

# Detecting anomalous sea-level states in North Sea tide gauge data using **an** autoassociative Neural Network

Kathrin Wahle<sup>1</sup>, Emil V. Stanev<sup>1,2,3</sup>, Joanna Staneva<sup>1</sup>

<sup>1</sup>Helmholtz Zentrum Hereon, Geesthacht, Germany

5 <sup>2</sup> Research Department, University of Sofia “St. Kliment Ohridski”, Sofia, Bulgaria,

<sup>3</sup> Department of Meteorology and Geophysics, University of Sofia “St. Kliment Ohridski”, Sofia, Bulgaria

*Correspondence to:* Kathrin Wahle (kathrin.wahle@hereon.de)

## Abstract.

The sea level in the North Sea is densely monitored by tide gauges. ~~These measurements are used for validation as well as for the detection of extreme events.~~ The data they provide can be used to solve different scientific and practical problems, including the validation of numerical models, as well as **and the** detection of extreme events. ~~Here, we focus on the detection of sea-level states with anomalous spatial correlations by means of autoassociative neural networks. An autoassociative neural networks. emulates the spatial correlations of gauges in a lower dimensional subspace. Anomalous sea-level states are defined by means of failure of the reconstruction model. Using spatially distributed data as input to the network thus reveals sea-level conditions with unusual spatial correlations.~~ This study focuses on the detection of sea-level states with anomalous spatial correlations using auto associative neural networks (AANNs), trained with different sets of observation- and model-based data. Such sea-level configurations are related to nonlinear ocean dynamics; therefore, neural networks appear to be the right candidate for their identification. The proposed network can be used to accurately detect such anomalies and localize them. We demonstrate that the ~~corresponding~~ atmospheric conditions ~~indicate~~ under which anomalous sea-level states occur are characterized by high wind tendencies and pressure anomalies. The results show the potential of AANNs in accurately detecting the occurrence of such events. We show that the method works with AANN trained on tide gauge records as well as with AANN trained with model-based sea surface height outputs. The latter can be used to enhance the representation of anomalous sea-level events in ocean models. Quantitative analysis of such states ~~might~~ **may** help assess and improve numerical model quality in the future as well as provide new insights into the nonlinear processes involved. The method has the advantage of being easily applicable to any tide gauge array without preprocessing the data or acquiring any additional information.

## 1 Introduction

The dynamics of sea level in tidal basins are one of the most addressed topics in physical oceanography. Theoretical prediction of tidal motion was pioneered by the application of Fourier analysis by Lord Kelvin (Thomson, 1880) and later improved by Doodson (1921), who developed the tide-generating potential in harmonic form. Analysis and interpretation of tidal observations by Proudman and Doodson (1924) enhanced the understanding of sea-level fluctuations due to winds and changes in atmospheric pressure. The development of numerical 2D storm surge models by Peeck et al. (1983) and Flather and Proctor (1983) led to early warning systems for coastal flooding. With increasing computational power and the availability of satellite data, sea-level predictions have been continuously improved. However, current model predictions are not always perfect (Stanev et al., 2015; Sandery and Sakov, 2017; Staneva et al., 2016; Ponte et al. 2019; Mey et al., 2019; Jacobs et al., 2021), ~~and thus~~ **which** emphasizes the need to ~~gain~~ further understanding of sea level.

A recent important evolution in predicting sea-level in the North Sea was achieved in the framework of the development of the Northwest European Shelf forecasting system (e.g. O'Dea et al., 2012, Tonani et al., 2019) **by enhancing the model**

40 resolution to 1.5 km. Thus, dynamical features such as coastal currents, fronts, and mesoscale eddies are better resolved, and the model results are improved, especially when compared to high spatial–temporal resolution observations.

Regional sea level variability and trends in the Baltic Sea have been largely studied using tidal gauge measurements. Satellite altimetry has added critical additional information in the last 30 years (Madsen et al., 2015). Notably, different measurement techniques have different advantages and disadvantages. Satellite-derived sea-level information, which has revolutionized oceanography and climate science, particularly in addressing global and large-scale change, is of limited use when addressing near-coastal short-periodic variability. However, advancements are underway, and new satellite missions characterized by better spatial and temporal sampling have paved the way for improvements in coastal sea-level research (e.g. Dieng et al., 2021; Prandi et al., 2021; Dodet et al., 2020; Sanchez-Arcilla et al., 2021a). In addition, they are not very precise in coastal areas and do not sufficiently resolve the short periodic variability. What is also currently available is data from operational models, which provide good temporal and spatial coverage.

50 Tide gauge stations operating along the North Sea coast provide high-quality records of sea-level observations over a long period (Wahl et al., 2013). Ponte et al. (2019), reviewing the state of science of coastal sea-level monitoring and prediction, outlined the importance of sea-level observations for studying sea-level variability. However, tidal gauges provide precise estimates with high temporal resolution in specific locations along the coast but do not provide information about the basin-wide patterns of sea level. Furthermore, some of these data are not continuous; different gauges do not always operate simultaneously, and there are gaps in many of the records. Therefore, a combination of numerical model results with tide gauge measurements would be beneficial. a question arises as to whether one can combine data from gauges and independent 2D maps of sea level (from models) to produce a consistent data set (covering all coastal locations at the same time and with high temporal resolution). A similar exercise was undertaken recently by Madsen et al. (2019) for the Baltic Sea and by Zhang et al. (2020) for the North Sea. Tide gauge stations operating along the North Sea coast provide high quality records of sea level observations over a long period (Wahl et al., 2013). Ponte et al. (2019), reviewing the state of science of coastal sea level monitoring and prediction, outlined the importance of sea level observations for studying sea level variability. Zhang et al. (2020) used machine learning to reconstruct the sea-level variability in the North Sea using observations from 19 coastal tide gauges and data from numerical models. Notably, they concluded that a relatively short-time record contains the most representative characteristics of sea-level dynamics in the North Sea. While this was clear for tides, it was not so obvious for changes in sea level caused by the atmosphere.

65

The North Sea (Figure 1a) is a shallow sea with an average depth of ~ 90 m located at the European continental shelf. The dynamics of sea level in the North Sea, which is the region of our study, can be considered a coupled response to different forcings, such as barotropic tides, wind and atmospheric pressure, as well as forcings from the open boundaries and rivers, including a thermohaline forcing. The coupling of the respective processes is, in most cases, nonlinear (Jacob et al., 2017), that is therefore, one cannot easily consider identify the response to individual drivers in isolation. This happens when either oscillatory motions have large amplitudes, e.g., tidal currents approaching 1 m/s, or when wind-driven current is of the same order. Thus, there is a need to use methods tailored to detect and reproduce nonlinear dynamics. The n Nonlinear processes are difficult to predict, even with sophisticated perfect models; therefore, one also has needs to identify situations in which predictions fail (Ponte et al., 2019). Furthermore, it is interesting to understand the reasons for the deviation in the model predictions from the observations. The detection of such situations in data from a network of tide gauges is the aim of the present study.

75

The spatial and temporal correlations of tide gauge data mirror the complex sea-level dynamics in the North Sea (Figure 2). First, tidal wave propagation is influenced by the topography of the North Sea. The north-south decrease in water depth shifts the central amphidromic point southwards (see Figure 1a for the oscillation pattern of the semidiurnal M2 tide). Two additional amphidromies are generated as a result of the superposition of incident and reflected Kelvin waves and cross-shore Poincare

80

waves with centres in the north, off the Norwegian coast and in the south, between Suffolk and Holland (Figure 1a). The amplitude of the tidal wave increases towards the British coast. Off the Danish and Norwegian coasts, the amplitude of the wave is significantly smaller due to dissipation by bottom friction in the shallow southern part of the North Sea (Figure 2a).

85 The circulation of the North Sea and thus the sea level is forced also by the atmosphere through either wind stresses (Figure 2c) or atmospheric pressure gradients (Figure 2d). Considering the length scales and depth of the North Sea, the wind has a dominant effect. The prevailing westerly winds result in a dominant cyclonic circulation. A reversal of circulation caused by easterly winds seldom occurs (Stanev et al., 2019). For northwesterly and southeasterly winds, circulation may stagnate. As shown by Jacob et al. (2017), the interaction between tidal and wind-driven currents is nonlinear in areas of strong currents,

90 particularly in the German Bight, and the atmospheric variability affects the spring-neap variability more strongly than the M2 tide. Additionally, shallow-water tides show significant small-scale spatial variability patterns in this region (Stanev et al. 2015). Thus, in the southern North Sea, tidal forcing strongly impacts the residual circulation (Figure 2b). In summary, the specific topography of the North Sea together with nonlinear effects lead to complex circulation patterns that hamper sea-level prediction.

95 Simple linear statistical methods, such as principal component analysis (PCA), fail to predict nonlinear dynamics because they represent linear combinations of mean states. Flinchem and Jay (2000) demonstrated that wavelet transforms provide an approach that is well applicable to suited for tidal phenomena that “deviate markedly from an assumed statistical stationarity or exact periodicity inherent in traditional tidal methods”. Wavelet analysis and that it can also reveal features that harmonic analysis could not elucidate.

100 ~~Here we use autoassociative neural nets (AANNs) as an alternative approach to detect such anomalous sea level states. Similar to PCA, this method is based on the assumption that the data can be compressed into a lower dimensional subspace. The combination of nonlinear components best describing the variability in the training data can be used to reconstruct the data themselves. Subsequently, data with correlations similar to those of the training are reproduced well.~~

105 ~~AANN provides such a nonlinear reconstruction model, with the reconstruction error as a measure of how well data are characterized by the model. Large errors in the reconstruction of independent data by the model signals situations, in which data do not belong to the same distribution the model was built on.~~

~~For sea level, the entirety of tide gauge measurements in the North Sea will show specific spatial correlations (changing in time with the tide). However, in anomalous situations (e.g. localized storms) these correlations may drastically change.~~

110 In ~~this~~ the same context, neural networks (NNs) have been used for (extreme) tidal surge prediction (French 2017, Tayel 2015, and Bruneau 2020). Similarly, Hieronymus et al. (2019) applied different machine learning techniques to the regression problem of time series data from tide gauges data, and Balogun and Adebisi (2021) investigated the impact of different ocean-atmosphere interactions on sea-level predictability using different NNs. However, all these applications ~~are limited to focus~~

115 ~~on data from~~ single tide gauge stations.

Horsburgh (2007) proved the importance of spatial and temporal correlations among tide gauge stations along the British coast for the distribution of surge residuals. However, their linear model is only capable of describing some of the possible interactions between tides and surges and was used to demonstrate the existence of critical spatiotemporal scales for surge development and decay. Wenzel (2010) used neural networks for the reconstruction of monthly regional mean sea-level anomalies from 59 tide gauges worldwide. Zhang et al. (2020) used generative adversal networks to reconstruct the sea level variability in the North Sea from an array of coastal tide gauge data using observations from 19 coastal tide gauges and data from numerical models. Notably, they concluded that a relatively short-term record contains the most representative characteristics of sea-level dynamics in the North Sea. While this was clear for tides, it was not so obvious for changes in sea level caused by the atmosphere. Similarly, Yang et al. (2021) used a deep learning approach to fuse altimeters data with tide

125 gauge data in the Mediterranean Sea [was described by Yang et al. \(2021\)](#), while Nieves (2021) et al. used open ocean temperatures to predict coastal sea-level variability [via machine learning](#).

In contrast to ~~all~~ these applications, we will focus on ~~the opposite opposite: we will analyse data from an array of tide gauges in the North Sea simultaneously to identify situations in which their spatial correlations deviate greatly from the mean correlations~~ [the identification of situations in which spatial correlations between tide gauge measurements deviate greatly from the principal correlations](#). Usually, tide gauge measurements in the North Sea show specific spatial correlations (changing in time with the tide). However, in anomalous situations (e.g., localized storms), these correlations may drastically change. We use autoassociative neural networks (AANNs, Kramer, 1992) to detect such sea-level states. Similar to PCA, the aim of this method is dimensionality reduction; however, the concept of orthogonal vectors is expanded to principal curves. The combination of nonlinear components best describing the variability in the training data can be used to reconstruct the data. Thus, AANN provides a nonlinear reconstruction model, and the reconstruction error is a measure of how well the data are characterized. Large errors in the reconstruction of data by the AANN represents situations in which the data do not belong to ~~the same distribution they were built from~~. Atmospheric conditions related to such situations might aid further understanding and future developments in sea-level prediction.

140 The paper is structured as follows: In section 2, we present ~~the method and data used~~ [observational and model data used throughout our study and introduce the concept of AANN](#). We then apply the AANN to the tide gauge array measurements in the North Sea. In section 3, ~~presents the extreme~~ [we use AANN to detect anomalous events and examine the dependence of identification of such events from the AANN training data used](#). Two events are studied in detail, including ~~and~~ their relationship with the atmospheric conditions. [This is followed by a discussion and conclusions](#).

## 2. Data and Methods

~~The spatial and temporal correlations of tide gauge data mirror the complex sea level dynamics in the North Sea, originating from the interactions among different forcings. First, the propagation of the tidal wave is influenced by North Sea topography. The north-south decline of water depth shifts the central amphidromic point southwards (see Figure 1a for the oscillation pattern of the semidiurnal M2 tide). Two additional amphidromies are generated as a result of a superposition of incident and reflected Kelvin waves and cross-shore Poincaré waves with centres in the north off the Norwegian coast, and in the south between Suffolk and Holland (Figure 1a). The amplitude of the tidal wave increases towards the British coast. Off the Danish and Norwegian coasts amplitudes are significantly smaller due to dissipation by bottom friction in the shallow southern part of the North Sea.~~

~~Second, the circulation of the North Sea and thus the sea level is forced by the atmosphere through the vertical flux of momentum due to either wind stresses or atmospheric pressure gradients. Considering the length scales and depth of the North Sea, the wind will have a dominant effect. The prevailing westerly winds result in a dominant cyclonic circulation, and reversed rotation (easterly winds) seldom occurs. For northwesterly and southeasterly winds, circulation might stagnate. As shown by Jacob et al. (2017), the interaction of tidal and wind-driven currents becomes nonlinear in areas of strong currents, particularly in the German Bight. They also demonstrated that the ocean response to atmospheric variability does not mimic that one, and the atmospheric variability affects the spring-neap variability more strongly than the M2 tide.~~

~~In summary, the specific topography of the North Sea together with nonlinear effects lead to complex circulation patterns that hamper sea level prediction.~~

### 2.1 Observational and Model Data

Observational sea-level data along the North Sea coast have been obtained from the historical and near real-time (NRT) dataset of the Copernicus Marine Environment Monitoring Services (CMEMS). ~~Altogether 14 gauge stations with hourly data are used.~~ The observations were taken from the respective in situ products for the Northwest Shelf area (Copernicus Marine In Situ TAC Data Management Team, 2020, <http://marine.copernicus.eu/>) with hourly resolution. The NRT *in situ* quality-controlled observations are updated hourly and distributed within approximately 24–48 h after acquisition. From these, we have chosen 14 gauge stations according to completeness of data time series. Their positions are shown in Figure 1a. ~~Figures 1b and d show the time series of measured and detided (residual) water levels.~~ Time versus position diagrams of measured and detided (residual) water levels are shown in Figure 2 a, b. ~~The former Figure 1a shows the propagation of the tidal wave along the English coast with water level amplitudes increasing southwards towards the English eChannel.), whereas the latter ones.~~ The slope of contours in Figure 2 gives the speed of propagation of the tidal wave. The specific feature between the Cromer and Vilssingen stations identifies the small amphidrome in front of the English Channel. This feature is not present in the detided data (Figure 2b); the latter resembles the atmospheric forcing (Figure 2 c, d). In the presence of large gradients in wind tendency (Figure 2c), the water surface tilts considerably (Figure 1c).

For the experiments discussed later in this study, we also use data from the GCOAST (Geesthacht Coupled cOASTal model SysTem) circulation, wave and ocean model (Madec, 2017, Staneva et al., 2017, 2021, Bonaduce et al., 2020). Wave-current interaction processes included in the model are momentum and energy sea state-dependent fluxes, wave induced mixing and Stokes-Coriolis forcing. ~~ocean circulation model is based on the Nucleus for European Modelling of the Ocean (NEMO v3.6, Madec, 2017, Staneva et al., 2017, 2021, Bonaduce et al., 2020) with 3.5 km horizontal resolution.)~~ The model area covers the Baltic Sea, the Danish Straits, the North Sea and part of the northeast Atlantic Ocean) with 3.5 km horizontal resolution. The data used in the present study are only for the North Sea region. ~~The model presents an ocean component of the Geesthacht Coupled cOASTal model SysTem (GCOAST, NEMO v3.6; Madec, 2017, Staneva et al., 2017, 2021, Bonaduce et al., 2020).~~ The ocean circulation model is based on the Nucleus for European Modelling of the Ocean (NEMO v3.6). The wave model is WAM (cycle 4.7), a third-generation wave model, that solves the action balance equation without any *a priori* restriction on the evolution of the spectrum. The two models are two-way coupled via the OASIS3-MCT version 2.0 coupler (Valcke et al., 2013). The NEMO setup used within the GCOAST system uses an explicit free surface formulation (Madec 2017; Staneva et al. 2021, Bonaduce et al., 2020). ~~The NEMO setup used within the GCOAST system uses an explicit free-surface formulation;~~ vertical discretisation of the NEMO setup uses 50 hybrid s-z\* levels with partial cells are used in the vertical. Atmospheric pressure and tidal potential are also included in the model forcings (Egbert and Erofeeva, 2002). Daily river run-off is based on river discharge datasets from the German Federal Maritime and Hydrographic Agency (Bundesamt für Seeschifffahrt und Hydrographie, BSH), Swedish Meteorological and Hydrological Institute (SMHI) and United Kingdom Meteorological Office (Met Office). Boundary conditions at the open boundaries (temperature, salinity, velocity and sea level) are taken from the AMM7 model (O’Dea et al., 2012) distributed by the Copernicus Marine Environment and Monitoring Service. They have a temporal resolution of one hour with a 7 km horizontal resolution. The model forcings for momentum, and heat fluxes are computed using bulk aerodynamic formulae and hourly data from atmospheric reanalyses of the European Centre for Medium Range Weather Forecasts (ERA5, Hersbach et al, 2020), the 5th generation of reanalysis of the ECMWF. ERA5 has horizontal resolutions of 28 km and 31 km, respectively. We use ERA5 1-hourly products provided by Copernicus Climate Center Service (C3S, 2021) available with 0.25° and 0.25° horizontal resolutions for the atmospheric and wave parameters, respectively.

Several studies have demonstrated the advantages of using the ERA5 reanalysis over its predecessor (ERA-INTERIM, e.g. Belmonte Rivas and Stoffelen, 2019).

## 2.2 Autoassociative Neural Network

Most environmental monitoring programs produce large sets of multivariate data, with the tide gauge data network in the North Sea being among them (CMEMS global ocean in situ near-real-time observations, <https://doi.org/10.48670/moi-00036>). Although these sea level data depend on ~~These multivariate data are determined by~~ a number of factors variables (here, tidal and atmospheric forcings, together with bathymetry, river input, etc.), this ~~Since the number of factors is often smaller than~~ the number  $n$  of tide gauges within the network. The observed data variables, the data are located within a subspace of the given  $n$ -dimensional data space and can thus be compressed. Different machine learning techniques, such as the k-nearest neighbours algorithm, ensemble-based methods, and support vector machine (SVM) algorithms, are used to predict the posterior probabilities of a given dataset and are optimal for data compression. The different techniques are not equally well suited for detecting outliers, i.e., identifying whether a given observation belongs to the same probability distribution. Outlier detection in high dimensions, or without any assumptions about the distribution function of the data is very challenging. SVM algorithms work well if training data are not contaminated by outliers. Ensemble and k-nearest neighbour methods perform well for multimodal datasets. Covariance estimators, in which category PCA also falls, degrade when the data are not unimodal.

This subspace can be described by its (linear or nonlinear) components. The components serve to construct a model, with the reconstruction error as a measure of how well data are characterized by this proxy. Thus, large reconstruction errors signal anomalous situations in the sense that data are not located in the spanned subspace. The model aims to provide a tight enclosure of the subspace spanned by the data under normal conditions. Here, anomalous applies to data not lying within the bounds of this model and not belonging to the same distribution the model was built on. Due to the absence of an appropriate model of the underlying processes, it is very often difficult to detect such situations hidden in the data.

In many such cases, an Autoassociative neural networks (AANNs) combine the robust performance of multimodal data with the geometric interpretability of PCA could be the proper tool to use to become aware of identify these situations.

An AANN is a reconstruction model based on a feed-forward neural network (Kramer 1992). In an AANN, it maps the  $n$ -dimensional input  $\mathbf{x}$  is mapped onto itself ( $\mathbf{x}'$ ) with a data compression step in between: The number of neurons in at least one hidden layer in an AANN is less than the dimension  $n$  of the input and output vectors  $\mathbf{x}$  and  $\mathbf{x}'$ . This layer is called the bottleneck-layer and is the key component of an AANN (Figure 3a). It provides data compression of the input with powerful feature extraction capabilities. The mapping part  $m(\mathbf{x})$  compresses the information of the  $n$ -dimensional vector to a smaller dimension subspace vector  $\mathbf{p}$ , whereas the demapping part  $\mathbf{x}'=d(\mathbf{p})$  uses compressed information to regenerate the original  $n$ -dimensional vector.

Thus, In the mapping part, ~~takes advantage of~~ the (nonlinear) correlations/functional relations existing among the input variables (here, the spatial correlations) are taken advantage of. Thus, an AANN can be considered to perform a nonlinear generalization of the ~~Principal Component Analysis~~ PCA (nonlinear principal component analysis, NLPCA (Kramer 1991)). Similar to PCA, NLPCA can also serve important purposes, e.g., filtering noisy data, feature extraction, data compression, outliers and novelty detection (Mori 2016).

During training, the AANN constructs a model, that captures the posterior probability distribution of a given dataset. At the beginning of the training, the outcome of the NN will differ largely from the target output. The mean squared relative error per neuron  $e$  is iteratively minimized during the training by backpropagating the error through the NN and adjusting free parameters according to a gradient descent scheme:

$$\mathbf{e} = \frac{1}{N} \sum_{tr=1}^N \mathbf{e}_{tr}, \quad \mathbf{e}_{tr} = \frac{1}{n} \sum_{i=1}^n \left( \frac{x'_i - x_i}{\max(x_i^{tr}) - \min(x_i^{tr}), tr \in [1, N]} \right)^2,$$

where  $N$  is the number of (training) input/output data pairs.

250

~~This model shows that the intrinsic dimensionality (= the number of independent components) of the data set is not higher than the number of neurons in the bottleneck layer.~~

~~Thus, an~~ An important step in the construction of an appropriate AANN is the choice of the number of neurons in the bottleneck layer, i.e., to determine the dimension of the subspace  $\mathbf{p}$ . The choice is based on the reconstruction abilities of the network:

255

The overall reconstruction error of the AANN decreases with an increasing number of neurons in the bottleneck layer. However, if the number of bottleneck neurons equals the intrinsic dimensionality, a further increase leads to ‘overfitting’ of the AANN, i.e., learning stochastic variations in the dataset rather than the underlying functions. As a result, the decrease in the overall error flattens (see Figure 3b).

### 260 2.3 Application of AANN to the North Sea Water Level from Tide Gauges

The proposed scheme was applied to data from 14 ( $\dim(\mathbf{n})=14$ ) tide gauges along the North Sea coast (see Section 2.1). ~~Data used in this study span from 1. January 2016 to 31. December 2018. The data set was~~ from 2016/17 were used to train the AANNs ~~spanning 2016 and 2017. D~~ data from 2018 were kept retained as an independent testing set. ~~The A~~ two-year training period was assumed to be long enough to capture the essential spatial correlations among observations at different gauges well. Thus, any dominant factor affecting the interrelationship between observations from individual stations (bathymetry, the distance between stations, dominant winds and atmospheric pressure, etc.) should be reflected by the AANN. The choice of tide gauge locations was a compromise between spatial coverage and data availability (since missing data from single gauge stations were not filled, e.g., by interpolation, and thus resulted in data loss). We applied no further preprocessing or quality control to the tide gauge data.

270

A sequence of AANNs with an increasing number of neurons in the bottleneck layer was trained to map the 14 water levels onto themselves. In Figure 23b, the relative reconstruction error  $\mathbf{e}$  for all training data is plotted versus the number of neurons in the bottleneck layer. ~~As expected, the e~~ Error decreases with an increasing number of bottleneck neurons; ~~the decrease starts to flatten~~ Two different regimes are visible: one regime is for the number of bottleneck neurons below approximately seven and the other regime is for those above this number. ~~Therefore, the smallest AANN belonging to the second regime with~~ at seven neurons ( $\dim(\mathbf{p})=7$ ) in the bottleneck layer is chosen. ~~Therefore, we decided to use seven bottleneck neurons and refer to this as AANN\_ref (reference AANN).~~

275

### 2.4 Quality of AANN Analysis

The performance of the AANN\_ref is shown in Figure 34. The scatterplots are used to compare the modelled vs. observed sea-level data. The distribution for the training phase (Figure 34a) closely follows the bisectrix and is very narrow. The standard deviation between the observed and modelled AANN reconstructed sea-level data is approximately 10 cm ~~well below 1 cm~~ (Figure 3e 4c). ~~During the validation~~ For the testing phase (Figure 3b, f 4b, d), the agreement between the observations and modelled data is almost as good as that for the training phase; however, some scatter is also clearly seen. Parts of this disagreement, with errors above the 99.9 percentile during the training phase, are marked in red and indicate possible candidates for anomalous situations, which will be further analysed. The distribution of errors is non-Gaussian during training and testing (Fig. 3e, f 4c, d). The long tail of the error histogram presents shows data where the reconstruction by AANN failed and where anomalous correlations among tide gauges are expected.

285

The non-Gaussian nature of the error distribution demonstrates the capability of the AANN to capture complex (usually nonlinear) processes. This is the fundamental difference from the case when PCA (using 7 PCs, analogous to the number of bottleneck neurons in AANN\_ref with the same subspace dimension) is used to reconstruct the observed data (Figure 3e, d and g, h 4e-h). In PCA, the results have approximately 3.5 times higher standard deviations. The latter model shows errors, which are two orders of magnitude larger than the errors in the AANN reconstruction, indicating non-linearity of the underlying physics. In this case, the linear statistical model PCA here leads to a short-tailed Gaussian error distribution (Figure 3g, h 4g, h), making it unsuitable for the desired task. Because the non-Gaussian distribution is a typical characteristic of nonlinear processes, and because linear methods (PCA) do not capture it, we conclude that the above comparison gives a clear demonstration of the power of the AANN method used in the present study to handle nonlinearities. Outliers are usually in the tail of the distribution, and their identification requires using nonlinear reconstruction methods.

### 3. Anomalous Sea-Levels and Their Relationship with the Atmospheric Conditions

To isolate events that strongly affect the spatial correlations among several tide gauges, further constraints are needed. High values of AANN\_ref reconstruction errors might originate from events on small scales, such as ocean response to wind gusts, as well as the erroneous measurement of a single tide gauge. Thus From the analysis of the training period, we postulate an ocean state as anomalous when the that a given threshold error  $\epsilon$  has to be exceeded 0.035 (see Figures 4c, d) by at for at least 3 gauges simultaneously for at least 3 hours.

To exclude the dependence of identification of such extreme events from the data used, we trained the alternative AANNs based on the following three datasets:

- trained with data extracted at the closest positions of gauge stations from the CMEMS operational model AMM15 (AANN\_NEMO);
- trained with a subset of gauge data from only 10 stations (with Dover, Oostende, Vlissingen and Europlatform in the southwest corner of the North Sea excluded) (AANN\_less); and
- trained with detided data from the 14 gauge stations (AANN\_resid). Software package T-TIDE (Pawlowicz et al., 2002) was used for this purpose.

As for AANN\_ref, training data span the two-year period 2016/17 and data from 2018 were kept for validation.

The first additional network AANN\_NEMO allows an intercomparison between observations and simulations. If the agreement is good, areal coverage of spatial data from the operational model will enable the analysis of horizontal patterns associated with the identified events. The second network, AANN\_less, identifies the possible occurrence of such anomalous events in the southwestern North Sea area, which is characterized by its complex dynamics and high amplitudes of tidal oscillations. This area is known to be dominated by a small amphidrome (Figure 1a) and is characterized by its complex dynamics and high amplitudes of tidal oscillations. The third network, AANN\_resid, reveals how strongly anomalous events are associated with the basic tidal dynamics, which are removed here from the training and control data sets.

The error distributions of AANN\_less, AANN\_resid and AANN\_NEMO are similar to those of AANN\_ref. The tail of the AANN\_NEMO error distribution is shortened, making modelling data less suitable for novelty detection (see Figure 5c,d). A plausible explanation is that model physics is constrained to Gaussian processes and is by definition consistent, whereas observational data have errors. This hypothesis is confirmed by a comparison of measured, NEMO modelled and AANN\_NEMO and AANN\_ref errors emulated water level residuals is shown in Figure 6. (Figure 4c). Generally, the three are in good agreement with each other, but an exceptionally high residual at station Cromer around June 7<sup>th</sup> is not reproduced

330 by the NEMO model and would thus result in a large reconstruction error. Anomalous situations detected with AANN\_ref (4.,  
11.-14. January 2017) reflect themselves in large (exceeding 25 cm) AANN\_NEMO errors but not necessarily vice versa.

335 **Table 1: Overview of the number of events anomalous sea states detected by the alternative training dataset AANN and the reference model AANN\_ref different approaches during the 2018 testing phase. based on anomalous sea states identified in the reference model during the validation period.**

	training dataset	Spring (MAM)	Summer (JJA)	Autumn (SON)	Winter (DJF)
AANN_NEMO	NEMO model (14 data points)	2	2	6	13
AANN_less	tide gauge data (raw, 10 gauges)	1	0	6	13
AANN_resid	tide gauge data (de-tided, 14 gauges)	1	1	6	13
<b>AANN_ref</b>	<b>tide gauge data (raw, 14 gauges)</b>	<b>2</b>	<b>2</b>	<b>6</b>	<b>13</b>

The proposed criterion was tested on the reference model AANN\_ref for the validation 2018 testing period. Several anomalous sea states were detected (Table 1). Events in the winter months occur more than twice as often as in other seasons. Afterwards, it was verified whether a specific event could be observed with AANNs trained on the alternative datasets approaches (see Table 1). Due to the short-tailed error distribution, all events can be detected with AANN\_NEMO (but as a stand-alone method, it gives a larger number of false alarms). Events from autumn and winter are also detected with the alternative AANNs (AANN\_resid and AANN\_less), which differs for This does not apply to spring and summer, where some events are not detected with AANN\_less and/or by AANN\_resid.

345 In the following section, we will analyse two such events: The first event from June 2017 (low pressure ‘Heinrich’) was chosen since it was detected by AANN\_resid but not by AANN\_less (Figure 5 a, b and d, e-7 and, 8c, d). This atmospheric system brought some unseasonably wet and windy weather from France to the British southeast coast through the eChannel. The rain was accompanied by very strong westerly winds (~15 m/s), with gusts of 20–25 m/s around the coast of England and Wales (see the snapshot in Figure 5 7a, b and Figure 8c,d). As a result, the measured sea-level (Figure 8a) exceeded the AANN\_ref modelled sea-level over several hours along the western Dutch coast and southeastern British coast (Figure 9a). The AANN\_ref and AANN\_NEMO (Figure 9b) model errors is the are largest at Oostende, where the measured low tide is higher than the low tide expected by the AANNs (Figure 9a upper panel). The occurrence of the largest error shifts to Dover for AANN\_resid and to Den Helder for PCA (Figure 9e, f). The NEMO model error (Figure 9c) is the largest for Lowestoft Vlissingen, where it under- and overestimates the measured values for both high and low tides, respectively. (Figure 9a lower panel).  
355 Consistently, AANN\_less does not detect the event (Figure 9d), indicating its origin in anomalous spatial correlations among gauges around the small amphidrome. Thus, the specific atmospheric forcing resulted in a localized (exceptional) increase in sea level at gauge station Den Helder. The snapshot in Figure 7c of NEMO water levels visualizes this finding in Figure 5 at the time the AANN largest errors occurred, 10 hours after the wind tendency. The water level residuals at tide gauge stations confirm this result (see inclination angle of plane fitted to residuals).

360 AANN\_resid (Figure 6b) also found the event. Even though residuals are largest at DenHelder, AANN\_resid is still largest for Oostende. However, without the stations around the first amphidrome (i.e., without Dover, Oostende, Europlatform and Vlissingen) AANN\_less (Figure 6c) could not detect the event.

As a second example, a typical winter storm was chosen: the deep depression ‘Burglind’ (Figures 7 10a, b and 11c, d), which  
365 was associated with high wind speeds (up to 20 m/s) along the English and the Dutch coasts but also with large gradients in the wind field (Figure 7c) with gusts exceeding 33 m/s in England and Wales. The storm track passed over Ireland, crossed the United Kingdom and then moved over Central Europe and dissipated.

As a result, the measured tide gauge values exceeded the AANN\_ref, AANN\_NEMO and AANN\_less modelled values at stations along the southern English and Dutch coasts (Figure 12a, b, d). An exception is Lowestoft and Cromer in the  
370 ‘shadow’ of the cyclone. The NEMO model results (Figure 12c) show good agreement during high tide, especially along the Scottish coast (Figure 9b, upper panel), but tend to overestimate the low tide values, especially along the Dutch coast (Figure 7e and 9b, lower panels).

This event also manifests in the error of AANN\_less (Figure 8c) since the event affects all stations along the English coast (not only the area around the first amphidrome). AANN\_resid and PCA (Figure 8b 12e, f) also detects the event. However,  
375 here, stations along the Dutch coast have the largest reconstruction errors (i.e., stations with the highest water level residuals, see Figure 7 11b). The snapshot in Figure 10c visualizes water level residuals at the time of the largest AANN errors, which occurred with a 7 hour delay to wind tendency.

To complete the analyses, Figures 6a and 8a show the PCA reconstruction error for the two time periods. The two events can  
380 be clearly identified. However, due to the Gaussian error distribution, PCA model reconstruction with only leading modes will result in high rates of false alarm.

On either occasion, a typical time delay of 5 to 15 hours between large tendencies in wind forcing and large reconstruction errors is observed, supporting the assumption of that the atmosphere plays the a key role in the detected events. However, their  
385 occurrence cannot be directly linked to atmospheric variables, such as large tendencies in the wind forcing. These tendencies occur often but seldom result in such exceptional anomalous correlations between tide gauge measurements. sea level states (see Figures 5c and 7c), where the water surface is unusually tilted.

## 390 Conclusions

We presented a tool for detecting anomalous water levels from a network of tide gauges in the North Sea by means of using AANN. Strong spatial correlations between gauges allow for a reconstruction model in a lower dimensional subspace. Combining a threshold value for reconstruction error with the requirement of its occurrence at three different gauges for at least three hours has been shown to be a valuable filter for such events. For the two events discussed in detail, atmospheric  
395 conditions showed high wind tendencies and pressure anomalies, with the piling-up accumulation of water masses in specific areas. Thus, the resultant sea states are assumed to be related to nonlinear interactions among the various atmospheric and oceanic forcings. Further analysis is needed to develop reveal a deeper understanding of the underlying processes. Ultimately, this understanding might improve numerical sea-level predictions. The AANN reveals an intrinsic dimensionality of  $p=7$ . Evident variable candidates that could be involved in further studies include location (latitude/longitude), time, wind tendency,  
400 pressure anomaly and information on water currents.

Otherwise In addition, the tool offers an inexpensive opportunity to monitor the tidal gauge array. The difference between using raw and detided data as input to the model was marginal. Thus, no further data preprocessing is needed. The AANN model reacts sensitively on-to the choice of tide gauge locations: AANN\_less, (where stations around the first amphidrome were removed from the training data, ) was not able to identify events resulting from storms along the channel.

Considering (The linear PCA reconstruction method reveals a high rate of false positive alarms, which indicates a multimodal probability distribution of the sea-level data. alarms because the underlying physics is nonlinear. For similar reasons, the detection of some of these events might fail for AANN\_NEMO where the fundamental underlying data are consistent throughout space and time and model physics is mainly restricted to Gaussian processes. On the other hand, this finding substantiates the hypothesis, that AANN is able to detect situations under which model physics need further improvement. The proposed method can be adapted easily to any tide gauge array. However, the intrinsic dimensionality of the constructed AANN might differ, as well as the involved variables forcings and underlying processes.

415 **Competing interests.** The contact corresponding author has declared that neither they nor their coauthors have any competing interests.

Figure 1: (a) Location of tide gauge stations used in this study and M2 cotidal chart (M2 amplitudes (colour scale) and phase lines (black isolines) reproduced from Jacob and Stanev, 2017). (b-d) Time series of measured water levels, water level residuals, wind tendency and pressure anomaly for January 2017 at tide gauge positions. (f, g) S. Additionally, snapshots of NEMO model water levels for two selected timepoints. Black bars show water level residuals at tide gauge locations together with a plane fitted linearly to these residuals (grey areas). times are shown together with water level residuals at tide gauge locations (black lines) and a plane fitted linearly to these residuals (grey areas): (b) 7. January 2017, 16 UTC and (c) 11. January 2017, 20 UTC.

425

Figure 2: (b-d) Time versus position diagrams series of (a) measured water levels, (b) water level residuals, (c) wind tendency (defined as change in wind speed within one hour) and (d) pressure anomaly for January 2017 at tide gauge positions.

Figure 23: (a) Autoassociative NN with bottleneck: the input is mapped by  $p = m(x)$  onto a lower dimensional space ( $\dim(p) < \dim(x)$ ) and is approximately reconstructed by the demapping NN part  $x' = d(p) \approx x$ . (b) Error measure from the training sample versus the number of neurons in the AANN 'bottleneck'-layer using the 2016/17 North Sea water level data from 14 tide gauge stations (Fig. 1a).

430

Figure 34: Performance of the reference AANN (a-d) and of the PCA (e-h) in reconstructing sea-level data. (a-d) Scatterplots of measured vs. modelled sea level data for the during training period (a, c, e, g) and testing period (b, d, f, h) periods. Points reconstructed with an error 99.9% above of all others during training of the AANN are plotted in red in either case. Red points indicate matchups with an absolute reconstruction error  $> 50.4$  cm. This limit was derived from the 0.01% of training data with the highest reconstruction error for the reference AANN. Histograms (e-h c, d, g, h) show the respective distributions of squared relative reconstruction errors e. The numbers are the mean error and standard deviation of the reconstruction. The insets represent show the tail of large errors the distributions, which are containing possible candidates for anomalous sea states. Additionally, absolute mean errors and standard deviations are given.

440

Figure 45: (a,b,d,e) Same as Fig. 34 but for AANN\_NEMO trained on NEMO modelled sea level data. (e) Time series of water level residuals at the 14 tide gauge locations as observed (red lines), as emulated by AANN (black lines) and as modeled by NEMO (blue lines) for June 2017.

Figure 46: (a,b,d,e) Same as Fig. 3 but for AANN\_NEMO trained on NEMO modeled sea level data. (e) Time versus position diagrams series of differences between measured water levels residuals at the 14 tide gauge locations as observed (red lines), as and NEMO modelled (a) and AANN\_NEMO emulated by AANN (black lines) and as modelled (b) ones by NEMO (blue lines) for June January 2017.

445

Figure 57: Hovmoeller diagrams of measured tide gauge data (a), its residuals (b) and atmospheric forcing (ERA 5 data, wind tendency (c) and pressure anomaly (f)) for 14 stations along the inner North Sea coast for an 14 day period in June 2017. The snapshots show the wind (a) and pressure (b) fields at the time of the largest wind tendency occurrence (6. June 2017, 12 UTC). The AANN reconstruction error is given in (d), where the period within the criteria for anomalous sea state distributions are met, is highlighted. (e) shows the NEMO model error. Relatedly, snapshot (c) shows NEMO ssh at the time of the largest

450

AANN error (6. June 2017, 22 UTC) together with tide gauge residuals (bars black lines). The grey area shows the orientation of a plane fitted linearly to these residuals.

455 Figure 58: ~~Hovmöller~~ Time versus position diagrams of measured tide gauge data (a), its residuals (b) and atmospheric forcing (ERA 5 data, wind tendency (c) and pressure anomaly (f d)) for 14 stations along the inner North Sea coast for an 14 day period in during 1.-15. June 2017.

460 Figure 69: ~~Hovmöller~~ Time versus position diagrams of different model reconstruction errors compared with respect to measured (detided) and NEMO modelled tide gauge data (stations and time period are the same as those in Figure 5): for 1.-15. June 2017: (a) PCA AANN\_ref, (b) AANN\_resid (trained on water level residuals), (c) AANN\_less (without 4 stations close to the first amphidrome) and (d) reconstruction error of AANN\_nemo (trained on modelled data) compared to NEMO data. (a) AANN\_ref, (b) AANN\_NEMO (trained on modelled data), (c) NEMO model, (d) AANN\_less (without 4 stations close to the first amphidrome) (e) AANN\_resid (trained on water level residuals), and (f) using PCA, respectively.

465 Figure 7 10: Same as Figure 5 7, but for January 2018 at the time of the largest wind tendency occurrence (3. January 2018, 6 UTC) and largest AANN reconstruction error (3. January 2018, 13 UTC) during storm 'Burglind'.

Figure 7 11: Same as Figure 5 8, but for 1.-15. January 2018 during storm 'Burglind'.

Figure 8 12: Same as Figure 6 9, but for 1.-15. January 2018 during storm 'Burglind'.