate when working with distances of this magnitude. Since the satellite altimetry files provide the latitudes and longitudes in degrees, these are first converted to radians before applying any of the functions (4.2), (4.3) or (4.4).

$$d_N = R * (\lambda_2 - \lambda_1) \tag{4.3}$$
$$d_E = R * (\varphi_2 - \varphi_1) \tag{4.4}$$

The final variables which are suited for machine learning are saved in a binary `numpy` file (`.npy`) (Table 4.2).

Table 4.2: Final satellite altimetry variables suited for machine learning.

Variable name	Description	Units
Time	Time stamp	yyyy-mm-dd hh:mm:ss
NTR	Non-tidal residual referenced to the geoid	Metres
d$_N$	Distance satellite observation to TG in lateral direction	Metres
d$_E$	Distance satellite observation to TG in longitudinal direction	Metres

4.2.2. ERA5

The assimilated datasets of ERA5 provide hourly gridded datasets of sea level pressure and wind speed, along with a yearly land-sea mask. These variables are downloaded from the Copernicus Climate Data Service (C3S) Climate Data Store (CDS) (Hersbach et al., 2018), after which they are clipped to the area of interest and masked for any pixels containing land (Figure 4.2).

The raw timestamps provided by the hourly datasets are given in integers, returning the number of hours since a reference date of 1900-01-01 at 00:00:00. Table 4.3 shows the variables that are saved in a binary `numpy` file (`.npy`). Every clipped hourly dataset contains 1498 pixels, and all pixels are compressed to form single arrays per hour.

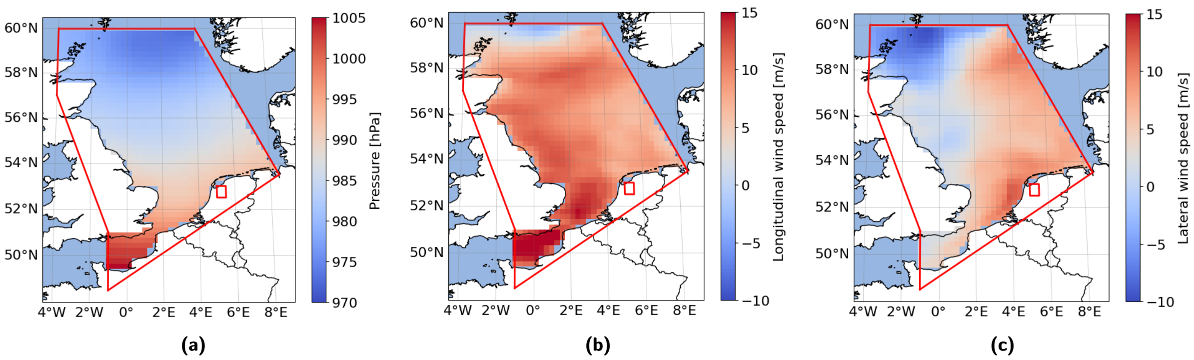


Figure 4.2: Examples of an ERA5 dataset on 2018-01-01 00:00:00, for (a) sea level pressure, (b) longitudinal 10-m wind speed component and (c) lateral 10-m wind speed component. The data is clipped to the area of interest and masked with a land-sea mask that masks all pixels which have a mask larger than 0.5. The outer red shape marks the edge of the area in which all gridded data is gathered, with the exception of the IJsselmeer.

Table 4.3: Final ERA5 variables suited for machine learning.

Variable name	Description	Units
Time	Time stamp referenced to 1900/01/01	Hours
p	Sea level pressure	Pascal
U10	10-m wind speed in longitudinal direction	Metres / second
V10	10-m wind speed in lateral direction	Metres / second

4.2.3. Tide gauge

Three tide gauges are considered to train and test the neural network developed in this study: Vlissingen, Scheveningen and Europlatform (Figure 3.1). The steps taken to pre-process the tide gauge records include the selection of hourly measurements and the tidal correction to account for harmonic tidal signals.

Step 1: selecting hourly measurements

The hourly values within the provided time series are selected by creating a separate time column and deleting all measurements before 1992-01-01. This time column has a frequency of ten minutes for Scheveningen and Europlatform and a frequency of one minute for Vlissingen. Next, the measurements taken on each full hour are selected to act as ground truth for the machine learning model (Figure 4.3). If there is no measurement available on the full hour, the method uses a nearest neighbour approach to find the next available measurement within the considered hour.

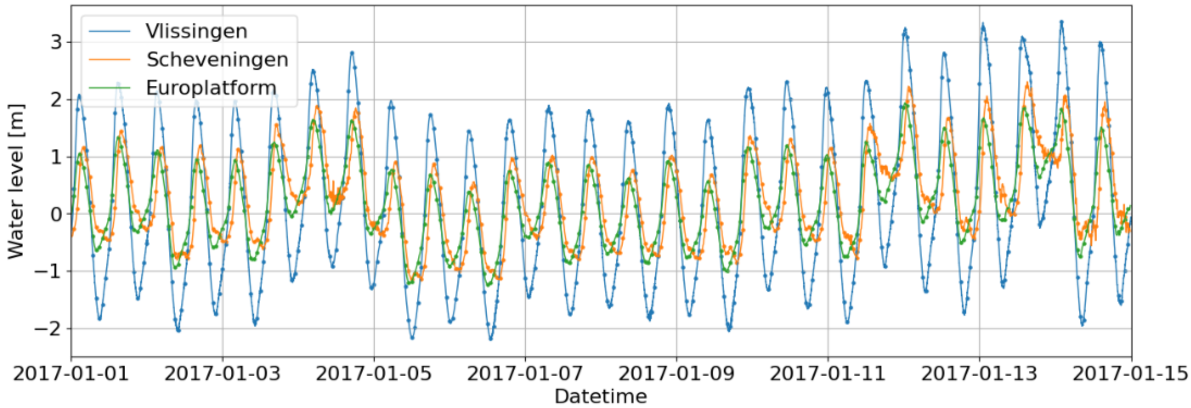


Figure 4.3: Water level measurements from tide gauges at Vlissingen, Scheveningen and Europlatform from 2017/01/01 to 2017/01/15. The dots represent the hourly values which are used as ground truth data in the machine learning model.

Step 2: tidal correction

To correct for the harmonic signals of the tides, the Hatyan method is applied (Veenstra & Kerkhoven, 2020). This method computes 95 harmonic components (see Appendix D), based on periods of one year. The original hourly measurements are corrected for the harmonic tidal signals by reconstructing a harmonic signal using the computed 95 components (Figure 4.4), and the resulting time series are saved as `.csv` files.

After pre-processing, the satellite altimetry observations and the tide gauge observations refer to the same vertical datum, namely the geoid, which is desirable when feeding the data into a machine learning model. Additionally, the satellite altimetry data is corrected for unwanted signals and clipped to the area of interest. The ERA5 datasets are also clipped and flattened into 1D arrays to allow easier shaping of input data for the developed neural network.

Figure 4.4: Water level measurements from tide gauges at Vlissingen, Scheveningen and Europlatform from 2017/01/01 to 2017/01/15, after applying the Hatyan method to correct for harmonic tidal signals. The values are hourly.

4.3. Input preparation

The desired input variables for the neural network include the non-tidal residuals (NTR), the distance components of the satellite observations to the tide gauge of interest (d_N and d_E), the time difference of the satellite observations to the time stamp of interest (dt), the sea level pressure (p), wind speeds in the longitudinal and lateral direction (U10 and V10), and the day of year (DOY) (Table 4.4). The DOY has been added to account for seasonal variability of the water level, for example the larger possibility of storm surges in winter months compared to summer months. The input preparation can be subdivided into five steps (Figure 4.5). The input preparation is done with the Python packages `pandas`, `numpy`, `tensorflow` and `scipy`.

Table 4.4: Input variables needed for machine learning.

Variable name	Description	Units	Source
NTR	Non-tidal residual	Metres	Satellite altimetry
d_N	Distance between satellite observation and tide gauge in lateral direction	Metres	Satellite altimetry
d_E	Distance between satellite observation and tide gauge in longitudinal direction	Metres	Satellite altimetry
dt	Time difference between satellite observation and desired time stamp	Seconds	Satellite altimetry
p	Sea level pressure	Pascal	ERA5
U10	Longitudinal 10-m wind speed component	Metres / second	ERA5
V10	Lateral 10-m wind speed component	Metres / second	ERA5
DOY	Day of the year	Days	-

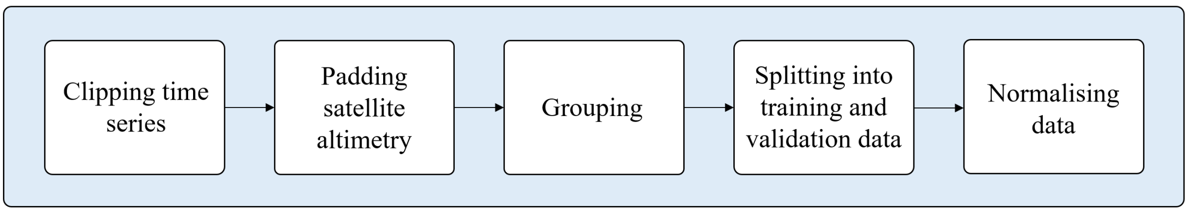


Figure 4.5: Step-wise overview of the method used to prepare the input data for developed machine learning model.

4.3.1. Clipping time series

Clipping all data to a preferable time series is the first step in creating the input data for the developed machine learning model. This is done by setting a *start* and *end* date, and then selecting all data within those two dates. For the pre-processed satellite altimetry dataset and ERA5 dataset, the start date is shifted with 48 hours, while the end date remains equal to the original. This is a crucial step, since the objective is to estimate hourly water levels based on data from the *previous* 48 hours. To make sure that the models for all three locations of interest can be compared, the start and end dates for all models must be the same. Additionally, for the best performance, it is important to have as much data as possible. These conditions result in a start date of 2001-07-01 at 00:00:00 and an end date of 2018-09-07 at 23:00:00.

4.3.2. Padding satellite altimetry

To be able to estimate water levels based on 48 hours of data, the input datasets are grouped into subsets of 48 hours. For satellite altimetry, an additional step is needed before the data can be grouped, because the neural network needs input data of a fixed length. This causes a complication for satellite altimetry, since not every subset of 48 hours contains the same number of satellite observations. This is caused by the fact that there are multiple satellite missions involved, which do not have the same repeat cycle and orbit. Some subsets have no observations, while others have over a thousand (Figure 4.6). This is dealt with by padding the satellite altimetry data with zeros (Figure 4.7). Consequently, the machine learning model must have a masking layer as the first layer, since these zeros cannot be taken into account when training the model. The largest number of observations within one hour is 514, meaning all hours are padded until they contain 514 values.

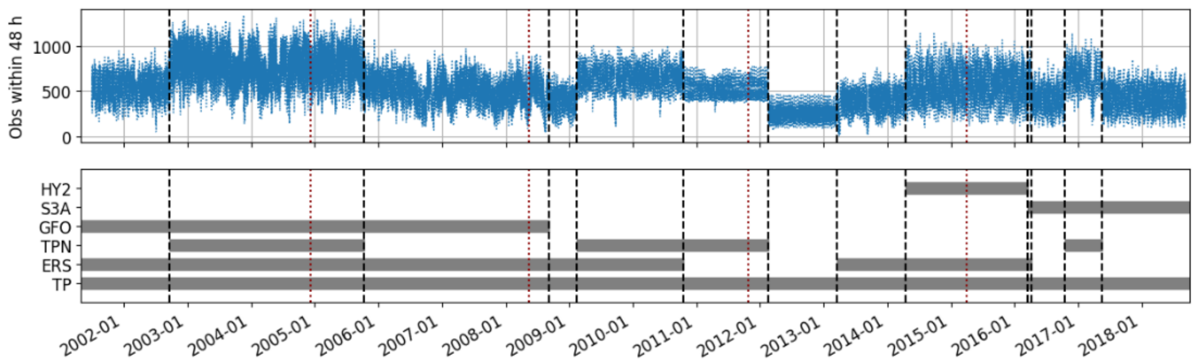


Figure 4.6: Data availability for the full period of interest. The upper graph shows the number of observations within the 48-hour subsets. The lower graph shows the period at which altimetry data from different missions is available. HY-2 refers to Haiyang-2, S3A to Sentinel-3A, GFO to GeoSat Follow-On, TPN to Topex/Poseidon and Jason-1/-2 on an interleaved track, ERS to ERS-1/-2, Envisat and SARAL and TP to Topex/Poseidon and Jason-1/-2/-3. The black dashed lines refer to the start/end of missions and the red dotted lines refer to the four validation datasets for K-fold cross-validation and the testing dataset.

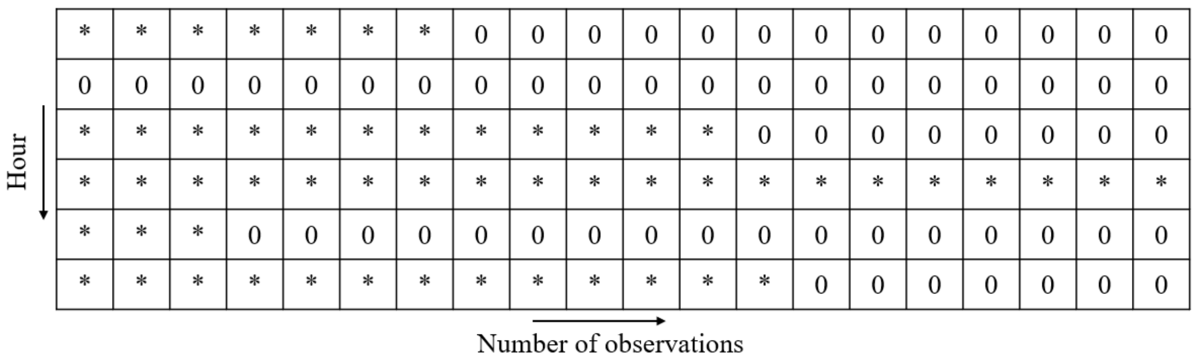


Figure 4.7: Example of padding satellite altimetry, where stars (*) refer to satellite observations and one row corresponds to one hour. The hour with the largest number of observations is taken (fourth row), after which all other hours are padded with zeros until every hour has the same shape, including hours which have no observations (second row). In reality, this matrix is much larger for each 48-hour subset (48 rows and 514 columns).

4.3.3. Grouping

To create the 48-hour subsets the neural network needs to estimate hourly water levels, the three datasets (satellite altimetry, ERA5 and tide gauge time series) are grouped. The group size depends on the task the data is to perform within the neural network. Since the tide gauge time series will be used as ground truth data, they need to be grouped with a group size of one hour, while the satellite altimetry dataset and ERA5 dataset need to be grouped with a group size of 48 hours. This results in one ground truth value for each 48 hours of input data. One exception is the DOY. For each ground truth value, the machine learning model has to know on which day of the year this value occurs to be able to link the 48 hours of input data to the correct physical processes (either winter or summer months). This means that of all the input datasets, only the DOY has a group size of one hour.

Table 4.5: Group size per data source.

Data source	Group size	Input shape
Satellite altimetry	48 hours	(4 * 48 * 514)
ERA5	48 hours	(3 * 48 * 1498)
TG time series	1 hour	(1)
Day of year	1 hour	(1)

4.3.4. Splitting into training and validation sets

Now that the datasets have been grouped into the correct subsets, they can be combined into one large input dataset. To enable K-fold cross-validation, the training, testing and validation data is created by splitting this dataset into five *folds* (Figure 4.8). Each fold has a size equal to 20% of the full input dataset.

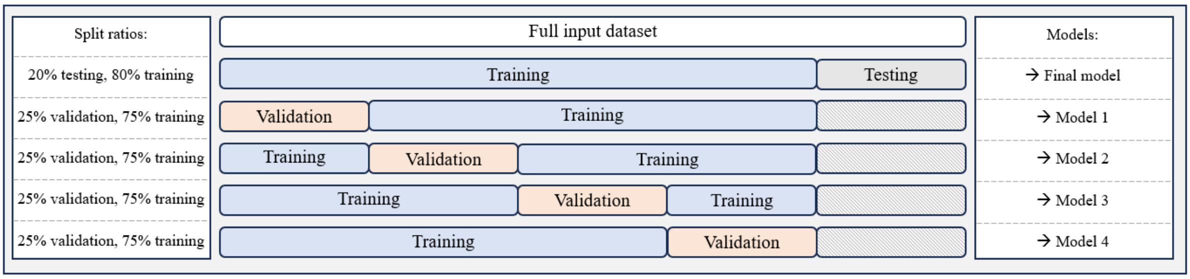


Figure 4.8: Overview of how the input data is split into testing, validation and training data for training and K-fold cross-validation. The folds are shown as orange boxes (validation data) or grey boxes (testing data).

The last fold will be used as a testing dataset, and will only be used to test the performance of the machine learning model in the final stage of this study. This means it will not be used for the K-fold cross-validation. The remaining data is split into 25% validation and 75% training data. Four different validation subsets can now be used for the K-fold cross-validation, resulting in four different training datasets, and ultimately, four different models. The time spans of the different validation subsets are presented in Table 4.6.

Table 4.6: Validation subsets used for the K-fold cross-validation.

	Start date	End date
Validation set 1	2001-07-01 00:00:00	2004-12-07 13:00:00
Validation set 2	2004-12-07 14:00:00	2008-05-16 03:00:00
Validation set 3	2008-05-16 04:00:00	2011-10-23 17:00:00
Validation set 4	2011-10-23 18:00:00	2015-04-01 09:00:00
Testing dataset	2015-04-01 10:00:00	2018-09-07 23:00:00

4.3.5. Normalising data

The next step in preparing the subsets for their use as input is to normalise the data. This helps the neural network to learn about the correlations between the input and the desired output, without having to worry about large differences in the magnitude of the different input parameters. A Z-Score normalisation method is applied for all ERA5 variables (p, U10, V10), as well as for three out of four satellite altimetry variables

(NTR, d_N, d_E), since they are assumed to be Gaussian distributed. This assumption is based on an analysis of the input variables (Appendix E), where the distribution of each variable is shown. The variables p, U10, V10 and NTR follow the Gaussian form better than d_E and d_N do. Some discussion on this is given in Section 6.6. The variables DOY and dt have been normalised with a min-max normalisation since they are assumed to be uniformly distributed.

Z-Score normalisation

A Z-Score normalisation is defined as a scaling of the dataset to change its distribution to one with a mean of zero and a standard deviation of one (Kreyszig, 1979). For each configuration of the data shown in Figure 4.8, the training data is taken to compute the mean and standard deviation of each variable. The Z-Score (z_i) is then computed by subtracting the mean (μ) from the input value and dividing it by the standard deviation (σ) as shown in (4.5). The training dataset as well as the validation dataset are normalised with the mean and standard deviation computed from the training dataset.

$$\mu = \frac{1}{m}\left(\sum_{i=1}^{m} x_i\right) = \frac{x_1 + x_2 + \cdots + x_m}{m}, \qquad \sigma = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(x_i - \mu)^2}, \qquad z_i = \frac{x_i - \mu}{\sigma} \tag{4.5}$$

Min-max normalisation

The final two variables (DOY and dt) are being normalised with the min-max normalisation. It takes the minimum and maximum values of the dataset, scaling the data according to (4.6).

$$x_i' = \frac{x_i - \min(\mathbf{x})}{\max(\mathbf{x}) - \min(\mathbf{x})} \tag{4.6}$$

For the DOY and dt, the maximum value is set on 365 days and 48*3600 seconds respectively, while the minimum value is set on 1 day and 0 seconds. This means the training and validation data will be scaled to a value between 0 and 1.

4.3.6. Resampling

The ground truth dataset used in this study is imbalanced, meaning that the distribution is skewed towards the mean of the dataset. Very high water levels or very low water levels are therefore underrepresented in the data, causing the neural network to focus more on the mean output values than on the high or low values, simply because they are more abundant in the data. To battle this, the training data is resampled to include more high non-tidal water levels.

Resampling of the training data happens based on the Cumulative Distribution Function (CDF) of the *absolute* ground truth data. A threshold of 0.8 has been used to determine which water levels should be sampled more often. The ratio between the water levels below the threshold and above the threshold is then set at 0.05. With this ratio and the true ratio found by dividing the number of water levels above the threshold by the number of water levels below the threshold, a constant is found according to (4.7). This constant is then used to repeat the selected high water levels.

$$q = \frac{0.05}{(\text{Number of observations with CDF} > 0.8)\,/\,(\text{Total number of observations})} \tag{4.7}$$

Finally, sample weights are applied to assign different levels of importance to individual samples during the training process. Each water level in the ground truth dataset is associated with a weight based on the CDF, reflecting its significance in the learning task. This means that higher water levels get higher weights. This weight is fed to the model alongside the input data and ground truth value associated with it. Higher weights indicate that a particular sample should contribute more to the models updates during training.

When the training dataset has been resampled, it is fed into the neural network. For each ground truth group containing one water level observation from the desired tide gauge, an input group of 48 hours containing three ERA5 variables, four satellite altimetry variables, and one additional variable is fed. Every single value within these input groups is referred to as an *input feature*, while the output values associated with them are called *sample values*.

4.4. Neural network design

The design for this neural network has been kept purposely shallow, meaning it has only one hidden layer, to minimise the level of complexity within the model. A second layer does not improve the results significantly based on intermediate results which are not included in this report. The design parameters that are non-trainable by the model itself are given in Table 4.7.

Table 4.7: Design parameters for the shallow neural network

Variable	Value	Variable	Value
Number of neurons	32	Batch size	32
Activation function hidden layer	ReLU	Loss function	MSE
Activation function output layer	Linear	Sample weights	CDF
Optimisation algorithm	Adam	Number of epochs	30
Learning rate	0.0001		

A number of 32 neurons is chosen, because it has been found that this number causes the model not to underfit or overfit on the data. The same holds for the number of epochs. After 30 epochs, the model has sufficiently converged without an increase in the validation loss. The only parameter of the Adam optimisation algorithm which is changed is the learning rate (α). It is set to 0.0001 to account for the large variability within the model. It has been found that the default of 0.001 results in less accurate results. The batch size for this method is a trade-off between fast training and accurate parameter updates and is chosen to be 32. These choices have been based on intermediate results during the development phase of this model which are not included in this report, because these results have not been saved. The neural network is built and trained with the Python package `tensorflow` and defined as shown in the code snippet below.

Model initialisation and training

Model architecture

```
model = tf.keras.Sequential([
  tf.keras.layers.Masking(mask_value=0.0, input_shape=input_shape),
  tf.keras.layers.Dense(32, activation="relu", dtype="float32"),
  tf.keras.layers.Dense(1, dtype="float32")
])
```

Set optimiser and learning rate

```
optimizer = tf.keras.optimizers.Adam(learning_rate=0.0001)
```

Compile the model with the right loss function

```
model.compile(optimizer=optimizer, loss="mean_squared_error",
  metrics=["MeanAbsoluteError"])
```

Batch the training dataset

```
train_batch = train_dataset.batch(32)
```

Fit the model

```
history = model.fit(train_batch, validation_data = val_dataset, epochs=30)
```

4.5. Performance analysis

After training the neural network on Scheveningen, Vlissingen and Europlatform, the performance is assessed by applying the model to the testing data. First, the general performance is assessed by looking at the mean squared error (MSE), the signal-to-noise ratio (SNR), the mean absolute error (MAE) and the standard deviation of the error (σ_e). The same metrics are used to assess the performance of DCSM for a similar period to allow for comparisons between the neural network and DCSM.

- **Mean Squared Error**
  The MSE is a popular metric for regression models. It penalises larger errors and since it has been used as the loss function in the neural network, gives a direct measure of the performance of the model.
- **Signal-to-noise ratio**
  The SNR is a measure of how the MSE of the model relates to the variability of the original signal. A higher SNR indicates that the model errors are small compared to the variance of the water level, while a lower SNR indicates that the model errors are too high for the output to be applicable. Therefore, it can be used to determine the physical applicability of the model.
- **Mean Absolute Error**
  The MAE computes how far off an estimation of the model is expected to be from the ground truth data.
- **Standard deviation of the error**
  The final metric σ_e is also known as the standard error and represents the spread of the error, which can be used to compute 95% or 99% confidence intervals.

The robustness and model variability are tested with a K-fold cross-validation with $K = 4$ for Scheveningen. The neural network has been trained on four different training subsets, and tested on four different validation subsets (Table 4.6). The same metrics have been computed for the four different models, creating an upper limit for the performance and model variability.

4.5.1. Performance on high water levels

In addition to the overall performance that has been computed with the metrics described above, the same performance is assessed during high water level events. The high water level events are selected with a peak-over-threshold (POT) method, for which the thresholds differ per location (Table 4.8). The POT method is applied to the observed water levels, after which the associated ML estimates and the corresponding DCSM values are obtained. The selection of high water levels is done using the Python package `pyextremes` (Bocharov, 2023). For more information on the threshold selection, see Appendix F.

Table 4.8: Selected thresholds for each location. The thresholds are selected with a POT method to assess the model performance on high water levels.

	Scheveningen	Vlissingen	Europlatform
Threshold [m]	1.70	2.70	1.50

To ensure that the performance metrics are valid, the high water level events need to be independent and identically distributed. This is achieved by applying a declustering window of 48 hours when selecting these events. For every observed event that exceeds the threshold, the method makes sure that this event is represented by its peak value within the time window (Figure 4.9).

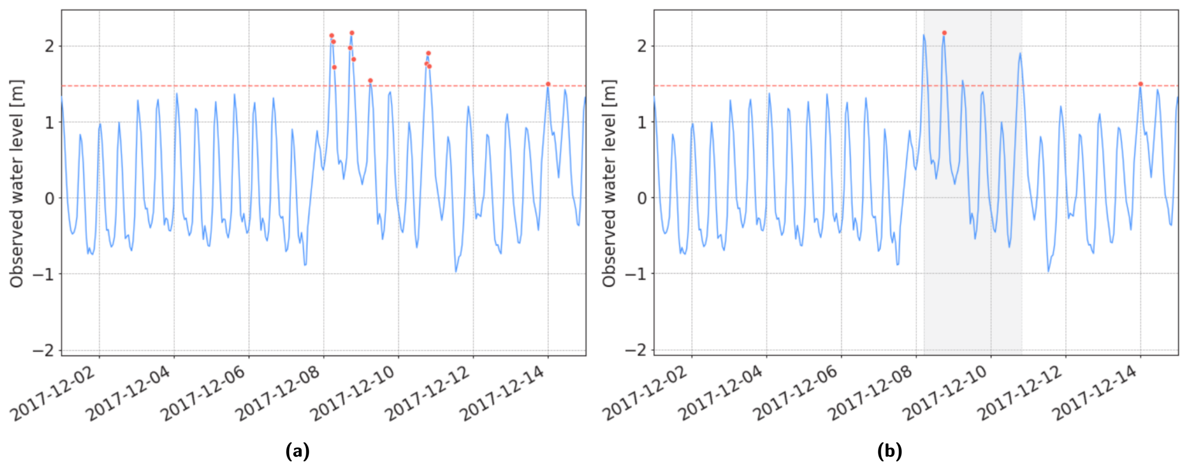


Figure 4.9: Observed water level with (a) no declustering and (b) declustering with a time window of 48 hours. On the y-axis, the water level observed by the tide gauge is shown, referenced to NAP. The dotted line refers to the threshold and the red dots to the selected peaks that exceed the threshold.

4.5.2. Weight analysis

Next to performance metrics, an additional analysis is done to assess the importance of different input variables of the neural network. The second-layer weights $[w_{n1}, \quad w_{n2} \quad \cdots \quad w_{nj}]$ with $j = 32$ (Figure 2.8) are used to look into the importance for each neuron. If a weight approaches zero ($w_{nj} \Rightarrow 0$), it means that any input into that neuron is negligible, meaning that their corresponding first-layer weights are also negligible. These first-layer weights (w_{11} to w_{mj}) in turn can be assessed by summing over either the 48 hours within the subsets, or over the 32 neurons. For ERA5, heat maps are created to show the importance of every input feature within each variable. For satellite altimetry, the weights are visualised as graphs and heat maps. This analysis gives a shallow approximation of feature importance for each input variable. Secondly, it enables linking poor performance parts to a possible lack of information given by the input variables.

5
Results

The results of the developed neural network are presented in this chapter. Starting at the objective of this study, which is **estimating hourly non-tidal water levels**, reconstructed time series for Scheveningen, Europlatform and Vlissingen are presented. Six months are extracted to visualise the differences between the ML estimates and the TG observations, which allows for the identification of high- and low-performance parts within the model. Next, the performance of each model and the results of the K-fold cross-validation are presented, giving an approximation for the robustness and spread of the error of the ML model. The following section presents the high water level estimates for each location, comparing them to the Dutch Continental Shelf Model, followed by the results of the K-fold cross-validation. Finally, the results of the weight analysis are presented to assess which features are most important when estimating non-tidal water levels.

5.1. Time series reconstruction

The developed neural network has been trained and tested for each location of interest, which is Scheveningen, Vlissingen and the Europlatform. Parts of the reconstructed time series are shown in Figures 5.1, 5.2 and 5.3. This reconstruction is based on the testing dataset, input data which the model has not seen before (see also Section 4.3.4), and therefore considered independent and a good measure of the model performance. The harmonic tidal signals are not included in this section, but will be added later on when the model performance is presented. This testing dataset applies to the period from 2015-04-01 at 10:00:00 to 2018-09-07 at 23:00:00. When zoomed into six months containing high-interest periods, such as periods with documented severe storms (Table 5.1), several important properties are discerned.

Table 5.1: Severe storms as registered by the Koninklijk Nederlands Meteorologisch Instituut (KNMI), flagged as a storm if one of the KNMI weather stations spread throughout the Netherlands records a wind speed higher than 24.5 m/s (KNMI, 2024). Shaded in grey are the storms that occurred during the ML model testing period. The highest wind speed recorded is averaged over one hour. The highest wind gust is instantaneous. The location refers to where these wind speeds and gusts were recorded. The last column indicates where the storm was coming from.

Date	Highest wind speed recorded [m/s]	Highest wind gust recorded [m/s]	Location	Coming from
2002-02-26	25	34	Vlieland/Vlissingen	England
2002-03-09	25	33	IJmuiden	Southern North Sea
2002-10-27	28	41	IJmuiden/Vlissingen	Ireland
2007-01-18	25	35	IJmuiden/Vlissingen	Southern North Sea
2013-10-28	29	42	Vlieland	England
2013-12-05	25	63	Vlieland	Scotland
2015-07-25	25	34	IJmuiden	English Channel
2016-11-20	26	37	IJmuiden	English Channel
2017-09-13	26	35	Vlieland	Scotland
2018-01-03	26	34	Vlieland	Scotland
2018-01-18	30	40	Hoek van Holland	Scotland

5.1.1. Scheveningen

Figure 5.1 shows the time series reconstruction for the testing dataset. The shaded areas correspond to the zoomed-in graphs below the full time series, showing the reconstruction in higher detail to analyse high- and low-variability periods.

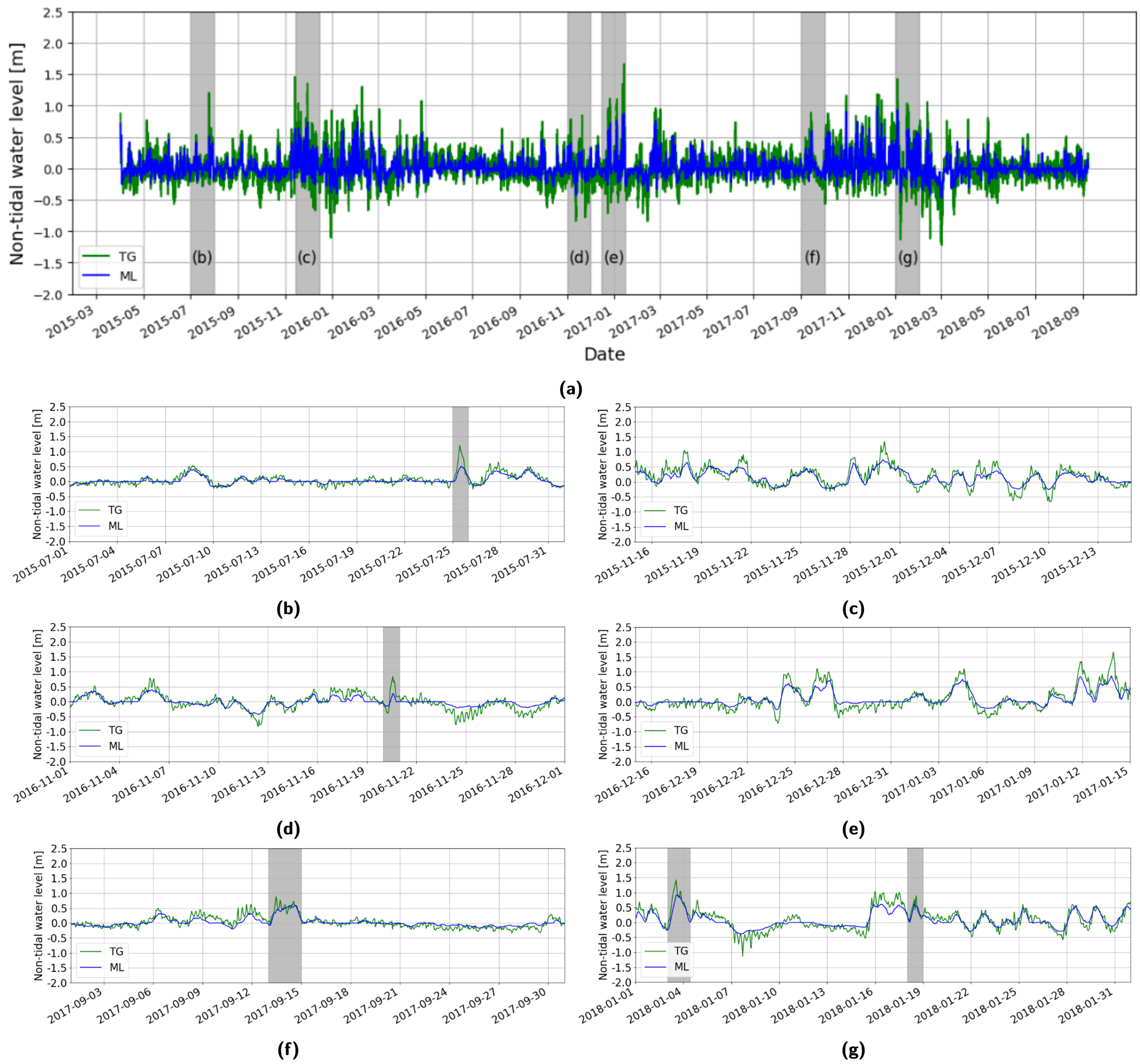


Figure 5.1: Time series reconstruction for the Scheveningen model. The shaded areas in (a) correspond to the months shown in (b), (c), (d), (e), (f) and (g). The green lines represent the TG observations, the blue lines refer to the ML model output. The shaded areas in the sub-graphs show severe storms catalogued by the KNMI (KNMI, 2024). The y-axis unit is in metres and depicts the non-tidal water level referenced to NAP.

The first observation is that the neural network provides a rather smoothed version of the non-tidal water level. High-frequency variations in the water level are not fully captured by the model, as opposed to most low-frequency variations that last over multiple hours. When these low-frequency variations are of small magnitude, the model quickly conforms to a constant value close to zero. Several high water level events are considered within the testing period that occurred on 2015-07-25, 2016-11-20, 2017-09-13, 2018-01-03 and 2018-01-18. Generally, the model underestimates the peaks in the non-tidal water levels during these storms (Table 5.2). However, one important characteristic of the model becomes apparent when analysing the sub-figures of Figure 5.1, 5.2 and 5.3. For storms coming from the south, meaning from the English Channel or southern North Sea, the model underestimates the high water level event with values between 0.243 m (Figure 5.3d) and 0.693 m (Figure 5.1b) and in some cases even misses the high water level event (Figure 5.2d). Conversely, storms with a longer duration that originate from Scotland or northern England are generally estimated better (Table 5.2). Some additional details and examples for Scheveningen are highlighted below.

Table 5.2: Error between machine learning estimates and tide gauge non-tidal water levels for severe storms as registered by the KNMI (KNMI, 2024) for the testing period. The last column indicates where the storm was coming from. Errors are given in metres and show the TG observations minus the ML estimates.

Date	Error Scheveningen [m]	Error Vlissingen [m]	Error Europlatform [m]	Coming from
2015-07-25	0.693	0.477	0.305	English Channel
2016-11-20	0.571	0.460	0.243	English Channel
2017-09-13	0.293	0.243	-0.087	Scotland
2018-01-03	0.496	0.287	-0.126	Scotland
2018-01-18	0.269	0.862	0.056	Scotland

- **Figure 5.1b:** July 2015 is presented, displaying how the model describes relatively high non-tidal water levels during summer storms. The storm on 2015-07-25 arrived from the south, travelling north along the Dutch coast in approximately eleven hours (InfoNu, 2016). The sharp peak this storm induced in the coastal non-tidal water levels in Scheveningen is underestimated by 0.693 m. Longer and lower peaks, such as the higher water levels between 2015-07-08 and 2015-07-10, are followed by the model with an underestimation of 0.124 m. When the observed non-tidal water levels are close to zero, meaning -0.192 $<$ non-tidal water level $<$ 0.235 m, the model quickly converges to give a constant output of -0.001 m.
- **Figure 5.1c:** The autumn and winter months such as November and December 2015 have much more variation in the non-tidal water level, which decreases the possibility of the model converging to a constant value. The peaks and troughs are followed, in some cases better than others, as can be viewed by comparing the peak on 2015-11-28 with an error of 0.174 m with the peak on 2015-11-30 with an error of 0.616 m.
- **Figure 5.1d:** 2016-11-20 contains another storm, caused by a depression that entered the North Sea through the English Channel, travelling north along the Dutch coast (KNMI, 2016). These south-westerly storms are severe, but short, causing a sharp peak in the non-tidal water level records such as the storm on 2015-07-25. This sharp peak is underestimated by 0.571 m.
- **Figure 5.1e:** December 2016 and January 2017 give examples of both the convergence to a constant value and the estimation of larger peaks. On days with little non-tidal water level variation, such as the period between 2016-12-16 and 2016-12-20 or between 2016-12-29 and 2017-01-01, the model output is constant. The large variations, such as the signals between 2016-12-26 and 2016-12-28 or the peak on 2017-01-04, are followed by the model, but underestimated systematically by a value of 0.395 and 0.348 m respectively. Sharper peaks are underestimated more than wider peaks, such as the peak on 2017-01-14 with an error of 0.791 m compared to the peak on 2017-01-04. The sharp peaks on 2017-01-12 with an error of 0.502 m and 2017-01-14 can be explained by a small depression that travelled from the English Channel northward, causing a lot of precipitation (KNMI, 2017b). The winds that accompanied this depression were not severe enough to be considered a storm, hence it is not shaded. Furthermore, higher winds were observed in the north of the Netherlands near the holidays of 2016, explaining the higher non-tidal water levels from 2016-12-24 to 2016-12-28 (KNMI, 2017a).
- **Figure 5.1f:** On 2017-09-13, a severe storm travelled over the North Sea towards Denmark, coming from Scotland (KNMI, 2017c), with the largest wind speeds recorded in the north of the Netherlands (Table 5.1). The model tracks these high non-tidal water levels better than those from the shorter, sharper peaks caused by storms coming from the south, underestimating this one with 0.293 m.
- **Figure 5.1g:** Two severe storms are recorded on 2018-01-03 and 2018-01-18, the first one longer and causing higher non-tidal water levels and the second one shorter, but stronger regarding wind speeds (Table 5.1). Both travelled from the coast of Scotland over the North Sea to the Dutch coast (KNMI, 2018a; KNMI, 2018b), and are captured by the model with errors of 0.496 m and 0.269 m respectively, and the higher non-tidal water levels between 2018-01-16 and 2018-01-18 are underestimated with an error of 0.429 m.

5.1.2. Vlissingen

As was mentioned in Section 3.2, Vlissingen lies enclosed in an estuary, experiencing large tidal amplitudes and land-sea interactions such as reflections. Figure 5.2 shows the full reconstruction of the non-tidal water level record, zoomed in to the same months as Figure 5.1 to discern the differences between this model and the model from Scheveningen.

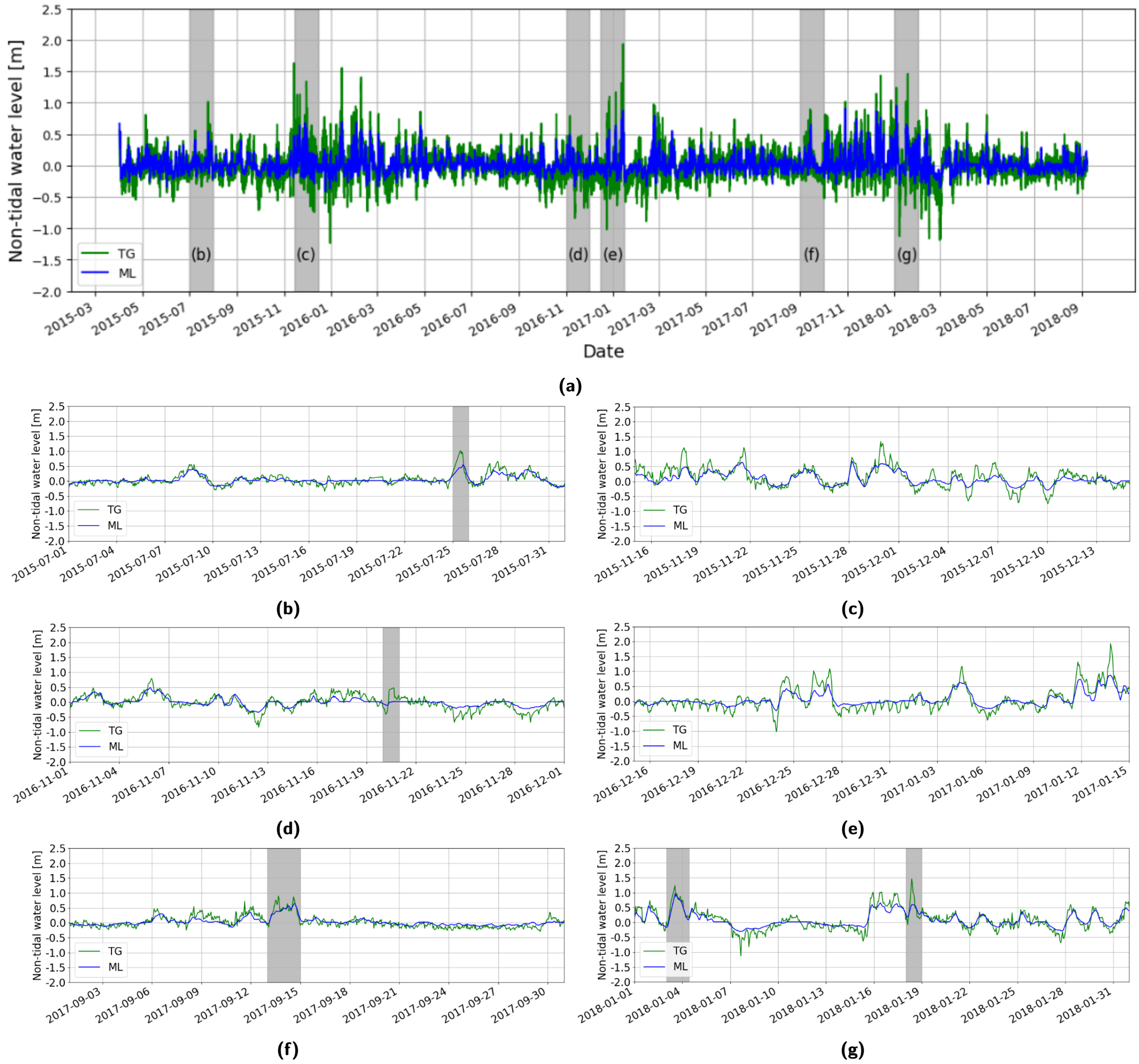


Figure 5.2: Time series reconstruction for the Vlissingen model. The shaded areas in (a) correspond to the months shown in (b), (c), (d), (e), (f) and (g). The green lines represent the TG observations, the blue lines refer to the ML model output. The shaded areas in the sub-graphs show storms catalogued by the KNMI (KNMI, 2024). The y-axis unit is in metres and depicts the non-tidal water level referenced to NAP.

The same characteristics that have been found for Scheveningen are also valid for Vlissingen, such as the smooth behaviour of the modelled non-tidal water level compared to the TG record and the general underestimation of high non-tidal water level events. Still, there are a few distinct differences, one of which relates to the convergence to a constant in e.g. Figure 5.2e compared to Figure 5.1e. Table 5.3 reports the values of the constants for each model, along with the total duration in hours where the models estimate this constant and in which 95% interval of non-tidal water levels this happens. The models do not consistently estimate this constant when the TG observation falls within this 95% interval, but they do so in 39.5% of cases for Scheveningen, 28.3% for Vlissingen, and 43.8% for Europlatform, while outside this range it only happens in 5.4% of cases for Scheveningen, 3.9% for Vlissingen and 4.6% for Europlatform. The total durations indicate that Scheveningen estimates a constant most frequently, followed by Europlatform, and then Vlissingen. This statistic correlates with the number of inactive neurons presented in Section 5.4.

Furthermore, the underestimation of 0.477 m of the summer storm on 2015-07-25 is smaller (Figure 5.2b) than for the Scheveningen model (Figure 5.1b) by 0.216 m. The Vlissingen model provides a similar water level (0.539 m) as the Scheveningen model (0.512 m), only differencing by 0.027 m, even though the Vlissingen TG recorded a lower peak (1.016 m) than the Scheveningen TG (1.205 m), which differs with 0.189 m. The storm on 2016-11-20 is skipped in its entirety (Figure 5.2d), while the Scheveningen model still captures some of it (Figure 5.1d). The severe storm on 2018-01-03 is followed by the Vlissingen model (Figure 5.2g), but the sharp peak on 2018-01-18 is underestimated by 0.862 m. This event is present in Scheveningen as well, though the peak is much smaller (Figure 5.1g).

5.1.3. Europlatform

Since the Europlatform tide gauge is located further offshore than Scheveningen or Vlissingen, the non-tidal water level is expected to be smoother, meaning less disrupted by land-sea interactions such as reflections from the coast. Figure 5.3 shows the reconstruction of the Europlatform model, zoomed in to the same months as the Scheveningen and Vlissingen model to enable comparisons between the three models.

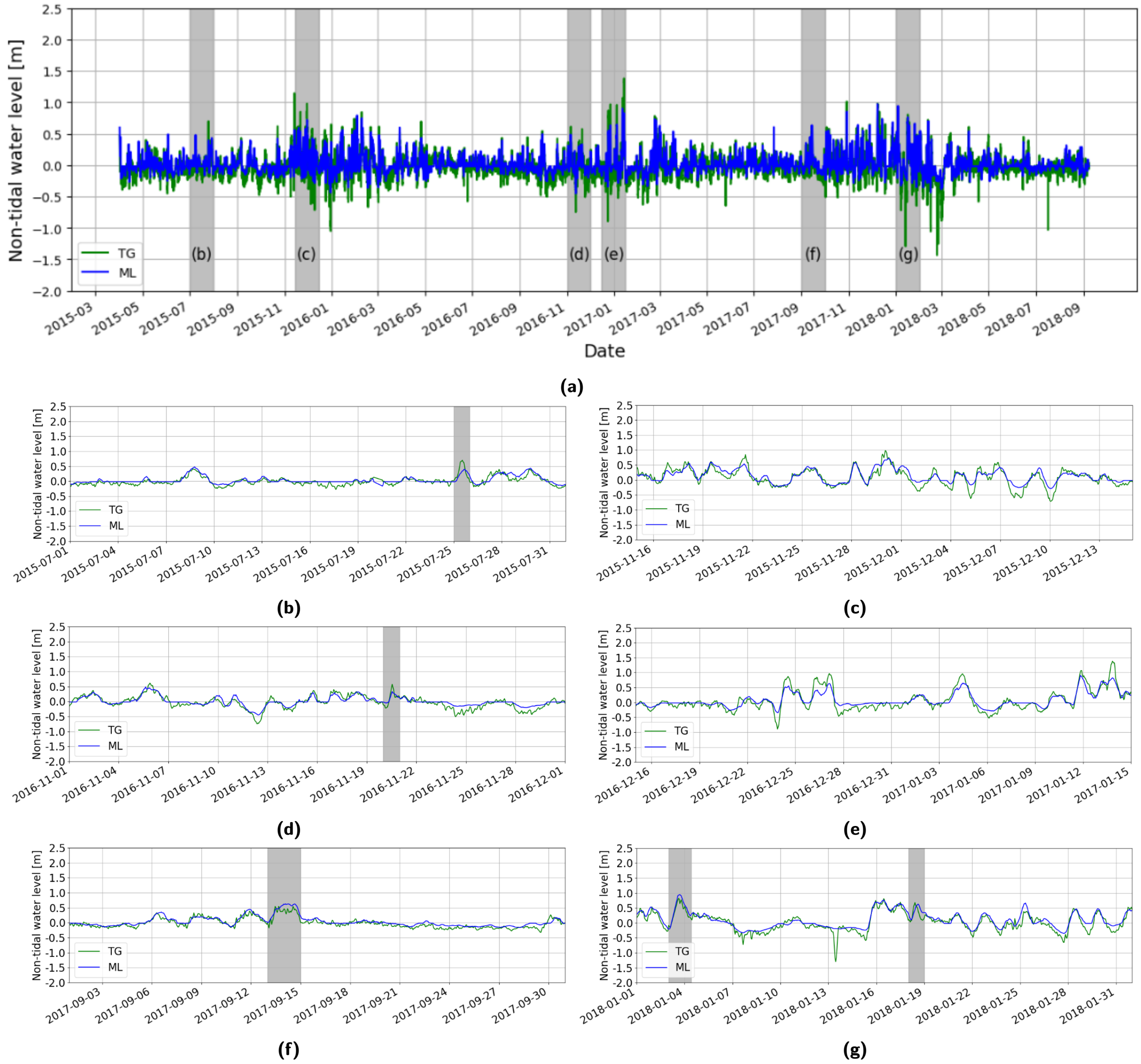


Figure 5.3: Time series reconstruction for the Europlatform model. The shaded areas in (a) correspond to the months shown in (b), (c), (d), (e), (f) and (g). The green lines represent the TG observations, the blue lines refer to the ML model output. The shaded areas in the sub-graphs show storms catalogued by the KNMI (KNMI, 2024). The y-axis unit is in metres and depicts the non-tidal water level referenced to NAP.

Table 5.3: Statistics on the constants that each model converges towards for the testing dataset. Given are the constants in metres, the total duration where the model output is equal to the constant in hours, the interval of the TG observations in which 95% of the constant output appears in metres and the percentage of constant outputs within this interval.

	Constant [m]	**Total duration [hr]**	**95% interval [m]**	**Percentage of 95% interval**
Scheveningen	-0.001	9021	[-0.192, 0.235]	39.5
Vlissingen	0.020	6463	[-0.160, 0.270]	28.3
Europlatform	-0.018	8952	[-0.195, 0.126]	43.8

The model for Europlatform follows high non-tidal water levels better than for Scheveningen or Vlissingen, overestimating high non-tidal water levels during the storms on 2017-09-13 (Figure 5.3f) and 2018-01-03 (Figure 5.3g) by 0.087 and 0.126 m respectively (Table 5.2). Additionally, the observed water level at Europlatform is smoother than at Scheveningen or Vlissingen, and the reconstruction at Europlatform conforms to the observations better. Still, large peaks such as those on 2015-07-25 and 2017-01-14 are underestimated (Figure 5.1b and 5.3e) by 0.305 m and 0.554 m respectively.

5.1.4. Scenario examples

Specific examples of the input that is fed into the model for three scenarios are given in Figure 5.4 and 5.5, for the ERA5 variables and satellite altimetry variables respectively. These scenarios occur on 2015-07-25 at 12:00, where all three models underestimate a summer storm, on 2016-12-31 at 00:00, where the Scheveningen and Europlatform model estimate a constant value, and on 2018-01-03 at 17:00, where all three models estimate a winter storm with better accuracy than the summer storm in 2015. The input variables are shown as histograms before they are normalised to give a better interpretation of the characteristics that cause good or bad performance of the model.

The ERA5 input variables corresponding to the bad performance scenario (2015-07-25 at 12:00) show a low-pressure field for a few hours before the high non-tidal water level occurs, though the spread is small compared to the good performance scenario (2018-01-03 at 17:00) shown at the right-hand side of Figure 5.4. The reported pressures are lower for this winter storm, pushing the output of the model up more. The sea surface pressure for the constant output scenario (2016-12-31 at 00:00) is high with little temporal variability.

The longitudinal and lateral wind speeds (U10 and V10) for the constant output scenario also show little temporal variation, suggesting stable, calm weather with wind speeds centred around 0 m/s. For the good performance scenario, U10 is centred around 9 m/s one hour before the timestamp of the output value, and centred around 5 m/s forty-eight hours before. This implies that the wind speeds are increasing within the time window considered for the input data. The bad performance scenario U10 contains a similar structure, but the wind speed itself is lower, centred around 5 m/s one hour before the timestamp of the output value and around 2 m/s forty-eight hour before.

The lateral wind speeds show less distinct patterns for both the bad and the good performance scenarios. Both are centred around 0 m/s and show a similar spread. For the bad performance scenario, V10 increases when going from -48 hours to -1 hour before the timestamp of the output value, visible as the histogram flattening and spreading. For the good performance scenario, V10 increases too, but then goes back to being centred around 0. This suggests that V10 does not play as large a role in the performance of the model for different scenarios as p and U10 do.

Finally, the number of satellite altimetry observations differs per scenario (Figure 5.5). The bad performance scenario contains 10 overpasses within the area of interest, the constant output scenario contains 9 and the good performance scenario contains 3. All overpasses have varying non-tidal residuals, ranging from -3 to 3 m. For the bad performance scenario, the NTR increases when going from -45 to -4 hours before the timestamp of the output value (Figure 5.5). The NTR for the constant output scenario stays centred around 0. From the good performance scenario, no time variability can be discerned due to the limited number of overpasses. No particularly clear patterns can be discerned otherwise from the observation that the model does not necessarily require a large amount of satellite altimetry data to make accurate estimations when looking at the good performance scenario.

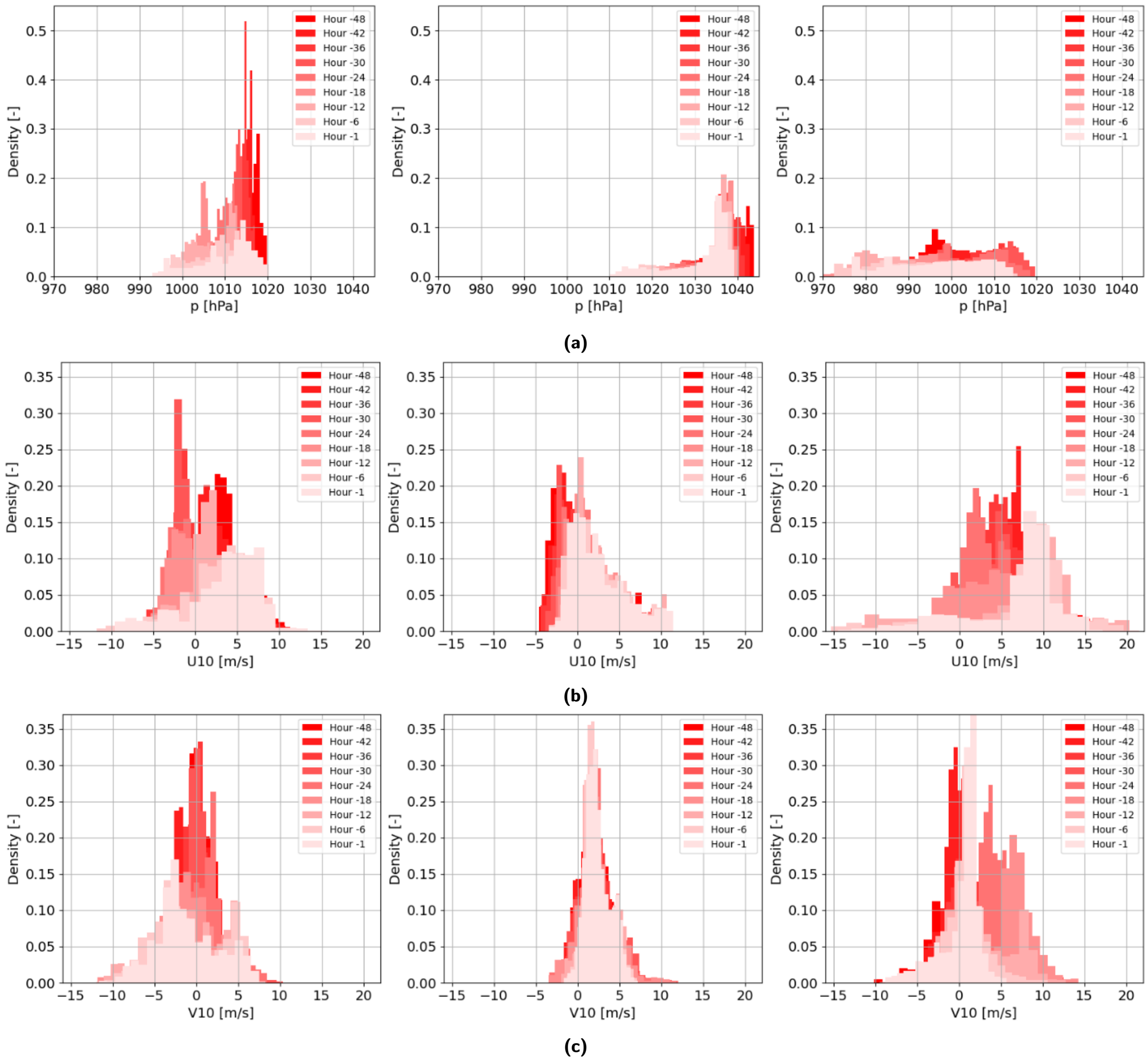


Figure 5.4: Histograms of (a) sea level pressure, (b) longitudinal wind speed and (c) lateral wind speed for (left) a good performance sample on 2015-07-25 at 12:00, (middle) a constant output sample on 2016-12-31 at 00:00 and (right) a good performance sample on 2018-01-03 at 17:00. The y-axis corresponds with the density of each bin.

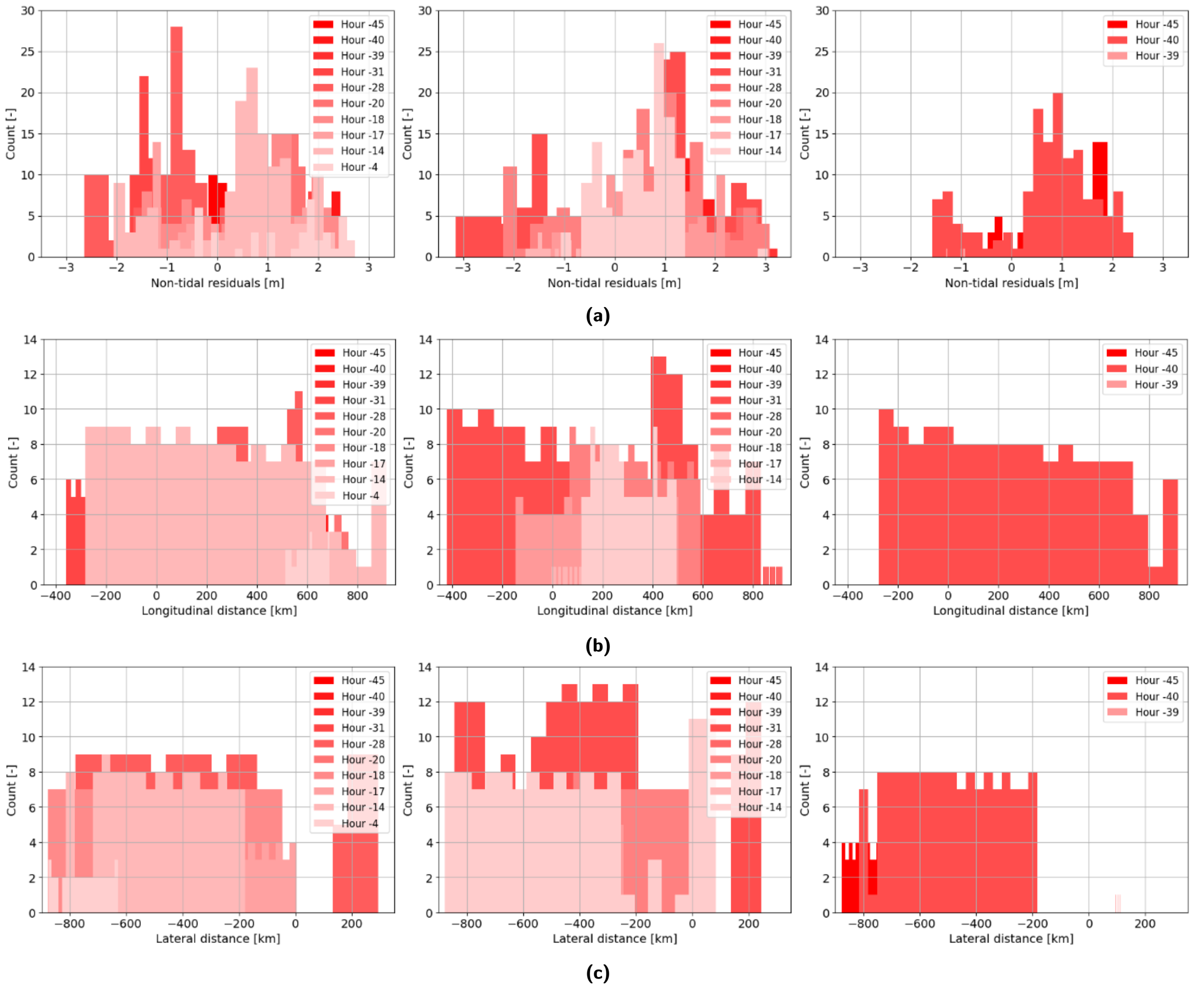


Figure 5.5: Histograms of (a) non-tidal residual, (b) longitudinal distance and (c) lateral distance for (left) a good performance sample on 2015-07-25 at 12:00, (middle) a constant output sample on 2016-12-31 at 00:00 and (right) a good performance sample on 2018-01-03 at 17:00. The y-axis corresponds with the number of observations in each bin.

5.2. Model performance

The performance of the three developed models is assessed with four performance metrics (see Section 4.5), computed with the observed water levels of the TG records and the ML model estimates. For these metrics, the harmonic tidal signal has been added to both the TG water levels and the ML model estimates. Only the signal-to-noise (SNR) metric is impacted by this step, since the mean squared error (MSE), the mean absolute error (MAE) and the standard error (σ_ε) rely on residuals in which the harmonic tidal signal cancels out. The SNR, however, is used to assess whether the model is useful for physical applications because it compares the variability of the errors to the variability of the water level itself. Including the harmonic tidal signals is essential here since this component induces the largest variability in the water level. These four metrics are compared to Dutch Continental Shelf Model (DCSM) estimates, of which the hourly water levels range from 2015-04-01 until 2018-01-01. The metrics for both the ML models and DCSM are presented in Table 5.4.

The ML models for Scheveningen and Vlissingen perform similarly, varying only in their MAE by a maximum of 0.002 m. The ML model for Europlatform performs slightly better, with an MSE of 0.011 m. Scheveningen, Vlissingen and Europlatform have an MAE of 0.101, 0.099 and 0.078 m respectively. The standard deviation of the error (σ_ε) is 0.134 m for Scheveningen and 0.135 m for Vlissingen, while Europlatform has a σ_ε of 0.100 m.

When looking at DCSM, the performance metrics appear to be approximately a factor two better than the ML models, with the exception of the signal-to-noise ratios (SNR). This metric differs greatly over the ML models and DCSM due to the large variability in tidal amplitudes over the three locations (see also Figure

4.3). The Scheveningen ML model has an SNR of 22.7, meaning that the MSE of the Scheveningen model is 22.7 times smaller than the variance of the TG water levels (including the tidal signal). This is considered sufficient for physical applications. Vlissingen has an SNR of 101.04 due to the large tidal variability and Europlatform of 37.4 due to the smaller MSE compared to Scheveningen. For DSCM, the SNR reports 106.1 for Scheveningen, 398.3 for Vlissingen and 135.6 for Europlatform, these differences having the same cause as the differences for the ML models.

In addition to the overall performance, a K-fold cross-validation (with $K = 4$) has been applied to the Scheveningen model. A rough estimate of the robustness and spread of the developed neural network is obtained by training and testing the model on different datasets. The period for each validation dataset is approximately 3.5 years, ranging from 2001-01-01 to 2015-04-01 (Table 4.6). The testing dataset has not been used for this validation.

Figure G.1, G.2, G.3 and G.4 show the time series reconstruction for all K-fold validation sets, each with four zoomed-in months to show the performance of the model in different conditions, especially during the severe storms mentioned in Table 5.1. For each validation dataset, the performance metrics are computed and presented in Table 5.4.

Table 5.4: Model performance metrics for each location. All metrics are based on data that the model has not used in the training stage. Shown are the mean squared error (MSE), the signal-to-noise ratio (SNR), the mean absolute error (MAE) and the standard error (σ_ε), also known as the standard deviation of the error. Units of MSE, MAE and σ_ε are in metres. The ML model metrics are shown above the DCSM metrics, followed by the K-fold cross-validation metrics.

ML model	**MSE [m]**	**SNR [-]**	**MAE [m]**	σ_ε **[m]**
Scheveningen	0.018	22.7	0.101	0.134
Vlissingen	0.018	101.0	0.099	0.135
Europlatform	0.011	37.4	0.078	0.100
DCSM	**MSE [m]**	**SNR [-]**	**MAE [m]**	σ_ε **[m]**
Scheveningen	0.004	106.1	0.049	0.062
Vlissingen	0.005	398.3	0.054	0.068
Europlatform	0.004	135.6	0.049	0.052
K-fold	**MSE [m]**	**SNR [-]**	**MAE [m]**	σ_ε **[m]**
Validation set 1	0.019	23.1	0.102	0.137
Validation set 2	0.016	27.8	0.097	0.126
Validation set 3	0.018	24.9	0.101	0.129
Validation set 4	0.015	30.1	0.090	0.120

The MSE for all K-fold validation sets is in the order of 0.015 to 0.019 m, with a σ_ε between 0.120 and 0.137 m. The MAE ranges between 0.090 and 0.102 m. The SNR of each validation set ranges between 23.1 and 30.1. Table 5.5 presents the sample means and the corresponding range of the K-fold validation metrics. The sample means show an expected MSE of 0.017 m, an SNR of 26.5, an MAE of 0.097 m and a σ_ε of 0.128 m when this model is trained. The largest difference between the K-fold validation sets is given as the range, reporting a range of 0.004 m for the MSE, 7.1 for the SNR, 0.012 m for the MAE and 0.017 m for the σ_ε. The corresponding metrics for the Scheveningen model differ from the sample means by 0.001 m for the MSE, 3.8 for the SNR, 0.004 m for the MAE and 0.006 m for the σ_ε. From these metrics, a high robustness is verified.

Table 5.5: Sample mean ($\bar{x}$) and range of the performance metrics for the Scheveningen model.

	MSE [m]	**SNR [-]**	**MAE [m]**	σ_ε **[m]**
$\bar{x}$	0.017	26.5	0.097	0.128
Range	0.004	7.1	0.012	0.017

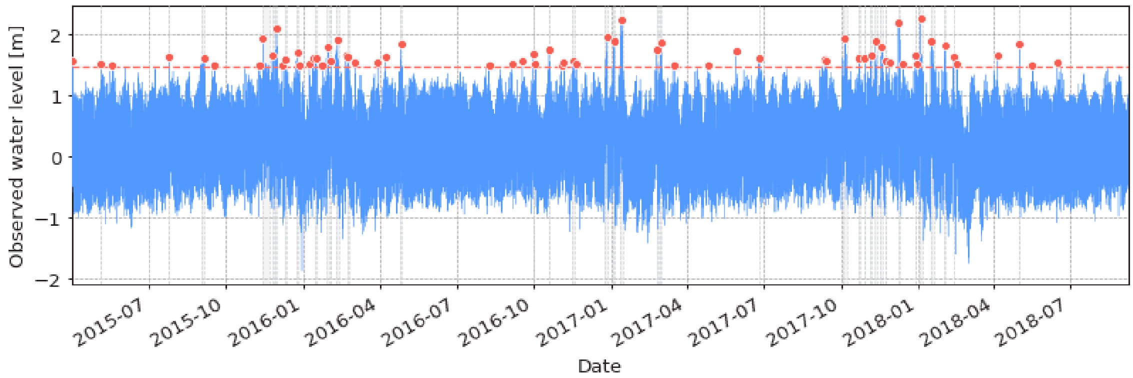
5.3. Estimation of high water levels

Since high water level events are used for making risk assessments and computing return periods, the performance of the model for high water levels is important to report. To assess these high water levels, a peak-over-threshold (POT) method is applied to find the relevant high water level events (Appendix F).

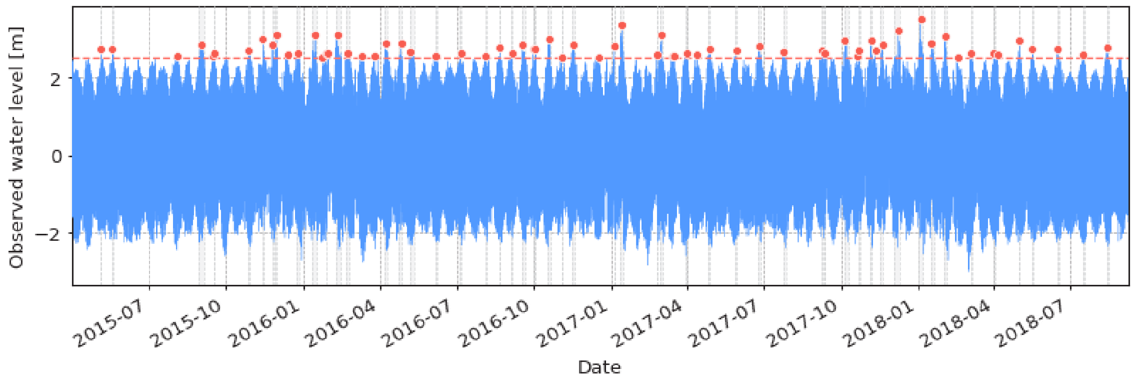
These results are based on the full water level signals and their reconstruction, meaning that the harmonic tidal signal has been added to the signals to account for high and low tidal periods.

The high water level events for each location are shown in Figure 5.6, where the red line refers to the threshold of the POT method and the red dots to the selected high water levels. Declustering of the events is applied with a window size of 48 hours (see Section 4.5). For each location, the corresponding ML output and the DCSM output have been collected, after which the performance metrics are again computed (Table 5.6). The number of high water levels found for Scheveningen, Vlissingen and the Europlatform is 69, 68 and 70 respectively.

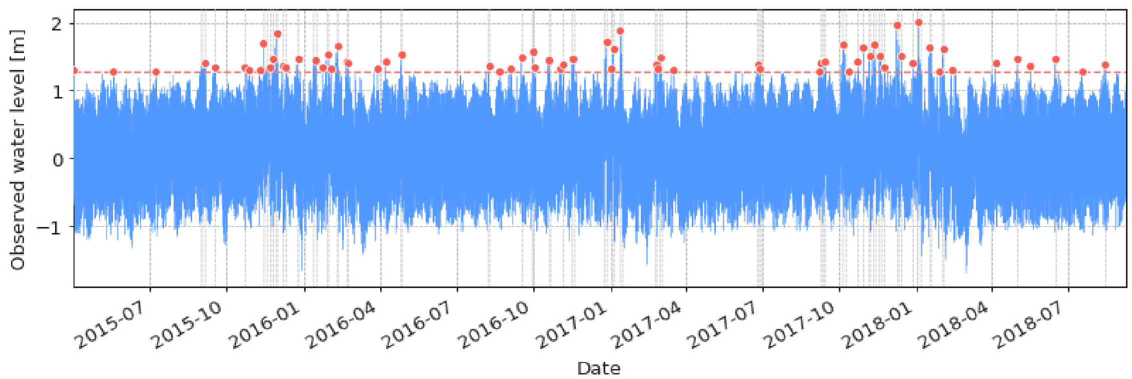
Similar to the overall performance analysis, the performance on the high water levels is also assessed with a K-fold cross-validation. The model of Scheveningen has been trained and validated for $K = 4$, the selected high water levels for each validation set presented in Figure 5.7. The performance metrics for each validation set are given in Table 5.6. The number of high water levels for each validation set is 70, 74, 57 and 66 respectively.



(a)

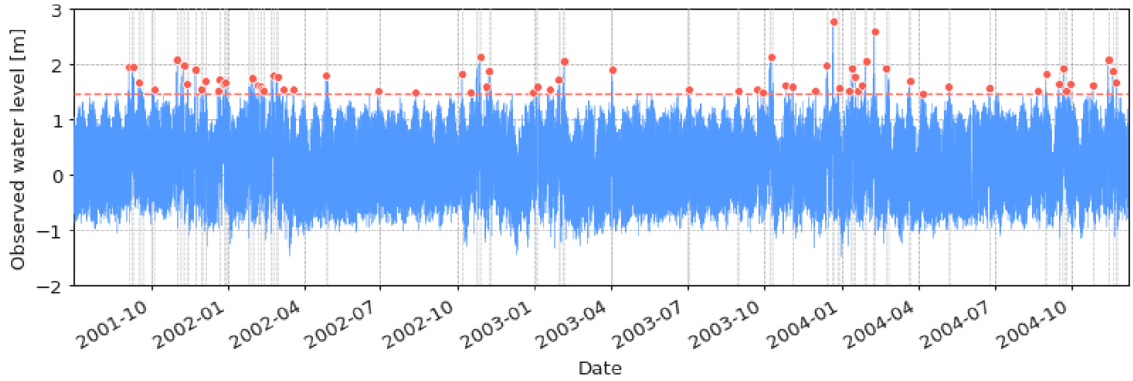


(b)

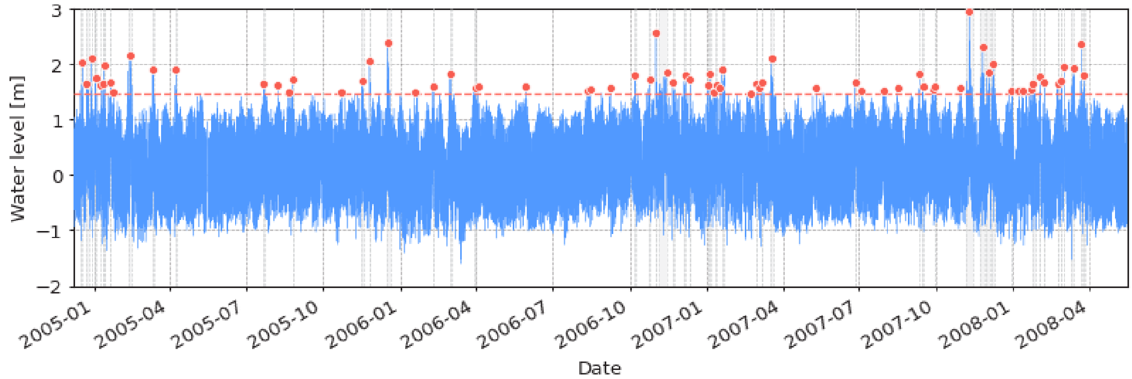


(c)

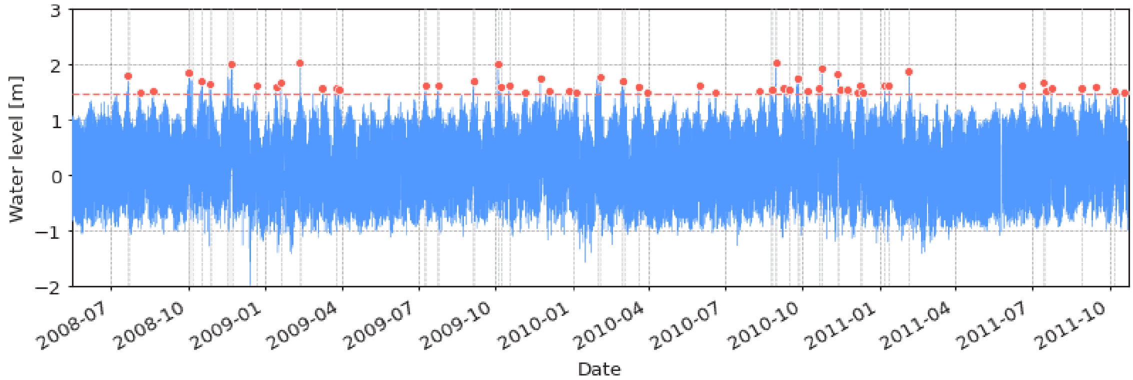
Figure 5.6: Selected high water levels from tide gauge data for (a) Scheveningen, (b) Vlissingen and (c) Europlatform. Thresholds are shown as red dashed lines and high water levels as red dots. A declustering window of 48 hours is used.



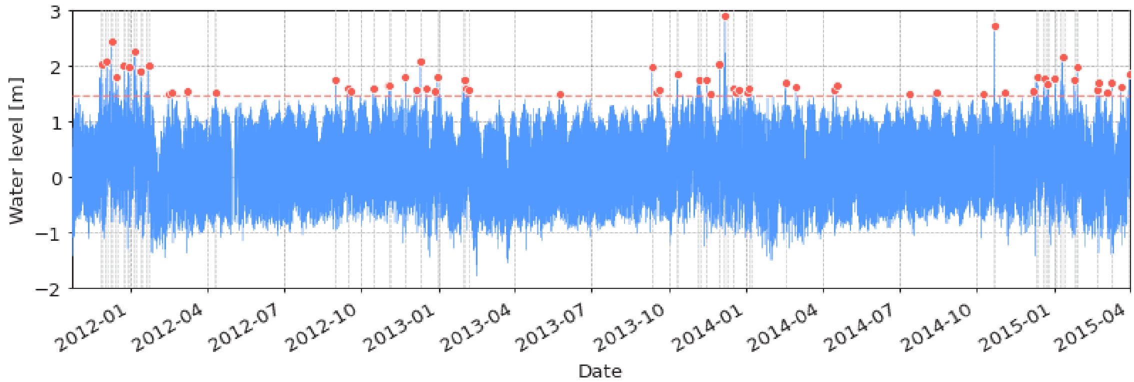
(a)



(b)



(c)



(d)

Figure 5.7: Selected high water levels from tide gauge data for (a) validation dataset 1, (b) validation dataset 2, (c) validation dataset 3 and (d) validation dataset 4. The threshold is shown as a red dashed line and high water levels as red dots. A declustering window of 48 hours is used.

Table 5.6: Model performance metrics for each location for high water levels, including the harmonic tidal signals. Shown are the mean squared error (MSE), the signal-to-noise ratio (SNR), the mean absolute error (MAE) and σ_ε. Units of MSE, MAE and σ_ε are in metres. The ML model metrics are shown above the DCSM metrics, followed by the K-fold cross-validation metrics.

ML model	**MSE [m]**	**SNR [-]**	**MAE [m]**	σ_ε **[m]**
Scheveningen	0.054	15.3	0.184	0.164
Vlissingen	0.016	133.8	0.099	0.117
Europlatform	0.013	29	0.090	0.113
DCSM	**MSE [m]**	**SNR [-]**	**MAE [m]**	σ_ε **[m]**
Scheveningen	0.019	42.3	0.103	0.098
Vlissingen	0.013	397.6	0.099	0.068
Europlatform	0.006	141.3	0.066	0.051
K-fold	**MSE [m]**	**SNR [-]**	**MAE [m]**	σ_ε **[m]**
Validation set 1	0.034	14.8	0.139	0.171
Validation set 2	0.044	13.2	0.148	0.183
Validation set 3	0.039	14.2	0.150	0.171
Validation set 4	0.034	15.6	0.142	0.166

From Table 5.6, it is concluded that the Europlatform model performs best when considering the performance metrics, for both the ML model and DCSM. The models for Vlissingen and Europlatform perform similarly on high water levels as they do for hourly water levels, with an MSE of 0.016 and 0.013 m respectively. Their MAE are 0.099 and 0.090 m and their σ_ε 0.117 and 0.113 m respectively, only differencing with a maximum of 0.002 m for the MSE, 0.012 for the MAE and 0.018 m for the σ_ε.

The model for Scheveningen performs significantly worse for high water levels with an MSE of 0.054 m, an MAE of 0.184 m and a σ_ε of 0.164 m. When looking at Figure 5.8, high water levels are being underestimated more for Scheveningen than for Vlissingen or Europlatform. This agrees with the earlier founding presented in Table 5.2, where Scheveningen underestimates four out of five storms more extremely than Vlissingen or Europlatform. Additionally, the SNR of 13.5 and 29 for Scheveningen and Europlatform respectively has decreased when compared to the performance on the hourly water levels. The Vlissingen model has an increased SNR of 133.8. These differences are caused by a difference in MSE compared to the model performance on high water levels.

For DCSM, the MSE and MAE for all locations have increased. The σ_ε for Scheveningen has increased, for Vlissingen has remained the same and for Europlatform decreased with 0.001 m. The MSE has most significantly increased, and consequently, the SNR has decreased. This is most noticeable in Scheveningen.

The K-fold cross-validation results show little variation in the different validation sets. Looking at Figure 5.7, it can be concluded that validation set 3 contains the smallest number of high water level events and the lowest maximum water level. Also, most high water level events occur in the winter months, as evident in Figure 5.7d. The mean value for each metric along with the range within the four values for each metric are given in Table 5.7.

Table 5.7: Sample mean ($\bar{x}$) and range of the performance metrics for the Scheveningen model for high water levels.

	MSE [m]	**SNR [-]**	**MAE [m]**	σ_ε **[m]**
$\bar{x}$	0.038	14.4	0.145	0.173
Range	0.010	2.4	0.011	0.017

The sample means show 0.038 m for the MSE, 14.4 for the SNR, 0.145 m for the MAE and 0.173 m for the σ_ε. The range for these metrics is 0.010 m for the MSE, 2.4 for the SNR, 0.011 for the MAE and 0.017 for the σ_ε. These ranges are sufficiently small to be able to consider this model robust, and the SNR is large enough to consider the ML model useful for practical application, though DCSM still outperforms the ML model in terms of accuracy.

The largest difference between the K-fold validation sets is given as the range, reporting a range of 0.004 m for the MSE, 7.1 for the SNR, 0.012 m for the MAE and 0.017 m for the σ_ε. The corresponding metrics for the Scheveningen model differ from the sample means by 0.001 m for the MSE, 3.8 for the SNR, 0.004 m for the MAE and 0.006 m for the σ_ε. From these metrics, a high robustness is verified.

Figure 5.8 shows the selected high water level events for all three ML models, where the dots represent the high water level events and the red line refers to the 1:1 line. Most high water level events are underestimated by all models, with higher events being underestimated more than lower ones. This effect is largest for Scheveningen, and smallest for Europlatform. The placement of the points, with Vlissingen positioned more to the top right than Scheveningen and Europlatform, is caused by the larger tidal signal in Vlissingen.

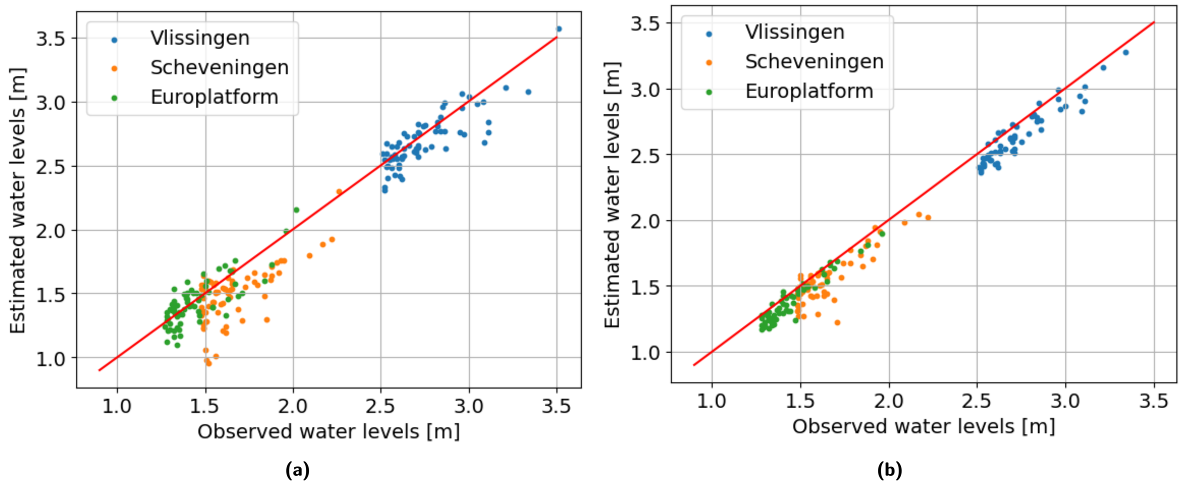


Figure 5.8: Observed high water levels (x-axis) vs their corresponding estimated water levels (y-axis) for (a) the machine learning model and (b) DCSM. The blue dots refer to the Vlissingen model, the orange dots to the Scheveningen model and the green dots to the Europlatform model.

5.4. Weight analysis

For further analysis of the results, the parameters trained by the model are presented. These parameters are defined as weights $[w_{11} \quad w_{w12} \quad \cdots \quad w_{mj}]$ and $[w_{n1} \quad w_{n2} \quad \cdots \quad w_{nj}]$ and biases $[b_{n1} \quad b_{n2} \quad \cdots \quad b_{nj}]$ and $[b_{\hat{y}}]$ (see Figure 2.8). For ERA5, the weights are shown as heat maps. For satellite altimetry, the weights are shown as graphs and heat maps, since the input corresponding to satellite altimetry is not linked to specific locations and/or time stamps.

The weights and biases that are not directly linked to input variables ($[w_{n1} \quad w_{n2} \quad \cdots \quad w_{nj}]$ and $[b_{n1} \quad b_{n2} \quad \cdots \quad b_{nj}]$) are presented in Figure 5.9. The bias of the output layer ($[b_{\hat{y}}]$) is -1.11*10^{-3} m for Scheveningen, 1.97*10^{-2} m for Vlissingen and -1.78*10^{-2} m for Europlatform, equal to the constant values the three models converge to when the observed non-tidal water level is close to zero (Table 5.3). This means that all neurons return a value of zero after the weighted sum of input data has been passed through the activation function. Regarding the shape of the activation function as described in Section 2.5, the weighted sums of all input features is < 0 for every neuron, which returns a 0 according to the ReLU function (Figure 2.10).

Figure 5.9 shows that the three models have different weight and bias configurations. All weights reside within a range of -0.4 to 0.4, but the Scheveningen model contains more "dead" neurons than Vlissingen or Europlatform. A dead neuron refers to a neuron with a corresponding weight close to zero. Any input into the neuron is nullified by this small weight, essentially rendering the neuron inactive. In Figure 5.9, these dead neurons and their corresponding biases are shown as red dots. Like the input features, the bias corresponding to a dead neuron is also nullified.

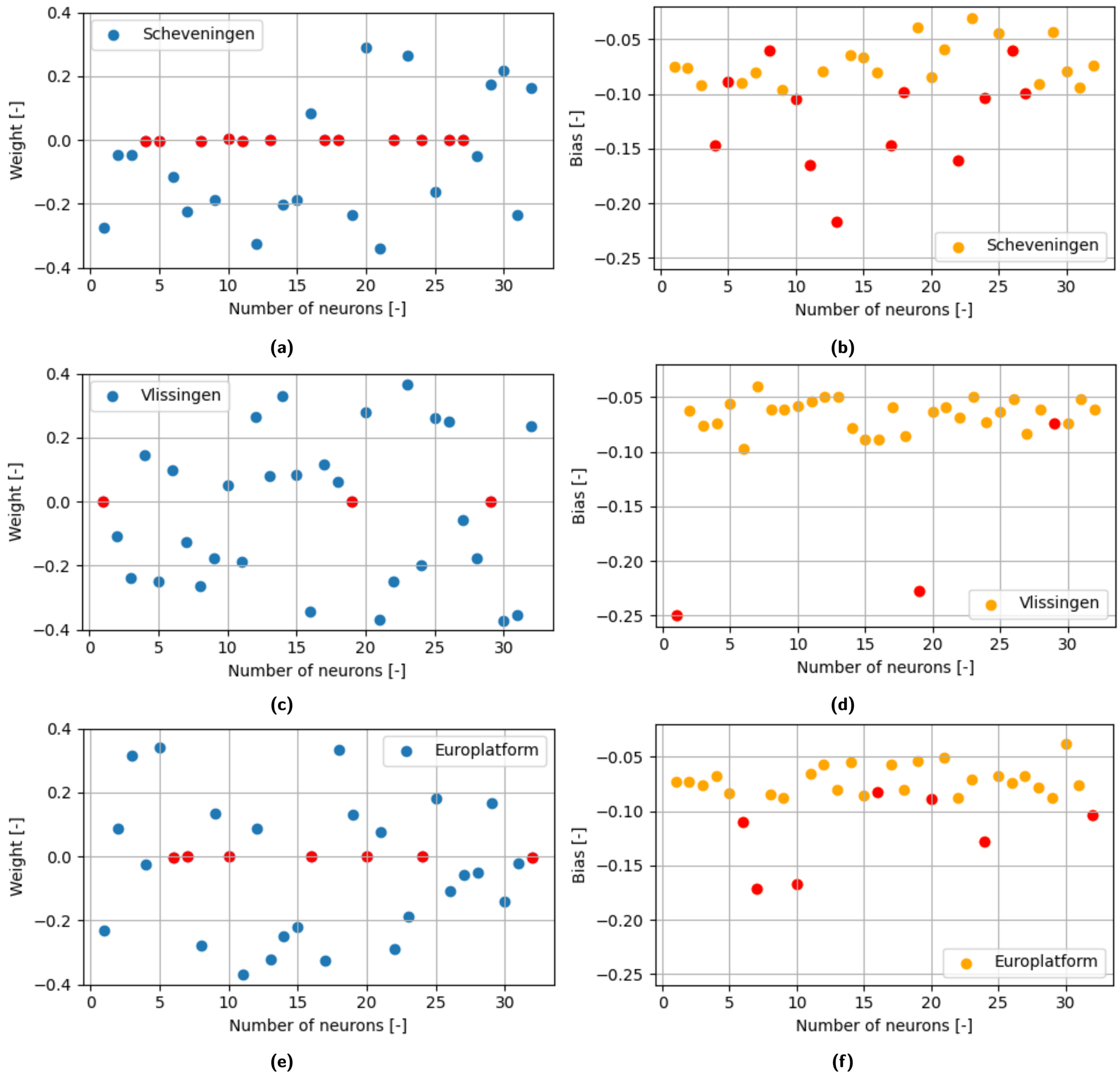


Figure 5.9: Second-layer weight $[w_{n1} \quad w_{n2} \quad \cdots \quad w_{n32}]$ and first-layer bias $[b_{n1} \quad b_{n2} \quad \cdots \quad b_{n32}]$ graphs for the (a, b) Scheveningen, (c, d) Vlissingen and (e, f) Europlatform model. Blue dots refer to the weights and orange dots to the biases. The red dots show the ones where the weight is so close to zero that their inputs are nullified.

For Scheveningen, neurons 4, 5, 8, 10, 11, 13, 17, 18, 22, 24, 26 and 27 are considered inactive. For Vlissingen, these neurons are 1, 19 and 29 and for Europlatform, these are 6, 7, 10, 16, 20, 24 and 32. From the corresponding biases, it can be concluded that the model stops training when these neurons are killed to speed up the training process.

The large number of dead neurons in the Scheveningen model agrees with the more frequent appearance of a constant value in the time series reconstruction compared to Vlissingen or Europlatform (Table 5.3). However, this does not necessarily imply that the performance of the model decreases if the number of dead neurons increases. While the Vlissingen model contains the smallest number of dead neurons, Europlatform performs better on all metrics (Table 5.4 and 5.6), even though the duration of a constant output for Europlatform is larger than for Vlissingen (Table 5.3).

5.4.1. Satellite altimetry features

The second-layer weights and biases that are considered to be non-significant (the red dots in Figure 5.9) have been excluded from the results for the first-layer weights ($[w_{11} \quad w_{w12} \quad \cdots \quad w_{mj}]$) corresponding to satellite altimetry. During the input preparation (Section 4.3), the satellite observations have been processed to form part of the input dataset. This is done by padding the data to consist of 48 one-hour arrays for every 48-hour subset. These arrays, counting 514 observations, are padded with zeros if they contain fewer satellite observations than 514.

The sum of the first-layer weights over the 48-hour time window for the Scheveningen model is presented in Figure 5.10, where the individual lines correspond to the individual neurons. The grey lines refer to the neurons whose weighted sum is nullified by the corresponding weight for the second layer.

The sea level observations from satellite altimetry and the longitudinal component of the distance between the tide gauge and the observations (d_E) vary between -2.5 and 2.5. The lateral component of the distance (d_N) varies between -1.5 and 1.5. The time difference between the observations and the time stamp of the desired output value (dt) is mostly negative, ranging between -4.2 and 0.5. For all variables, the weights converge to zero when the number of observations increases. For all neurons, the weights are in the same order of magnitude for each corresponding observation, meaning that all lines within the sub-graphs follow a similar pattern. The same graphs for the Vlissingen and Europlatform models are shown in Appendix H.

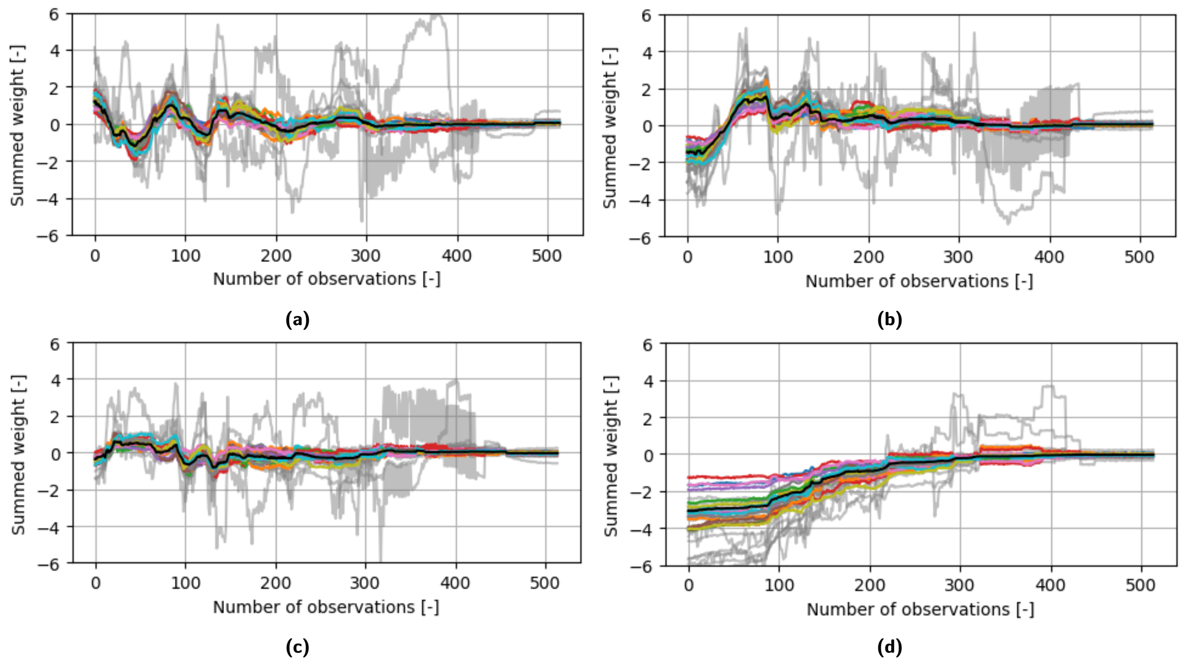


Figure 5.10: First-layer weights summed over the 48-hour time window for the Scheveningen model for (a) the non-tidal water level, (b) the longitudinal component of the distance d_E, (c) the lateral component of the distance d_N and (d) the time difference between observation and tide gauge dt. Each coloured line corresponds to a neuron. The black lines show the mean of all 32 lines.

The convergence phenomenon causes many satellite observations to be less important than earlier ones. Adding more observations does not provide additional information to the model because their weight is considered negligible. This phenomenon stems from the processing method and availability of the satellite altimetry data, where the padding method renders many observations redundant. The lateral distance component (d_N) also has smaller weights compared to the other variables, making it less important. Furthermore, weights associated with dead neurons (grey lines) differ significantly from other weights. This happens because the model stops training these neurons once they are killed, preventing their weights from converging. Lastly, the largest differences between neurons are found in the time difference component (dt) within the first 200 observations, ranging from -4.2 to 0.5, while other variables show smaller variations.

Instead of presenting the weights as graphs, the same weights can be presented as heat maps. These maps are created by taking the latitudes and longitudes of the satellite altimetry observations, and computing the mean weight per location over the *active* neurons for the different variables. Figure 5.11 shows the mean weight per location for the NTR observed by the satellites, Figure 5.12 shows the mean weight for d_E, Figure 5.13 shows the mean weight for d_N. The weights corresponding to the dead neurons (see Figure 5.9) have been excluded to prevent adding noise (see the grey lines in Figure 5.10).

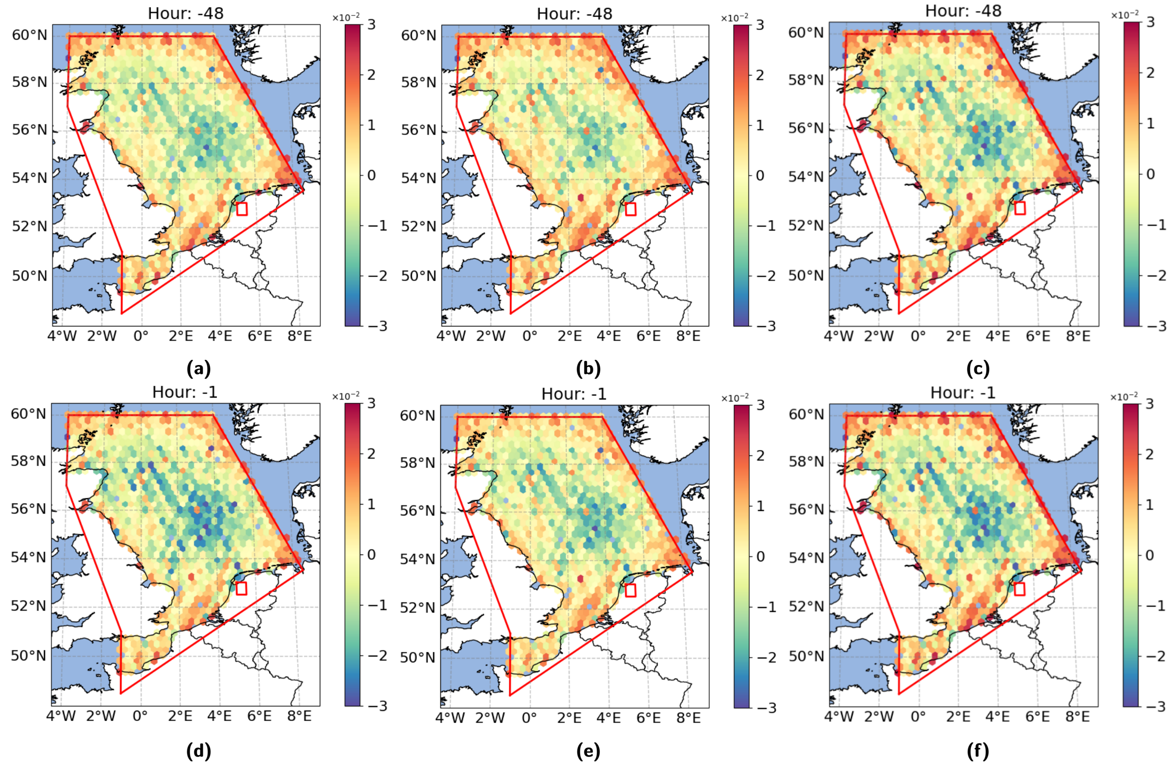


Figure 5.11: First-layer mean weights of the non-tidal water level observations from satellite altimetry. The first and last hour of the 48-hour time window is shown for (a, d) Scheveningen, (b, e) Vlissingen and (c, f) Europlatform. High positive weights are shown in red, high negative values in blue and weights close to zero in yellow.

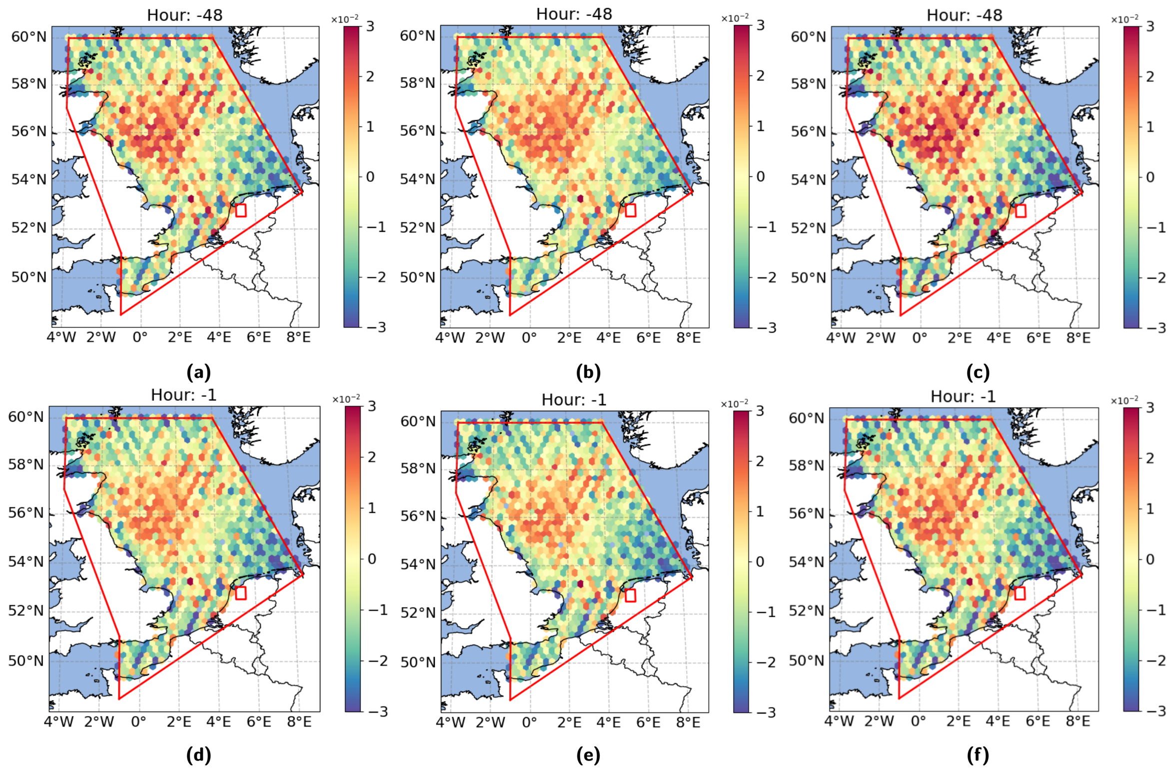


Figure 5.12: First-layer mean weights of the longitudinal distance between the satellite altimetry observations and the tide gauge. Only the first and last hour of the 48-hour time window is shown for (a, d) Scheveningen, (b, e) Vlissingen and (c, f) Europlatform. High positive weights are shown in red, high negative values in blue and weights close to zero in yellow.

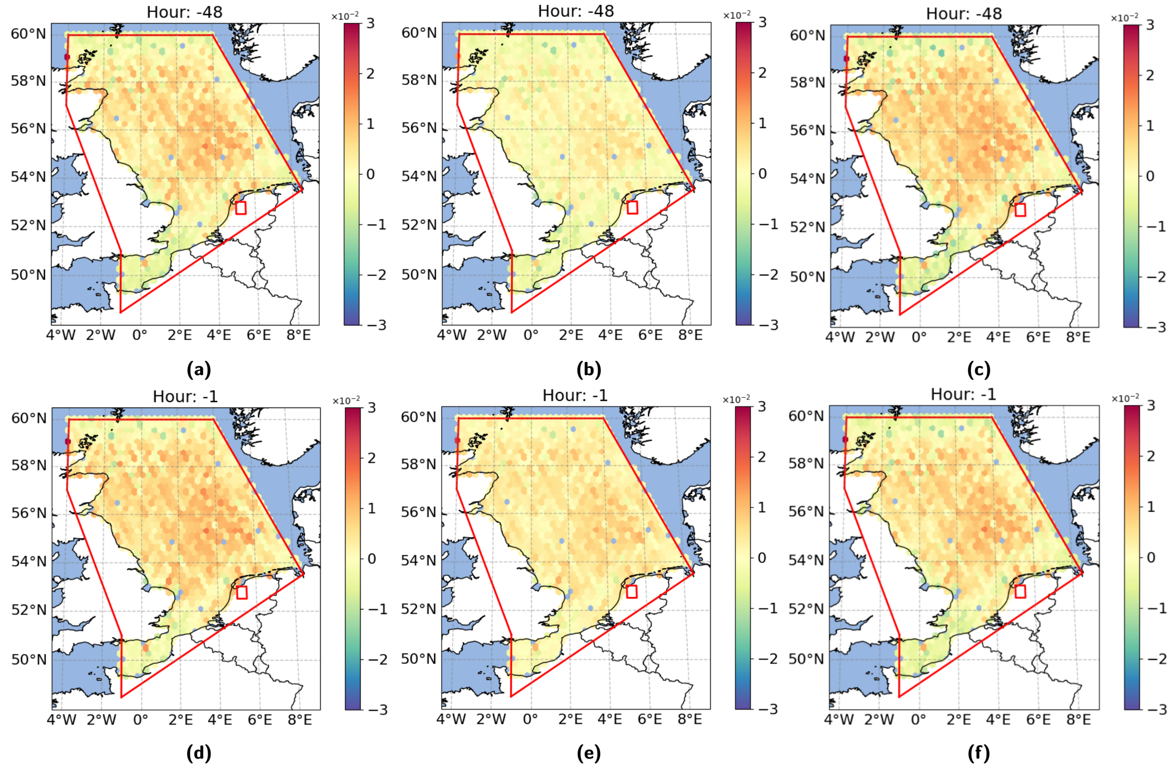


Figure 5.13: First-layer mean weights of the lateral distance between the satellite altimetry observations and the tide gauge. The first and last hour of the 48-hour time window is shown for (a, d) Scheveningen, (b, e) Vlissingen and (c, f) Europlatform. High positive weights are shown in red, high negative values in blue and weights close to zero in yellow.

Notably, the largest positive weights corresponding to the observed non-tidal water level are found along the Dutch coast and the northern North Sea (Figure 5.11). The middle of the North Sea reports the highest negative weights. This general pattern is valid for all three models. The differences between the first and last hour of the 48-hour time window show that the features 1 hour before the time stamp of the output value have larger weights than 48 hours before for the middle of the North Sea, and smaller ones near the borders of the area of interest (see Appendix H).

The large importance at the borders of the area of interest is caused by the padding of satellite altimetry data in the processing method. As concluded from Figure 5.10, the weights gradually converge to zero when the number of observations increases within an hour. Processing is done by storing all available satellite observations within each hour as arrays in chronological order. This means that, in most cases, the start of every array contains observations at the border of the area of interest. Due to the padding, the remaining observations quickly become redundant.

Additionally, d_E shows a similar but inverse pattern, with high positive weights in the middle of the North Sea and negative weights along the border of the area of interest (Figure 5.12). The high negative weights east of the tide gauges are caused by d_E being negative in this area. The differences between the first and last hour of the 48-hour time window are mostly negative for Scheveningen and Europlatform in the middle of the North Sea (Appendix H), except the region east of the tide gauges. This means that d_E is considered more important 48 hours before the time stamp of the output value than 1 hour before. For Vlissingen, most differences are positive, meaning that d_E is considered more important 1 hour before the time stamp of the output value than 48 hours before.

Figure 5.13 shows weights close to zero, with faint patterns showing positive weights over the North Sea and negative ones at the borders of the area of interest and from the Strait of Dover southwards for all locations. These negative weights are caused by d_N being negative there. The differences between the first and last hour of the 48-hour time window are mostly positive for Scheveningen and Vlissingen (Appendix H), meaning that 1 hour before the time stamp of the output value is more important than 48 hours before. Europlatform shows slightly negative differences, suggesting that 48 hours before the time stamp of the output value is more important than 1 hour before.

5.4.2. ERA5 features

The first-layer mean weights corresponding to ERA5 variables are shown in Figure 5.14, 5.15 and 5.16, where dark blue or red refers to higher weights, either positive or negative, and yellow refers to weights closer to zero. Figure 5.14 visualises the first and last hour of the input features for sea level pressure (p), meaning the 48th hour and the 1st hour before the time stamp of the output value. For all models, the area with the largest weights is to be found at the Dutch coast, reaching as high as $1.9*10^{-2}$ in some locations. The northern North Sea is also of importance, though the weights are inversely proportional to the weights along the Dutch coast. Furthermore, the weights 48 hours before the time stamp of the output value are higher at the Dutch coast than the weights 1 hour before the time stamp of the output value (see also Appendix H). The opposite is true in the northern North Sea, where the weights 1 hour before the time stamp of the output value are higher.

For the longitudinal wind speed component (U10) (Figure 5.15), the area along and below the Dutch coast contains the highest weights, with some negative weights in two areas across the middle of the North Sea. For the Dutch coast, these weights are higher 48 hours before the time stamp of the output value compared to 1 hour before. The northern North Sea contains two line-shaped areas that contain high weights, negative as well as positive. These are also higher 48 hours before the time stamp of the output value compared to 1 hour before. The middle of the North Sea also contains an area with higher weights 48 hours before the time stamp of the output value, but their surroundings experience the opposite, having larger weights 1 hour before the time stamp of the output value.

Those line-shaped areas are more distinguishable when looking at the lateral wind speed component (V10) (Figure 5.16), where they are noticeable from both 48 hours to 1 hour before the time stamp of the output value, and for all three models. In addition, those weights are mostly negative. The area around and below the Dutch coast contains the highest weights, which are increasingly positive for 1 hour before the time stamp of the output value in comparison to 48 hours before. The opposite is true for the two line-shaped areas in the northern North Sea. Their importance has decreased for 1 hour before the time stamp of the output value in comparison to 48 hours before. In general, the weights for the lateral wind speed component over the entire North Sea grow more positive when for hours closer to the time stamp of the output value.

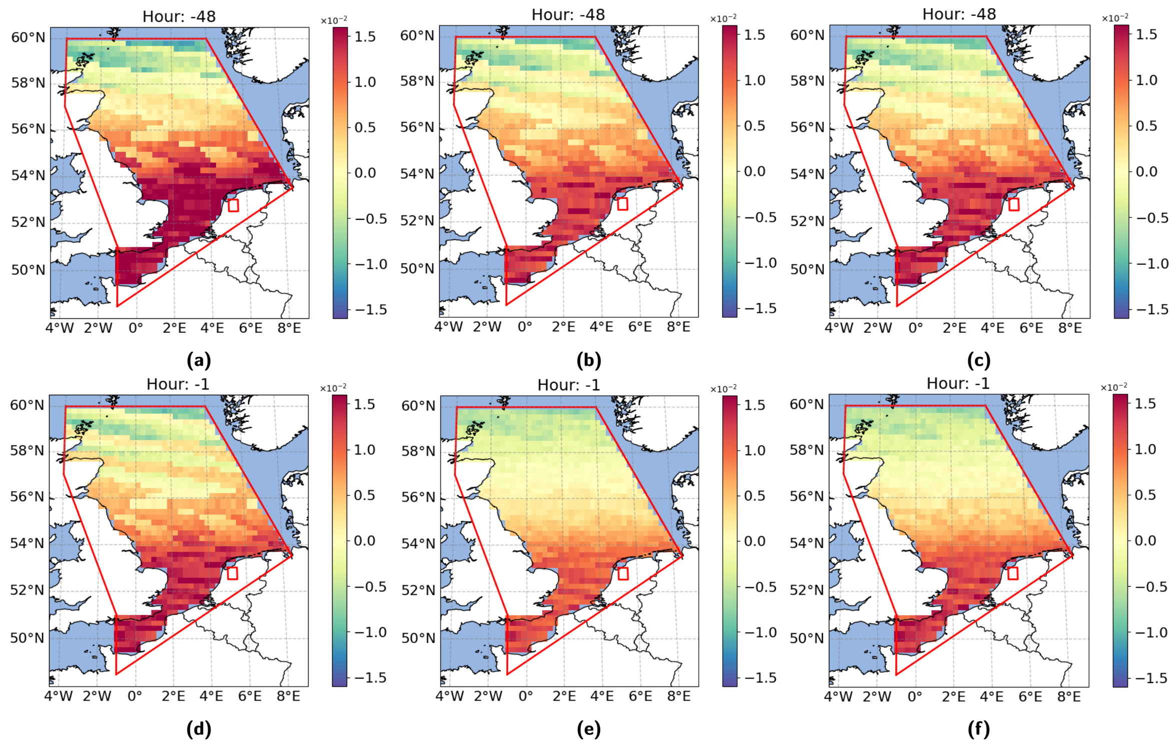


Figure 5.14: First-layer mean weights of ERA5 sea surface pressure. The first and last hours of the 48-hour time window is shown for (a, d) Scheveningen, (b, e) Vlissingen and (c, f) Europlatform. High positive weights are shown in red, high negative values in blue and weights close to zero in yellow.

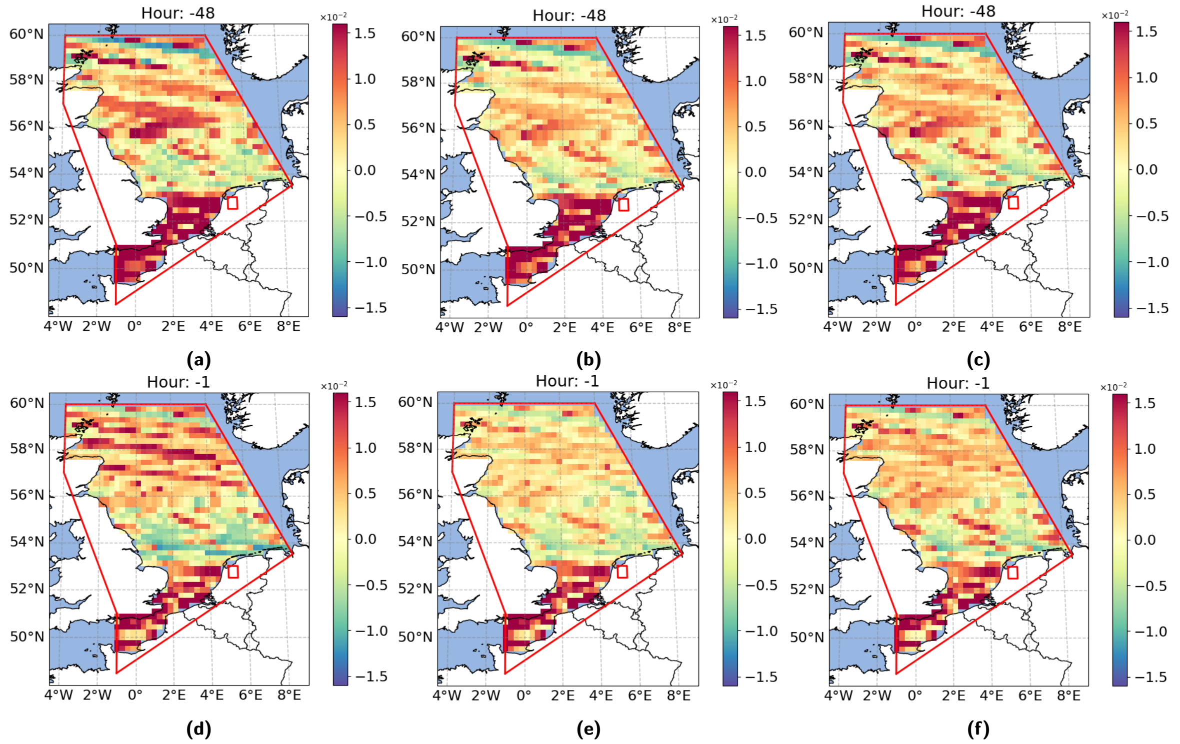


Figure 5.15: First-layer mean weights of ERA5 longitudinal wind speed. The first and last hours of the 48-hour time window is shown for (a, d) Scheveningen, (b, e) Vlissingen and (c, f) Europlatform. High positive weights are shown in red, high negative values in blue and weights close to zero in yellow.

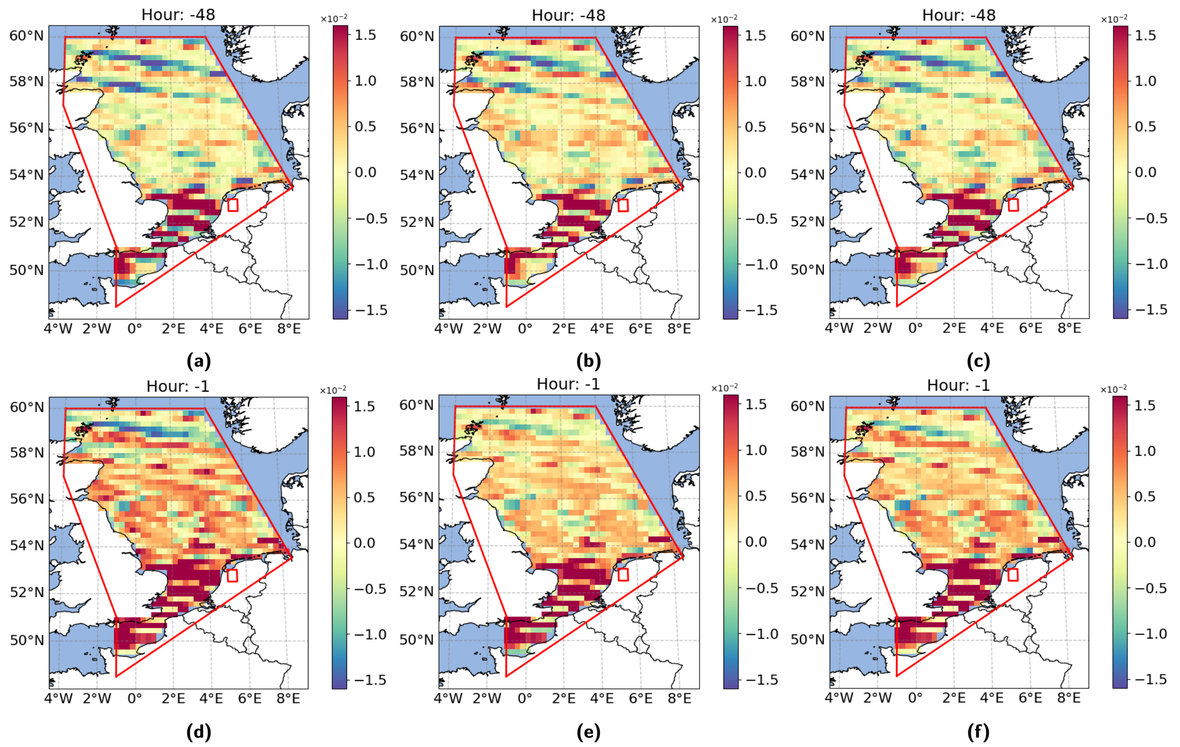


Figure 5.16: First-layer mean weights of ERA5 lateral wind speed. The first and last hours of the 48-hour time window is shown for (a, d) Scheveningen, (b, e) Vlissingen and (c, f) Europlatform. High positive weights are shown in red, high negative values in blue and weights close to zero in yellow.

In Figure 5.14, 5.15 and 5.16, the mean of all active neurons is computed to get the weight per hour for every variable. The maps to show the differences between 1 hour and 48 hours before the time stamp of the output value are shown in Appendix H.

5.4.3. Weight importance

To assess which variables contribute most to the estimation of non-tidal water levels, the corresponding first-layer weights are summed over the full 48-hour time window and over all active neurons. Table 5.8 presents the percentages of how much each variable contributes, defined as the sum of the absolute first-layer weights for the corresponding variable divided by the sum of all absolute first-layer weights.

Table 5.8: Weight contributions of input variables in percentages. The weights are summed over the 48-hour time window and all active neurons for every model.

		Scheveningen [%]	Vlissingen [%]	Europlatform [%]
ERA5	p	17.37	16.38	16.49
	U10	18.28	18.18	17.95
	V10	15.71	16.31	15.99
Satellite altimetry	NTR	9.36	9.54	9.76
	d_N	7.62	8.27	8.28
	d_E	10.75	10.72	11.23
	dt	20.93	20.60	20.29
	DOY	$2.54*10^{-3}$	$2.43*10^{-3}$	$2.02*10^{-3}$

The variable with the largest weights is dt, which gives the time difference between the satellite altimetry observations and the time stamp of the output value. This variable accounts for $\pm$20% of the weights for all models. The next most important variables are the three ERA5 variables, U10 being the most crucial one. d_E accounts for $\pm$11% of the weights, NTR for $\pm$9.5% and d_N for $\pm$8%. DOY, which is only a single value in an input dataset of 314401 features, has such a small weight that it only accounts for $\pm$0.002% of the weights.

Overall, the variables corresponding to satellite altimetry observations, meaning NTR, d_N, d_E and dt, account for 48.66% of the total weights for Scheveningen, 49.13% for Vlissingen and 49.56% for Europlatform. Consequently, ERA5 variables, which consist of p, U10 and V10, make up 51.33% of the total weights for Scheveningen, 50.87% for Vlissingen and 50.43% for Europlatform.

5.5. Computation time

One of the advantages of using machine learning over elaborate numerical models is the efficient computational cost. Table 5.9 presents the computation times for the estimation of one year of coastal water levels compared to DCSM. For DCSM, the 2022 release of DCSM-FM 100m (flexible mesh with a highest resolution of 100 metres) (Zijl, Groeneboom, et al., 2022) has been used to compare the ML model to, but DCSM-FM 0.5nm (flexible mesh with a highest resolution of 0.5 nautical miles) (Zijl, Zijlker, Laan, & Groenenboom, 2022) is also presented to highlight the differences between the different DCSM models.

Table 5.9: Computation time for one year for the ML model developed in this study vs DCSM-FM 100m and DCSM-FM 0.5nm. Computation times are shown in seconds. Computational power is shown in the number of available cores. The time step refers to the time step of the estimates.

	Computation time [hours]	Available computational power [number of cores]	Time step [seconds]	Computation time per time step [seconds]
ML model	0.06	4	3600	0.09
DCSM-FM 100m	52.8	20	35.4	4.27
DCSM-FM 0.5nm	8.3	20	118.7	2.25
DCSM-FM 0.5nm	66	1	118.7	0.89

For DCSM, the computation times are extracted from the documentation and are based on runs on a computational cluster with different computational power settings. For DCSM-FM 100m, one year of model output with a time step of 35.4 seconds is generated with 20 computational cores. For DCSM-FM 0.5nm, two runs are reported, both with a times step of 118.7 seconds. One run is done with 20 computational cores and one with 1 core. When a model is run on more cores, the computations are done more in parallel, speeding up the process. To relate the computational efficiency to that of the ML model, Table 5.9 reports the computation time in seconds per time step by multiplying the original computation time by the number of cores and dividing it by the number of estimates in the reference period of one year. According to this metric, the ML model takes 0.09 seconds to estimate one output value, while DCSM-FM 100m takes 46 times as long with 4.27 seconds.

6 Discussion

This chapter highlights the possibilities and challenges of the model developed in this study. The potential of the model to estimate coastal water levels is discussed, as well as the results of this study within the context of environmental monitoring of high water levels. Additionally, the discussion addresses the assumptions and decisions that have been made in this study.

6.1. Potential of machine learning

The model developed in this study shows that machine learning is a powerful tool in coastal non-tidal water level estimation. It creates opportunities to use globally available data in a simple neural network, and integrates multiple data sources into one data-driven model. An analysis of the trained weights of the model is possible, which can be used to explain the main drivers of high water level events.

The neural network developed in this study is one of the simplest machine learning models available, yet it is still able to estimate coastal non-tidal water levels with an MSE between 0.011 and 0.018 m, an MAE between 0.078 and 0.101 m and a σ_ε between 0.100 and 0.134 m. In comparison, the same metrics for DCSM are between 0.004 and 0.005 m for the MSE, between 0.049 and 0.054 m for the MAE and between 0.052 and 0.068 m for the σ_ε. While the ML model does not reach the same performance, the SNR that lies between 22.7 and 101.0 suggests that the model is still applicable in practice, since the variability of the water level itself is 22.7 to 101.0 times larger than the MSE of the ML model. Despite the better performance of DCSM, the neural network possesses several advantages relevant to the development of coastal water level research.

Firstly, the computation time for the ML model is much smaller than for a local numerical model such as DCSM (Table 5.9). The neural network takes 3 minutes and 26 seconds to estimate one year of coastal water levels, where DCSM-FM 0.5 nm takes 8.3 hours (Zijl, Zijlker, Laan, & Groenenboom, 2022) and DCSM-FM 100m even takes 52.8 hours (Zijl, Groeneboom, et al., 2022). Furthermore, the neural network is applied on a computer with a computational power of 4 cores, while DCSM is executed on a computational cluster with a power of 20 cores, divided over 5 nodes. When using only a single core, DCSM-FM 0.5 nm takes 66 hours to estimate one year. One should take into account, however, that DCSM produces a grid of estimates, as well as other variables aside from water level. Additionally, DCSM computes water level estimates with a time step of 35 seconds to 2 minutes, while the ML model computes water levels with an interval of 1 hour. To truly be able to compare the two models, DCSM has to be calibrated for 1-hour estimates and applied on a computer with 4 cores. Unfortunately, this was not possible. Therefore, the computation time per time step has been reported as well. These results show that the ML model is able to compute a single output value in 0.09 seconds, while DCSM takes 0.89 to 4.27 seconds. So, for the specific application of estimating coastal water levels at one location, the ML model is still assumed to be much faster. This difference is partly due to the simplicity of the neural network, even though it takes a dataset of 314401 input features to estimate one non-tidal water level.

Another advantage of this ML model is its adaptability. Traditional numerical models such as DCSM rely on predefined parameters and equations, whereas a ML approach allows for continuous refinement and adaptation. Naturally, DCSM also undergoes many updates as the knowledge about the physics and hydrodynamics on the North Sea and Waddensea increases, which hopefully in the future will also contain

ML or satellite data, but for the ML model, this property is more flexible and fast-paced. Adapting the model can be done by e.g. adding neurons or hidden layers, changing the activation function, the optimising algorithm or the other design parameters mentioned in Section 4.4. This allows for a more comprehensive understanding of seasonal variations, long-term trends, and episodic events by the model, contributing to a more robust and possibly more accurate predictive method.

Ultimately, the goal of this study is to effectively apply a machine learning model such as a neural network in regions lacking observational ground stations and local models, especially when global models perform only marginally. The main limitation of this developed neural network is its dependence on ground-truth data. This means that a tide gauge is needed for training the model. Subsequently, its applicability is confined to regions which have the same physical and hydrodynamic properties, meaning that the patterns and correlations between the data and the coastal water levels should align with those of the training region. Moreover, the input data needs to have the same shape, which is challenging when the area of interest is different. Despite these constraints, this study proves that machine learning methods like a neural network have potential, because of their flexibility, computational efficiency and room for refinement.

6.2. Assumptions and uncertainties

Within this study, a couple of uncertainties have to be highlighted. These consist of choices and assumptions that have been made throughout the development of the neural network, either based on visual inspection of intermediate results or associated literature.

Firstly, the harmonic tidal analysis of the tide gauge time series is done with the Hatyan method, which implements 95 harmonic components (Veenstra & Kerkhoven, 2020). The satellite altimetry reprocessing algorithm X-TRACK uses FES2014b to compute the ocean tide, which implements 34 harmonic components (Birol et al., 2021). This means that there are harmonic components, however small, still present in the satellite altimetry data that are not present in the tide gauge time series. The components with the largest amplitudes within the TG record (M2, M4, S2, MS4, N2, MU2, NU2, O1, SA, K1 and MN4, see also Appendix D) are applied in both tidal analyses, reducing the differences between the two harmonic corrections significantly. Additionally, the uncertainties in FES2014b differ from those in Hatyan due to differences in the harmonic analysis. This can cause discrepancies between the satellite altimetry observations and the TG record. While the ML model can theoretically recognise and manage these differences, it is recommended to look into them or apply the same harmonic tidal analysis on both data sources to reduce discrepancies and thus simplify things for the ML model.

Secondly, these tidal models only include harmonic tidal signals. However, non-linear interactions between tides and other processes such as wind speed and atmospheric pressure are still present in both the TG time series and the satellite altimetry observations. These signals add noise to the non-tidal water levels, which means that the ML model is expected to estimate them as well. To improve these estimates, it is recommended to include the tidal signals in the model.

When comparing the performance of the ML model with DCSM, they are both evaluated against tide gauge time series, which serve as the ground truth. However, DCSM relies on tide gauge data from 211 stations for the calibration of bottom roughness, including those used in this study (Zijl, Groeneboom, et al., 2022). This calibration used data from 2017, a year included in the validation period of the ML model. However, once the bottom friction is determined, the model no longer uses any tide gauge data for the estimation of water levels. Moreover, DCSM itself is validated against tide gauges, which is considered valid. Therefore, this dependency is not considered significant.

During the performance analysis of the ML model on high water levels, the thresholds based on the POT are chosen carefully (see Appendix F). Since POT is generally considered a sensitive method, one has to take into account that the performance metrics on high water levels are sensitive as well and prone to changes when the threshold is adjusted.

Within the developed neural network, assumptions about the number of neurons, the choice of activation function and the resampling method have been made. The number of neurons is set to 32, based on visual inspection of results after trying a number of 8, 16, 32, 64 and 128 neurons. A number of 32 was found that converged the model without overtraining it. The choice of the activation function has been based on literature and was chosen to be the ReLU function. A ReLU function accounts for any non-linearities in the data while simultaneously decreasing the risk of a vanishing gradient problem (Glorot et al., 2011; Tan and Lim, 2019). Many variations of the ReLU function are tested nowadays to find the best suitable version (Z. Hu et al., 2021; Javid et al., 2021; Mirzadeh et al., 2023; Oh et al., 2023). Finally, the constant which states how many times the high water level events should be repeated within the training data is based on an empirically defined number, namely a CDF > 0.8 and a ratio of 0.05. These values are chosen based on visual inspection of intermediate results, where a ratio of 0.02, 0.05 and 0.1 have been tested. It has been found that a ratio of 0.05 results in a better performance than the other ratios. A ratio of 0.1 resulted in a degraded general performance and a ratio of 0.02 did not improve the model's performance on high water level events significantly. The threshold of 0.8 has not been tested but should be when continuing this research.

6.3. Estimation of coastal water levels

Several details about the developed ML model are worth interpreting in further detail. Firstly, the performance metrics of the ML model were computed after the harmonic tidal signals were restored. The MSE, MAE and σ_ε remain unaffected by this step. However, the SNR is impacted, which is crucial for assessing the model's applicability as it compares the water level variability to the variability of the model's error. Since tidal signals account for the majority of the water level variations, it is essential to understand how the model's error relates to this.

Comparing the three models developed in this study, the model for Europlatform performs best on all metrics (Table 5.4), even though the input into the model is exactly the same. Vlissingen and Scheveningen perform similarly, though Vlissingen performs a little better when looking at the high water levels (Table 5.6). These performances align with the differences in environmental factors between the locations. Europlatform, located offshore, experiences minimal coastal interference such as reflection, refraction or wind set-up. On the other hand, both Vlissingen and Scheveningen are located within harbours (Figure 3.2a and 3.2b), where the presence of passing ships, refraction and/or reflection of water, potential wind set-up and other factors induce strong coastal interference. Given that Scheveningen is located deeper within a harbour with shallower water, the model's poorer performance is expected.

In general, the produced time series of the coastal water levels appears smoothed compared to the ground truth data (Figure 5.1, 5.2 and 5.3). This differs from findings reported in the literature (Bruneau et al., 2020; Passaro and Juhl, 2023; Xie et al., 2023). However, Bruneau et al. (2020) use tide gauge time series to train their global neural network, which contain high-frequency temporal variability. Passaro and Juhl (2023) use satellite altimetry observations to estimate daily SLAs, looking at a circular area around the location of interest with a radius of 300 km and a time window of 15 days. They reach an MSE between $4*10^{-4}$ and $1.4*10^{-2}$ m. Though the Random Forest Regression method appears promising, it has not been tested on hourly SLAs, nor does it include sea surface variability caused by atmospheric disturbances. The storm surge forecasting model developed by Xie et al. (2023) uses a convolutional neural network with a varying time window to predict storm surges up to 24 hours beforehand, also using tide gauge time series as input into the model. The model reaches an MSE of $2.89*10^{-2}$ m. The neural network developed in this study uses a time window of 48 hours and ERA5 gridded data of 31 km spatial resolution, without any high-frequency information from tide gauge time series. To minimise the MSE, the neural network trains the weights and biases to fit most situations. Due to the lack of high-frequency information, this possibly causes the smoothing. Still, an MSE of 0.017 m is considered satisfactory in line with the literature, with a SNR between 22.7 and 101.0 sufficient for practical applications such as extracting trends and return periods.

In addition to the lack of high-frequency temporal variability in the input data, the tide gauge non-tidal water levels show significant non-linear interactions. The ML model misses these signals, despite their presence in both tide gauge and satellite altimetry data. These interactions, influenced by local environmental factors like wind speed and atmospheric pressure, are too complex for the model to estimate accurately, possibly because they occur on a local scale not captured by satellite data. As a result, the output of the ML model looks smoothed compared to the tide gauge non-tidal water levels.

6.4. Estimation of high water levels

Since this study aims to estimate high water level events, it is worth noting that these events are generally underestimated by the ML model. Similar results have been found by Bruneau et al. (2020), Gharineiat and Deng (2015) and Xie et al. (2023), who all underestimate high water level events with their respective ML models. This can occur due to several factors, such as the underestimation of extreme winds by ERA5 (Haakenstad et al., 2021) and the underestimation of extreme water levels by satellite altimetry due to undersampling (Darko et al., 2023). The ML model assumes stationary relationships between input variables and output, but when both ERA5 and satellite altimetry underestimate extreme conditions, these relationships change. As a result, high water level events are underestimated more compared to regular water levels (see also Figure 5.8).

Another reason for the underestimation of high water level events is the imbalanced training data. High water levels are underrepresented, and the events that are included are skewed towards storms from the northwest. This bias is due to the area of interest, which includes almost the entire North Sea but only the Strait of Dover and part of the English Channel to the south. Consequently, the model can anticipate storms from the northwest up to 48 hours in advance, whereas storms from the south appear more suddenly. Additionally, northwest storms, typically occurring in winter, last longer, while southern storms, often in summer, are shorter. This imbalance explains why summer storms from the south are more frequently underestimated than those from the north.

To battle this imbalance, a resampling of the training data is applied to include more high non-tidal water level events during training, which improves the performance. Additionally, sample weights are added to the training samples, linking each sample to a weight that increases proportionally to the increase in non-tidal water level. This enables the ML model to estimate high water level events coming from the west or north better (Figure 5.1g, 5.2g and 5.3g), though events that originate from the south are still captured marginally (Figure 5.1a, 5.2a and 5.3a).

6.5. K-fold cross-validation results

Regardless of the different performances at different locations, the K-fold cross-validation proves that the model is robust. Training and validating the model on different datasets confirms that the performance metrics will not differ much from the sample mean. To elaborate, the range of the error metrics MSE, MAE and σ_ε are all between 0.004 and 0.017 m (Table 5.5). In addition, the metrics of the Scheveningen model (Table 5.4) are all within this range. Lastly, the K-fold cross-validation is applied to Scheveningen, which performs poorest. This suggests that the models for Vlissingen and Europlatform are equally or more robust, though this has not been verified.

For high water level events, the performance of the K-fold models is better than the full Scheveningen model. The sample means for MSE and MAE and are all smaller than those of the full model (Table 5.6). Contrarily, the σ_ε's of the K-fold models are larger than the one of the full model. Still, the range of the error metrics for high water level events ranges between 0.010 and 0.017 (Table 5.7), which means the model is only slightly less robust when looking at the MSE, but equally robust when looking at the MAE or σ_ε. The sample mean for the SNR is decreased from 26.5 to 14.4, the range decreased as well from 7.1 to 2.4, also indicating a lesser robustness. Improving the SNR, and thus the MSE, would be beneficial here.

6.6. Weight analysis

Since the ML model developed in this study is relatively simple compared to many other machine learning methods, it is possible to distinguish several characteristics of the trained model. Firstly, the second-layer weights (Figure 5.9) for all three models fall within the range of -0.4 to 0.4, but some of these weights are sufficiently close to zero, that they become ineffective, meaning that they do not transmit any signal to the output layer. These inactive neurons, commonly referred to as "dead neurons", produce values that are close to zero.

For the ML models developed for the three locations, the Scheveningen model has the highest number of dead neurons, while the Vlissingen model has the least. There are a couple of explanations for this discrepancy. Firstly, the high complexity of the Scheveningen area means the model struggles to find significant relationships between input and output when non-tidal water levels are small (between -0.192 and 0.235 m). In these cases, the Scheveningen model estimates 39.5% of the cases as a constant value, indicating that the input provides insufficient information under nearly stationary conditions (Table 5.3). However, this is not necessarily decreasing performance, as the Europlatform model performs better than the Vlissingen model despite having more dead neurons and more frequently giving a constant output.

The range in which constant outputs are estimated is smallest for Europlatform, with 95% of the constants falling between -0.195 and 0.126 meters, a range of 0.321 meters (Table 5.3). In contrast, this range is 0.430 meters for Vlissingen and 0.517 meters for Scheveningen. Additionally, Europlatform's constants are more concentrated within this range, with 43.8% of outputs being constant compared to 28.3% for Vlissingen. This can be an explanation as to why Europlatform, despite having more dead neurons, performs better than Vlissingen. This indicates that the number of dead neurons is not directly proportional to the performance.

Furthermore, data redundancy is a factor. The ML model may deactivate neurons deemed redundant, as they do not contribute significant information to the output. Vlissingen may have less redundant data due to slightly greater complexity, possibly resulting from increased land-sea interactions.

6.6.1. Satellite altimetry

The first-layer weights for the input features corresponding to satellite altimetry observations (WL, d_E, d_N, and dt) account for about 49% of the first-layer weights of the active neurons, indicating equal importance of satellite altimetry and ERA5 in estimating non-tidal water levels. Among these, the weights for dt contribute the most with approximately 20%. However, since dt is min-max normalised, it always ranges between 0 and 1. The other variables (WL, d_E and d_N) are Z-score normalised, meaning that they often exceed an absolute value of 1. Therefore, these variables require a smaller weight to have the same impact as dt. This indicates that dt has a smaller contribution to the estimation of non-tidal water levels than implicated in Table 5.8, and the other three variables have more.

The results show that d_E has a larger contribution than d_N, as shown by their weight contributions (Table 5.8) and spatial distributions (Figures 5.12 and 5.13). Both are individually normalised with a Z-score normalisation because they have different distributions (see Appendix E). However, both variables convey the same kind of information, which is distances between the satellite observations and the tide gauge. Furthermore, the distribution of d_N (Figure E.5) does not look as normally distributed as the distribution of d_E does (Figure E.4). This indicates that Table 5.8 might report a larger contribution of d_E than preferable for the model. Z-score normalisation might thus reduce the usefulness of d_E and d_N, especially when normalised separately. It is recommended to normalise both variables with the same parameters or use a different normalisation method. This can change the configurations of the weights during the training stage to allow more contribution from d_N and possibly a better estimation.

Weights associated with inactive neurons are excluded from the analysis to avoid noise, as these weights are nullified once their corresponding neurons become inactive. These weights have an anomalous shape compared to the first-layer weights connected to active neurons, as shown in Figure 5.10 as grey lines, implying that the model does not spend its resources training first-layer weights once their corresponding neurons are killed.

Figure 5.10 also shows that the first-layer weights seem to converge to zero when the number of observations increases. As mentioned in Section 5.4, this phenomenon is caused by the processing method and availability of satellite altimetry data. The processing method applies a padding to deal with the inhomogeneous nature of the satellite altimetry data due to their availability (Figure 4.7). Consequently, many input datasets of 48 hours contain many zeros at the end of their one-hour arrays. To get the best performance, the model trains a certain configuration of weights that minimises the overall loss, which results in less weights on these observations, as they become redundant.

Another key observation is the high weights at the edges of the area of interest for observed non-tidal water levels. This is likely due to the processing method as well, but it is unclear if these observations are genuinely the most important or just prioritised because they are fed to the model first. Restructuring the processing method could help, such as feeding observations closest to the output timestamp into the model first or prioritising the observations closest to the tide gauge.

6.6.2. ERA5

The first-layer weights connected to the ERA5 input variables are presented as heat maps shown in Figure 5.14, 5.15 and 5.16, their differences shown in Appendix H. Their weight contributions are shown in Table 5.8, which shows that the longitudinal wind speed (U10) is the most important variable for estimating non-tidal water levels, accounting for about 18% of the total weights. The lateral wind speed (V10) and sea level pressure (p) follow closely with around 16% and 17% respectively. Since the Dutch coast is perpendicular to the direction of U10, this causes a larger wind set-up compared to V10. The model has effectively recognised this correlation, giving U10 larger weights.

As mentioned in Section 5.4, a large and prominent characteristic of all three ERA5 variables is that the most important features are located along the Dutch coastal region and decrease in importance when travelling north. When the northern North Sea is reached, this importance becomes inversely proportional to the ones at the Dutch coast. This is most evident in the sea level pressure heat maps (Figure 5.14), but also in the lateral wind speed maps (Figure 5.16). It implies that active pressure systems have the most influence on the estimates of the non-tidal water level at the Dutch coast and in the northern North Sea. Similarly, large lateral wind speeds along the southern Dutch coast and in the northern North Sea signal important characteristics for the estimation of coastal non-tidal water levels. When looking at the lateral wind speed component, the overall importance increases as time progresses from -48 hours to -1 hour before the estimated water level.

6.6.3. Regular events over high water level events

The ML model adjusts its parameters (weights and biases) to minimise the overall loss across all samples, ensuring good average performance. However, this generalisation means the model might not be optimised for extreme situations like high water level events, where a different configuration of parameters could be more effective. For example, since we know U10 and V10 are underestimated in extreme conditions, the current parameters might not be optimal for the performance on high water level events. Solving this problem within the current model by changing the parameters will decrease the performance on regular conditions.

Differences between bad and good performance parts are explained using three examples in Figures 5.4 and 5.5. These confirm that poor performance during high water level events is attributed to the short duration of storms and their direction, characterised by brief periods of high U10 and p. Conversely, good performance is linked to higher wind speeds and longer storm durations. Here, the model shows improved water level estimation even without satellite altimetry compared to the poor performance during the short southern storm. Interestingly, V10 does not show significant differences between poor and good performance scenarios, indicating that it is less crucial. This is also proven by Table 5.8.

Identifying which features matter most in specific conditions is challenging, as the trained parameters only provide a general view, and some characteristics or examples might be coincidental. This characteristic of ML is both a risk and an advantage. The model can uncover patterns and correlations that might otherwise go unnoticed, yet these patterns can be hard to explain. Though, in a way, this makes working with machine learning methods all the more fun.

7 Conclusion and outlook

The model developed in this study presents a method to estimate hourly coastal non-tidal water levels at a location of a tide gauge, combining machine learning with satellite altimetry and assimilated ERA5 data. This chapter presents the key findings of the sub-questions listed in Chapter 1 and determines if the research objective has been reached. The outlook shares some suggestions for improvements and further research, either short- or long-term.

7.1. Conclusion

The main research objective is **to develop a machine learning model that produces hourly non-tidal water level estimates at the location of a tide gauge, based on satellite altimetry observations and ERA5 pressure and wind fields within a 48-hour time window and across an area including the North Sea and the Strait of Dover**. Three locations have been chosen in this study to train and test the developed neural network on, to assess whether this neural network is able to estimate non-tidal water levels in locations with different physical and hydrodynamical processes. This study has proven that it is possible to build a neural network that estimates hourly coastal non-tidal water levels with a mean squared error between 0.011 and 0.018 m and a standard deviation between 0.100 and 0.134 m.

How does the performance of the ML model compare to tide gauge observations and a regional numerical model?

The developed neural network has been trained and tested on tide gauge time series from Scheveningen, Vlissingen and Europlatform. The performance of the model has been determined with four metrics, namely MSE, SNR, MAE and σ_ε, which have been applied to the ML model outputs as well as DCSM outputs for similar time intervals. Furthermore, the robustness of the model has been tested with a K-fold cross-validation with $K = 4$.

With an MSE of 0.011 m, an MAE of 0.078 m and a σ_ε of 0.100 m, Europlatform outperforms Vlissingen and Scheveningen, which both have an MSE of 0.018 m, an MAE close to 0.100 m and a σ_ε of 0.134 m. The robustness of the neural network is high according to the K-fold cross-validation. The SNR differs per location, depending on the amplitude of the tidal signals, ranging from 22.7 for Scheveningen to 101.0 for Vlissingen, with Europlatform in between with 37.4. These high SNR values mean that the variability of the original signal is larger than the error spread of the model, which suggests that the models are physically applicable, though this depends on the application. When comparing these metrics to DCSM, the local numerical model outperforms the neural network for all locations. This agrees with the expectation related to the performance, though several advantageous characteristics are highlighted.

The computation time for the trained neural network is much faster than for DCSM. Where the neural network takes 0.06 hours to compute one year of coastal non-tidal water levels, DCSM-FM 100m takes 52.8 hours, even with a computational power that is five times what the neural network has used. This corresponds to 0.09 seconds per time step for the neural network and 4.27 seconds for DCSM. Furthermore, the neural network is adaptable, allowing for continuous refinement and fast updates compared to traditional numerical models such as DCSM. However, the neural network's main limitation is its reliance on ground-truth data, restricting its applicability to regions without tide gauges. Despite these constraints, the study demonstrates the potential of ML models in coastal non-tidal water level estimation due to their flexibility, computational efficiency, and capacity for refinement.

How well does the ML model estimate high water levels?
Next to testing the model on hourly water levels, an analysis of high water levels is conducted through a POT method. As expected, the Europlatform performs best, followed by Vlissingen and Scheveningen. The neural network at Europlatform has an MSE of 0.013 m, a MAE of 0.090 m and a σ_ε of 0.113 m. At Vlissingen, the MSE equals 0.016 m, the MAE is 0.099 m and the σ_ε is 0.117 m. For Scheveningen, the model performs worst, with an MSE of 0.054 m, an MAE of 0.184 m and a σ_ε of 0.164 m.

The high water levels are consistently underestimated in the neural network, and some events are missed when storms enter the North Sea from the south. Explanations for this include the underestimation of extreme winds and water levels by ERA5 or satellite altimetry and the imbalanced training data that is skewed towards general conditions. The neural network is equally robust for high water levels as it is for hourly non-tidal water levels, with an equal range for σ_ε, a smaller range for SNR, a similar range for the MAE and a larger range for the MSE.

Improving the model for high water levels is challenging, yet rewarding, because that is why these kinds of models essentially exist. Recommendations include revising the data processing, extending the area of interest to the south, adjusting the loss function or adding layers and/or neurons to the network.

Which factors affect the estimation of non-tidal water levels estimated with the ML method?
Next to general performance metrics, it is also analysed which variables or environmental factors affect the estimation of water levels. This is done by reconstructing time series and zooming into a few bad or good performance parts to analyse the input data that the model receives. A weight analysis is also done to determine which features contribute more to the estimations.

From the reconstructed time series, one can conclude that short high water level events coming from the south are hard to estimate by the model. This can be explained by the area of interest not spanning the English Channel and the time window of 48 hours that is used as input data. Short storms are harder to capture in this sense, and the model does not see them coming when they originate from the south.

The availability of satellite observations influences the model's ability to estimate water levels due to the way the satellite observations are structured within the training data. The weights given to the observations decrease as the number of observations increases. Still, the model can perform well without much satellite data if the ERA5 data provides enough information.

Finally, the importance of the ERA5 variables is subject to some spatial variation. In general, the data close to the Dutch coast are most important, which aligns with the expectation. The importance of data originating from the northern North Sea is inversely proportional to the importance of the data from the Dutch coast, meaning that the northern North Sea is also significant in the estimation of coastal water levels at the locations of interest.

7.2. Outlook

The key areas for improvement in this study are the estimation of high water levels and the ML models applicability to different regions. The ML model currently underestimates high water level events due to several factors, which are discussed in Chapter 6. However, accurate estimation of high water level events is crucial, as they can lead to coastal hazards such as floods and erosion, which can cause damage to important infrastructure and population displacement (Almar et al., 2021; Hallegatte et al., 2011; Nicholls, 2011; Parise et al., 2009). Recommendations to improve high water level estimation are divided into two categories: input data and model design.

The second important issue is the applicability of the model to different regions. Currently, the ML model is only applicable within the area of interest, and only for locations with similar physical and hydrodynamical processes as the three locations on which the model has been trained. For the Dutch coastal zone, many tide gauges already exist, making the ML model superfluous. It would be beneficial to know if this method can be used to create models in more remote locations where tide gauges are sparse and where there is no access to elaborate, accurate numerical models.

7.2.1. Estimation of high water levels

To improve the performance of the ML model during high water level events, the input data can be refined. Firstly, it is known that ERA5 underestimates strong wind speeds and satellite altimetry undersamples extreme water levels. Therefore, creating a separate model that has been trained specifically on high water level events can be beneficial here. This can be done by applying the POT method to the TG data, taking only the high water level events with a time window of 48 hours for training. Comparing this model to the original ML model for hourly non-tidal water levels will help to identify which input features in which regions are more crucial for the estimation of high water levels compared to regular water levels.

Input data

Another recommendation is to extend the area of interest, particularly towards the English Channel. Figures 5.14, 5.15 and 5.16 show a clear pattern of high weights near the Strait of Dover for all three models, suggesting that this area is important for estimating coastal water levels at the locations of interest. The same holds for the northern border of the area of interest. This will be a trade-off between computational cost and desired performance, as the number of input features will increase rapidly. Another addition would be to include more satellite altimetry missions, such as Cryosat-2, Sentinel-3B, Sentinel-6, SWOT and future missions HY-2E/-2F/-2G/-2H and CRISTAL (Figure 2.3). These have to be reprocessed by XTRACK first before they can be included.

Including the tidal harmonics in the training data also gives the model a better overview of the processes that happen inside the area of interest, especially on the non-linear interactions that are prominent in the ground-truth data, but not in the ML model estimates. This has already been applied by multiple studies that predict storm surges with ML models (Jia and Taflanidis, 2013; Sahoo and Bhaskaran, 2019; Xie et al., 2023). Moreover, using the same tidal model (FES2014b) as XTRACK, or the same tidal harmonics, reduces any discrepancies between the ground-truth data and the satellite altimetry data. Additional information on sea surface temperatures (Hersbach et al., 2018), wave climates Jia and Taflanidis, 2013, salinity (Antonov et al., 2002) and freshwater influxes (Lombard et al., 2009) would also help.

Another recommendation is to look into the redundancy that the variable satellite data availability and processing method causes in the ML model. This includes looking into whether this redundancy is an issue, for example by looking if low-availability periods have a different accuracy than high-availability periods, or finding a way to get rid of the padding that adds many zeroes to the input data from satellite altimetry.

Continuing on the processing method, the normalisation of d_N and d_E should be improved. Currently, they are normalised individually based on a Z-score normalisation, which assumes a Gaussian distribution. However, the distribution of d_N and d_E differ from the Gaussian (see Appendix E). Moreover, d_N contributes less to the estimates than d_E does according to the weight analysis presented in Section 5.4. A more in-depth feature analysis is recommended here, where different distributions such as the Gaussian, uniform, Poisson or exponential distributions can be fitted to determine which one fits best on the input variables. Then, the variables can be normalised accordingly.

This analysis will also address dataset imbalances, which cause the model to estimate general conditions well but high water events poorly. Applying stratified sampling and performing a sensitivity analysis on resampling parameters can further improve high water level estimates. This analysis will also address the issue of data imbalances. Imbalances are induced by the underrepresentation of high water level events, which causes the ML model to skew towards estimating general conditions well and extreme conditions poorly. A resampling method has been applied in this study, which results in some improvements, but the high events are still underestimated. To improve them, stratified sampling is recommended, which means binning the ground-truth non-tidal water levels and selecting an equal number of non-tidal water levels from each bin. Another recommendation is to perform a sensitivity analysis on the resampling parameters (threshold and ratio) to determine which values result in the most accurate high water level estimates.

Model design

Next to improvements on the data and their processing method, several recommendations for the model design are also identified. Firstly, the choices on the design parameters, such as the number of neurons and layers, the activation function, the loss function, the number of epochs and the batch size, can be changed one by one to find the most optimal combination of design parameters. Adjusting the loss function could help to focus more on the estimates that underestimate high water levels rather than overestimate them. One possible improvement could be to use the "Pinball Loss" function which is used in ML as a replacement for the MSE as a loss function, for example in forecasting energy usage of households (Wang et al., 2019). The choice of activation function could help in reducing the probability of a vanishing gradient problem (Glorot et al., 2011). The number of neurons, hidden layers, epochs and the batch size are determined empirically, and can be optimised by trying different configurations.

Testing different ML methods would provide a better understanding of the potential of using ML in the estimation of coastal water levels. Popular methods for estimating storm surges include Kriging models (Kyprioti et al., 2021), Support Vector Regression (Awad et al., 2015) and Recurrent Neural Networks (Qin et al., 2023). One recommendation to build further on the current model is to combine the shallow neural network with a convolutional neural network (CNN) which is suitable for spatial data. This has been proven to be effective by Park et al. (2022) and Vincent et al. (2022). The ERA5 data can be passed through a CNN before feeding the output into the neural network (see Figure 4 from Vincent et al., 2022). Finally, ensemble learning could be a powerful tool to improve the model (Mohammed & Kora, 2023). The easiest example to achieve this would be to let all K-fold models estimate hourly non-tidal water levels and take the mean of all four estimates. Alternatively, a sequence of neural networks could be considered, each one focussing on the largest errors made by the previous network. Note that this approach will significantly increase the computational cost.

7.2.2. Applicability to different regions

Section 6.1 mentions that the largest limitation of the current model is its applicability to different regions. Passaro and Juhl (2023) suggests an approach that takes all data within a range of 300 km around the tide gauge as input, making it independent of the shape of the region. Combining this with the shallow neural network and the input variables considered in this study will increase the ML models general applicability and reduce the dependence on a pre-defined area of interest.

Another recommendation is to apply one of the three trained models to a different location containing a TG, for example IJmuiden or Den Helder, and assess the performance at those locations. Seeing how the performance of the model changes when travelling further from the ground-truth location is useful to know, especially if the model will be applied to locations without a TG.

Given the desire to apply this ML model in regions with few or no tide gauges, the final recommendation of this study is to train the model in a different region and evaluate its performance there. The XTRACK algorithm has processed coastal satellite altimetry data from 27 geographical zones, spanning most of the world's coasts (Aviso+, 2023), and with the global coverage of ERA5, the model can be trained on virtually any coast. This approach will test the methods general applicability across various regions. However, two challenges remain: initial training still requires a tide gauge, and design parameters must be re-optimized for each new location. Overcoming these obstacles will pave the way for robust, globally applicable coastal water level estimations.

References

Adebisi, N., Balogun, A. L., Min, T. H., & Tella, A. (2021). Advances in estimating Sea Level Rise: A review of tide gauge, satellite altimetry and spatial data science approaches. *Ocean & Coastal Management*, *208*, 105632. https://doi.org/https://doi.org/10.1016/j.ocecoaman.2021.105632

Almar, R., Ranasinghe, R., Bergsma, E. W., Diaz, H., Melet, A., Papa, F., Vousdoukas, M., Athanasiou, P., Dada, O., Almeida, L. P., et al. (2021). A global analysis of extreme coastal water levels with implications for potential coastal overtopping. *Nature communications*, *12*(1), 3775.

Alves, J. H. G., & Young, I. R. (2003). On estimating extreme wave heights using combined Geosat, Topex/-Poseidon and ERS-1 altimeter data. *Applied Ocean Research*, *25*(4), 167–186.

Andersen, O. B., Cheng, Y., Deng, X., Steward, M., & Gharineiat, Z. (2015). Using satellite altimetry and tide gauges for storm surge warning. *Proceedings of the International Association of Hydrological Sciences*, *365*, 28–34.

Andersen, O. B., & Scharroo, R. (2010). Range and geophysical corrections in coastal regions: And implications for mean sea surface determination. In *Coastal altimetry* (pp. 103–145). Springer.

Antonov, J. I., Levitus, S., & Boyer, T. P. (2002). Steric sea level variations during 1957–1994: Importance of salinity. *Journal of Geophysical Research: Oceans*, *107*(C12), SRF–14.

Apel, J. R. (1987). *Principles of Ocean Physics*. Academic Press.

AVISO+. (2022). *X-TRACK-L2P: sea level anomalies dedicated to coastal applications (2022 version)*. Retrieved March 5, 2024, from https://www.aviso.altimetry.fr/en/data/products/sea-surface-height-products/regional/x-track-sla/x-track-l2p-sla-version-2022.html

Aviso+. (2022). *Along-Track Sea Level Anomalies regional products (X-TRACK-L2P) User Handbook* (tech. rep.). https://doi.org/10.24400/527896/A01-2022.020

Aviso+. (2023). *Home*. Retrieved November 1, 2023, from https://www.aviso.altimetry.fr/en/home.html

Aviso+. (2024). *Timeline of modern radar altimetry missions*. Retrieved April 15, 2024, from https://doi.org/10.24400/527896/A02-2022.001

Awad, M., Khanna, R., Awad, M., & Khanna, R. (2015). Support Vector Regression. *Efficient learning machines: Theories, concepts, and applications for engineers and system designers*, 67–80.

Badulin, S. I., Grigorieva, V. G., Shabanov, P. A., Sharmar, V. D., & Karpov, I. O. (2021). Sea state bias in altimetry measurements within the theory of similarity for wind-driven seas. *Advances in Space Research*, *68*(2), 978–988.

Bao, L., Lu, Y., & Wang, Y. (2009). Improved retracking algorithm for oceanic altimeter waveforms. *Progress in Natural Science*, *19*(2), 195–203.

Bij de Vaate, I. (2023). *Spatiotemporal Variability in Global Storm Surge and Tidal Water Levels from Satellite Radar Altimetry* (Doctoral dissertation). Delft University of Technology. https://doi.org/https://doi.org/10.4233/uuid:c9aa7e4e-5425-49cc-93ff-e53382402549

Birol, F., & Passaro, M. (2020). *ESA Sea Level CCI+, Algorithm Theoretical Basis Document* (tech. rep.).

Birol, F., Fuller, N., Lyard, F., Cancet, M., Nino, F., Delebecque, C., Fleury, S., Toublanc, F., Melet, A., Saraceno, M., et al. (2017). Coastal applications from nadir altimetry: Example of the X-TRACK regional products. *Advances in Space Research*, *59*(4), 936–953.

Birol, F., Léger, F., Passaro, M., Cazenave, A., Niño, F., Calafat, F. M., Shaw, A., Legeais, J.-F., Gouzenes, Y., Schwatke, C., et al. (2021). The X-TRACK/ALES multi-mission processing system: New advances in altimetry towards the coast. *Advances in Space Research*, *67*(8), 2398–2415.

Bocharov, G. (2023). *pyextremes* (Version 2.3.2). https://github.com/georgebv/pyextremes

Bosboom, J., & Stive, M. J. (2021). *Coastal Dynamics*. TU Delft Open.

Brown, G. (1977). The average impulse response of a rough surface and its applications. *IEEE transactions on antennas and propagation*, *25*(1), 67–74.

Bruneau, N., Polton, J., Williams, J., & Holt, J. (2020). Estimation of global coastal sea level extremes using neural networks. *Environmental Research Letters*, *15*(7), 074030.

Cabanes, C., Cazenave, A., & Le Provost, C. (2001). Sea level rise during past 40 years determined from satellite and in situ observations. *Science*, *294*(5543), 840–842.

Carrère, L., Faugère, Y., Lyard, F., Allain, D., Dibarboure, G., & Picot, N. (2019). Last TUGO model simulations and perspectives of evolution of the Dynamic Atmospheric Correction for altimetry.

Carrère, L., & Lyard, F. (2003). Modeling the barotropic response of the global ocean to atmospheric wind and pressure forcing-comparisons with observations. *Geophysical Research Letters*, *30*(6).

Cartwright, D., & Edden, A. C. (1973). Corrected tables of tidal harmonics. *Geophysical journal international*, *33*(3), 253–264.

Cartwright, D., & Tayler, R. (1971). New computations of the tide-generating potential. *Geophysical Journal International*, *23*(1), 45–73.

Castillo, E., & Hadi, A. S. (1997). Fitting the generalized pareto distribution to data. *Journal of the American Statistical Association*, *92*(440), 1609–1620.

CEOS and ESA. (2024). *Missions, Instruments, Measurements and Datasets*. Retrieved April 10, 2024, from https://database.eohandbook.com/index.aspx

Church, J. A., Hunter, J. R., McInnes, K. L., & White, N. J. (2006). Sea-level rise around the Australian coastline and the changing frequency of extreme sea-level events. *Australian Meteorological Magazine*, *55*(4), 253–260.

COASTALT. (2023). *Altimetry sails closer to the coast*. https://coastalt.eu/coastalt-short-web-summary

Darko, S., Adjei, K. A., Gyamfi, C., & Odai, S. N. (2023). Statistical evaluation of Jason-2 satellite altimetry products in a trans-boundary river basin; the case study of the White Volta River basin in Ghana. *Modeling Earth Systems and Environment*, 1–13.

Davis, C. (1995). Growth of the Greenland ice sheet: A performance assessment of altimeter retracking algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, *33*(5), 1108–1116.

Davis, C. (1997). A robust threshold retracking algorithm for measuring ice-sheet surface elevation change from satellite radar altimeters. *IEEE Transactions on Geoscience and Remote Sensing*, *35*(4), 974–979.

Davis, J., Herring, T., Shapiro, I., Rogers, A., & Elgered, G. (1985). Geodesy by radio interferometry: Effects of atmospheric modeling errors on estimates of baseline length. *Radio science*, *20*(6), 1593–1607.

Dee, D. P., Uppala, S. M., Simons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., . . . Vitart, F. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597.

Den Bieman, J. P., de Ridder, M. P., Irías Mata, M., & van Nieuwkoop, J. C. (2023). Hybrid modelling to improve operational wave forecasts by combining process-based and machine learning models. *Applied Ocean Research*, *136*, 103583. https://doi.org/https://doi.org/10.1016/j.apor.2023.103583

Deng, X., & Featherstone, W. (2006). A coastal retracking system for satellite radar altimeter waveforms: Application to ERS-2 around Australia. *Journal of Geophysical Research: Oceans*, *111*(C6).

Denker, H. (2013). Regional gravity field modeling: theory and practical results. In *Sciences of geodesy-ii: Innovations and future developments* (pp. 185–291). Springer.

Desai, S., Wahr, J., & Beckley, B. (2015). Revisiting the pole tide for and from satellite altimetry. *Journal of Geodesy*, *89*(12), 1233–1243.

Deutscher Wetterdienst. (2002). *Storm 'Anna' over northern Germany in February 2002*. Retrieved April 2, 2024, from https://rcccm.dwd.de/DWD-RCCCM/EN/products/significant/national/Germany/evdl_Anna_200202.html

Dong, X., Xu, K., Liu, H. G., & Jiang, J. (2004). The radar altimeter and scatterometer of China's HY-2 satellite, 1703–1706 vol.3. https://doi.org/10.1109/IGARSS.2004.1370659

Dullaart, J. C., Muis, S., Bloemendaal, N., & Aerts, J. C. (2020). Advancing global storm surge modelling using the new ERA5 climate reanalysis. *Climate Dynamics*, *54*, 1007–1021.

Dullaart, J. C., Muis, S., Bloemendaal, N., Chertova, M. V., Couasnon, A., & Aerts, J. C. (2021). Accounting for tropical cyclones more than doubles the global population exposed to low-probability coastal flooding. *Communications Earth & Environment*, *2*(1), 135.

Eleveld, M. A., Van der Wal, D., & Van Kessel, T. (2014). Estuarine suspended particulate matter concentrations from sun-synchronous satellite remote sensing: Tidal and meteorological effects and biases. *Remote Sensing of Environment*, *143*, 204–215.

ESA. (2023a). *eoPortal*. Retrieved November 8, 2023, from https://www.eoportal.org/

ESA. (2023b). *Home*. Retrieved November 1, 2023, from https://eoatsee.eu/

Fenoglio-Marc, L., Fehlau, M., Ferri, L., Becker, M., Gao, Y., & Vignudelli, S. (2010). Coastal Sea Surface Heights from Improved Altimeter Data in the Mediterranean Sea. https://doi.org/10.1007/978-3-642-10634-7_33

Fernandes, M. J., Lázaro, C., Ablain, M., & Pires, N. (2015). Improved wet path delays for all ESA and reference altimetric missions. *Remote Sensing of Environment*, *169*, 50–74. https://doi.org/https://doi.org/10.1016/j.rse.2015.07.023

Fernandes, M. J., Lázaro, C., Nunes, A. L., & Scharroo, R. (2014). Atmospheric Corrections for Altimetry Studies over Inland Water. *Remote Sensing*, *6*(6), 4952–4997. https://doi.org/10.3390/rs6064952

Fernandes, M. J., Lázaro, C., & Vieira, T. (2021). On the role of the troposphere in satellite altimetry. *Remote Sensing of Environment*, *252*, 112149.

Fu, L. L., Christensen, E. J., Yamarone Jr, C. A., Lefebvre, M., Menard, Y., Dorrer, M., & Escudier, P. (1994). TOPEX/POSEIDON mission overview. *Journal of Geophysical Research: Oceans*, *99*(C12), 24369–24381.

Fukushima, K. (1969). Visual feature extraction by a multilayered network of analog threshold elements. *IEEE Transactions on Systems Science and Cybernetics*, *5*(4), 322–333.

Gaspar, P., Ogor, F., Le Traon, P. Y., & Zanife, O.-Z. (1994). Estimating the sea state bias of the TOPEX and POSEIDON altimeters from crossover differences. *Journal of Geophysical Research: Oceans*, *99*(C12), 24981–24994.

Gerritsen, H., de Vries, H., & Philippart, M. (1995). The Dutch continental shelf model. *Quantitative skill assessment for coastal ocean models*, *47*, 425–467.

Gertsenshtein, M. (1962). Wave resonance of light and gravitional waves. *Sov Phys JETP*, *14*, 84–85.

GESLA. (2024). *Downloads*. Retrieved April 9, 2024, from https://gesla787883612.wordpress.com/downloads/

Gharineiat, Z., & Deng, X. (2015). Application of the multi-adaptive regression splines to integrate sea level data from altimetry and tide gauges for monitoring extreme sea level events. *Marine Geodesy*, *38*(3), 261–276.

Ghavidel, A., Schiavulli, D., & Camps, A. (2015). Numerical computation of the electromagnetic bias in GNSS-R altimetry. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(1), 489–498.

Ghosh, S., & Resnick, S. (2010). A discussion on mean excess plots. *Stochastic Processes and their Applications*, *120*(8), 1492–1517.

Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep sparse rectifier neural networks. *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, 315–323.

Gommenginger, C., Thibaut, P., Fenoglio-Marc, L., Quartly, G., Deng, X., Gómez-Enri, J., Challenor, P., & Gao, Y. (2011). Retracking altimeter waveforms near the coasts: A review of retracking methods and some applications to coastal waveforms. *Coastal altimetry*, 61–101.

Guo, J., Di, W., Yuan, J., Niu, Y., & Li, Z. (2019). Waveform retracking of satellite radar altimeter based on singular spectrum analysis. *IOP Conference Series: Earth and Environmental Science*, *237*(3), 032024.

Guo, L., Zhu, C., Cai, H., Wang, Z. B., Townend, I., & He, Q. (2023). River flow induced nonlinear modulation of M4 overtide in large estuaries. *Estuaries and Coasts*, *46*(4), 925–940.

Haakenstad, H., Breivik, Ø., Furevik, B. R., Reistad, M., Bohlinger, P., & Aarnes, O. J. (2021). NORA3: A nonhydrostatic high-resolution hindcast of the North Sea, the Norwegian Sea, and the Barents Sea. *Journal of Applied Meteorology and Climatology*, *60*(10), 1443–1464.

Haigh, I. D., Marcos, M., Talke, S. A., Woodworth, P. L., Hunter, J. R., Hague, B. S., Arns, A., Bradshaw, E., & Thompson, P. (2022). GESLA Version 3: A major update to the global higher-frequency sea-level dataset. *Geoscience Data Journal*.

Hallegatte, S., Ranger, N., Mestre, O., Dumas, P., Corfee-Morlot, J., Herweijer, C., & Wood, R. M. (2011). Assessing climate change impacts, sea level rise and storm surge risk in port cities: a case study on Copenhagen. *Climatic change*, *104*, 113–137.

Hanson, H., & Larson, M. (2008). Implications of extreme waves and water levels in the southern Baltic Sea. *Journal of Hydraulic Research*, *46*(S2), 292–302.

Harris, D. L. (1963). *Characteristics of the hurricane storm surge*. Department of Commerce, Weather Bureau.

Hayne, G. (1980). Radar altimeter mean return waveforms from near-normal-incidence ocean surface scattering. *IEEE Transactions on Antennas and Propagation*, *28*(5), 687–692.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., et al. (2018). ERA5 hourly data on single levels from 1979 to present. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*, *10*(10.24381).

Hieronymus, M., & Hieronymus, F. (2023). A Novel Machine Learning Based Bias Correction Method and Its Application to Sea Level in an Ensemble of Downscaled Climate Projections. *Tellus A: Dynamic Meteorology and Oceanography*. https://doi.org/10.16993/tellusa.3216

Hu, S., Liu, B., Hu, M., Yu, X., Deng, Z., Zeng, H., Zhang, M., & Li, D. (2023). Quantification of the nonlinear interaction among the tide, surge and river in Pearl River Estuary. *Estuarine, Coastal and Shelf Science*, *290*, 108415.

Hu, Z., Zhang, J., & Ge, Y. (2021). Handling vanishing gradient problem using artificial derivative. *IEEE Access*, *9*, 22371–22377.

Hwang, C., Guo, J., Deng, X., Hsu, H. Y., & Liu, Y. (2006). Coastal gravity anomalies from retracked Geosat/GM altimetry: improvement, limitation and the role of airborne gravity data. *Journal of Geodesy*, *80*, 204–216.

Idier, D., Bertin, X., Thompson, P., & Pickering, M. D. (2019). Interactions between mean sea level, tide, surge, waves and flooding: Mechanisms and contributions to sea level variations at the coast. *Surveys in Geophysics*, *40*, 1603–1630.

Idris, N. H., Maharaj, A., Abdullah, N. N., Deng, X., & Andersen, O. (2014). A comparison of Saral/Altika coastal altimetry and in-situ observation across Australasia and maritime continent.

Iglesias, I., Lorenzo, M. N., Lázaro, C., Fernandes, M. J., & Bastos, L. (2017). Sea level anomaly in the North Atlantic and seas around Europe: Long-term variability and response to North Atlantic teleconnection patterns. *Science of the total environment*, *609*, 861–874.

InfoNu. (2016). *De zomerstorm van 25 juli 2015*. Retrieved March 12, 2024, from https://wetenschap.infonu.nl/weer/160303-de-zomerstorm-van-25-juli-2015.html

Ishii, M., Kimoto, M., Sakamoto, K., & Iwasaki, S.-I. (2006). Steric sea level changes estimated from historical ocean subsurface temperature and salinity analyses. *Journal of oceanography*, *62*, 155–170.

Izaguirre, C., Méndez, F. J., Menéndez, M., & Losada, I. J. (2011). Global extreme wave height variability based on satellite data. *Geophysical Research Letters*, *38*(10).

Javid, A. M., Das, S., Skoglund, M., & Chatterjee, S. (2021). A ReLU dense layer to improve the performance of neural networks. *2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2810–2814.

Ji, T., & Li, G. (2020). Contemporary monitoring of storm surge activity. *Progress in Physical Geography: Earth and Environment*, *44*(3), 299–314.

Ji, T., Li, G., & Zhang, Y. (2019). Observing storm surges in China's coastal areas by integrating multi-source satellite altimeters. *Estuarine, Coastal and Shelf Science*, *225*, 106224.

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

Jiang, H. (2020). Evaluation of altimeter undersampling in estimating global wind and wave climate using virtual observation. *Remote Sensing of Environment*, *245*, 111840.

Jones, J. E., & Davies, A. M. (2008). On the modification of tides in shallow water regions by wind effects. *Journal of Geophysical Research: Oceans*, *113*(C5).

Jongman, B., Ward, P. J., & Aerts, J. C. (2012). Global exposure to river and coastal flooding: Long term trends and changes. *Global Environmental Change*, *22*(4), 823–835.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

KNMI. (2002). *Storm van 27 oktober 2002 was zwaarste in twaalf jaar*. Retrieved March 14, 2024, from https://www.knmi.nl/over-het-knmi/nieuws/storm-van-27-oktober-2002-was-zwaarste-in-twaalf-jaar

KNMI. (2007). *De zware storm Kyrill van 18 januari 2007*. Retrieved March 14, 2024, from https://www.knmi.nl/over-het-knmi/nieuws/de-zware-storm-kyrill-van-18-januari-2007

KNMI. (2013a). *De zware storm van 5 december*. Retrieved March 14, 2024, from https://www.knmi.nl/over-het-knmi/nieuws/de-zware-storm-van-5-december

KNMI. (2013b). *Weeralarm voor zware herfststorm*. Retrieved March 14, 2024, from https://www.knmi.nl/over-het-knmi/nieuws/weeralarm-voor-zware-herfststorm

KNMI. (2016). *November 2016*. Retrieved March 12, 2024, from https://www.knmi.nl/nederland-nu/klimatologie/maand-en-seizoensoverzichten/2016/november

KNMI. (2017a). *December 2016*. Retrieved March 12, 2024, from https://www.knmi.nl/nederland-nu/klimatologie/maand-en-seizoensoverzichten/2016/december

KNMI. (2017b). *Sneeuw op 12 en 13 januari 2017*. Retrieved March 12, 2024, from https://www.knmi.nl/kennis-en-datacentrum/achtergrond/sneeuw-op-12-en-13-januari-2017

KNMI. (2017c). *Zeer zware windstoten 13 september 2017*. Retrieved March 12, 2024, from https://www.knmi.nl/kennis-en-datacentrum/achtergrond/zeer-zware-windstoten-13-september-2017

KNMI. (2018a). *Code oranje voor zeer zware windstoten op 3 januari 2018*. Retrieved March 12, 2024, from https://www.knmi.nl/kennis-en-datacentrum/achtergrond/Code-oranje-voor-zeer-zware-windstoten-op-3-januari-2018

KNMI. (2018b). *Zeer zware storm van 18 januari*. Retrieved March 12, 2024, from https://www.knmi.nl/over-het-knmi/nieuws/zeer-zware-storm-van-18-januari

KNMI. (2024). *Zware stormen in Nederland sinds 1910*. Retrieved March 12, 2024, from https://www.knmi.nl/nederland-nu/klimatologie/lijsten/zwarestormen

Komjathy, A., & Born, G. H. (1999). GPS-based ionospheric corrections for single frequency radar altimetry. *Journal of Atmospheric and Solar-Terrestrial Physics*, *61*(16), 1197–1203.

Komjathy, A., Sparks, L., Wilson, B. D., & Mannucci, A. J. (2005). Automated daily processing of more than 1000 ground-based GPS receivers for studying intense ionospheric storms. *Radio Science*, *40*(06), 1–11.

Kreyszig, E. (1979). *Advanced engineering mathematics* (4th ed.). John Wiley & Sons US.

Kyprioti, A. P., Taflanidis, A. A., Nadal-Caraballo, N. C., & Campbell, M. (2021). Storm hazard analysis over extended geospatial grids utilizing surrogate models. *Coastal Engineering*, *168*, 103855.

Leadbetter, M. R. (1991). On a basis for Peaks over Threshold modeling. *Statistics & Probability Letters*, *12*(4), 357–362.

Lee, H., Shum, C., Yi, Y., Braun, A., & Kuo, C. Y. (2008). Laurentia crustal motion observed using TOPEX/-POSEIDON radar altimetry over land. *Journal of Geodynamics*, *46*(3-5), 182–193.

Lobeto, H., Menendez, M., & Losada, I. (2018). Toward a methodology for estimating coastal extreme sea levels from satellite altimetry. *Journal of Geophysical Research: Oceans*, *123*(11), 8284–8298.

Lombard, A., Garric, G., & Penduff, T. (2009). Regional patterns of observed sea level change: insights from a 1/4 global ocean/sea-ice hindcast. *Ocean dynamics*, *59*, 433–449.

Lyard, F. H., Allain, D. J., Cancet, M., Carrère, L., & Picot, N. (2021). FES2014 global ocean tide atlas: design and performance. *Ocean Science*, *17*(3), 615–649.

Marcos, M., Calafat, F. M., Berihuete, Á., & Dangendorf, S. (2015). Long-term variations in global sea level extremes. *Journal of Geophysical Research: Oceans*, *120*(12), 8115–8134.

Martin, T. V., Zwally, H. J., Brenner, A. C., & Bindschadler, R. A. (1983). Analysis and retracking of continental ice sheet radar altimeter waveforms. *Journal of Geophysical Research: Oceans*, *88*(C3), 1608–1616.

Mertz, F., Mercier, F., Labroue, S., Tran, N., & Dorandeu, J. (2005). ERS-2 OPR data quality assessment Long-term monitoring–particular investigation. *Annual report*.

Mirzadeh, I., Alizadeh, K., Mehta, S., Del Mundo, C. C., Tuzel, O., Samei, G., Rastegari, M., & Farajtabar, M. (2023). Relu strikes back: Exploiting activation sparsity in large language models. *arXiv preprint arXiv:2310.04564*.

Moftakhari, H., Muñoz, D., Akbari Asanjan, A., AghaKouchak, A., Moradkhani, H., & Jay, D. A. (2024). Nonlinear Interactions of Sea-Level Rise and Storm Tide Alter Extreme Coastal Water Levels: How and Why? *AGU Advances*, *5*(2), e2023AV000996.

Mohammed, A., & Kora, R. (2023). A comprehensive review on ensemble deep learning: Opportunities and challenges. *Journal of King Saud University-Computer and Information Sciences*, *35*(2), 757–774.

Muis, S., Apecechea, M. I., Dullaart, J., de Lima Rego, J., Madsen, K. S., Su, J., Yan, K., & Verlaan, M. (2020). A high-resolution global dataset of extreme sea levels, tides, and storm surges, including future projections. *Frontiers in Marine Science*, *7*, 263.

Mullan, B., Salinger, J., Thompson, C., Ramsay, D., & Wild, M. (2005). Chatham Islands Climate Change.

Needham, H. F., Keim, B. D., & Sathiaraj, D. (2015). A review of tropical cyclone-generated storm surges: Global data sources, observations, and impacts. *Reviews of Geophysics*, *53*(2), 545–591.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The computer journal*, *7*(4), 308–313.

Nicholls, R. J. (2011). Planning for the impacts of sea level rise. *Oceanography*, *24*(2), 144–157.

Nicholls, R. J., & Cazenave, A. (2010). Sea-level rise and its impact on coastal zones. *science*, *328*(5985), 1517–1520.

Nicholls, R. J., & Tol, R. S. (2006). Impacts and responses to sea-level rise: a global analysis of the SRES scenarios over the twenty-first century. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *364*(1841), 1073–1095.

Oh, J., Kim, S., Lee, C., Cha, J. H., Yang, S. Y., Im, S. G., Park, C., Jang, B. C., & Choi, S. Y. (2023). Preventing Vanishing Gradient Problem of Hardware Neuromorphic System by Implementing Imidazole-Based Memristive ReLU Activation Neuron. *Advanced Materials*, *35*(24), 2300023.

Oppenheimer, M., Glavovic, B., Hinkel, J., van de Wal, R., Magnan, A. K., Abd-Elgawad, A., Cai, R., Cifuentes-Jara, M., Deconto, R. M., Ghosh, T., et al. (2019). Sea level rise and implications for low lying islands, coasts and communities. *The Intergovernmental Panel on Climate Change (IPCC)*.

Parise, C. K., Calliari, L. J., & Krusche, N. (2009). Extreme storm surges in the south of Brazil: atmospheric conditions and shore erosion. *Brazilian Journal of Oceanography*, *57*, 175–188.

Park, Y., Kim, E., Choi, Y., Seo, G., Kim, Y., & Kim, H. (2022). Storm surge forecasting along Korea strait using artificial neural network. *Journal of Marine Science and Engineering*, *10*(4), 535.

Passaro, M., Cipollini, P., Vignudelli, S., Quartly, G. D., & Snaith, H. M. (2014). ALES: A multi-mission adaptive subwaveform retracker for coastal and open ocean altimetry. *Remote Sensing of Environment*, *145*, 173–189.

Passaro, M., & Juhl, M. C. (2023). On the potential of mapping sea level anomalies from satellite altimetry with Random Forest Regression. *Ocean Dynamics*, *73*(2), 107–116.

Passaro, M., Nadzir, Z. A., & Quartly, G. D. (2018). Improving the precision of sea level data from satellite altimetry with high-frequency and regional sea state bias corrections. *Remote Sensing of Environment*, *218*, 245–254.

Prandle, D., & Wolf, J. (1978). The interaction of surge and tide in the North Sea and River Thames. *Geophysical Journal International*, *55*(1), 203–216.

Qin, Y., Su, C., Chu, D., Zhang, J., & Song, J. (2023). A review of application of machine learning in storm surge problems. *Journal of Marine Science and Engineering*, *11*(9), 1729.

Resti, A., Benveniste, J., Roca, M., Levrini, G., & Johannessen, J. (1999). The Envisat radar altimeter system (RA-2). *ESA bulletin*, *98*(8), 94–101.

Ries, J. (2017). Conventional model update for rotational deformation. *UT Faculty/Researcher Works*.

Rijkswaterstaat. (2024). *Home - Waterinfo Extra*. Retrieved April 9, 2024, from https://waterinfo-extra.rws.nl/

Roblou, L., Lamouroux, J., Bouffard, J., Lyard, F., Le Hénaff, M., Lombard, A., Marsaleix, P., De Mey, P., & Birol, F. (2011). Post-processing altimeter data towards coastal applications and integration into coastal models. *Coastal altimetry*, 217–246.

Sahoo, B., & Bhaskaran, P. K. (2019). Prediction of storm surge and inundation using climatological datasets for the Indian coast using soft computing techniques. *Soft Computing*, *23*(23), 12363–12383.

Scharroo, R., & Smith, W. (2010). A GPS-based climatology for the total electron content in the ionosphere. *Journal of Geophysical Research: Space Physics*, *115*(A10).

Scharroo, R., & Lillibridge, J. (2005). Non-parametric sea-state bias models and their relevance to sea level change studies. *Envisat & ERS Symposium*, *572*.

Schwatke, C., Dettmering, D., Bosch, W., & Seitz, F. (2015). DAHITI – an innovative approach for estimating water level time series over inland waters using multi-mission satellite altimetry. *Hydrology and Earth System Sciences*, *19*(10), 4345–4364.

Shah, E. (2023). *Regularization in Machine Learning*. Retrieved April 8, 2024, from https://www.scaler.com/topics/regularization-in-machine-learning/

Shu, S., Liu, H., Beck, R. A., Frappart, F., Korhonen, J., Lan, M., Xu, M., Yang, B., & Huang, Y. (2021). Evaluation of historic and operational satellite radar altimetry missions for constructing consistent long-term lake water level records. *Hydrology and Earth System Sciences*, *25*, 1643–1670. https://doi.org/10.5194/hess-25-1643-2021

Stammer, D., Ray, R., Andersen, O. B., Arbic, B., Bosch, W., Carrère, L., Cheng, Y., Chinn, D., Dushaw, B., Egbert, G., et al. (2014). Accuracy assessment of global barotropic ocean tide models. *Reviews of Geophysics*, *52*(3), 243–282.

Tan, H. H., & Lim, K. H. (2019). Vanishing gradient mitigation with deep learning neural network optimization. *2019 7th international conference on smart computing & communications (ICSCC)*, 1–4.

Tijssen, A., & Diermanse, F. (2010). *Storm surge duration and storm duration at Hoek van Holland* (tech. rep.). Deltares.

Timmermans, B., Shaw, A. G., & Gommenginger, C. (2020). Reliability of extreme significant wave height estimation from satellite altimetry and in situ measurements in the coastal zone. *Journal of Marine Science and Engineering*, *8*(12), 1039.

Tran, N., Labroue, S., Philipps, S., Bronner, E., & Picot, N. (2010). Overview and update of the sea state bias corrections for the Jason-2, Jason-1 and TOPEX missions. *Marine Geodesy*, *33*(S1), 348–362.

Tran, N., Philipps, S., Poisson, J., Urien, S., Bronner, E., & Picot, N. (2012). Impact of GDR_D standards on SSB corrections. *Ocean Science Topography Science Team Meeting, Venice*, *2229*.

u0192. (2002). *Het weer van zaterdag 9 maart 2002*. Retrieved April 2, 2024, from https://weerverleden.nl/20020309

Veenstra, J., & Kerkhoven, D. (2020, December 17). *Gebruikershandleiding hatyan getijanalyse in Python*. Version 1.0. Deltares.

Vincent, G., Kimble, R., Bhattacharya, S., & Blanton, B. (2022). Hybrid Machine Learning Models for Storm Surge Prediction.

Wahr, J. M. (1985). Deformation induced by polar motion. *Journal of Geophysical Research: Solid Earth*, *90*(B11), 9363–9368.

Walker, D., Crowley, R., & Dobrzanski, J. (1993). The GEOSAT Follow-on (GFO) Altimeter.

Wang, Y., Gan, D., Sun, M., Zhang, N., Lu, Z., & Kang, C. (2019). Probabilistic individual load forecasting using pinball loss guided LSTM. *Applied Energy*, *235*, 10–20.

Wingham, D., Rapley, C., & Griffiths, H. (1986). New techniques in satellite altimeter tracking systems. *Proceedings of IGARSS*, *86*, 1339–1344.

Wunsch, C., & Stammer, D. (1997). Atmospheric loading and the oceanic inverted barometer effect. *Reviews of Geophysics*, *35*(1), 79–107.

Xie, W., Xu, G., Zhang, H., & Dong, C. (2023). Developing a deep learning-based storm surge forecasting model. *Ocean Modelling*, *182*, 102179.

Zijl, F., Groeneboom, J., Laan, S., & Zijlker, T. (2022). *DCSM-FM 100m: a sixth-generation model for the NW European Shelf* (tech. rep.). Deltares.

Zijl, F., Zijlker, T., Laan, S., & Groeneboom, J. (2022a). *3D DCSM FM: a sixth-generation model for the NW European Shelf* (tech. rep.). Deltares.

Zijl, F., Zijlker, T., Laan, S., & Groeneboom, J. (2022b). *DCSM-FM 0.5nm: a sixth-generation model for the NW European Shelf* (tech. rep.). Deltares.

Zijl, F., Zijlker, T., Laan, S., & Groenenboom, J. (2022). *DCSM-FM 0.5nm: a sixth-generation model for the NW European Shelf* (tech. rep.). Deltares.

Zwally, H. J., & Brenner, A. C. (2001). Ice sheet dynamics and mass balance. In *International geophysics* (pp. 351–xxvi). Elsevier.

A

Satellite altimetry corrections

Atmospheric range corrections

The atmosphere causes the satellite altimetry radar signal to refract due to the presence of water vapour, dry gases and free electrons in the atmosphere. The atmospheric range corrections Δh_{dry}, Δh_{wet} and Δh_{iono} correspond to the dry tropospheric correction for dry gases, wet tropospheric correction for water vapour and ionospheric correction for free electrons in the atmosphere respectively.

Dry tropospheric correction

The dry tropospheric correction is considered to be the largest range correction in the order of two metres (Fernandes et al., 2014), accounting for about 90% of the total path delay caused by the troposphere (Andersen & Scharroo, 2010), and can be computed with an accuracy of a few millimetres according to the Saastamoinen model (J. Davis et al., 1985). The model depends on the sea surface pressure, the geodetic latitude and the surface height z_s with respect to the geoid. The correction Δh_{dry} can be described by

$$\Delta h_{dry} = -\frac{0.0022768 P_s}{1 - 0.00266\cos 2\varphi - 0.28 \cdot 10^{-6} z_s} \tag{A.1}$$

where P_s is the total atmospheric pressure in hPa and φ is the geodetic latitude. For ocean points, $P_s = P_0$ and $z_s = 0$, meaning that only the sea surface pressure (P_0) and the geodetic latitude are needed to compute the correction (Fernandes et al., 2021). The most frequently used method to compute Δh_{dry} is to apply (A.1) with the sea level pressure provided by sea level pressure grids. Often these grids are provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) or the United States National Centers for Environmental Prediction (NCEP) (Birol et al., 2017). For coastal applications of satellite altimetry, the accuracy of this method is not expected to decrease, since the spatial and temporal variability of the surface pressure do not significantly change on the land-sea transition border, therefore not making land contamination an issue (Andersen & Scharroo, 2010).

Wet tropospheric correction

Besides the dry tropospheric correction (also often referred to as hydrostatic troposphere correction or zenith hydrostatic delay), the water vapour present in the troposphere also introduces a significant delay in the altimetric signal. This delay does not have the same magnitude as the dry tropospheric correction (less than fifty centimetres), but due to the large spatial and temporal variability in water vapour concentration, the estimation of this wet tropospheric correction is harder to compute (Fernandes et al., 2015). One way to compute the water vapour correction in centimetres is given by

$$\Delta h_{\text{wet}} = -0.64 \int_{z_s}^{z_{\text{sat}}} \rho_w(z) dz \tag{A.2}$$

where $\int_{z_s}^{z_{\text{sat}}} \rho_w(z) dz$ is the total mass of water vapour in the atmospheric column from the satellite to the water surface with a cross-section of 1 m^2, given in millimetres (Fernandes et al., 2021). This mass of water can accurately be computed with measurements from onboard microwave radiometers. The microwave bands in which the radiometers operate are used to observe the brightness temperature of the ocean and to monitor the water vapour absorption line, which is centred at 22.235 GHz (Andersen & Scharroo, 2010). The correction typically ranges from a few millimetres in cold, dry air to more than thirty centimetres in hot, humid air. In coastal areas, the microwave signal can be contaminated by land, because the algorithms to retrieve the water vapour density assume a constant surface ocean emissivity, and the emissivity of land is higher than that of ocean. Moreover, the footprint of a radiometer typically ranges between 20 and 30 kilometres (Figure A.1). This, coupled with the typically warmer nature of land, causes a degradation in the method used to measure humidity. As a result, this affects the range correction for water vapour.

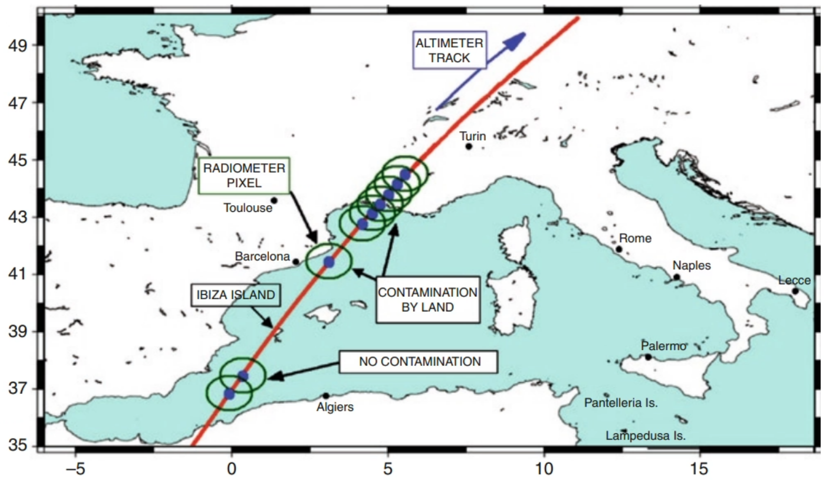


Figure A.1: An example of a Jason-1 track crossing the western Mediterranean Sea. Blue dots indicate the footprint of the altimeter and green circles show the extent of the main beam of the radiometer. The figure illustrates where the radiometer observations are contaminated by land (Andersen & Scharroo, 2010).

Ionospheric correction

The number of free electrons in the atmosphere increases when one travels further from the Earth's surface. The speed of the altimetric signal slows down due to the relation between electromagnetic waves and the ions in the ionosphere (Andersen & Scharroo, 2010), which can be quantified as a range correction with an expression dependent on the electron density in the ionosphere.

$$\Delta h_{\text{iono}} = -k\mathrm{TEC}/f^2 \tag{A.3}$$

In (A.3), TEC equals the Total Electron Content, which is defined as the number of electrons per unit area in a column extending from the Earth's surface to the satellite (Andersen & Scharroo, 2010), and one unit (referred to as TECU) equals 10^{16} electrons/m^2. The factor k is a constant of 0.40250 m GHz^2/TECU and f is the radar frequency of the altimeter. The TEC can be approximated using a dual-frequency altimeter. The travel time of both signals will differ according to (A.3), and this difference can be used as a measure for the TEC. The variability of the ionospheric path delay depends primarily on the season of the year, the time of day and the solar activity. As an alternative to using dual-frequency altimeters, it is possible to use observational models that produce global GPS-derived ionosphere maps (GIM), which can be interpolated to the location of the satellite tracks. These maps are derived from observations of more than one hundred GPS stations within the IGS network (Komjathy and Born, 1999; Komjathy et al., 2005; Scharroo and Smith, 2010). Noticeably, the ionosphere is not affected by land, and the dual-frequency altimeters or the global models can be applied to coastal areas as well as open ocean (Andersen & Scharroo, 2010).

Geophysical corrections

To correct for the physical behaviour of the ocean's surface, a sea state bias (SSB) correction has been applied to the observed range. This correction is a sum of different components, representing different physical characteristics of the ocean's surface that influence the signal reflection.

$$\Delta h_{\text{ssb}} = \mathrm{SWH}\left(a_1 + a_2 U + a_3 U^2 + a_4 \mathrm{SWH}\right) \tag{A.4}$$

In (A.4), U refers to the wind speed derived from the backscatter coefficient (see Figure 2.5), which can be computed from the waveform (Figure 2.4b) by taking the amplitude of the received signal. The parameter SWH refers to the significant wave height, which is derived from the slope of the leading edge of the waveform. The other four parameters a_1, a_2, a_3 and a_4 differ per altimeter, representing an electromagnetic bias (EM), a skewness bias and an instrument tracker bias. The EM bias is a result of the mean sea surface being underestimated due to the smaller reflectivity of the wave troughs compared to the wave crests (Ghavidel et al., 2015). The skewness bias corrects for the non-linear dynamics of ocean waves (Badulin et al., 2021), since the skewness of the sea surface distribution causes the altimeters to make use of a sea surface tracker that estimates the median, not the mean sea surface (Andersen & Scharroo, 2010). Finally, the tracker bias is a sum of smaller errors linked to the method of how the altimeter tracks the returning signals. Often this is estimated empirically (Passaro et al., 2018). The parameters for eight different altimeters are shown in Table A.1.

Table A.1: The four parameters for the BM4 model for eight different altimeters. The a_0 parameter is added to ensure that $\Delta h_{\mathrm{ssb}} = 0$ when SWH = 0 (Scharroo & Lillibridge, 2005)

Altimeter	a_0	a_1	a_2	a_3	a_4
ERS-1 (OPRv6)	0.054265	-0.075043	0.001413	-0.001790	0.000098
ERS-2	0.107618	-0.068219	0.001465	-0.001701	0.000082
GFO	0.092034	-0.055742	0.002743	-0.003756	0.000153
Poseidon	0.015731	-0.062778	0.001894	-0.001194	0.000057
TOPEX (Side A)	0.012450	-0.030578	0.002776	-0.002962	0.000127
TOPEX (Side B)	0.028889	-0.032113	0.002992	-0.002780	0.000101
Jason-1	0.110106	-0.034376	0.001145	-0.001969	0.000083
Envisat	0.026530	-0.052849	0.001746	-0.001713	0.000068

B

Adaptive Leading-Edge Subwaveform retracker

Retracking algorithm

The ALES retracking algorithm is subdivided into two phases that fit waveforms to the original Brown model (Figure B.1). First, the leading edge of the incoming waveform is detected, which is referred to as the subwaveform. The gate at which the subwaveform starts is referred to as the startgate, and is defined for each mission, depending on the onboard processing of the waveforms. The subwaveform is then fitted to the Brown model with an unweighted least-square estimator, after which the stopgate is updated to include a wider part of the original waveform into the final subwaveform. This is a trade-off between including as much information from the original waveform as possible and the noise that the trailing edge often induces in the waveform due to artefacts such as ships or other bright target responses. Fitting the second subwaveform to the Brown model then gives the final estimates for the tracking gate, SWH and backscatter coefficient.

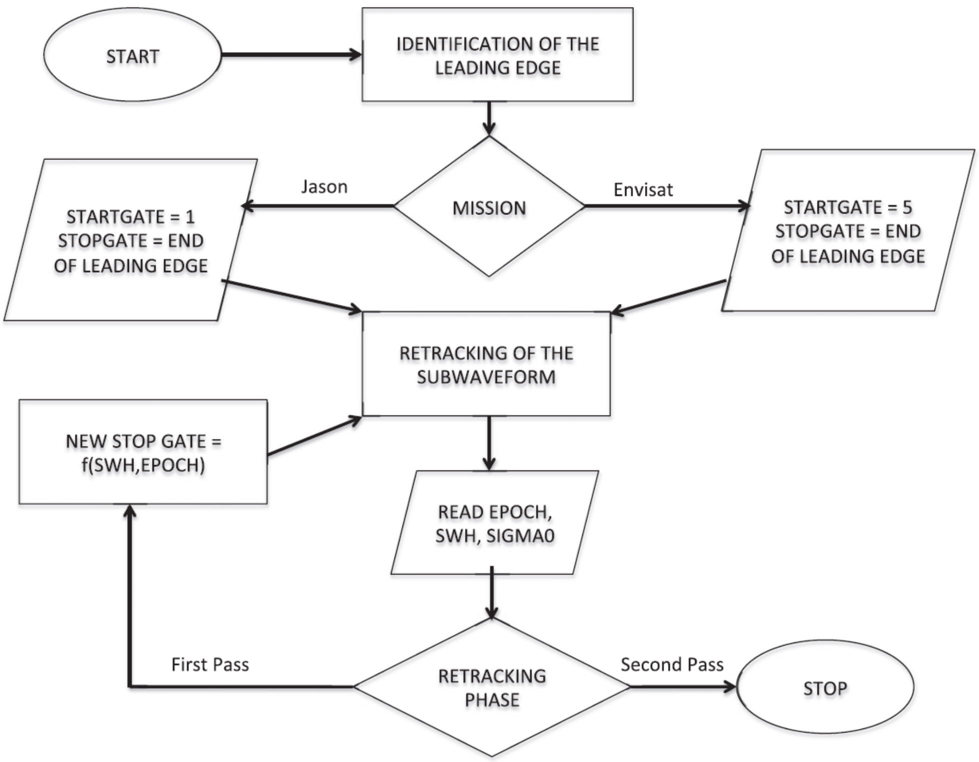


Figure B.1: Flow diagram of ALES retracking algorithm for each waveform (Passaro et al., 2014).

Phase one
The first step in detecting the leading edge is to remove the thermal noise (see Figure 2.5). This is done by computing the average of the first few gates and removing this from the signal. For Envisat, this is done with gates 5 to 10, as the first few gates are subject to aliasing. For Jason-1 and Jason-2, gates 1 to 5 are used. After removing the thermal noise, the difference between consecutive gates is computed, starting at the startgate, which differs per mission (5 for Envisat and 1 for Jason). The start of the leading edge is detected where the difference between two consecutive gates is positive and larger than 1% of the normalisation factor (see also (B.1)). The vector that holds the differences between consecutive gates is referred to as Dwf, and is expressed in normalised power units.

$$Dwf > 0.01 \qquad\qquad Dwf < 0 \tag{B.1}$$

The end of the leading edge is found where the first following gate has a negative difference (see also (B.1)). Due to some remaining noise, one gate is added to the stopgate. The selected leading edge is now fitted to the Brown model with an unweighted least-square estimator, converged with the Nelder-Mead algorithm (Nelder & Mead, 1965). When convergence is not reached within $Nmax$ iterations, with $Nmax = 600$ for ALES, one more gate is added to the stopgate. Convergence in this case refers to an unweighted least-square estimator smaller than $1 * 10^{-10}$. When the final leading edge is fitted, referred to as the subwaveform, the tracking gate and SWH are found.

Phase two
Using the SWH and the tracking gate from the first phase, the width of the final subwaveform is estimated. The relationship between the stopgate and the SWH and the tracking gate is determined through Monte Carlo simulations. For Envisat and the Jason missions, the following relationship was found:

$$\text{Stopgate} = \text{Ceiling}\ (\text{Tracking point} + 2.4263 + 4.1759 * SWH)\ \text{for Envisat} \tag{B.2}$$
$$\text{Stopgate} = \text{Ceiling}\ (\text{Tracking point} + 1.3737 + 4.5098 * SWH)\ \text{for Jason-1 and Jason-2} \tag{B.3}$$

where *Ceiling* refers to the rounded-up integer. The final estimations for the tracking point, SWH and backscatter coefficient are derived from this final subwaveform (Figure B.2).

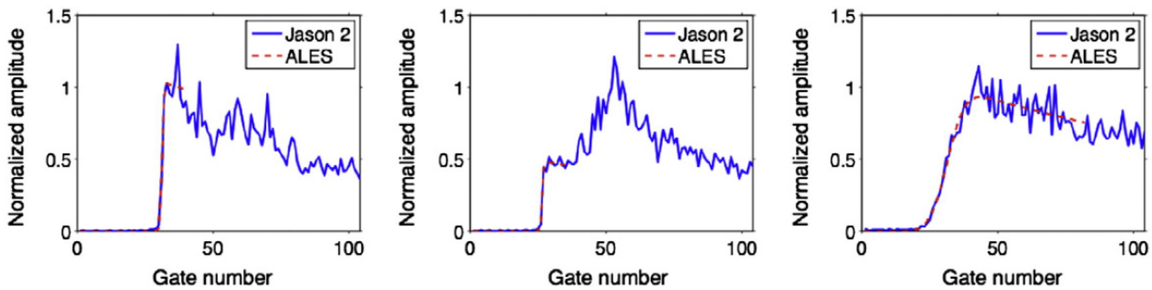


Figure B.2: Examples of real recorded waveforms with their ALES retracked waveforms for (left) open ocean with SWH = 0.75 m, (middle) coastal ocean with corrupted trailing edge and SWH = 1.65 m and (right) open ocean with SWH = 9.448 m (Passaro et al., 2014).

Reprocessing algorithm

In addition to the retracked SSH that is computed using the ALES retracker, the X-TRACK algorithm also improves coastal water level estimates by providing refined atmospheric and geophysical corrections and applying editing and filtering methods for the correction terms (Birol et al., 2017). These filtering methods prevent a lot of data from being lost due to land contamination by interpolating the flagged corrections.

X-TRACK uses the Geophysical Data Records (GDR) to assess the corrective terms, where abrupt changes in geophysical corrections are associated with erroneous data based on thresholds (Table B.1), and consequently flagged.

Table B.1: X-TRACK thresholds for detecting erroneous data in corrective terms (Roblou et al., 2011).

Parameter	Min. threshold	Max. threshold
Backscatter coefficient (Ku band)	1.0 dB	30.0 dB
Backscatter coefficient (C, S band)	7.0 dB	30.0 dB
Wet tropospheric path delay	-0.5 m	0.0 m
Sea state bias	N/A	0.0 m
Ionospheric path delay	N/A	0.0 m

The dry tropospheric correction is not considered here, since only few values are discarded by the editing method. After the outliers are flagged, a second filter is applied to the corrective terms, flagging them when they exceed a threshold of 3σ or 4σ, with σ being the standard deviation of the along-track record. All the flagged corrections are then interpolated from a Bezier curve created from the edited data. This method results in a better coastal estimate for the corrective terms, as can be seen in Figure B.3.

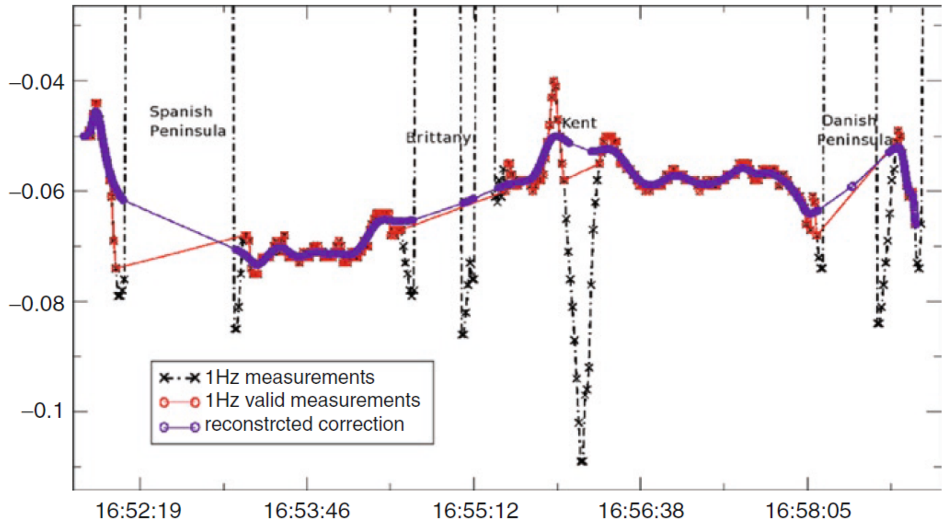


Figure B.3: Wet tropospheric correction for TOPEX/Poseidon pass 137, cycle 200. Black represents the raw correction from the GDR. Red represents the valid correction points after the second filter. Purple represents the data after interpolation. The x-axis represents the time on 1998-02-22. The y-axis corresponds to the wet tropospheric correction in metres (Roblou et al., 2011).

Besides the editing and filtering process, the wet tropospheric correction, sea state bias and ionospheric path correction are also updated in the X-TRACK software. For the wet tropospheric correction, a Global Navigation Satellite System (GNSS) derived Path Delay algorithm was used (Fernandes et al., 2015). This method improves the estimates for coastal observations of the wet troposphere correction by combining estimates from three sources: Zenith wet delay from nearby GNSS ground/offshore stations, water vapour from atmospheric models, and microwave radiometer measurements.

X-TRACK uses an improved estimation of the sea state bias in coastal areas with methods developed in Gaspar et al. (1994), Mertz et al. (2005), Tran et al. (2010) and Tran et al. (2012). They are combined in the ALES retracking method described in Passaro et al. (2018) and Aviso+ (2022).

The ionospheric correction is improved with a median absolute deviation (MAD) threshold filter, with $MAD = \frac{1}{l}\sum_i^l |x_i - R_{med}(X)_i|$, where l is the number of data points in the record and $R_{med}(X)$ is the respective running median value. The ionospheric correction (Figure B.4a) and wet tropospheric correction (Figure B.4b) are provided as additional data to the SSH within the X-TRACK product.

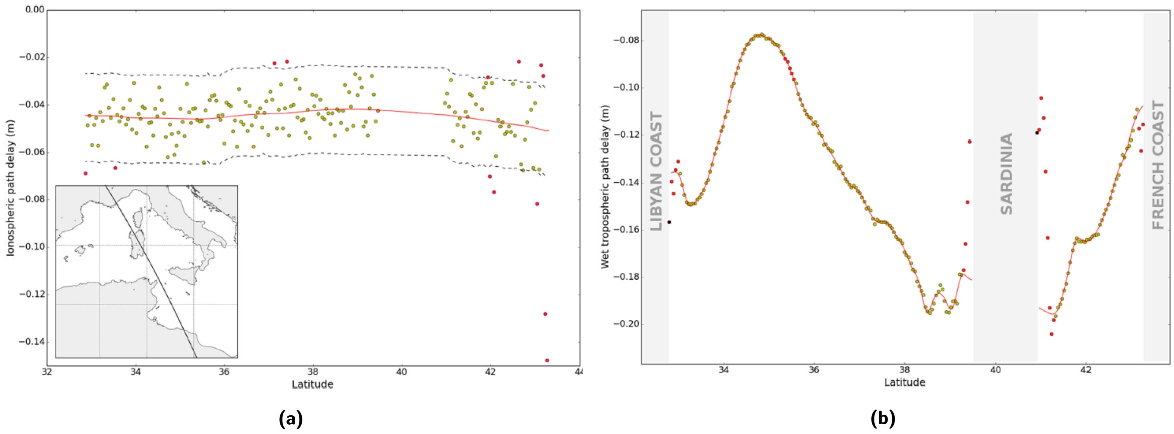


Figure B.4: Examples of (a) ionospheric correction for Jason-1, track 222, cycle 10 and (b) wet tropospheric correction for cycle 8 in the Mediterranean Sea. The yellow circles indicate the corrections before and the red line indicates the corrections after X-TRACK processing (Birol et al., 2017).

C

Additional information on satellite altimetry missions

ERS-2

Both ERS-1 and ERS-2 missions have the same active microwave Radar Altimetry (RA-1) instrument on board meant to observe the surface in one of two modes: ocean or ice. The instrument contains a Ku-band (13.8 GHz) nadir-pointing sensor that operates in bandwidths of 330 MHz for ocean surfaces and 82.5 MHz for ice surfaces. ERS-1 was launched in 1991 and ended in 2000. ERS-2 was launched in 1995 and ended in 2011. Both missions shared the same orbit, with an orbit height between 782 to 785 km and an inclination angle of 98.52 degrees, covering the Earth from a latitude of 82°N and S. The pulse repetition frequency is 1020 Hz and chirp pulse length is 20 μs.

Table C.1: Technical Details of ERS-1 and ERS-2 Radar Altimeters (RA-1) (ESA, 2023a)

Parameter	Value
Orbit Height	782 - 785 km
Inclination Angle	98.52°
Radar Frequency	13.8 GHz (Ku-band)
Pulse Length	20 μs
Bandwidth	330 MHz
Pulse Repetition Frequency	1020 Hz

Envisat

The European Earth Observation satellite Envisat, which was in orbit from 2002 to 2012, carried ten instruments for Earth Observation, of which this research uses data from the Radar Altimeter 2 (RA-2). The instrument had a dual-frequency, nadir-pointing, pulse-limited radar sensor. It operated at Ku-band (13.575 GHz) and S-band (3.2 GHz) and was the successor of the Radar Altimeter (RA-1) used on the ERS-1 and ERS-2 missions. Its orbit height ranged between 764 and 825 km, with a similar inclination angle as the ERS-1 and ERS-2 missions. The instrument uses three different bandwidths so the sensor can be adapted for different scenarios (ocean, coastal zones, ice, ice sheet, sea ice and land), namely 320 MHz, 80 MHz and 20 MHz. The chirp pulse length of 20 μs is also similar to RA-1 from ERS-1 and ERS-2.

Table C.2: Technical Details of Envisat Radar Altimeter 2 (RA-2) (Resti et al., 1999)

Parameter	**Value**
Orbit Height	764 - 825 km
Inclination Angle	98.55°
Radar Frequency	13.575 GHz (Ku-band) - 3.2 GHz (S-band)
Pulse Length	20 μs
Bandwidth	320 MHz - 80 MHz - 20 MHz
Pulse Repetition Frequency	1795.33 Hz (Ku-band) - 448.83 Hz (S-band)

SARAL

The SARAL satellite was specifically launched in 2013 for altimetric purposes meant to bridge the gap between the missions Envisat and Sentinel-3. Therefore, it has the same orbit specifications as Envisat. It is the first satellite that uses a frequency of 35.75 GHz (Ka-band) for its active radar altimetry sensor, and utilises additional passive bands to correct for the errors caused by the wet troposphere. Several advantages of Ka-band over Ku-band can be identified. Ka-band can travel through the ionosphere mainly unaffected, which nullifies the need for a dual-frequency altimeter. Additionally, the footprint size decreases from 15 km on Envisat to 8 km on SARAL, creating a finer horizontal resolution. The same holds for the vertical resolution, since Ka-band can use a larger bandwidth than Ku-band altimeters. Ka-band also permits a higher PRF (Pulse Repetition Frequency) than Ku-band, allowing for better along-track sampling. Finally, the wavelength of Ka-band is better suited to observe small signals from the ocean's surface (e.g. capillary waves), which gives an improved estimate of the sea surface roughness. Drawbacks of using Ka-band include the low radar penetration of snow and ice compared to Ku-band, and the theoretical attenuation due to water in the troposphere during increased rainfall conditions.

Table C.3: Technical Details of SARAL radar altimeter (AltiKa) (ESA, 2023a)

Parameter	**Value**
Orbit Height	800 km
Inclination Angle	98.538°
Radar Frequency	35.75 GHz (Ka-band)
Pulse Length	110 μs
Bandwidth	480 MHz
Pulse Repetition Frequency	3800 Hz

GFO

The GeoSat Follow-On mission was launched in 1998 and ended its life cycle in 2008. Its radar altimeter (GFO-RA) used a Ku-band sensor with a centre frequency of 13.5 GHz and a bandwidth of 320 MHz. It orbits the Earth in a non-sun-synchronous polar orbit with an inclination of 108.04°and an orbit height of 784 km. This mission is the first of its kind where a microwave radiometer and a radar sensor share the same antenna.

Table C.4: Technical Details of GeoSat Follow-On radar altimeter (GFO-RA) (Walker et al., 1993)

Parameter	**Value**
Orbit Height	784 km
Inclination Angle	108.04°
Radar Frequency	13.5 GHz (Ku-band)
Pulse Length	102.4 μs
Bandwidth	320 MHz
Pulse Repetition Frequency	1020.4 Hz

HY2

The Haiyang-2 satellite series consists of eight satellites, of which four are currently in orbit (HY-2A, HY-2B, HY-2C, HY-2D) and four are planned to be launched by 2025 (HY-2E, HY-2F, HY-2G, HY-2H). The satellites contain a Radar Altimeter (RA/HY-2) operating on a dual frequency signal of 13.58 GHz (Ku-band)

and 5.25 GHz (C-band) (Dong et al., 2004). The C-band has a bandwidth of 160 MHz. The Ku-band has bandwidths of 320 MHz over open ocean, 80 MHz on coastal oceans and 20 MHz on land or ice. The pulse length is 102.4 μs, the PRF ranges between 1 and 4 kHz, and the footprint size is 16 km. HY-2A has a near sun-synchronous frozen orbit with an inclination of 99.3°and an orbit altitude of 971 km (ESA, 2023a). HY-2B and HY-2D both are in sun-synchronous orbits at 963 km altitude and an inclination of 66°. HY-2C has a non-sun-synchronous orbit with the same inclination (66°) and an altitude of 957 km. The planned satellites (HY-2E, HY-2F, HY-2G and HY-2H) are planned to be in a sun-synchronous orbit with an altitude of 963 km and an inclination of 99.3°.

Table C.5: Technical Details of HY-2 Radar Altimeters (RA/HY-2) in orbit (Dong et al., 2004)

Parameter	**Value**
Orbit Height	971 km (HY-2A) - 963 km (HY-2B/HY-2D) - 957 km (HY-2C)
Inclination Angle	99.3°(HY-2A) - 66°(HY-2B/HY-2C/HY-2D)
Radar Frequency	13.58 GHz (Ku-band) - 5.25 GHz (C-band)
Pulse Length	102.4 μs
Bandwidth	320 MHz - 80 MHz - 20 MHz (Ku-band) - 160 MHz (C-band)
Pulse Repetition Frequency	1-4 kHz

Sentinel-3A

Copernicus Sentinel-3A contains a nadir-looking altimeter instrument (SRAL) which operates in Ku-band and C-band. The Ku-band has a centre frequency of 13.575 GHz with a bandwidth of 350 MHz. The C-band has 5.41 GHz as the centre frequency and a bandwidth of 320 MHz and is used for ionospheric corrections. The pulse length is 50 μs. It has two measurement modes: LRM and SARM, corresponding to Low Resolution Mode with a PRF of 1.92 kHz and SAR Mode with a PRF of 17.8 kHz respectively. The LRM is similar to a conventional altimeter pulse-limited mode, which has a pulse pattern of 3 Ku / 1 C / 3 Ku. The SARM has a high along-track resolution, with a pulse pattern of 64 Ku-band pulses surrounded by two C-band pulses.

Table C.6: Technical Details of Sentinel-3A Radar Altimeter (SRAL) (ESA, 2023a)

Parameter	**Value**
Orbit Height	814.5 km
Inclination Angle	98.65°
Radar Frequency	13.575 GHz (Ku-band) - 5.41 GHz (C-band)
Pulse Length	50 μs
Bandwidth	350 MHz (Ku-band) - 320 MHz (C-band)
Pulse Repetition Frequency	1.92 kHz (LRM) - 17.8 kHz (SARM)

TOPEX/Poseidon

The TOPEX/Poseidon mission is a successor of the GEOS-3, SeaSat and GeoSat missions launched by the US, launched in 1992 and ending in 2006. Its orbit has an altitude of 1336 km with an inclination of 66.039°and a repeat orbit of 10 days. It contains a dual-frequency NASA Radar Altimeter (NRA) using Ku-band and C-band to observe the sea surface height and estimate the ionospheric correction. The bandwidth for the Ku-band is 320 MHz, and for the C-band 100 MHz. The PRF for Ku-band is 2400 Hz, and for C-band 1220 Hz. The pulse duration is 102.4 μs for Ku-band and 102.4 or 32 μs for C-band. Next to NRA, the satellite also carries a single-frequency altimeter (SSALT), with a frequency of 13.65 GHz (Ku-band). It was installed as an experiment to demonstrate a low-cost, low-power, low-mass and low-data rate solution for future altimetry missions. Successfully, the altimeter proved to be able to estimate sea surface heights with an accuracy of 2.5 cm. The experimental altimeter was 4 times as light and consumed 5 times less power than the NRA.

Table C.7: Technical Details of TOPEX/Poseidon Radar Altimeter (NRA) (Fu et al., 1994)

Parameter	Value
Orbit Height	1336 km
Inclination Angle	66.039°
Radar Frequency	13.575 GHz (Ku-band) - 5.3 GHz (C-band) - 13.65 GHz (SSALT, Ku-band)
Pulse Length	102.4 μs (Ku-band) - 102.4 or 32 (C-band) - 105.6 μs (SSALT)
Bandwidth	320 MHz (Ku-band) - 100 MHz (C-band) - 300 MHz (SSALT)
Pulse Repetition Frequency	2400 Hz (Ku-band) - 1220 Hz (C-band) - 900 MHz (SSALT)

Jason-1, Jason-2 and Jason-3

The Jason series is a mission collection to continue the successful work of TOPEX/Poseidon. Jason-1, Jason-2 and Jason-3 were launched in 2001, 2008 and 2016 respectively. All of them share the same orbit, similar to the TOPEX/Poseidon orbit. All of them are or have been in an orbit with an altitude of 1336 km and an inclination angle of 66°. The development of the SSALT instrument on TOPEX/Poseidon has been implemented in all satellites, resulting in an improved version of the instrument in Jason-1 (Poseidon-2), Jason-2 (Poseidon-3) and Jason-3 (Poseidon-3B). All of them operate in dual-band frequency, with Ku-band and C-band, and have the same pulse length. The bandwidth for Poseidon-2 is 320 MHz for Ku-band and 320 or 100 MHz for C-band. For Poseidon-3 and Poseidon-3B, the bandwidth for both Ku-band and C-band is 320 MHz. The PRF for Poseidon-2 is 1800 Hz for Ku-band and 300 Hz for C-band. The Poseidon-3 and Poseidon-3B altimeters have an interlaced PRF of 2060 Hz, which equals 3Ku-1C-3Ku-band signals.

Table C.8: Technical Details of Jason Radar Altimeters (Poseidon-2/-3/-3B)

Parameter	Value
Orbit Height	1336 km
Inclination Angle	66°
Radar Frequency	13.575 GHz (Ku-band) - 5.3 GHz (C-band)
Pulse Length	105.6 μs
Bandwidth	320 MHz (Ku-band) - 320/100 MHz (C-band)
Pulse Repetition Frequency	1800 Hz (Ku-band) - 300 Hz (C-band) / 2060 Hz interlaced (3Ku-1C-3Ku)

D

Hatyan tidal analysis

This appendix presents the tidal analysis used to correct the TG data for any harmonic tidal signals. The Python package `Hatyan` is used to compute 95 tidal constituents (Veenstra & Kerkhoven, 2020), which are summed and subtracted from the TG time series to compute the non-tidal residuals used in the neural network as ground truth data. The main expression used for this tidal analysis is

$$h(t) = A_0 + \sum_{i=1}^{N} \gamma_i A_i \cos\left\{\omega_i t + (v_0 + u)_i - g_i\right\} \tag{D.1}$$

where

- $h(t)$ = water level at time t
- A_0 = the mean of the time series
- N = the number of constituents
- γ_i = a correction term for the amplitude of the 18.6-year cycle
- A_i = the amplitude of component i
- ω_i = the angular velocity of component i
- $v_0 + u$ = the phase of the equilibrium tide at $t = 0$, with correction term u for the 18.6-year cycle
- g_i = improved kappa-number (phase) of component i

The corrections for the 18.6-year cycle (γ_i and u) are needed here because the tidal analysis is based on time series of one year, which means it is otherwise impossible for the algorithm to obtain estimates for an 18.6-year cycle. The equilibrium tide in this context refers to the tidal variation that would occur if the Earth was completely covered in deep ocean. The phase difference between the equilibrium tide and the actual tide is given with g_i. For this study, a number of $N = 94$ components is used, which are listed in Table D.1.

A0	SA	SM	Q1	O1	M1C	P1	S1	K1
3MKS2	3MS2	OQ2	MNS2	2ML2S2	NLK2	MU2	N2	NU2
MSK2	MPS2	M2	MSP2	MKS2	LABDA2	2MN2	T2	S2
K2	MSN2	2SM2	SKM2	NO3	2MK3	2MP3	SO3	MK3
SK3	4MS4	2MNS4	3MS4	MN4	2MLS4	2MSK4	M4	3MN4
MS4	MK4	2MSN4	S4	MNO5	3MK5	2MP5	3MO5	MSK5
3KM5	3MNS6	2NM6	4MS6	2MN6	2MNU6	3MSK6	M6	MSN6
MKNU6	2MS6	2MK6	3MSN6	2SM6	MSK6	2MNO7	M7	2MSO7
2(MN)8	3MN8	M8	2MSN8	2MNK8	3MS8	3MK8	2(MS)8	2MSK8
3MNK9	4MK9	3MSK9	4MN10	M10	3MSN10	4MS10	2(MS)N10	3M2S10
4MSK11	M12	4MSN12	5MS12	4M2S12				

Table D.1: Constituents used in the tidal analysis. A_0 is not a tidal constituent, but the mean of the time series (Veenstra & Kerkhoven, 2020).

The `Hatyan` algorithm estimates the amplitude (A_i) and phase (g_i) for every tidal component along with the mean (A_0) mentioned in Table D.1 with a least-squares adjustment. The nodal factors (γ_i and u) are estimated for all time steps, and v_0 is computed for the start of the time series. The angular velocity (ω_i) is pre-defined for every component. The amplitudes and phases for the Scheveningen TG are shown in Figure D.1.

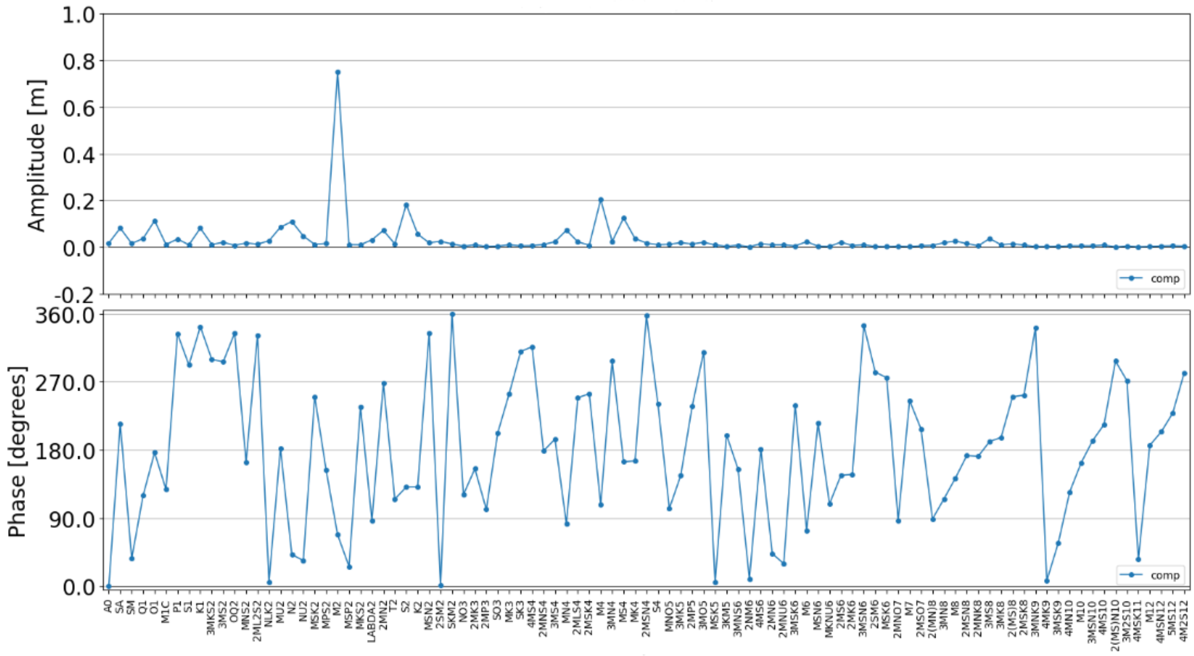


Figure D.1: Amplitude and phase of all considered tidal constituents for Scheveningen. Amplitude is given in metres and phase in degrees.

The largest component, which has a period of 12.4206 hours, is M2. The components M4 and S2 follow with much smaller amplitudes, and have a period of 6.2103 and 12 hours respectively. Other significant components are, MS4, N2, MU2, NU2, O1, SA, K1, 2MN2 and MN4, of which their respective period and angular velocity are shown in Table D.2.

Component	**Period [hours]**	ω_i **[rad/hr]**
M2	12.4206	0.5059
M4	6.2103	1.0117
S2	12.0000	0.5236
MS4	6.1033	1.0295
N2	12.6583	0.4964
MU2	12.8718	0.4881
NU2	12.6260	0.4976
O1	25.8193	0.2434
SA	8765.8128	0.0007
K1	23.9345	0.2625
2MN2	12.1916	0.5154
MN4	6.2692	1.0022

Table D.2: Components with significant amplitudes according to `Hatyan` with their respective period in hours and angular velocity in radians per hour (Veenstra & Kerkhoven, 2020). The components with the largest amplitudes are shown in grey.

Summing all tidal components according to (D.1) results in a non-tidal residual as shown in Figure 4.4, which is then used as ground truth data to train the neural network.

E

Input variables

This appendix shows the distribution of the datasets used in this study. Tide gauge non-tidal residuals are presented, which act as ground truth data, along with the input variables of the neural network. The input variables from satellite altimetry (NTR, dt, d_N and d_E) are shown, as well as the ERA5 input variables (p, $U10$ and $V10$) and the DOY. For interpretability, ERA5 p is divided by 100 to convert pascals to hectopascals and satellite altimetry dt is divided by 3600 to present the data in hours instead of seconds.

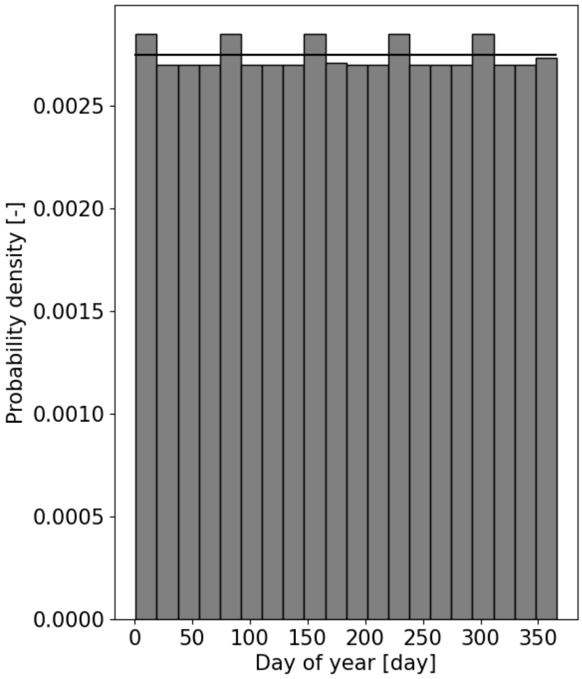
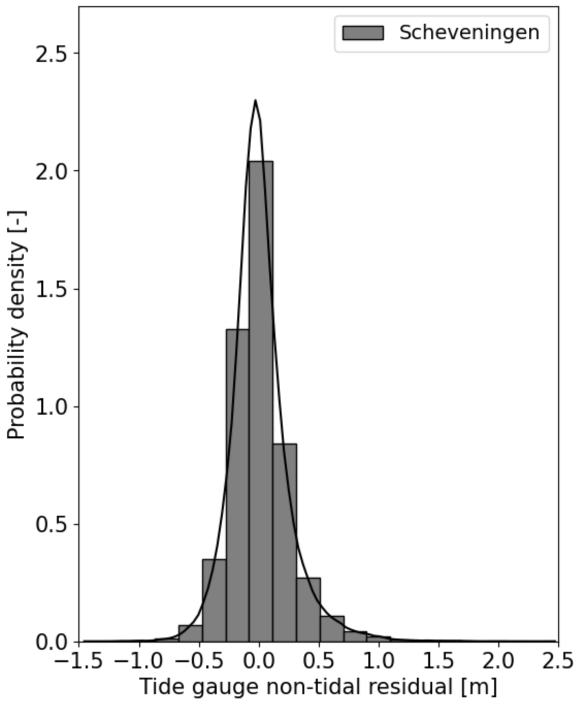
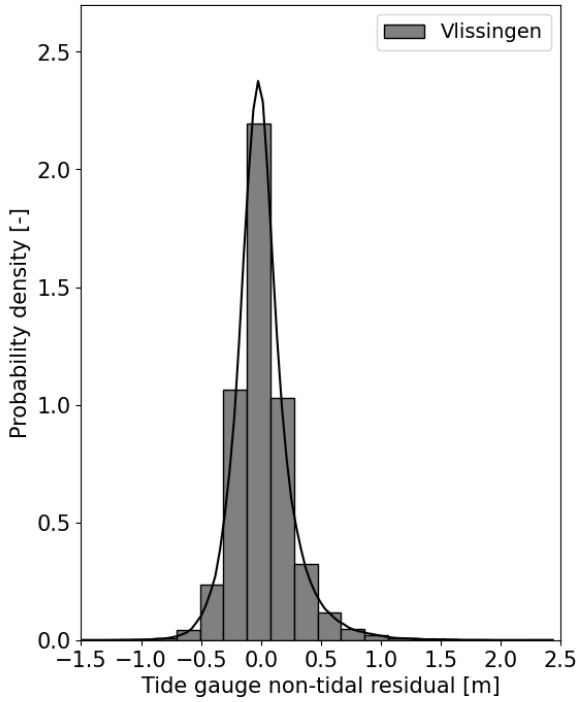


Figure E.1: Distribution of the day of the year

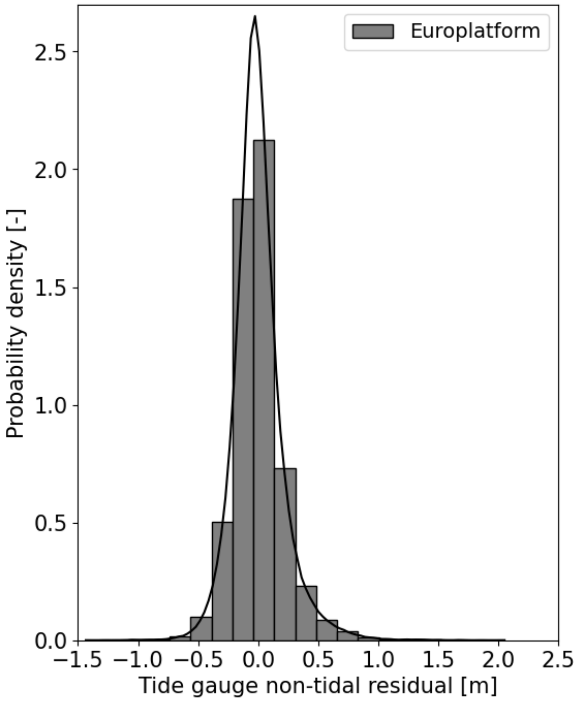
Tide gauge non-tidal residuals



(a)



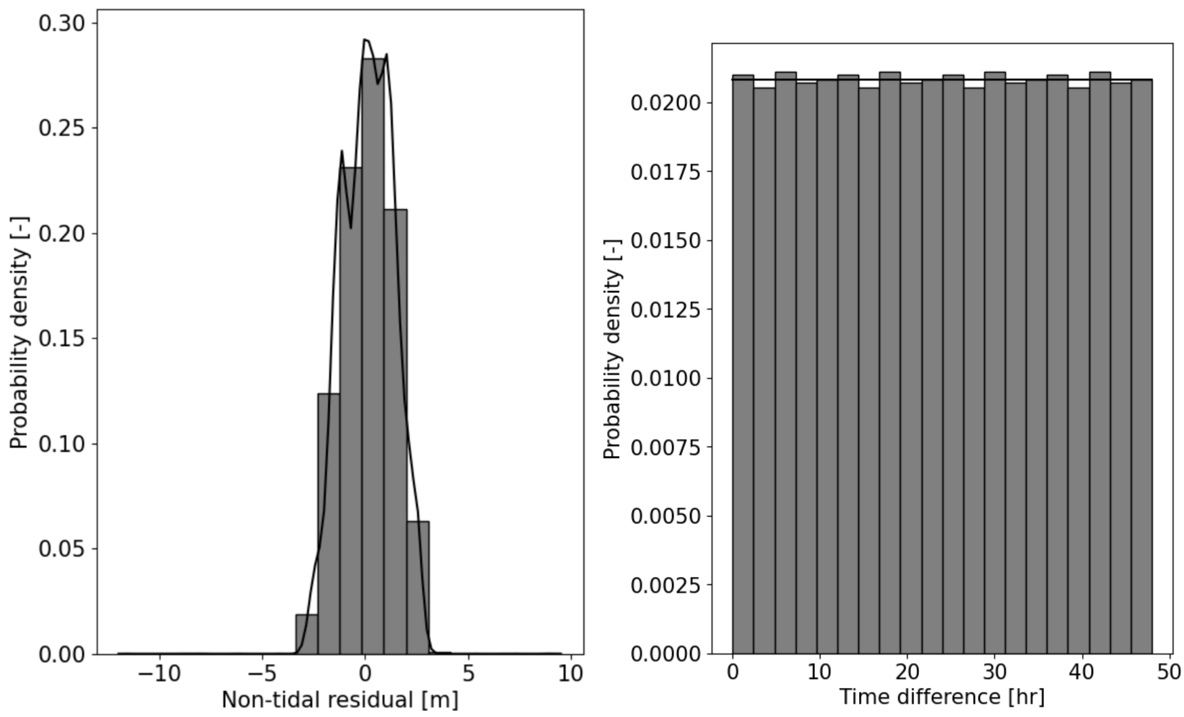
(b)



(c)

Figure E.2: Distribution of the non-tidal residuals for (a) Scheveningen, (b) Vlissingen and (c) Europlatform.

Satellite altimetry features



(a)

(b)

Figure E.3: Distribution of (a) non-tidal residuals for satellite altimetry observations (b) the time difference between each satellite altimetry observation and the time stamp of interest. Note that each satellite altimetry observation appears 48 times in 48 successive input datasets.

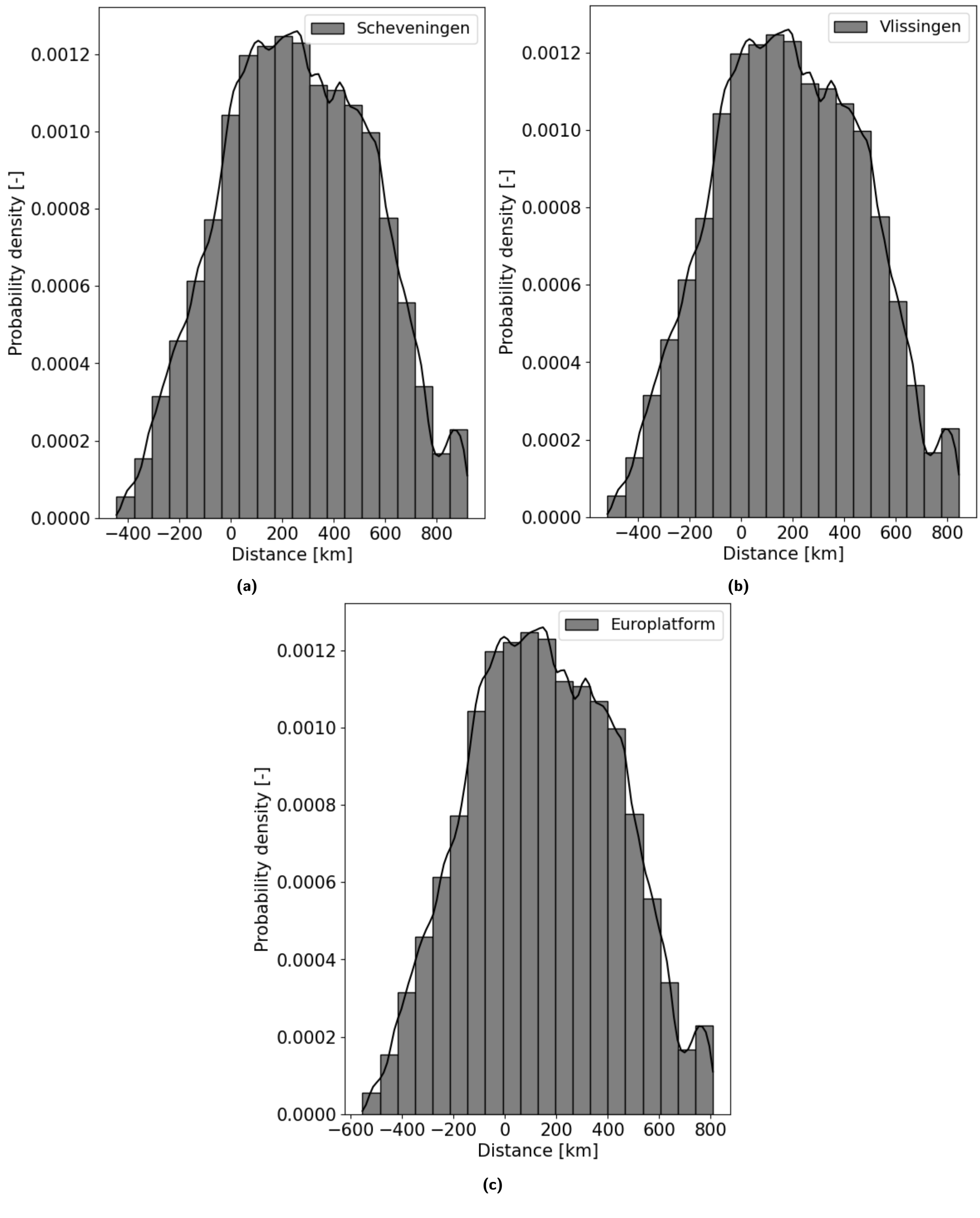
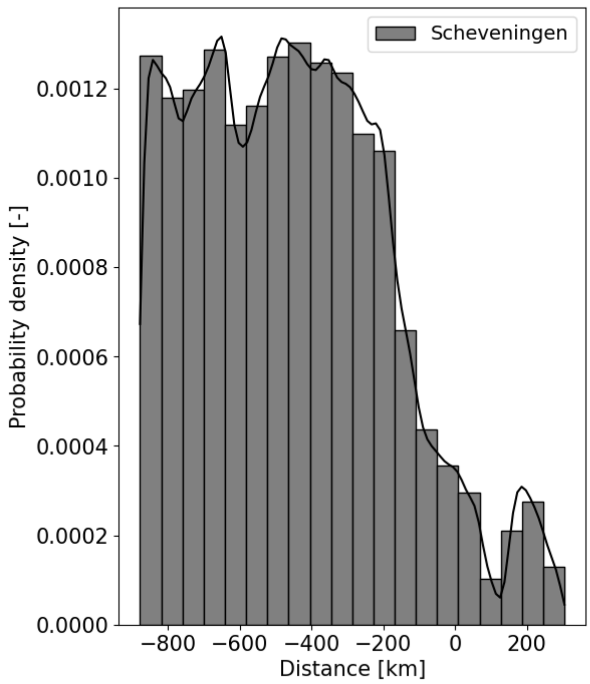
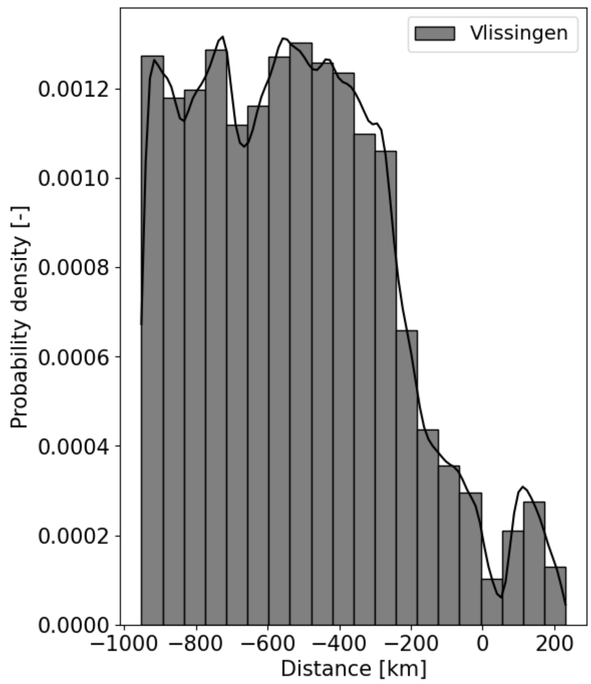


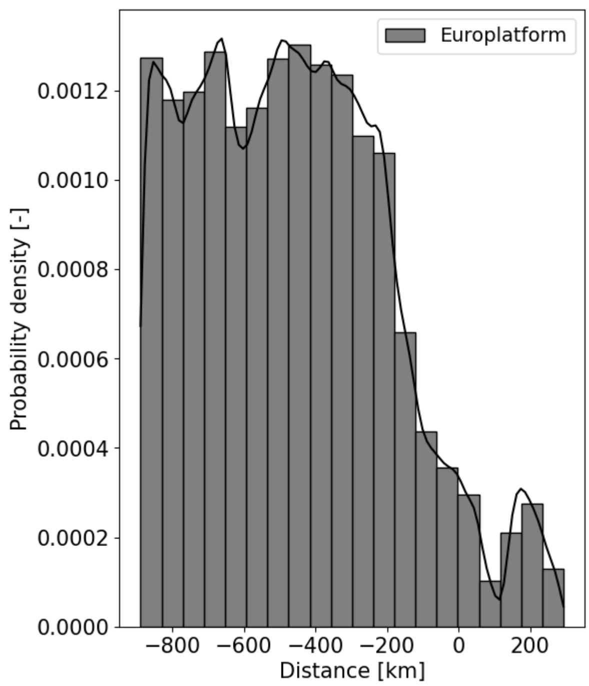
Figure E.4: Distribution of the distance between satellite altimetry observations and the tide gauge of interest in longitudinal direction for (a) Scheveningen, (b) Vlissingen and (c) Europlatform. While the distribution is the same, the three histograms are slightly shifted with respect to one another, depending on the longitude of the tide gauge.



(a)



(b)



(c)

Figure E.5: Distribution of distance between satellite altimetry observations and the tide gauge of interest in lateral direction for (a) Scheveningen, (b) Vlissingen and (c) Europlatform. While the distribution is the same, the three histograms are slightly shifted with respect to one another, depending on the latitude of the tide gauge.

ERA5 features

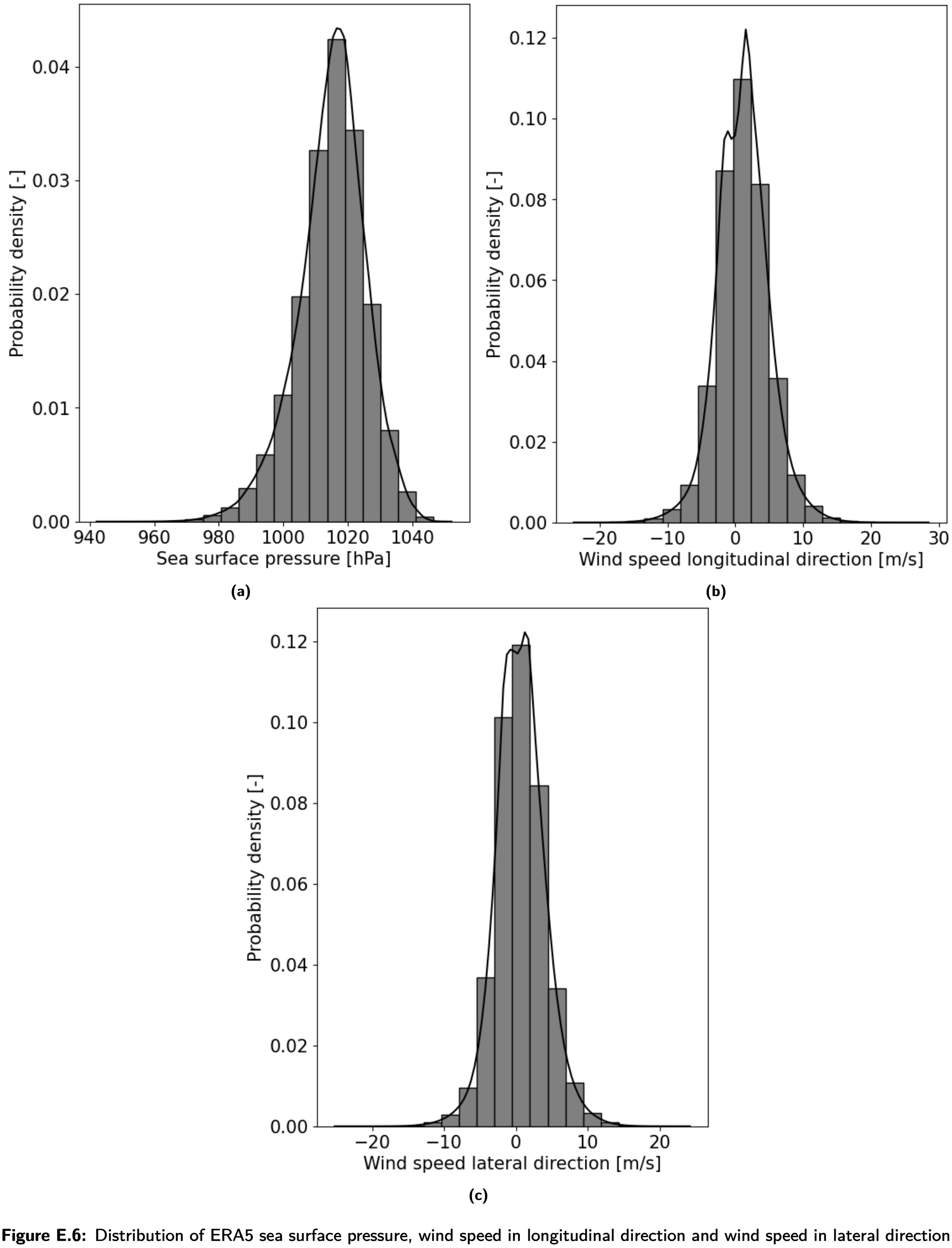


Figure E.6: Distribution of ERA5 sea surface pressure, wind speed in longitudinal direction and wind speed in lateral direction.

F

POT threshold selection

The peak-over-threshold method requires a manual selection of a threshold, with the requirement that a generalised Pareto distribution (GPD) has to be valid over the thresholded data (Leadbetter, 1991). The GPD can be fitted on the data according to Equation F.1, where T represents the threshold, ξ and η are the shape and scale parameters of the GPD, and x the data (Castillo & Hadi, 1997).

$$\mathrm{GPD}_{\xi,T,\eta}(x) = \begin{cases} 1 - \left[1 + \xi\left(\frac{x-T}{\eta}\right)\right]^{-\frac{1}{\xi}} & \text{if } \xi \neq 0 \\ 1 - \exp\left[-\frac{x-T}{\eta}\right] & \text{if } \xi = 0 \end{cases} \tag{F.1}$$

When $\xi \geq 0$, $\eta > 0$ and $x - T \geq 0$. If $\xi < 0$, then $0 \leq x - T \leq -\eta/\xi$. To find suitable thresholds T for every location, mean excess plots are used that can give an approximation for the threshold by visual inspection of the graphs (Ghosh & Resnick, 2010). The mean excess (ME) of a dataset is computed with Equation F.2, where $\hat{M}(T)$ refers to the empirical ME function, with X_i the data values, T the threshold, and $I_{[X_i>u]}$ an indicator function that equals 1 if $X_i > T$ and 0 otherwise. The ME plot is then given as a function of T, and shows the average excess of the data in metres for all data points above the threshold (Figure F.1). When the ME plot is linear within a certain range of thresholds, a GPD is considered to be valid for the thresholded data.

$$\hat{M}(T) = \frac{\sum_{i=1}^{n} (X_i - T)\, I_{[X_i>T]}}{\sum_{i=1}^{n} I_{[X_i>T]}}, \quad T \geq 0 \tag{F.2}$$

Figure F.1 shows the mean excess plots for Scheveningen, Vlissingen and Europlatform, based on a threshold range of 100 values between the 90th percentile and the 10th highest value of the water levels, including the harmonic tidal signals.

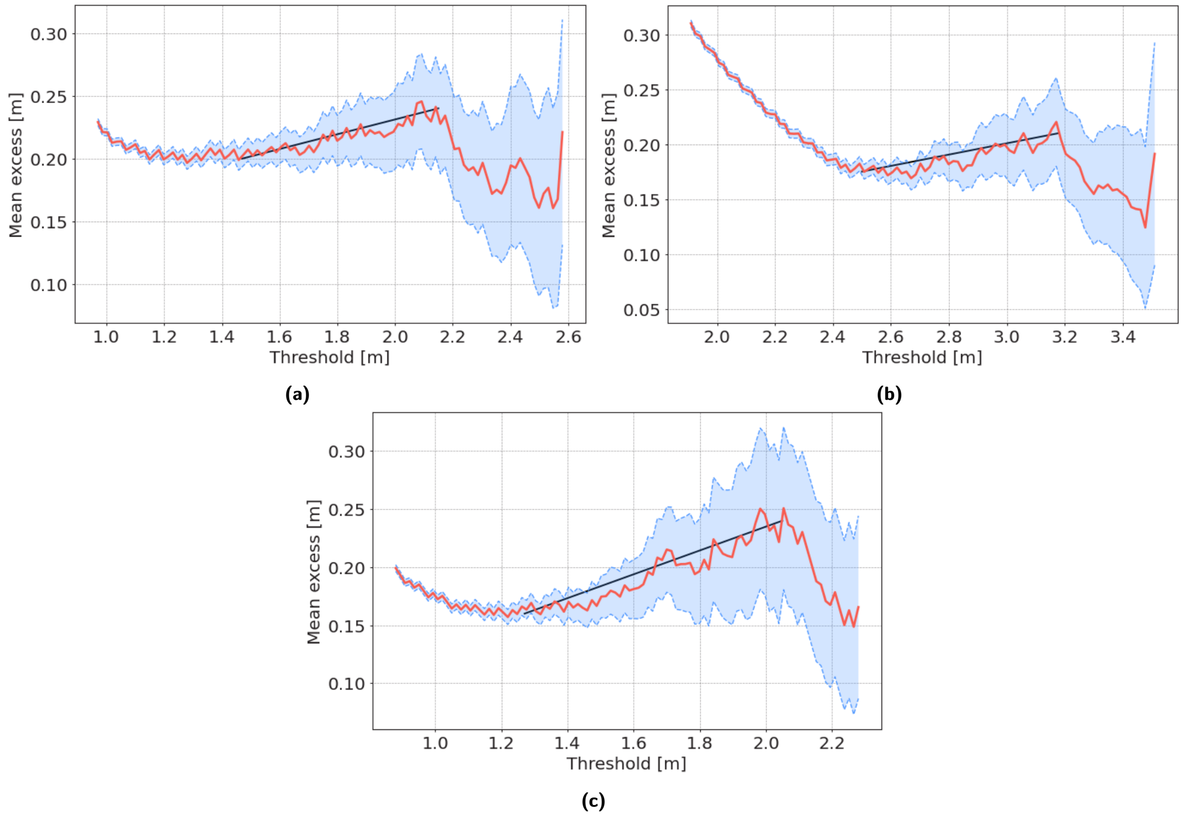


Figure F.1: Mean excess plots for (a) Scheveningen, (b) Vlissingen and (c) Europlatform. The black lines indicate where the mean excess plot is approximately linear. The shaded area refers to the 95% confidence interval.

From Figure F.1, the valid threshold range is $[1.47 \text{ to } 2.15]$ metres for Scheveningen, $[2.50 \text{ to } 3.18]$ metres for Vlissingen, and $[1.27 \text{ to } 2.05]$ metres for Europlatform. To validate the right threshold range, the shape (ξ) and scale (η) parameters of the GPD (Equation F.3 and F.4) are plotted against the selected threshold range. If these parameters are seemingly stable within the selected range, the selected range is assumed to be valid (Bocharov, 2023).

$$\hat{\xi} = \frac{1}{2}\left[1 - \frac{(\hat{\mu} - T)^2}{s^2}\right] \tag{F.3}$$

$$\hat{\eta} = \left[\frac{\hat{\mu} - T}{2}\right]\left[\frac{(\hat{\mu} - T)^2}{s^2} + 1\right] \tag{F.4}$$

In Equations F.3 and F.4, $\hat{\mu}$ refers to the empirical mean of the values above the threshold T, and s^2 to the empirical variance. Figure F.2 shows the parameter stability as a function of the threshold T, for the same range of thresholds as the ME plots in Figure F.1. The grey shaded areas indicate the threshold range defined by the ME plots, and are adjusted for every location based on the stability. For Scheveningen, the range is modified to $[1.47 \text{ to } 2.00]$ metres. Vlissingen is adjusted to $[2.50 \text{ to } 3.15]$ metres, and Europlatform to $[1.27 \text{ to } 1.90]$ metres.

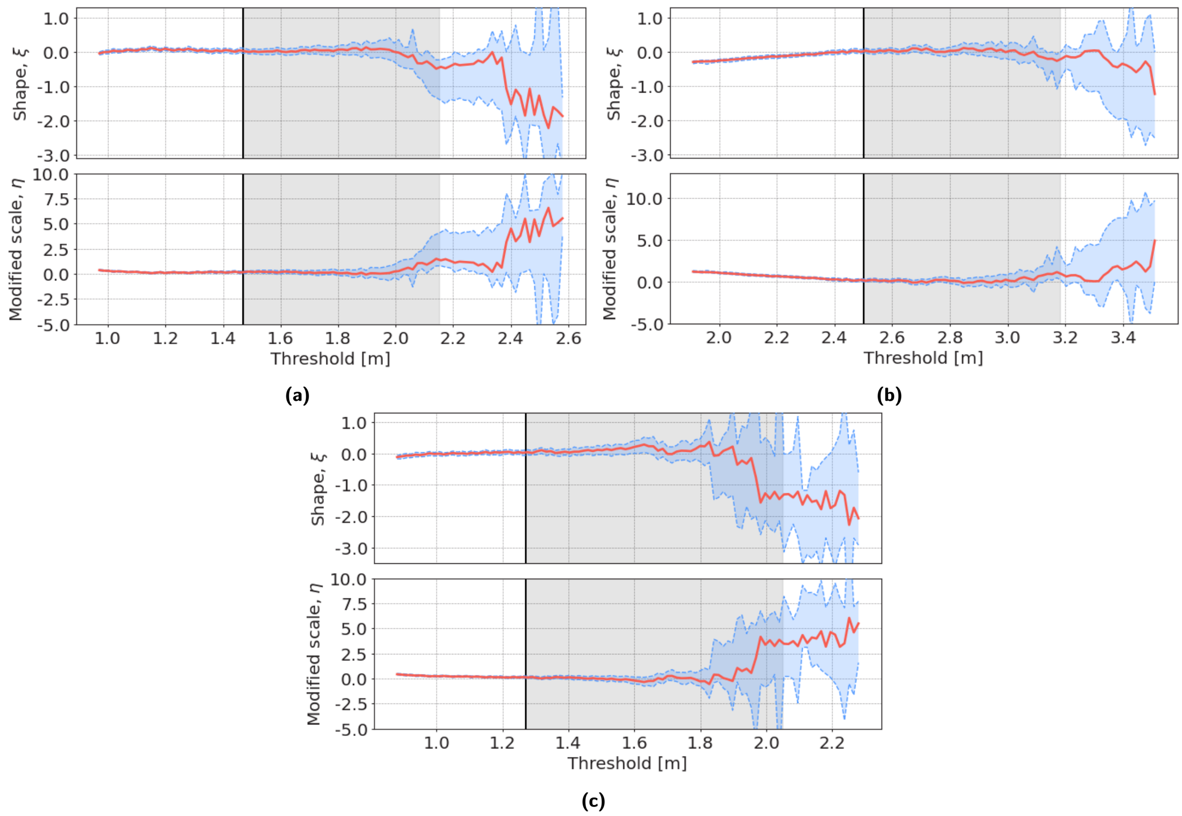
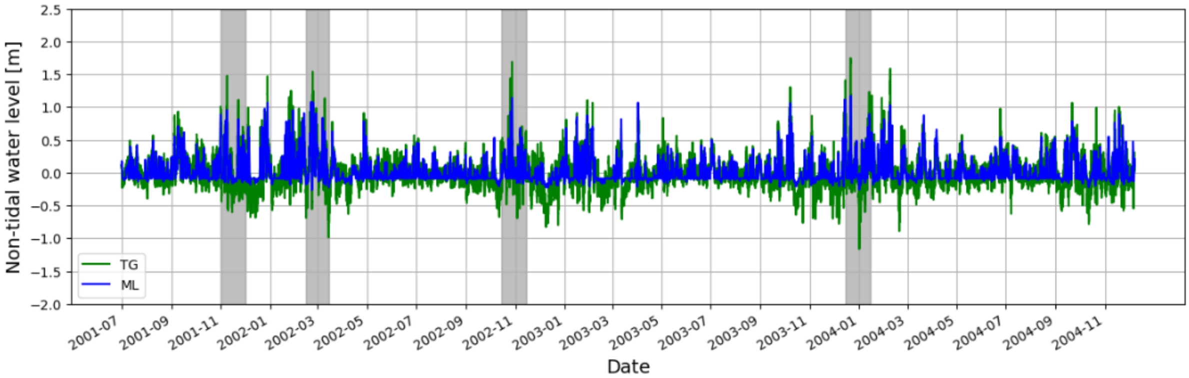


Figure F.2: Parameter stability plots of the shape ξ and scale η parameters of the GPD for (a) Scheveningen, (b) Vlissingen and (c) Europlatform. The blue shaded area refers to the 95% confidence interval and the grey shaded area to the threshold range defined with the ME plots. The black line refers to the selected thresholds for each location.

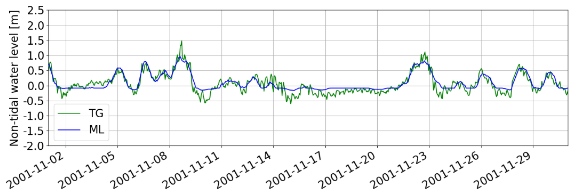
Based on Figure F.1 and F.2, appropriate thresholds are chosen for each location, presented as black lines in Figure F.2. These thresholds are used in the POT method to select high water levels within the water level observations from the tide gauges and their reconstruction from the machine learning models. They are chosen to be as low as possible for the GPD to still be valid, since the testing period of the ML model is only 3.5 years, which means a high threshold might result in not finding enough high water levels to test the model performance on.

G
K-fold reconstructions

The Figures presented in this appendix cover all results from the K-fold cross-validation not shown in Section 5.2. The full time series reconstructions are shown, with zoomed-in months of interest that include severe storms as catalogued by the KNMI (KNMI, 2024). These storms are recorded on the 2002-02-02 (Deutscher Wetterdienst, 2002), 2002-03-09 (u0192, 2002), 2002-10-27 (KNMI, 2002), 2007-01-17 (KNMI, 2007), 2013-10-28 (KNMI, 2013b) and 2013-12-05 (KNMI, 2013a).

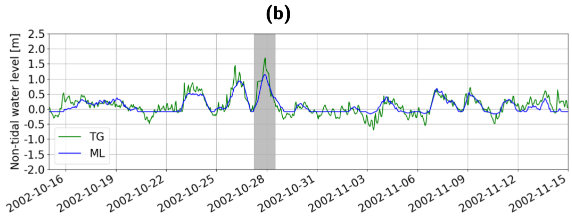


(a)

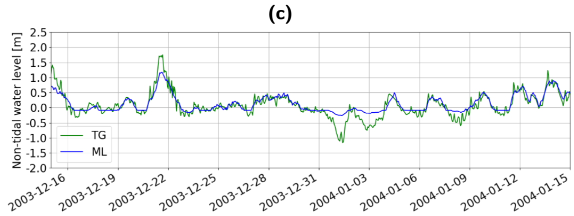


(b)

(c)



(d)



(e)

Figure G.1: Time-series reconstruction for validation set 1 of the Scheveningen model. The shaded areas in (a) correspond to the months shown in (b), (c), (d) and (e). The green lines represent the tide gauge observations, the blue lines refer the machine learning model output. The shaded areas in the sub-graphs show storms catalogued by the KNMI (KNMI, 2024). The unit of the y-axis is in metres and depicts the non-tidal water level referenced to NAP.

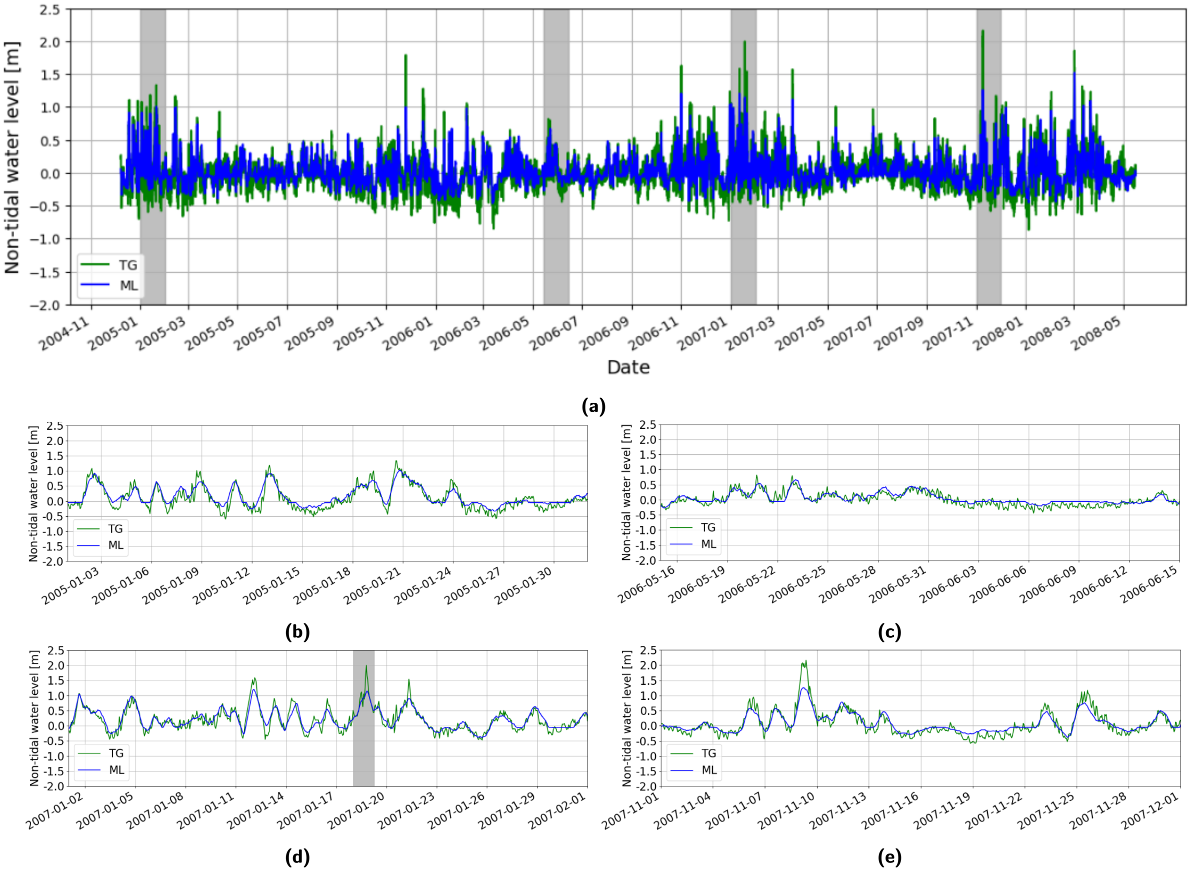


Figure G.2: Time-series reconstruction for validation set 2 of the Scheveningen model. The shaded areas in (a) correspond to the months shown in (b), (c), (d) and (e). The green lines represent the tide gauge observations, the blue lines refer the machine learning model output. The shaded areas in the sub-graphs show storms catalogued by the KNMI (KNMI, 2024). The unit of the y-axis is in metres and depicts the non-tidal water level referenced to NAP.

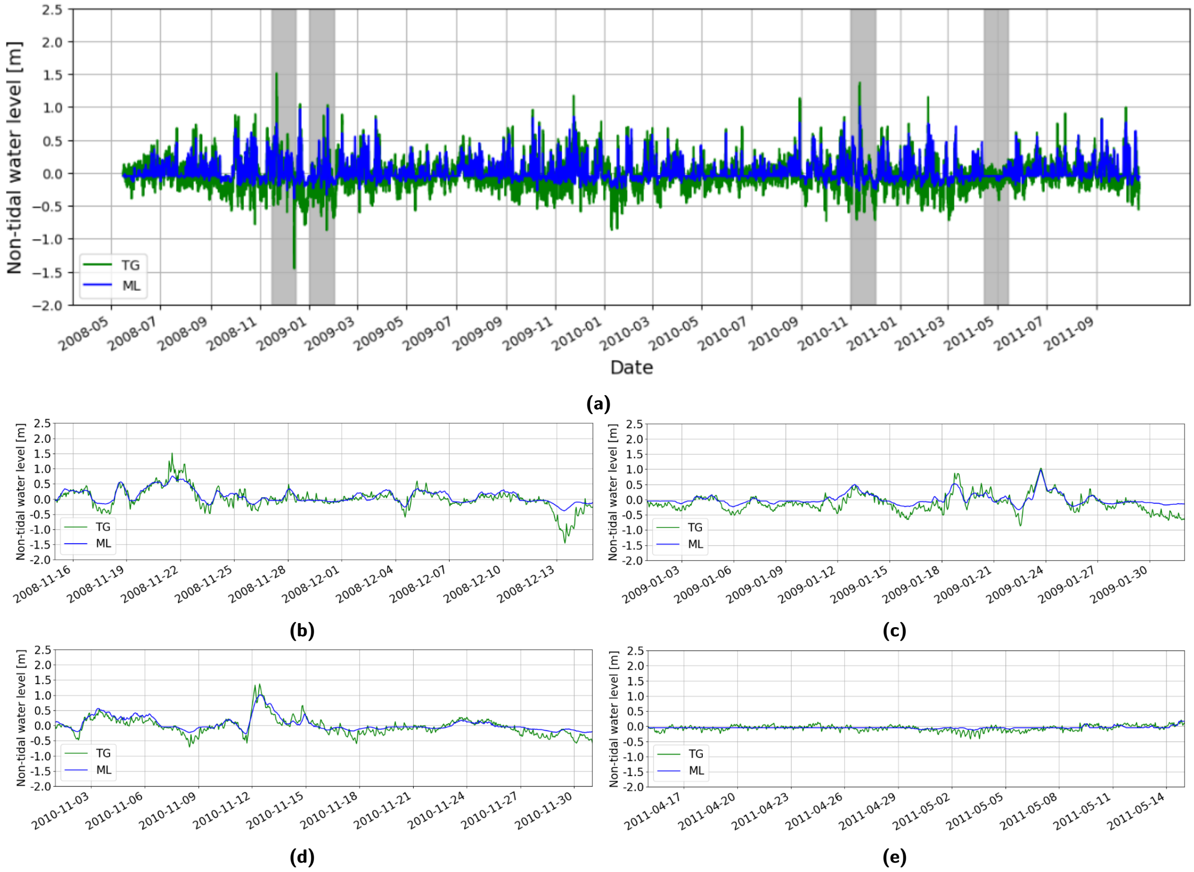


Figure G.3: Time-series reconstruction for validation set 3 of the Scheveningen model. The shaded areas in (a) correspond to the months shown in (b), (c), (d) and (e). The green lines represent the tide gauge observations, the blue lines refer the machine learning model output. The shaded areas in the sub-graphs show storms catalogued by the KNMI (KNMI, 2024). The unit of the y-axis is in metres and depicts the non-tidal water level referenced to NAP.

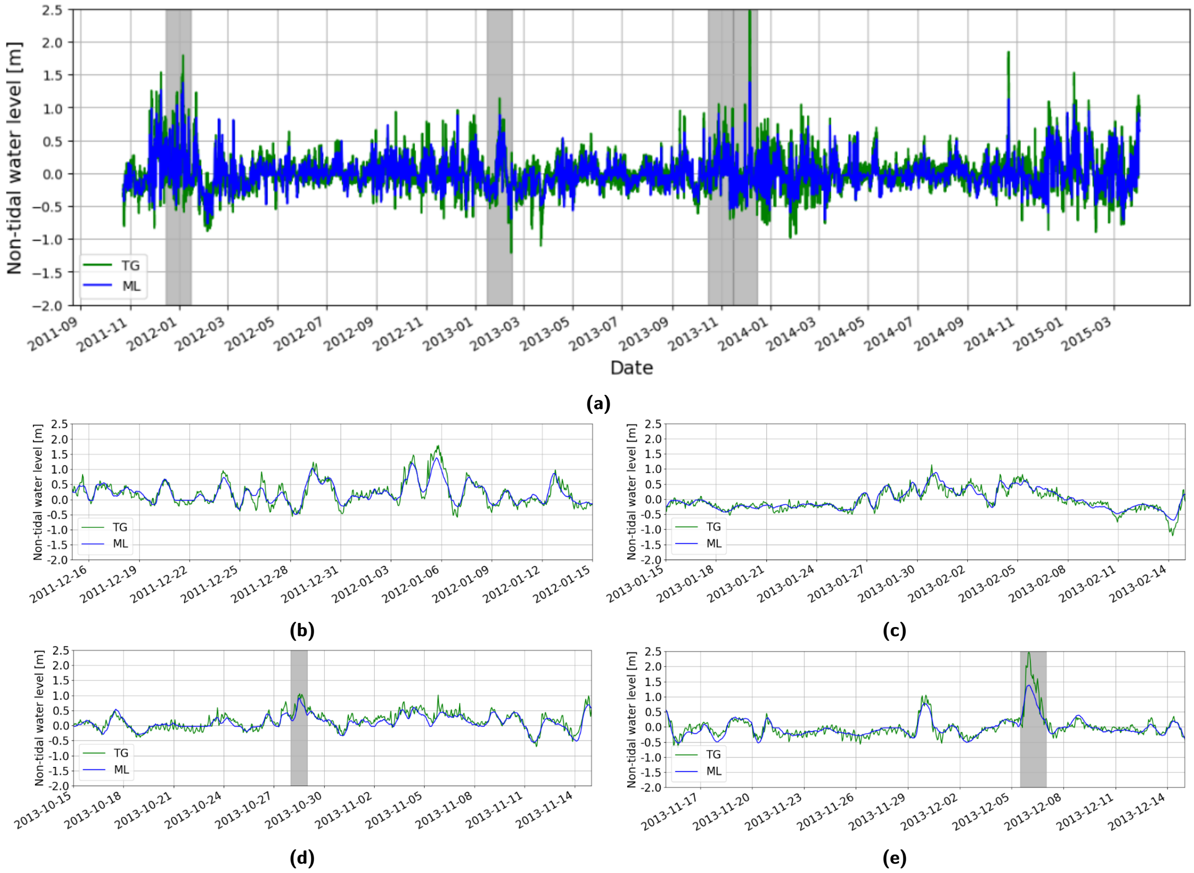


Figure G.4: Time-series reconstruction for validation set 4 of the Scheveningen model. The shaded areas in (a) correspond to the months shown in (b), (c), (d) and (e). The green lines represent the tide gauge observations, the blue lines refer the machine learning model output. The shaded areas in the sub-graphs show storms catalogued by the KNMI (KNMI, 2024). The unit of the y-axis is in metres and depicts the non-tidal water level referenced to NAP.

H

Weight analysis

For every model, the model weights have been visualised. Section 4.4 explains the inner workings of the neural network applied in this study, of which the trained parameters (weights and biases) are presented here. The weights for the first layer are linked directly to the input features. This enables the p, U10 and V10 from ERA5 as well as the NTR, d$_E$ and d$_N$ from satellite altimetry to be shown as individual maps. Each map shows the difference between the first and the last hour within the 48-hour time window used as input for the model, which ranges from 48 to 1 hour prior to the time stamp of the target value. Presented in this appendix are the first-layer weight differences of the satellite altimetry variables and the ERA5 variables for each model. If the weight increases between hour -48 and hour -1 within the input dataset, the differences are shown in red. If the weights decrease, the differences are shown in blue. Additionally, the first-layer weights of the satellite altimetry features are shown for Europlatform and Vlissingen, since the weights for Scheveningen have already been presented in Chapter 5.

H.1. Satellite altimetry

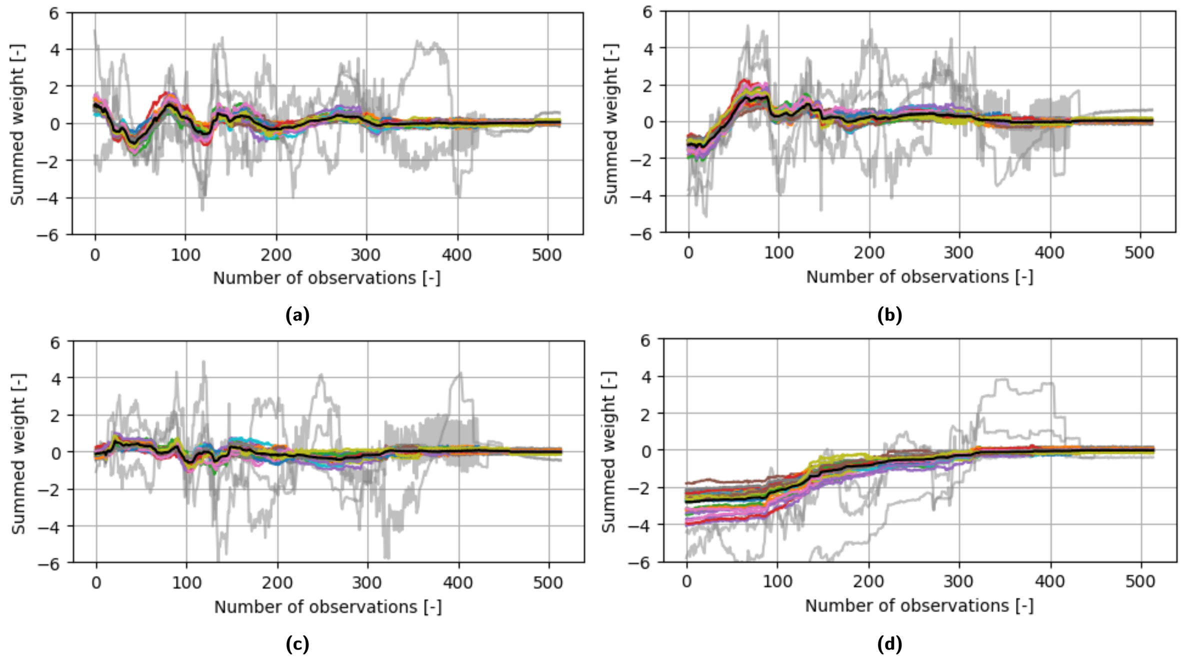


Figure H.1: First-layer weights summed over the 48-hour time window for the Vlissingen model for (a) the NTR, (b) the longitudinal component of the distance d$_E$, (c) the lateral component of the distance d$_N$ and (d) the time difference between observation and tide gauge dt.

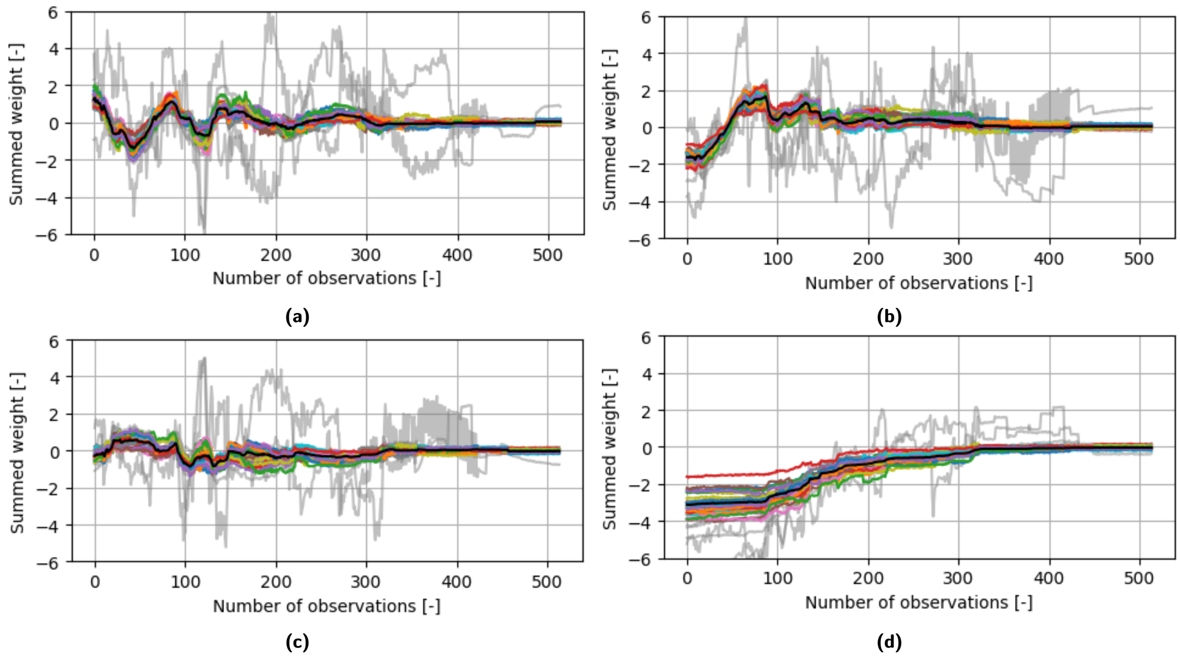


Figure H.2: First-layer weights summed over the 48-hour time window for the Europlatform model for (a) the NTR, (b) the longitudinal component of the distance d_E, (c) the lateral component of the distance d_N and (d) the time difference between observation and tide gauge dt.

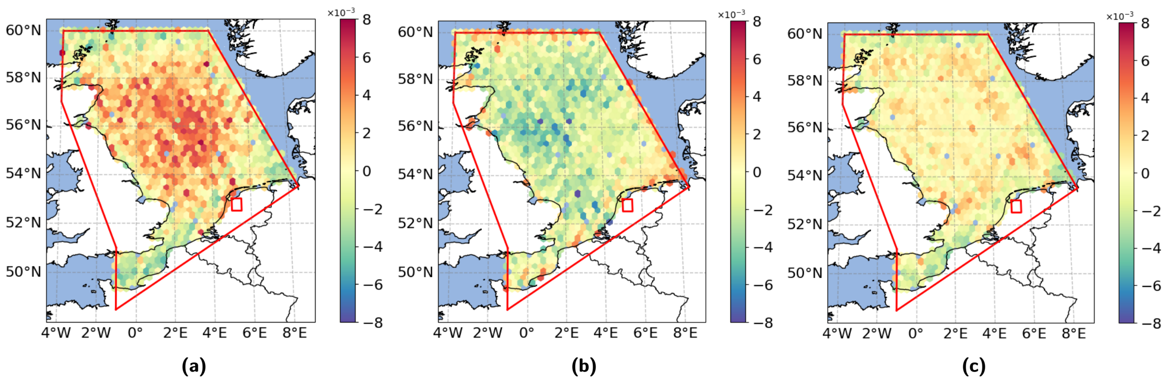


Figure H.3: Difference between hour -48 and -1 of satellite altimetry first-layer weights corresponding to (a) non-tidal water level, (b) longitudinal distance and (c) lateral distance for Scheveningen.

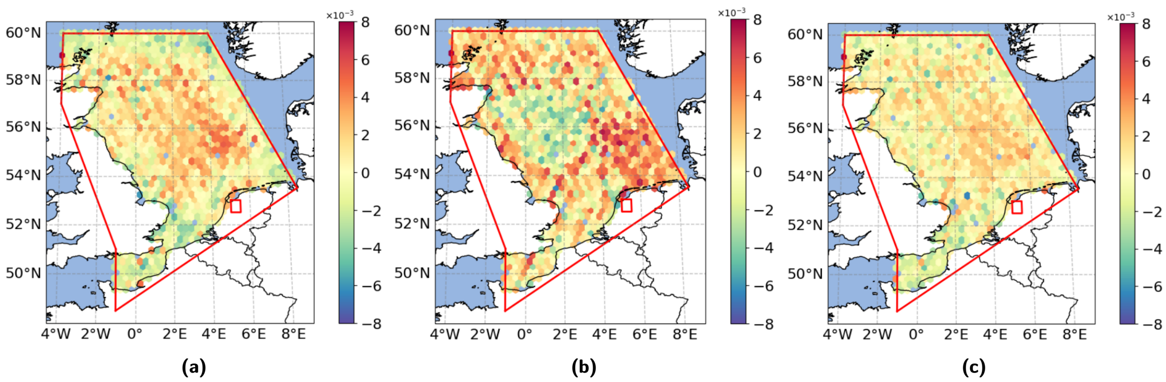


Figure H.4: Difference between hour -48 and -1 of satellite altimetry first-layer weights corresponding to (a) non-tidal water level, (b) longitudinal distance and (c) lateral distance for Vlissingen.

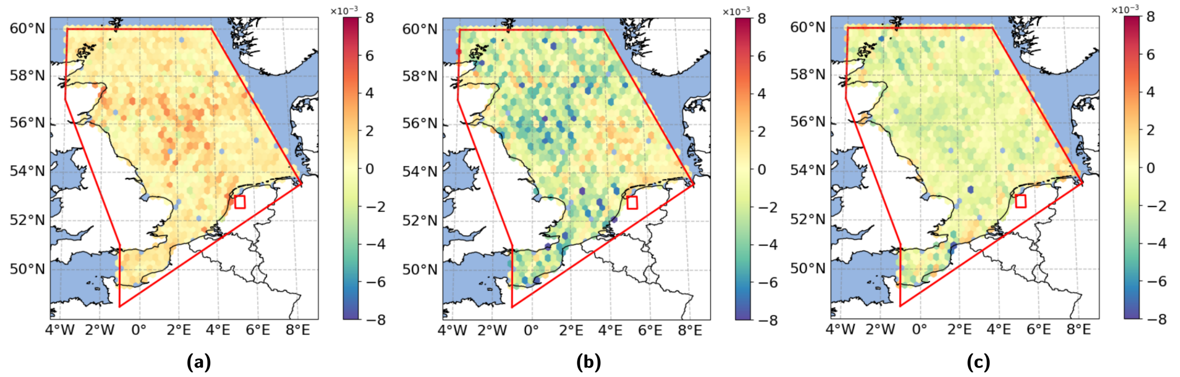


Figure H.5: Difference between hour -48 and -1 of satellite altimetry first-layer weights corresponding to (a) non-tidal water level, (b) longitudinal distance and (c) lateral distance for Europlatform.

H.2. ERA5

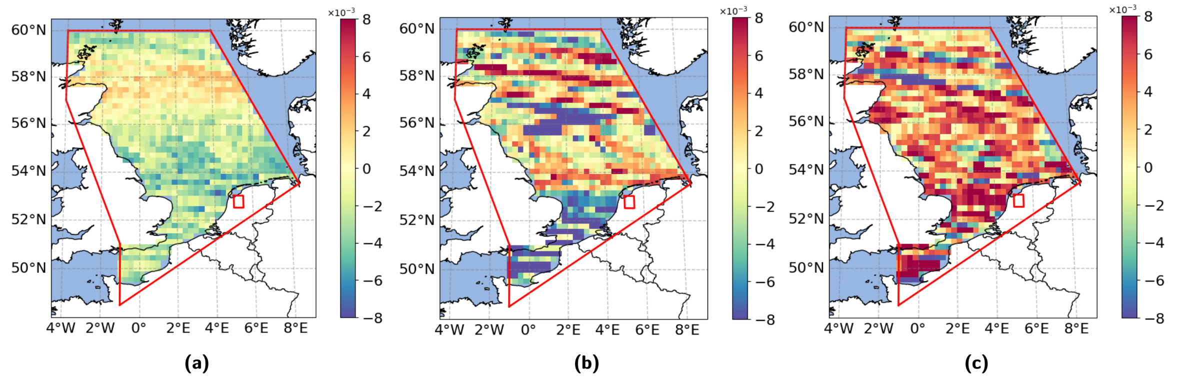


Figure H.6: Difference between hour -48 and -1 of ERA5 first-layer weights corresponding to (a) sea surface pressure, (b) longitudinal wind speed and (c) lateral wind speed for Scheveningen.

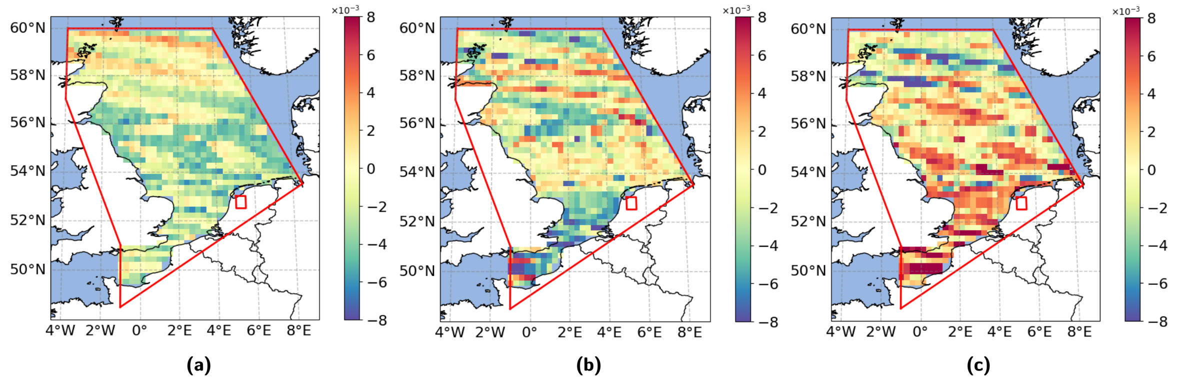


Figure H.7: Difference between hour -48 and -1 of ERA5 first-layer weights corresponding to (a) sea surface pressure, (b) longitudinal wind speed and (c) lateral wind speed for Vlissingen.

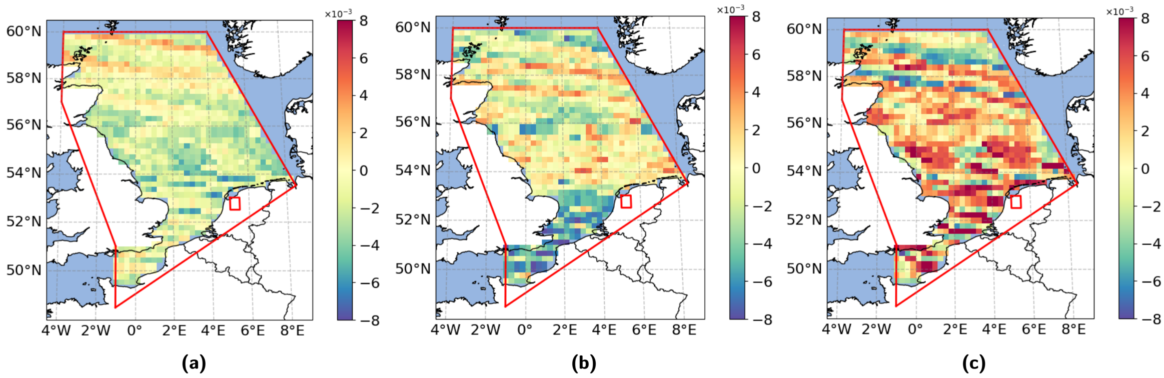


Figure H.8: Difference between hour -48 and -1 of ERA5 first-layer weights corresponding to (a) sea surface pressure, (b) longitudinal wind speed and (c) lateral wind speed for Europlatform.