Faculty of Science

Drivers of sea level variability using neural networks

Linn Carlstedt

Degree of Master of Science (120 credits)
with a major in Earth Sciences
60 hec

Department of Earth Sciences
University of Gothenburg
2023 B1218

UNIVERSITY OF GOTHENBURG

UNIVERSITY OF GOTHENBURG
Department of Earth Sciences
Geovetarcentrum/Earth Science Centre

Drivers of sea level variability using neural networks

Linn Carlstedt

ISSN 1400-3821

B1218
Master of Science (120 credits) thesis
Göteborg 2023

Mailing address
Geovetarcentrum
S 405 30 Göteborg

Address
Geovetarcentrum
Guldhedsgatan 5A

Telephone
031-786 19 56

Geovetarcentrum
Göteborg University
S-405 30 Göteborg
SWEDEN

Abstract

Understanding the forcing of regional sea level variability is crucial as many people all over the world live along the coasts and are endangered by extreme sea levels and the global sea level rise. The adding of fresh water into the oceans due to melting of the Earth's land ice together with thermosteric changes has led to a rise of the global mean sea level with an accelerating rate during the twentieth century. However, this change varies spatially and the dynamics behind what forces sea level variability on a regional to local scale are still less known, especially the high-frequency variability causing extreme sea levels. Finding a straightforward approach to understand the dynamics behind local variations is beneficial for decision makers who want to mitigate and adapt with appropriate strategies. Here I present a novel approach using machine learning to identify the dynamics and determine the most prominent drivers forcing coastal high-frequency sea level variability. I use a recurrent neural network called Long Short-Term Memory (LSTM) network, with the ability of learning data in sequences and capable of storing long memory and finding temporal dependencies in the data. As input data in the model I use hourly ERA5 10-m wind, mean sea level pressure, sea surface temperature, evaporation and precipitation data between 2009-2017 in the North Sea region. I use data from the entire North Sea basin, to be able to capture the larger climatic patterns forcing the sea level variability. The target data in the model are hourly in-situ sea level observations from West-Terschelling in the Netherlands. My results show that the dominant pressure pattern over the North Sea, coupled with a zonal wind pattern, is the most important driver of high-frequency sea level variability in my location of interest. The model also found a strong relationship between the sea level variability and another zonal wind pattern that could not be detected by classical correlation analysis, which indicates that the LSTM network has the ability to capture more complex relationships. This approach shows great potential and can easily be applied to any coastal zone and is thus very useful for a broad body of decision makers all over the world. Identifying the cause of local high-frequency sea level variability will also enable the ability of producing better models for future predictions, which is of great importance and interest.

Table of contents

1 Introduction

1.1 Background

The climate is changing. There is consensus among the scientific community that the global warming is linked to the modern human society and its increased use of fossil fuels since pre-industrial times (IPCC, 2021). Climate change is affecting the whole planet, in different ways and in different magnitudes. It is both spatially and temporally dependent, which makes the future outcomes hard to predict. Global warming has led to a shrinking of the Earth's cryosphere with ice mass losses from glaciers, ice sheets and sea ice, melting permafrost and decreased snow cover (Oppenheimer et al., 2019). Simultaneously, the warming leads to an increase in ocean heat content and thus a thermosteric change, and an intensification of extreme climate and weather events (Pörtner et al., 2022; Arias et al., 2021). These changes affect the hydrosphere and the entire climate system in multiple ways. Previous studies show that the global mean sea level (GMSL) has risen with an accelerating rate during the twentieth century (Dangendorf et al., 2019; Nerem et al., 2018; Chen et al., 2017; Dieng et al., 2017). According to IPCCs *Special Report on the Ocean and Cryosphere in a Changing Climate* (2019), part of the 6th assessment cycle, GMSL has risen 0.16 m (0.12-0.21 m) between 1902-2015 and has now reached a rate of 3.6 mm yr-1 (3.1-4.1 mm yr-1) (Oppenheimer et al., 2019). This sea level rise has globally contributed to an increased frequency of extreme sea levels since 1960 together with extreme climate and weather events such as wind-induced storm surges, tropical cyclones and extreme precipitation (Arias et al., 2021).

What has driven the sea level rise in the last decades is well known on a global scale, where thermal expansion explains 50.4%, mass loss from glaciers 22.2%, mass loss from Greenland Ice Sheet 12.6%, mass loss from Antarctic Ice Sheet 7.1%, and changes in land-water storage explains 7.7% of the global rise between 1971-2018 (Fox-Kemper et al., 2021). There are however regional deviations from the rising GMSL caused by contributing components that act locally (Slangen et al., 2014). What determines the sea level is a result of many interacting components, with different spatial and temporal scales, and depends on their local dynamics. These components are for example atmospheric forcings such as surface wind and air pressure, oceanic forcings such as temperature and salinity structure, astronomical tides and geomorphology (Oppenheimer et al., 2019; Sterlini et al., 2017; Chen et al., 2014). The

sea level variability (SLV) is the variation in sea level through time, and what causes the variation on different time scales and spatially can be studied by linking the interacting components' influence to the variability. The high-frequency variability causing extreme sea levels is driven by a combination of the relative sea level change together with tidal cycles and fluctuations in the water level due to physical forcings of e.g. waves and storm surges (Pörtner et al., 2022; Vousdoukas et al., 2017).

1.2 Motivation and aim

Understanding coastal sea level, as well as high-frequency sea level variability, is crucial as many people all over the world are living along the coasts and thus are endangered by the sea level change. Since what forces SLV varies a lot spatially, it is of great value to develop a simple stepwise approach of finding the drivers behind local high-frequency sea level variability, as it allows local decision makers to mitigate and adapt with appropriate strategies.

This Master's project is part of the FORMAS funded project *Would the Northern European Enclosure Dam really protect Sweden from sea level rise?* (NEEDS). NEEDS aims to identify the dynamics of regional sea level at multiple time-scales, by a novel approach using machine learning techniques to determine what forces sea level locally, and use that information to predict future sea levels and flood risks for the next 30 years. The NEEDS study area is Northern Europe with a main focus on Sweden.

In this project I will focus on the North Sea region and the dynamics of the climatic forcing of high-frequency sea level variability off the Dutch coast (**Figure 1.1**). I will use hourly observed sea level data from a tide-gauge station at West-Terschelling, the largest village on Terschelling, together with hourly ERA5 data, to 1) train a LSTM network to predict high-frequency sea level variability, and 2) determine the importance of the drivers forcing the sea level variability at the location. The results will then contribute to the next phase of the project NEEDS, where sea level drivers are needed to predict future sea level changes.

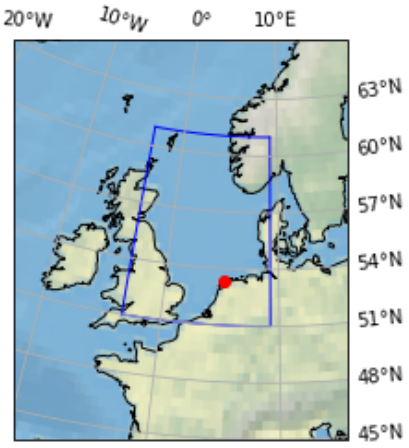

Figure 1.1 Overview map over study region. The blue area defines the studie region of the drivers forcing the sea levels at West-Terschelling (red dot).

1.3 Study area

Terschelling belongs to the Frisian Islands which stretch from the northwest of the Netherlands up to Denmark and together create an archipelago that encloses the Wadden Sea from the North Sea. South of Terschelling is Amsterdam, which has been protected from flooding since 1932 when the Afsluitdijk dam closed the inlet from the North Sea (Vogt, 2019). The Dutch society has historically faced many water-related challenges, one of the worst being the impacts of the North Sea Flood in 1953, causing severe flooding of the delta south of Rotterdam. It was caused by a storm surge, associated with a high spring tide and a European windstorm (extratropical cyclone), and in the Netherlands almost 2000 lives were lost (Hall, 2013). This tragedy led to the development of an extensive flood protection program, the Delta Plan, which today is one of the largest water management projects in the world (VanKoningsveld et al., 2008). In the future the Netherlands, with half of the population actually living below the sea level, will have to face challenges that are not yet clear, with estimates of future sea level rise reaching 30-105 cm by the end of the century, being a threat to both flooding and fresh water supply (Kwadijk et al., 2010).

Looking at the wider North Sea region, previous studies have shown that what forces SLV differs spatially (Sterlini et al., 2016; Chen et al., 2014; Dangendorf et al., 2014). On shorter (few years) time scales the local barometric pressure has been found to be the most important influence at the British and Norwegian coastlines in the northwest (Dangendorf et al., 2014), while the sea level along the southeast coastlines, including the Netherlands, have strong

evidence for zonal wind forcing being the most important factor, with strong eastward winds being correlated to higher sea levels (Sterlini et al., 2016; Chen et al., 2014; Dangendorf et al., 2013). Sterlini et al. (2016) examined the spatial pattern of drivers forcing daily SLV in the North Sea and used a multiple linear regression model to reconstruct monthly SLV at an off-shore area north of Terschelling. They showed that the southeast coastline of the North Sea, including Denmark's west coast, the German, Dutch, and Belgian north coasts, and the English channel, have faced the highest SLV in the last decades, and that the regional differences and coherence in what forces SLV depend on several things, such as the basin shape, bathymetry and coastal formation (Sterlini et al., 2016). The spatial pattern of coherence in the North Sea sea levels in relation to climatic variability has been shown earlier by Plag & Tsimplis (1999), who identified a distinct pattern between a maritime region and a continental region and found strong correlations to atmospheric forcings.

Teleconnection patterns of the North Atlantic Oscillation (NAO) are known to impact vast land- and ocean masses in the mid-high Northern latitudes and alter both atmospheric and oceanic modes such as: wind direction, wind speed, heat and moisture transport, storm strength and frequency, sea surface temperature and oceanic circulation (Chen et al., 2014; Gupta & McNeil, 2012). During a positive NAO phase there is a steep pressure gradient between the subpolar low-pressure system centered over Iceland and the subtropical high-pressure system over the Azores. The subtropical anticyclones, Subtropical Highs, are semi-permanent high pressure systems that connect the mid-latitude westerly and tropical easterly circulation regimes (Fahad et al., 2020). During a positive NAO phase, the westerlies, which are driven by the subtropical anticyclone, enhance with increased moisture fluxes towards the north and east onto northern Europe and Scandinavia, while the mid-latitude cyclones shift northward (Hurrell, 1995). In north Europe this positive mode is associated with stormy and wet weather with warmer-than-average temperatures during winter (Ibebuchi, 2022; Plag & Tsimplis, 1999). In particular during the winter months December, January and February (DJF), the climate in northern Europe is mainly driven by the effects of the NAO system and thus has a strong linkage to winter SLV in the North Sea, especially in the southeast region (Dangendorf 2013; Chen et al., 2014). The pressure mode of the NAO thus impacts the climate in the North Sea region by steering the pressure-induced winds and circulation patterns of water fluxes.

Looking into the future, climate projections are expected to cause an increase in the frequency and severity of coastal flooding in low-lying areas, due to the continuing sea level rise, compounded with extreme events such as storm surges, heavy rain and tropical cyclones, which also are expected to increase in intensity and frequency (Pörtner et al., 2022; Arias et al., 2021). Vousdoukas et al. (2017) estimated the future projections of extreme sea levels and flood risks along the European coast, and found that the North Sea region will meet the highest increase in extreme sea levels by the end of the century under RCP8.5, attributed to the change in mean sea level, astronomical tides and episodic fluctuations in the water level caused by waves and storm surges. Under the worst RCP8.5 scenario 5 million Europeans could be at risk from annual flooding, with impacts such as morphological erosion, failure of coastal protection and overwash (Vousdoukas et al., 2017; Larson et al., 2009). While the effect of tidal variability, vertical crustal movements and to some extent the seasonal variability on SLV are rather well known (Plag & Tsimplis 1999), the additional dynamics of what forces high-frequency sea level change on a regional to local scale are less understood. Indeed, regional sea level variability has been studied prior to this thesis, and in particular in the North Sea as it provides long sea level observation records (Iglesias et al., 2017; Chen et al., 2014; Dangendorf et al., 2014; Dangendorf et al., 2013; Plag & Tsimplis, 1999). However, all above referred to studies are using monthly resolution, which do not capture the high-frequency variability related to severe impact on human societies by floodings, whereas others use altimeter data which are known to have difficulties in representing coastal zones due data degradation (Iglesias et al., 2017; Sterlini et al., 2016).

1.4 Machine learning

The implementation of machine learning is now seen in a vast area of both high-tech product development and sciences, where computers can learn from data without being explicitly programmed for the task. They use algorithms to recognize patterns and relationships in the data to be able to make predictions on unseen data (Géron, 2017). The technique of Deep Learning was introduced by Hinton et al. (2006) when they trained a deep neural network that could recognize handwritten digits with very high accuracy. Deep learning is a subset of machine learning, where the algorithms learning the data have a structure of the human brain, called artificial neural networks (ANNs). Feed forward ANNs are particularly useful on static tabular data, while for other types of domains more sophisticated network architectures have been developed (Géron, 2017). In the domain of image- and speech recognition,

Convolutional Neural Networks (CNN) are mainly used (LeCun & Bengio, 1995), while the recent advances in Natural Language Processing (NLP), e.g., with the chatbot GPT-3 (Brown et al., 2020) and its alikes, are based on the Transformer architecture (Vaswani et al., 2017). For sequential data tasks the Recurrent Neural Network (RNN) architecture is most common (Karpathy et al., 2015). One of the most popular types of RNN is the Long Short-Term Memory (LSTM) network that has the ability to learn long-term dependencies, which is why it is a powerful tool in various time series predictions (Géron, 2017).

Previous studies using Machine Learning in the domain of climate research have shown good potential in predicting sea levels (Balogun et al., 2021, Nieves et al., 2021; Ishida et al., 2020; Winona & Adytia, 2020; Roshni et al., 2019; Hieronymus et al., 2019). Hieronymus et al. (2019) investigated the performance of three different machine learning models in comparison with a linear regression model to predict hourly sea levels along the Swedish coast. They demonstrate that neural networks work excellent in predicting sea levels when local and remote sea level data are used together with meteorological drivers, compared with the more conventional use of linear regression models (Hieronymus et al., 2019). LSTM models have shown to perform particularly well, using both oceanic- and atmospheric climatic drivers as predictors to local sea level (Balogun et al., 2021; Nieves et al., 2021; Ishida et al., 2020). Balogun et al. (2021) compared three different neural networks to predict monthly SLV in Malaysia, using both atmospheric and oceanic altimeter data as input to the model. They found that the atmospheric variables had the greatest influence on the model prediction and that the LSTM model got the highest accuracy. Ishida et al. (2020) used a LSTM network to predict the sea levels at Osaka tide-gauge station, using ERA5 wind, sea level pressure and air temperature data together with the relative position of the Sun and Moon (as a proxy for the tidal effect). Their model could reconstruct the tidal effect with high accuracy, whereas it had some problem with underestimating the sea level during strong winds and low pressure.

2 Data processing

Here I present the data used in the model and how they are processed, where section **2.1** describes the processing of the target data, i.e. the observed sea levels, including tidal-removal (**2.1.1**), seasonal adjustment and detrending (**2.1.2**). Section **2.2** outlines the process of the data used as input in the model, i.e. the climate data, including seasonal adjustment and detrending (**2.2.1**), dimensionality reduction (**2.2.2**) and feature scaling (**2.2.3**). All processing is done using Python (Van Rossum & Drake, 2009) in the integrated development environment (IDE) Jupyter Notebook (Kluyver et al., 2016).

2.1 Sea level data

With the aim of finding the most important drivers of SSV at my location of interest I use sea level observations from the station at West-Terschelling as the target data in my model. The data are downloaded from the Global Extreme Sea Level Analysis (GESLA) project, which provides high-frequency tide gauge observations all over the globe (Haigh et al., 2022). The first GESLA dataset was established in 2009 and contained 675 records spanning over 21 years, while today the 3d edition GESLA-3 consists of 5 119 records covering 90 years (Haigh et al., 2022). It is a widely used dataset as it is unique in its spatial and temporal coverage. In this thesis I use tide gauge data from 2009-2017 that originally have a very high frequency of 10 min. Using in-situ observations compared to altimeter data enhances the ability of making good predictions in coastal zones as altimeter data has a tendency to degrade along the coasts. Since the input data of the climatic drivers I use have hourly temporal resolution, I first calculate hourly means of the GESLA data which then consists of a sea level record with 78 888 data points (time steps).

2.1.1 Tidal removal

The hourly tide gauge data captures both high-frequency variability, e.g. the tidal signal and the daily cycle, and low-frequency variability, e.g. the seasonal variations and the trend. Looking at the topical region, the Dutch north coast together with the adjacent West Frisian Islands had a mean tidal range of 2.33 m between 1958-2014 (Jänicke et al., 2021), and examine the record of West-Terschelling it measures a typical tidal range of approximately 2 m (**Figure 2.1**), while the maximum range in the dataset is 4.81 m. The tidal signal, which in many cases is the most pronounced variability of the signal, is very clear in this dataset and

needs to be removed to get an insight into the physical information of the non-tidal variability of the data. Data cleaning, as it is usually referred to in machine learning modeling, is also important for the neural network algorithms which perform better if the data has been prepared properly according to the objective, which in this case is not finding tidal signals. To remove the tidal signal I use a Python re-implementation of the MatLab package UTide (Codiga, 2011). UTide uses harmonic analysis to detect the tidal and sub-tidal energy constituents of the sea level variability. It finds the most significant constituents and calculates their frequencies, amplitude and Greenwich phase. The significance level is determined by the signal-to-noise ratio, and constituents are considered significant with a default number of 2. Complementary, the function ranks the constituents by their percent energy to get the relative importance regardless of amplitude size (Codiga, 2011). The reconstruction of the tidal-signal is then subtracted from the original data. Residual data still show some sub-tidal signals left as seen in **Figure 2.1**, but clearly the most significant ones are gone.

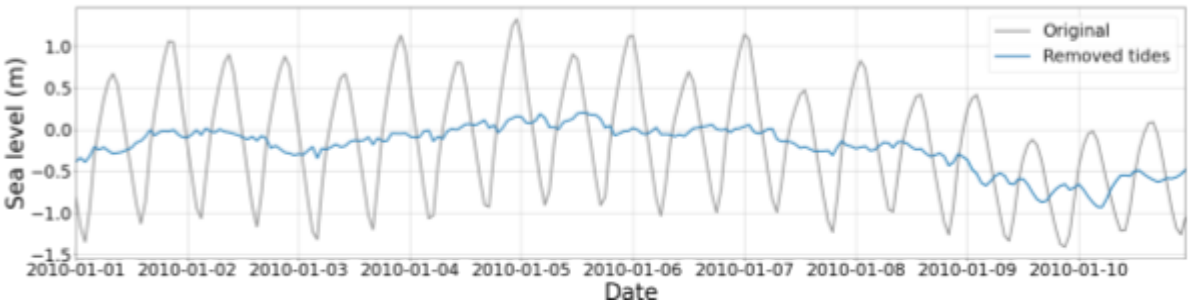

Figure 2.1 A subset of the original tide gauge data from West-Terschelling in comparison with the residual data from the tidal removal.

2.1.2 Seasonal adjustment and detrending

Furthermore, there are signals with lower frequencies found that are intended to be removed: the seasonal variation and the trend. To correct for seasonality, I calculate and subtract the climatology over the whole time period (9 years) based on the hourly averages. As my dataset includes two leap years in 2012 and 2016, I use data from the 1st of March in all non-leap years together with the data from February the 29th in the leap years to calculate the averages of those hours. Computing averages of only two values would not represent the total time period as good, and this is a way of solving that problem and still be able to preserve all the

data. After the seasonal adjustment I detrend the data using a tool from SciPy (Virtanen et al., 2020). The final preprocessed data of SLV (**Figure 2.2**) are further on referred to as the target data for the approaching machine learning modeling.

Figure 2.2 The data process of seasonal adjustment and detrending which results in the final data used in the model.

2.2 ERA5 data

As input data, or predictors in the model, I use ERA5 data as potential drivers of SLV at the location of Terschelling. ERA5 is the European Centre for Medium-Range Weather Forecasts (ECMWF) fifth generation of reanalysis, which is an assimilation of modeled and observed data (Hersbach et al., 2023). In this study I use a lat-lon grid of 0.25 degrees of hourly data between 2009-2017 in the North Sea region, 3.25E-9.25W, 51N-61N (**Figure 1.1**). The potential drivers I use in this study are listed in **Table 2.1.** The choice of potential drivers are based upon previous studies mentioned in section **1.3**, where the atmospheric parameters of wind (u10 and v10) and barometric pressure (MSLP) have been found to be of most importance in forcing SLV in this area (Sterlini et al., 2016; Dangendorf et al., 2014). Even though the steric contribution has been found not to be that important for regional SLV in this area (Dangendorf et al., 2014; Plag & Tsimplis, 1999), I include sea surface temperature (SST) as proxy for the thermal expansion. Precipitation (TP) and evaporation (eva) are added as factors of extreme weather events.

Table 2.1 ERA5 data used as potential drivers of the sea surface variability at West-Terschelling.

Driver	Description
10m u-component of wind (u10)	The horizontal speed of zonal flow towards east at 10 m above surface.
10m v-component of wind (v10)	The horizontal speed of meridional flow towards north at 10 m above surface.
Evaporation (eva)	The accumulation of evaporated and transpired water from the Earth's surface.
Mean sea level pressure (MSLP)	The pressure of the atmosphere at mean sea level.
Total precipitation (TP)	The accumulated rain and snow that reaches Earth's surface.
Sea surface temperature (SST)	The temperature of sea water near the surface.

2.2.1 Seasonal adjustment and detrending

All drivers contain some sort of seasonal and/or diurnal variability, which I reduce by calculate and subtract the climatology over the whole time period (9 years) based on the hourly averages, as I did for the tide gauge data (section **2.2**). I also detrend the data with a tool from SciPy (Virtanen et al., 2020). In addition I need to remove the NaN values corresponding to the land-laying grids in the SST dataset, prior to analysis ahead.

2.2.2 Dimensionality reduction

The ERA5 data covering the North Sea region, are 3-dimensional with 78888 time steps, 41 latitudes and 51 longitudes. To speed up the training and thus be able to have more time for tuning and hence optimize the model, I reduce the dimensionality substantially. Doing that not only saves time, but it can also improve the modeling performance since too many input features may be subject to the so-called curse of dimensionality (Géron, 2017). The idea is that with more dimensions, the bigger the space and the sparser the data. In machine learning this means that the model trained with such data generalizes worse on new data as there will be larger extrapolations with sparse data and hence the risk of overfitting increases (Géron, 2017). In theory one can solve that problem by adding more data points, but the problem is that the amount of data points required grows exponentially to the number of dimensions. That is why dimensionality reduction in many cases is necessary.

I use Principal Component Analysis (PCA), an algorithm that projects the multidimensional data onto lower-dimensional axes. It finds the axis that explains the largest amount of variance in the dataset, where the i^{th} axis is called the i^{th} principal component (PC). The $i^{th}+1$ axis is orthogonal to the i^{th} axis and accounts for the largest amount of remaining variance. Scikit-Learn provides a PCA class that uses a standard matrix factorization technique called Singular Value Decomposition (SVD) to decompose the dataset matrix into the dot product of one matrix holding the PCs (eigenvectors) and one holding the explained variances (eigenvalues) of the corresponding PC (Pedregosa et al., 2011).

The number of PCs of each driver that are selected for further analysis are chosen with the aim of describing as much variance as possible upon following criteria: maximum of 6 PCs and no explained variance lower than 0.4. Regarding MSLP there is no need for additional PCs as only the two first describes almost 90% (**Table 2.2**). These arbitrary assumptions are taken upon the context of not having too large input data but still preserving as much information as possible. Then, the selected PCs are cross-correlated to eliminate collinearity in the dataset which could have a negative influence on the modeling. If two PCs have a correlation greater than 0.40, the one explaining the most of the variance are selected and further used in the model. The remaining PCs (highlighted in green in **Table 2.2**), further also referred to as features, are then concatenated to one matrix i.e. the input data for the model.

Table 2.2 Explained variance for the chosen PCs of all drivers. In green are the final PCs used after cross-correlation.

Driver	***PC 1***	***PC 2***	***PC 3***	***PC 4***	***PC 5***	***PC 6***	**Total explained variance**
MSLP	0.58	0.31					88 %
u10	0.39	0.23	0.11	0.05			78 %
v10	0.39	0.23	0.11	0.05			78 %
TP	0.13	0.07	0.06	0.05	0.04	0.04	39 %
SST	0.35	0.19	0.10	0.04			68 %
eva	0.30	0.17	0.13	0.05			65 %

Reducing the dimensionality is also very useful for data visualization, as the high-dimensional data that covers 41x51 time series per driver gets compressed into a small number of time series, that can conveniently be plotted in 2D-graphs. I also produce Empirical Orthogonal Function (EOF) maps by fitting and transforming the PCA into the spatial variability of the PCs for further insight of the data. In this way I can display patterns that represent the entire North Sea region in just a few figures.

Note that in machine learning, dimensionality reduction is commonly performed on the final input data (Géron, 2017), which for me would be to conduct PCA upon the matrix consisting of the PCs representing the drivers. Doing that would probably improve the model performance, as mentioned above concerning the curse of dimensionality, but as my aim is to identify the importance of the drivers, it is not applicable for this task.

2.2.3 Feature scaling

One last process for the dataset is to scale the data, as machine learning algorithms perform poorly when the input numerical attributes (PCs) have different scales. I use a Scikit-Learn transformer to operate min-max scaling (often referred to as normalization) (Pedregosa et al., 2011). This scaler rescale the values to a range from 0-1 by the following equation:

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

where x is the unscaled time series, x_{min} is the minimum value of the data set and x_{max} is the maximum value of the data set. I use this scaler as neural networks generally expect the input values ranging from 0-1 (Géron, 2017). The dataset is then split into training, validation and test sets before the model operation.

Utilizing transformations, e.g. dimensionality- and seasonality reduction, on the entire dataset during data cleaning will be subject to leakage from the test data into the model, and thus bias the results, which is why it is common practice in machine learning to split the dataset before transforming the data. To generalize even better upon unseen data, the validation set should also be transformed solely as the model tuning is directed based on the results of the evaluation of the validation set. The idea is, once again, to optimize the model's ability to

generalize the best upon new data, which could be the, e.g, predicting future sea levels. But as my aim is not to make a model that generalizes the best upon new data, I choose to use the entire dataset in unison for the preprocessing, as the types of transformations I conduct include calculating averages which would be less effective with smaller amounts of data. Using climate data in machine learning models is still a rather novel approach, so there is sparse guidance in how to preprocess the data properly. In many papers it is even hard to find out how they actually dealt with such obstacles or decisions (Nieves et al., 2021; Liyew & Melese, 2021; Roshni et al., 2019; Hieronymus et al., 2019).

3 Machine learning

Here I outline the machine learning procedure, where first there is a theory part describing neural networks in general and LSTM in particular (section **3.1**), followed by an explanation on how to initialize and train a model (section **3.2**). Section **3.3** outlines the hyperparameter optimization with associated results. Finally, the procedure of predicting the sea level (section **3.4**) and finding the most important drivers using permuted feature importance (section **3.5**) are described. For the machine learning modeling I use Tensorflow (Abadi et al., 2015), through Keras interface (Chollet et al., 2015), a Machine Learning library developed by the Google Brain team. It provides various algorithms, including LSTM (Hochreiter & Schmidhuber, 1997), which I am using.

3.1 Machine learning theory

In a neural network, the artificial neurons (or cells or units) are connected to each other where each neuron has one or more inputs, and one output. Looking at a basic feed forward ANN, consisting of only one neuron cell, each input is associated with a weight (**Figure 3.1**). The weighted sum is computed by multiplying each input to the corresponding weight and added together. Then an activation function is applied to the sum, which determines the magnitude of the output. A deeper feed forward network consists of layers of neurons with one (passthrough) input layer, meaning there are no computations performed, followed by one or more hidden layers and one output layer. The output of the first hidden layer is thereby taken as the input to the second hidden layer and so on. The number of neurons in the output layer is dependent on the task for which the network is trained to solve. E.g. in the case of a regression problem the output layer may consist of one neuron with a linear activation function and in the case of a categorization problem the output layer may consist of a number of neurons equal to the number of categories to distinguish between.

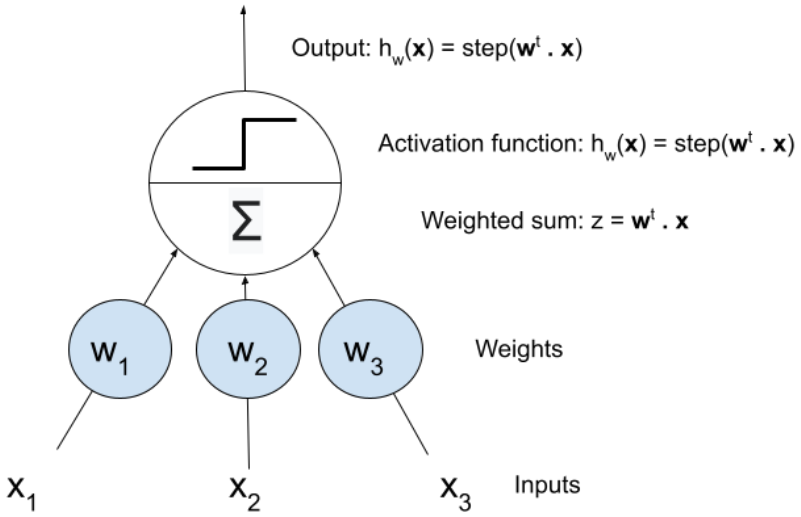

Figure 3.1 A simple feed forward neuron cell.

In comparison with the feed forward ANN, the neural network that I will use, LSTM, is a more sophisticated algorithm that is based upon the architecture of a recurrent neural network, as mentioned earlier in section **1.4**. A basic RNN is explained by the following equations

$$h_t = tanh(W_{hx}x_t + W_{hh}h_{t-1} + b_h)$$
$$y_t = W_{yh}h_t + b_y$$

where h_t is an internal state corresponding to the t^{th} time step of the sequence, commonly referred to as the *hidden state,* and y_t is the output vector at time step t. The computations of h_t are done using two separate weight matrices where W_{hx} is applied to the current input and W_{hh} is applied to the hidden state vector stored from the previous time step. The output vector y_t is then computed using a third set of weights W_{yh} taking the current hidden state as input. b_h and b_y are bias terms. The idea is that the hidden state is continuously updated during the processing of the input sequence and that each output is not only based on the current time step's input but also on information from previous time steps stored in the hidden state (Karpathy et al., 2015). Deep RNN networks are formed by stacking RNN units taking one layer's unit outputs as the next layer's unit inputs.

Since the hidden state in a basic RNN is overwritten at each time step of the input sequence the information from early time steps in the sequence gets lost if the sequences are too long. This can be crucial as the first time step *t* in a sequence can influence time step *t+n*, so one wants an architecture that can preserve a longer memory, which is true for the climate system, where previous climatic states affect the current state. For this reason I use a Long Short-Term Memory (LSTM) network, first presented in 1997 by Hochreiter and Schmidhuber. Compared to a basic RNN cell, the LSTM cell builds upon the same idea to store information from previous time steps internally, but it is more sophisticated. Parts of the computations done in a LSTM cell are exactly the same as for the basic RNNs but the flow of information is regulated by a set of *gates* and the memory from previous times steps is managed differently. The equations for LSTM computations are stated below

$$i_t = \sigma(W_{ix}\, x_t + W_{ih} h_{t-1} + b_i)$$

$$f_t = \sigma(W_{fx}\, x_t + W_{fh} h_{t-1} + b_f)$$

$$o_t = \sigma(W_{ox}\, x_t + W_{oh} h_{t-1} + b_o)$$

$$g_t = tanh(W_{gx}\, x_t + W_{gh} h_{t-1} + b_g)$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes g_t$$

$$y_t = h_t = o_t \otimes tanh(c_t)$$

where *i, f* and *o* denotes the *input gate*, *forget gate* and *output gate* respectively, g is the output of the main layer (same as RNN hidden state), c denotes the *cell state* and y is the output which is also taken as the LSTM hidden state h (Géron, 2017). Different weight matrices are used for the different computations and σ is the sigmoid activation function. Referring to the above statement about the computations being the same as for the RNNs we see that the equation of g is equivalent to the equation of h for RNN. But instead of setting g as the hidden state directly, it is regulated by the input gate and then added to the previous cell state regulated by the forget gate and the sum of these is set to be the updated cell state. The output y (and hidden state h) is then computed by regulating the cell state by the output gate. Conceptually the LSTM unit holds two inner states, the hidden state as for the RNN unit and the cell state. The forget gate and input gate regulate what information to remove from and add to the cell state, respectively. The output gate then regulates what information from

the cell state should contribute to the output. Intuitively, the hidden state corresponds to LSTM short-term memory and the cell state to long-term memory. Thus, the LSTM has the advantage of the ability of considering the temporal dependencies, which is known to be an important factor of the forcing of SLV (Sterlini et al., 2016; Dangendorf et al., 2014).

3.2 Initialize and train a model

Tom Mitchell (1997) defines learning in the context of machine learning as "a computer is said to learn from experience E with respect to some task T and performance measure P, if its performance on task T, as measured by P, improves with experience E". To put in the context of my thesis, the task of the learning is to predict the sea level given the PCs of potential climatic drivers. The experience is historical samples of the PCs paired with the observed sea level at the time of sampling. The measure of performance is a function of the difference between the prediction and the observed sea level. In machine learning this function of performance is defined as the loss or cost function (Géron, 2017).

In my experiments the LSTM weights are initialized based on random numbers drawn from a normal distribution and biases set to zero (Chollet et al., 2015). To go from a network initialized with random weights to a model that, given some input data, is able to predict the sea level it has to be trained (Goodfellow, 2016). This is done by exposing it to sample input data with associated sea level and to update its parameters to improve its predictive ability. Ideally, the number of input data samples should be large but as collection and labeling of sample data is generally costly, the number in practical applications is limited. To accommodate for the network to see large amounts of data during training, the input sample set is therefore reused where an epoch defines one complete pass through the data.

Another important aspect to take into account during training is that the real goal is usually not to predict the training data, but for the network to generalize its ability to predict other data from the current domain. To monitor the training, the set of input samples are therefore divided into three parts. The first and largest part is the training set. The model updates its parameters based on the loss value from predictions of the training set only. The frequency of the updates are defined by the batch size, which is a subdivision of the training set into smaller parts, whereas the network updates its parameters after each batch. The second part is the validation set. Predictions on the validation set are done during training but these are not a

basis for parameter (weights and biases) updates. Instead, they are used to check that the network does not learn just to memorize the training data but that it retains its generalization ability for unseen data. The third part is the test set on which no predictions are made during training. When the training is considered complete, predictions are made on the test set and these results then represent the model performance on the given task.

The most common learning algorithms for training neural networks are all based on the gradient descent. It measures the local gradient of the loss function and goes in the direction of descending gradient with the aim of converging to a minimum (Reddi et al., 2019; Géron, 2017; Kingma & Ba, 2014). In other words, the model output gets evaluated using the loss function and through backward propagation the gradients of the loss function, with respect to the weights and biases, are measured when propagated backwards through the network. The model parameters (weights and biases) are then updated to minimize the error based on the gradients. In LSTM the backpropagation goes through time as LSTMs allow sequence data as input, making the gradients a function of all the previous time steps in a sequence. The model parameters then get updated based on the summarization of the gradients. This is very beneficial for analyzing time series as it enables the network's ability to learn the temporal dependencies (Nieves et al., 2021; Ishida et al., 2020).

3.3 Hyperparameter optimization

For measuring the performance of the model, I use a loss function that measures the mean squared error (MSE) between the output value and the target value. To minimize the MSE I use an iterative optimization algorithm from Tensorflow called Adam optimization (Abadi et al., 2015; Kingma & Ba, 2014), utilizing gradient descent. After each epoch, the model predicts on the unseen validation set and measures the MSE between the predicted output value and the target value.

To optimize the model performance I create a Grid Search function, with a method for finding the optimal set of certain hyperparameters, which are the settings that must be determined outside the algorithm itself (Goodfellow, 2016). It explores all possible combinations of chosen hyperparameter values, by training each model on the training set and evaluating on the validation set. In the grid search I include the following hyperparameters; sequence length, number of layers, number of neurons, batch size and

learning rate, with a heuristic choice of values (**Table 3.3**). The sequence length is defined as the number of previous + current time steps to be included in each input to predict the current output. Number of layers refers to the number of layers with LSTM cells to be included in the model, excluding the input- and output layer. The batch size determines after how many input sequences the network should update its parameters and the learning rate specifies the size of the update step during convergence by gradient descent. The evaluation of the validation set is given in MSE in the model but presented in root mean squared error RMSE, as this can be interpreted as the mean distance in centimeters from the observed sea level. The number of epochs are determined by an early-stopping method where the model stops training if the validation loss has not improved in 10 previous epochs (Chollet et al., 2015)

In parallel I use the Keras tuner (O'Malley et al., 2019) HyperBand algorithm, which is based on an early-stopping strategy that allows the algorithm to speed up the configuration evaluation, to search for the optimal combination of hyperparameters (Li et al., 2017). Here my objective is to find the best combination of learning rate and number of units, i.e., number of neurons in each layer, as a complement to the results from the Grid Search. The results from the hyperband tuner are presented in **Table 3.4**.

The Grid Search found the best result to be a combination of a sequence of 48 hours, 2 hidden layers with 250 LSTM cells, a learning rate of 0.001 and a batch size of 10, giving a RMSE of 11.72 cm (**Table 3.3**). Other promising results are shown in **Table 3.3** with various hyperparameter combinations. Additionally, the HyperBand tuner found the best combination of number of units and learning rate to be 50 and 0.01 respectively with a RMSE of 12.80 cm (**Table 3.4**), which is not consistent with the Grid Search result. But comparing the results from the Grid Search and the HyperBand tuner, the RMSE differs by 1.08 cm, which is not much considering the different search methods and that the Grid Search function includes many more hyperparameters. The conclusion here would be that changing the hyperparameters within this range does not affect the model performance substantially. It rather emphasizes that the network performs well on the task and seems stable.

Table 3.3 Results from the Grid Search.

Hyperparameters	**Grid Search values**	**Best result**	**Other results from Grid Search**		
Sequence length	24, 48, 72	48	48	48	48
No. of layers	1, 2	2	1	1	1
No. of units	50, 250	250	250	50	50
Learning rate	0.001, 0.01, 0.1	0.001	0.001	0.001	0.01
Batch size	10, 50, 200	10	10	10	200
RMSE (cm)		**11.72**	**11.74**	**11.72**	**12.24**

Table 3.4. Results from the HyperBand tuner.

Hyperparameters	**HyperBand Objectives**	**Best result**
No. of units	50, 100, 150, 200, 250, 300	50
Learning rate	0.0001, 0.001, 0.01	0.01
RMSE (cm)		**12.80**

In addition to finding the best model based on the lowest RMSE, I need to find a model with the best balance between the time vs result aspect, as deeper neural nets take longer time to train. Finding a model with a good balance would make the final results more robust as there will be more time for training several models. To speed up the training I reduce the amount of input data to only two continuous years, as LSTMs have shown good performance even with less data (Balogun et al., 2021; Winona & Adytia, 2020). I choose two separate 2-year segments, one with a generally negative phase of NAO (years 2010-2011) and one in the positive phase of NAO (years 2016-2017), to enable further exploration of the influence of NAO to sea surface variability at my location of interest (NOAA, accessed on 8/09/22). As the amount of data decreases, the yearly split of the training- validation- and test set is no longer feasible. With only two years of data I decide to split them randomly to preserve the diversity of data in all sets. Shortening the input data to two years and using a randomly selected validation set reduces the loss by over 100% when training *Model 1* with the best hyperparameters found in the Grid Search (**Table 3.5**). This could be due to the fact that different years have different distributions regarding climatic state and therefore, using a

completely different year for validation, the model generalizes worse. As I do not feel the need of using a whole consecutive year for the test set, which in other cases could be of interest (e.g. predicting the future), it seems reasonable to proceed with this amount of data and with a random selection of datasets.

However, having the combination of a learning rate of 0.001, a batch size of 10 and 2 hidden layers with 250 neurons unfortunately makes the network way too slow even with less data (approximately 2.5 hours per model run), which is why I need to adjust the parameters additionally to find a reasonable compromise between the time vs result aspect. So, looking at the other promising results from the Grid Search actually reveals that having 2 hidden layers with 250 neurons compared to only one layer with 50 neurons does not seem to make much difference (**Table 3.3**). Also, the model run in the Grid Search with a learning rate of 0.01 and a batch size of 200 does not worsen the result substantially, but speeds up the training significantly, which is further verified by the HyperBand tuner (**Table 3.4**). Based on these findings I train *Model 2-4* with the shorter data set of two years, with a sequence of 48 hours, 1 hidden layer, 50 LSTM cells, a learning rate of 0.01 and various batch sizes (**Table 3.5**). Having a larger batch size speeds up the training significantly, but looking at the results it also increases the error. Based on that, I choose a batch size that satisfies the result vs time aspect. The final choice of hyperparameters is highlighted in green in **Table 3.5**.

Table 3.5 Results from the fine-tuning where I only use two continuous years with random data as validation set for the evaluation.

Hyper-parameters	***Model 1***	***Model 2***	***Model 3***	***Model 4***	***Model 5***
Sequence. length	48	48	48	48	48
No. of layers	2	1	1	1	1
No. of units	250	50	50	50	50
Learning rate	0.001	0.01	0.01	0.01	0.01
Batch size	10	50	100	150	200
Validation loss (MSE)	28.19	49.46	40.36	42.94	49.68
RMSE score (cm)	**5.31**	**7.03**	**6.35**	**6.55**	**7.05**

3.4 Sea level prediction

The results from the hyperparameter optimization (seen in previous section **3.3**) ended up in a final model architecture consisting of one input layer, one hidden layer with 50 LSTM-cells and one output layer (**Figure 3.4**). The input data has a 3D structure of a matrix **X** consisting of the 20 PCs (time series) as features ordered in sequences of 48 hours, where $\overline{x}_t$ is a vector with the current time step of all features and $\overline{x}_{t-n}$ is the vector consisting of the n^{th} previous time step in the sequence. Having a output layer with only one neuron gives an output $\hat{y}$ of only one value. Thus, in my model the output value $\hat{y}_t$ for time step t is modeled by feeding the network a sequence of inputs $\overline{x}_{t-47}$,... , $\overline{x}_{t-1}$, $\overline{x}_t$. This type of network has a sequence-to–vector structure, and gives me one modeled sea level value based on the history data of 20 different PCs the last 48 hours.

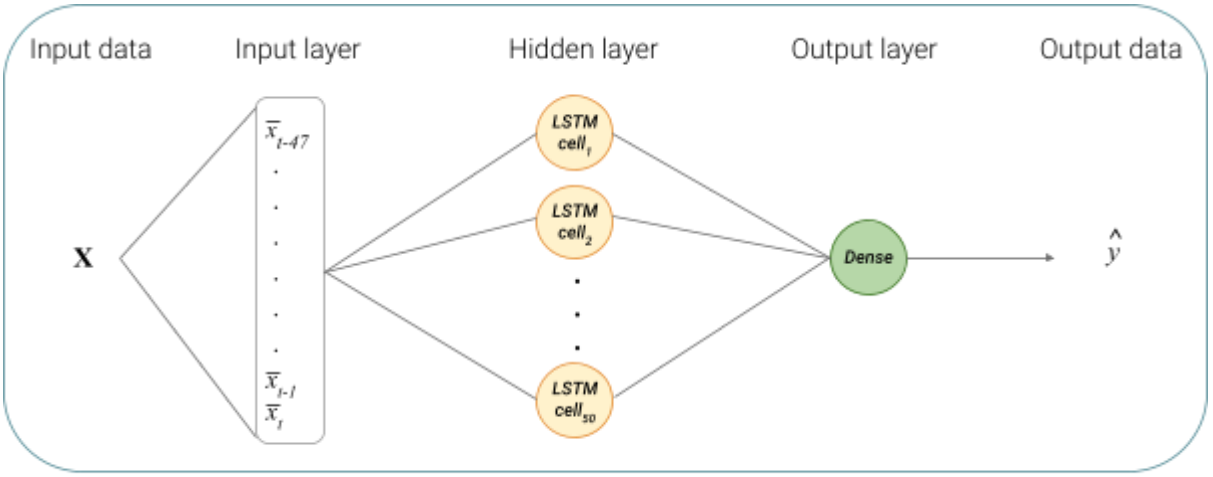

Figure 3.2 An illustration of the LSTM model architecture used.

The next step is to train several models with this architecture and predict the sea levels on the unseen test set. For training I use two 2-year segments, where one dataset consists of the data from 2010-2011 and one dataset with data from 2016-21017, as described in the previous section. Both datasets are further split into randomly selected subsets, whereas 70% of the data are assigned to the training set, and 15% are assigned to the validation and test set respectively. I then initialize 100 models per time period and train them on the training set, whereafter they are saved. The repetition of the training is carried out to be able to verify the performance of the model properly, since the network is randomly initialized and thus performs differently on each model run. The choice of how many times to train is once again based on the time vs result aspect. For this I am satisfied that the reduced amount of data to

only two years showed to perform better, as it requires less time for training, and hence gives the possibility of an increased number of training repetitions, which strengthens the robustness of the model performance result. To finally evaluate the model performance, I predict the sea level with all models on the unseen test set for each time period.

3.5 Feature importance

To determine the importance of each input feature I use a method where I permute one feature at a time and evaluate the model performance (Breiman, 2001). This method reveals the model's reliance on the permuted feature and is seen in the measured loss, where an increased loss indicates that the permuted feature is of relevance for the model prediction. To do so I repeat the procedure seen in previous section **3.4**, when predicting the sea level on the test set consisting of the 20 PCs, but this time I permute one feature (PC) at a time. The current feature is set to a vector with randomized numbers between 0-1, to keep the same scale as the processed input data. Then I evaluate the model performance on the test set by measuring the loss on all models for each excluded feature. Hence, every combination of excluding features is evaluated on 100 models for each time period. In this way I can determine the order of importance the features have on the task of predicting sea level.

Given that the results of the permutation feature importance reflect not only the model dependencies but also the physical dependencies, I examine the most influential PCs found. This is partly done by looking into the EOF maps corresponding to the PCs and trying to attribute the modes to physical phenomena. The model predictions where these PCs were permuted, and thus not influencing the prediction anymore, are also analyzed to further look into the potential influence the PCs have on the sea level. Also, previously removed PCs due to high correlations found in the cross-correlation (section **2.2.2**) will be considered in the analysis.

4 Results & Discussion

Here I present my results with the associated discussion. First I present the results from the model predictions using all data (section **4.1**), followed by the results from the permutation feature importance analysis (section **4.2**). Finally, the features found to be of highest importance in the model are attributed to physical forcings of sea level at West-Terschelling (section **4.3**).

4.1 Predicting sea level

The 100 base models trained with the final architecture exhibit a median RMSE of 7.21 cm for the time period 2010-2011 and 7.36 cm for 2016-2017, when predicting the sea levels on the test set (**Figure 4.1**). There are some model outliers present, with a maximum of 11.23 cm at time period 2016-2017, although most of the models (50%) for each time period perform within the range of 6.92-7.76 cm (demonstrated by the gray boxes in **Figure 4.1**).

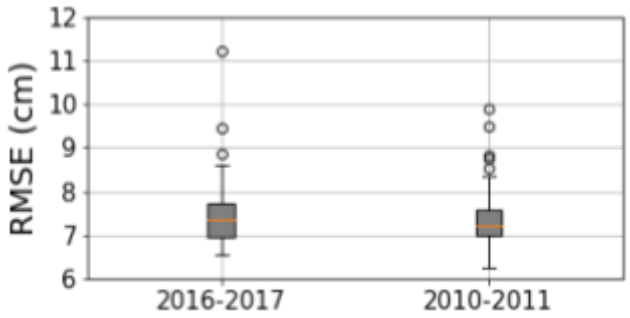

Figure 4.1 The model evaluation from 100 base model runs for each time period. The median models are highlighted in orange.

Looking at the predictions on the test set of the median models for both time periods, with a mean distance to the target of 7.21 cm and 7.36 cm respectively, the network seems to follow the data variability, even during extreme sea level events (e.g. November), with a correlation of $r = 0.91$ for both time periods. However, the model does not capture the highest-frequency variability in the SLV, which could be described as none of the features include that signal which most likely is the remaining tidal noise discussed in the method section **2.1.1**. The model prediction thus appears as a smoothed sea level, which also affects the RMSE being higher. This makes it interesting to question if this approach could operate as an effective

tide-removal function, which has been acknowledged by Ishida et al. (2020). To put this result in comparison with their study of hourly sea level predictions in Osaka using LSTM models (mentioned in the introduction section **1.3**), Ishida et al. (2020) got a mean RMSE of 9.33 cm.

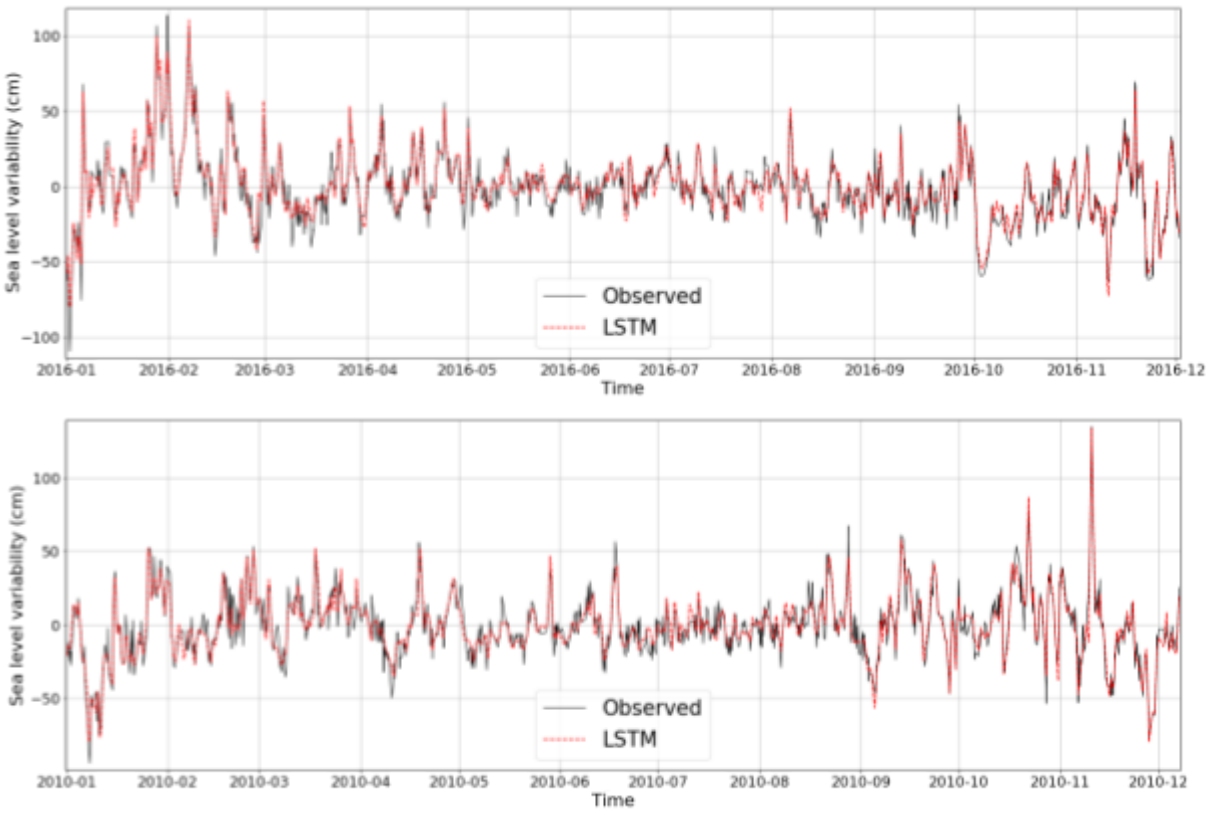

Figure 4.2 Comparison between the predicted SLV on the test set versus the observed SLV i.e. the target data for both time periods. Here subsets of the data (2010 and 2016 respectively) are represented for easier visualization.

4.2 Driver importance

4.2.1 Permutation feature importance

The results from the permutation feature importance reveal the magnitude of the different PCs' importance to the model performance, based on the RMSE from 100 model runs per excluded driver (**Figure 4.3**). It shows that PC1 of MSLP is the most important feature with a median RMSE of 18.00 cm and 20.32 for time periods 2010-2011 and 2016-2017 respectively when the feature is permuted, followed by PC1 of zonal wind component u10 with a median RMSE of 16.15 cm and 19.37 cm and PC1 of meridional wind component v10 with a median RMSE of 13.79 cm and 13.71 cm. The exclusion of other drivers did not produce such large differences, but it always increased the RMSE compared to the base model with all PCs included, with the third PC of evaporation being one of the higher among those. This indicates that all PCs somewhat contribute to the observed sea level variability. Generally low RMSEs are seen in the exclusion of the PCs of TP and SST, which indicates a rather low dependency on those features. Though, excluded PCs at an earlier stage will be considered in the following section when the forcing of SLV will be discussed more deeply.

Comparing the two time periods it appears that PC1 of both MSLP and zonal wind are of higher importance during 2016-2017, shown in a higher RMSE when those features are excluded (**Figure 4.3**). Comparing the results of the base models, the difference in performance between the two time spans is not much (**Figure 4.1**). Hence, this could be related to the differences in NAO phases where the time period 2016-2017, earmarked a positive phase, could be expected to be more strongly influenced by MSLP and wind, which will be further discussed in next section.

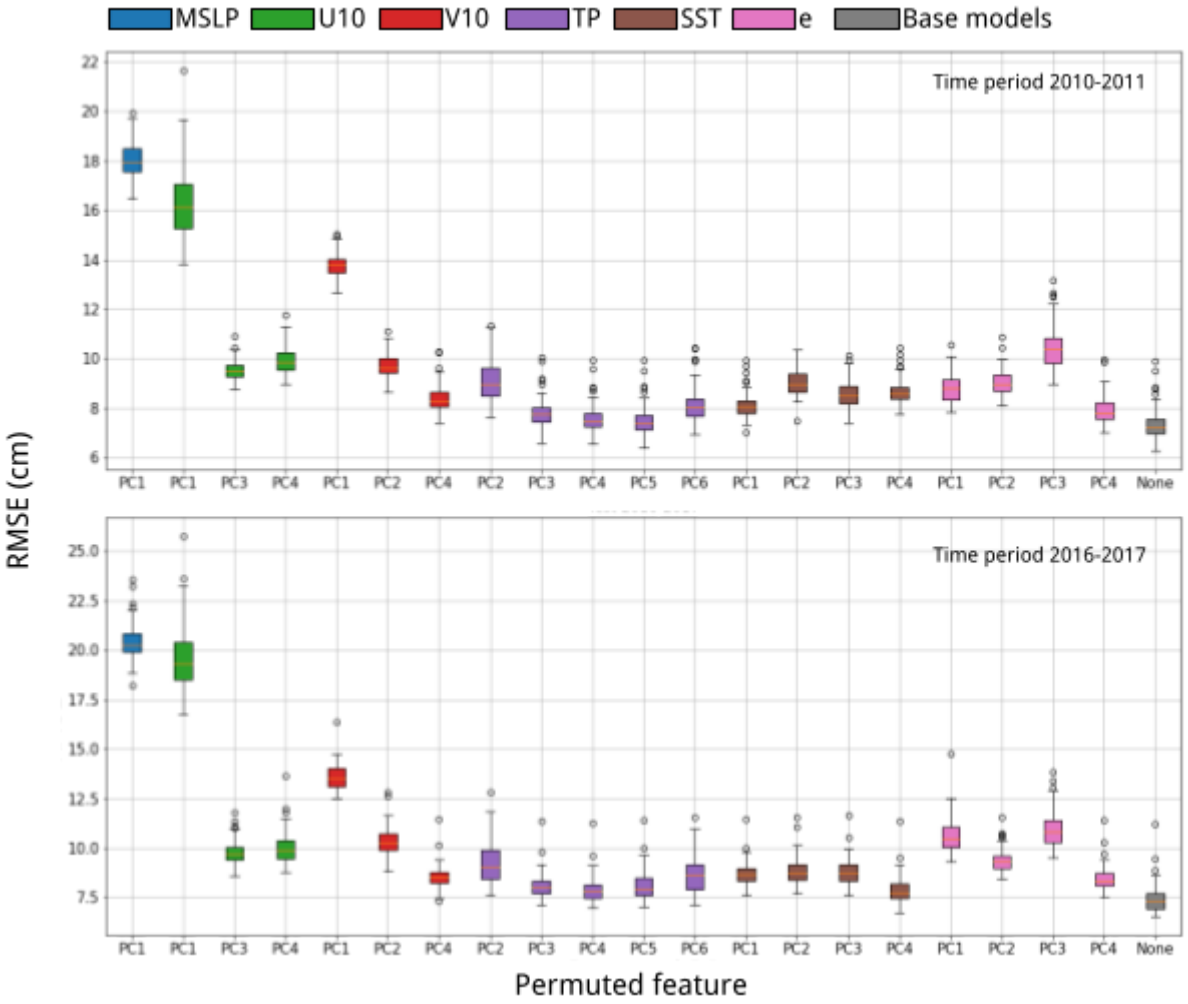

Figure 3.3 Results of the prediction on the test set of 100 different models with 21 different combinations of PCs for each time period.

4.2.2 Analyzing the drivers of sea level variability

Given that the results of the permutation feature importance reflect not only the model dependencies but also the physical dependencies, I will now analyze the three most important PCs and the associated EOF maps. To interpret the physical meaning of the relationship between the atmospheric drivers and the sea level variability, one needs to look at the corresponding PC time series. Also, a clarification of the EOF analysis should be outlined, as this is a mathematical procedure and not based on physics. The signs, described as positive in red areas and negative in blue areas are arbitrary, meaning that the EOF analysis can yield different signed EOFs, i.e. the signs can flip, for slightly different inputs. If the sign were to be flipped, also the corresponding time series would be flipped accordingly, thus making the signs unimportant in that sense (Dommenget & Latif, 2002). For interpretations though, the signs can be of importance as they can make the interpretation more or less intuitive.

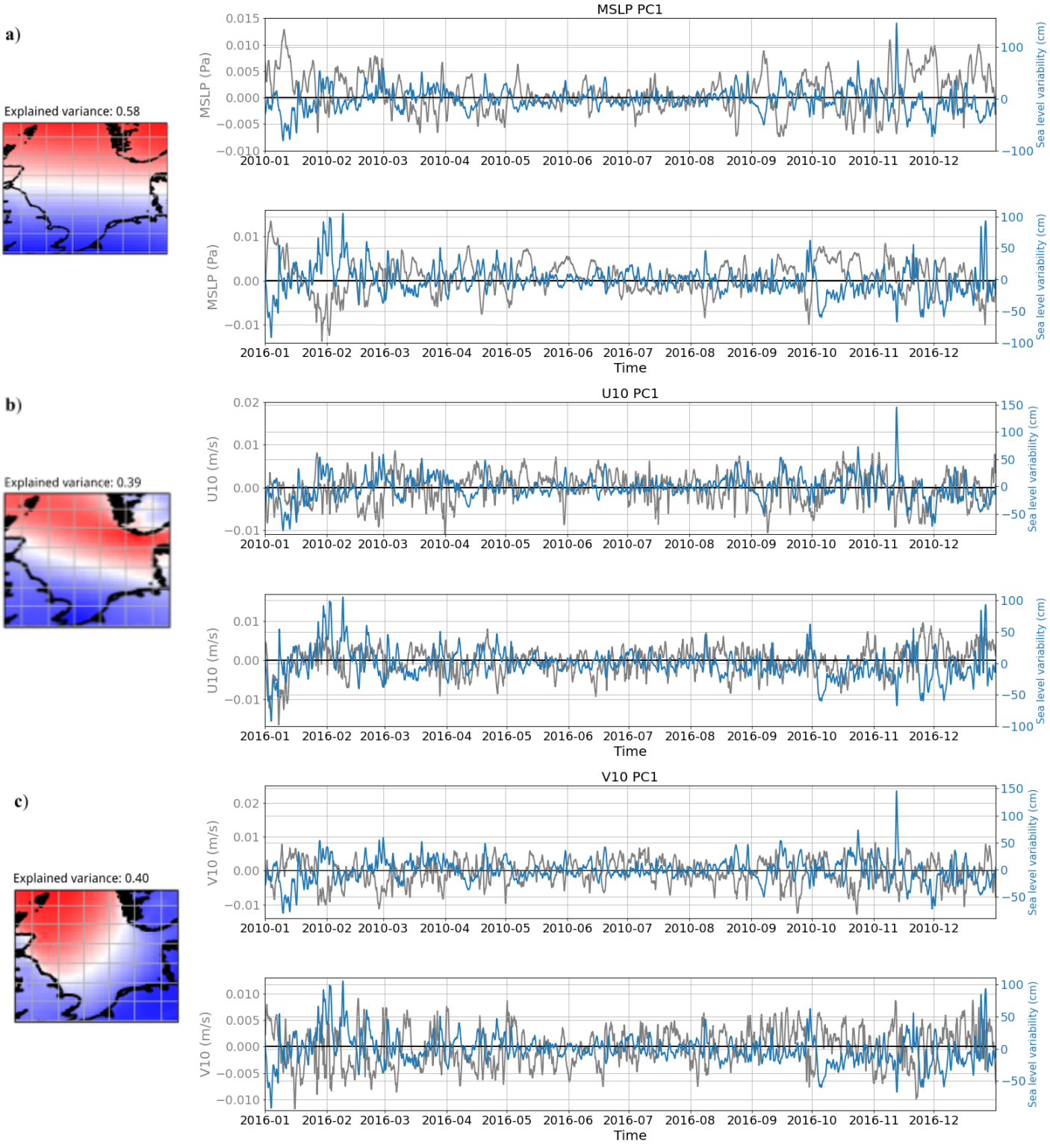

Figure 4.4 The EOF maps of the three most important drivers with corresponding PC time series, where **a)** is the first PC of MSLP explaining 58% of the variance, **b)** is the first PC of zonal wind component u10 explaining 39% of the variance and **c)** is the first PC of meridional wind component v10 explaining 40% of the variance. Here subsets of the two time periods are shown for easier visualization (2010 and 2016 respectively) together with the observed sea level variability at West-Terschelling for comparison, smoothed with a 12 hour running mean.

The mode of the first EOF and associated PC1 time series of the MSLP is found as the most important driver of SLV at West-Terschelling (**Figure 4.4 a**). It resembles the NAO, with a pronounced south-north pressure gradient. The increasing magnitude of importance of the PC, shown in the strength of the EOF sign (red and blue colors), towards both south and north could be described by the proximity to the pressure centers who are located outside my study region (Hurrell, 1995). The positive EOF sign for the MSLP, shown in red color, is associated with an increasing sea level pressure when the corresponding PC time series increases. This time series is negatively correlated with the observed SLV at West-Terschelling with an r of -0.53 for 2010-2011 and -0.46 for 2016-2017. In a physical sense that means that increasing sea level pressure in the northern part of the study region (red area) is associated with lower sea levels at West-Terschelling. Or, when the mode shifts its phase with increasing sea level pressure in the south the sea level increases at West-Terschelling. This is expected since higher sea level pressure in the south would imply a stronger south-north pressure gradient, i.e. a positive NAO phase, which is associated with higher sea levels in the North Sea, and are consistent with previous findings (Chen et al., 2014; Dangendorf 2013).

The result of MSLP being the most important driver of sea level at West-Terschelling seems at first glance to disagree with previous studies of this region, where the zonal winds have been found to be the most prominent factor (Sterlini et al., 2016; Chen et al., 2014; Dangendorf et al., 2014). According to Dangendorf et al (2014) the SLV of West-Terschelling can be described dominantly by the zonal wind with just a minor fraction of MSLP when comparing the local atmospheric forcings on sea level. Sterlini et al. (2016) neither did find a significant correlation between local barometric pressure and SLV but stated that there is a strong linkage to a remote large-scale pressure system which in turn drives the wind patterns. Consequently, this means that my results are most likely not contradictory, as there is a strong link between the pressure system examined here and zonal winds (Chen et al., 2014; Hurrell et al., 1995). Rather, it empathizes that the use of PCA in my study enables the ability of capturing large-scale pattern relationships to sea level variability.

This leads me to draw attention to the previously conducted cross-correlation, whereas the second PC of zonal wind was excluded prior to the modeling, as it had a strong positive correlation of 0.80 to PC1 of MSLP. Examining this EOF and corresponding time series, it

reveals a negative correlation of −0.60 to the SLV in 2010-2011 and −0.53 to the SLV in 2016-2017, where stronger eastward winds over the entire basin are related to higher sea levels (**Appendix 1**). This driver should therefore be taken into consideration as being an important forcing of SLV at West-Terschelling, and being strongly linked to the MSLP pattern found here, assigned the most important driver, resembling the NAO.

The second in the sequence of important drivers belongs to the first PC of the zonal wind component u10 (**Figure 4.4 b**). Looking at the EOF mode there is a pronounced contrast between the southwest and northeast part of the region, which could be explained by the prevailing westerly-southwesterly winds that dominate the North Sea region (Sušelj et al., 2010), where the British island hinders the westerly winds from approaching the southern part of the north sea, making them weaker in this area. Examining the time series of the PC compared to the sea level at West-Terschelling there is no obvious relationship, although at some times the weaker eastward winds in the northeastern part are related to increasing sea level, while at other times the sea level rises jointly with enhancing eastward winds. The inconsistency could be described by the fact that the sea level at West-Terschelling is influenced by the entire wind system of the basin, which is in line with previous research findings by Sterlini et al. (2016). That is, when there are enhanced eastward winds in the southeastern part, the sea level at West-Terschelling rises due to direct transport of water towards the coast, and contrary when the wind blows towards the west it transports water away from the coast which results in a sea level drop. Also, eastward winds blowing in the northern North Sea tend to transport water southeast due to Ekman transport, which piles up water along the coasts of the Netherlands and Germany, which also increases the sea levels at West-Terschelling (Sterlini et al., 2016; Dangendorf 2014).

The third most prominent driver according to the model results is attributed to the first PC of meridional wind component v10 (**Figure 4.4 c**). This mode of variability has a distinct pattern, separating the northwestern part of the water basin from the rest of the study area, which has been found in previous studies and related to the geomorphology, where land obstructs the wind development while the ocean allows for amplification (Plag & Tsimplis, 1999). The corresponding time series has a negative correlation of −0.50 for both time periods to the sea level at West-Terschelling, which is visible in the time series comparison. In a physical sense that could be interpreted as enhancing northward winds being related to a

lowering in sea levels, driving water masses out of the basin, while a weakening in northward winds are related to increasing sea levels by driving water masses towards the south. Consistent with Sterlini et al. (2016), the meridional wind component is not as influential as that of the zonal component. Which of the components is most important for sea level is thus strongly dependent on the coastal geometry (Sterlini et al., 2016; Sušelj et al., 2010).

With the aim of investigating the effect of NAO on the local sea level, the winter months (DJF) will be examined in detail. First, looking at the variability in **Figure 4.4 a**, both the sea level and the PC1 of MSLP display larger variations during the winter months, which is consistent with previous findings as mentioned earlier (Dangendorf et al., 2014: Sterlini et al., 2016). Examining the time periods in **Figure 4.5** reveals that the sea level anomalies are more frequently positive with higher levels in the winter of 2016-2017, corresponding to a decrease in MSLP in the north and a positive NAO phase according to indices published by National Oceanic and Atmospheric Administration (NOAA) (accessed on 8/09/22). The monthly NAO indices in 2010 are all negative (NOAA, accessed on 8/09/22), which is consistent with the increased pressure in the north seen here in the winter months, with an associated drop in sea levels. However, in the beginning of January there is rather an opposite situation of a lowering in the MSLP causing high sea levels in 2011, and an increasing MSLP in 2017 causing sea level drop.

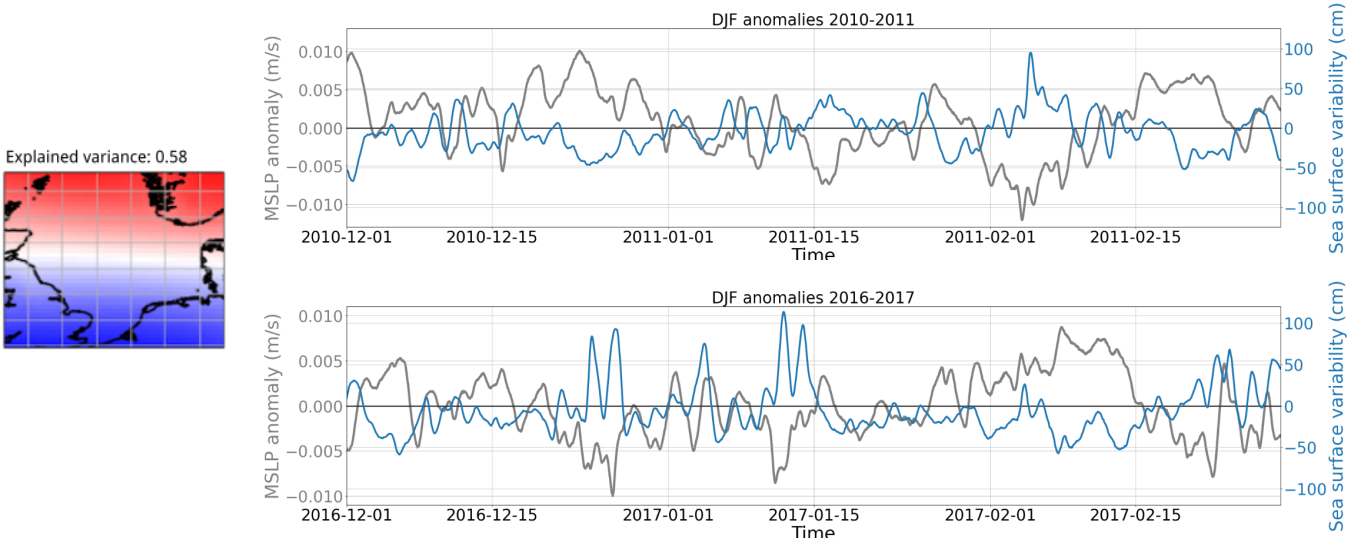

Figure 4.5 Winter (DJF) SLV in 2010-2011 and 2016-2017 in comparison with PC1 of MSLP.

The worst predictions made by the model are the ones where the most important features are permuted. To investigate the dependencies of these features related to the model performance, the predictions on the test set of these features being permuted are compared with the base prediction of all features included, together with the observed SLV (**Figure 4.6**). Looking at the top graph, the prediction without PC1 of MSLP fluctuates more, whereas the base prediction appears a lot smoother. The variability without PC1 of MSLP is more similar to that seen in the observed sea level. This could maybe be attributed to MSLP having a lower-frequency variability compared to e.g. winds, which can be examined in **Figure 4.4**. At many times the prediction without MSLP also results in higher peaks and lower throughs, resulting in a similar mean, which indicates that MSLP lowering the amplitudes and smoothing the network predictions.

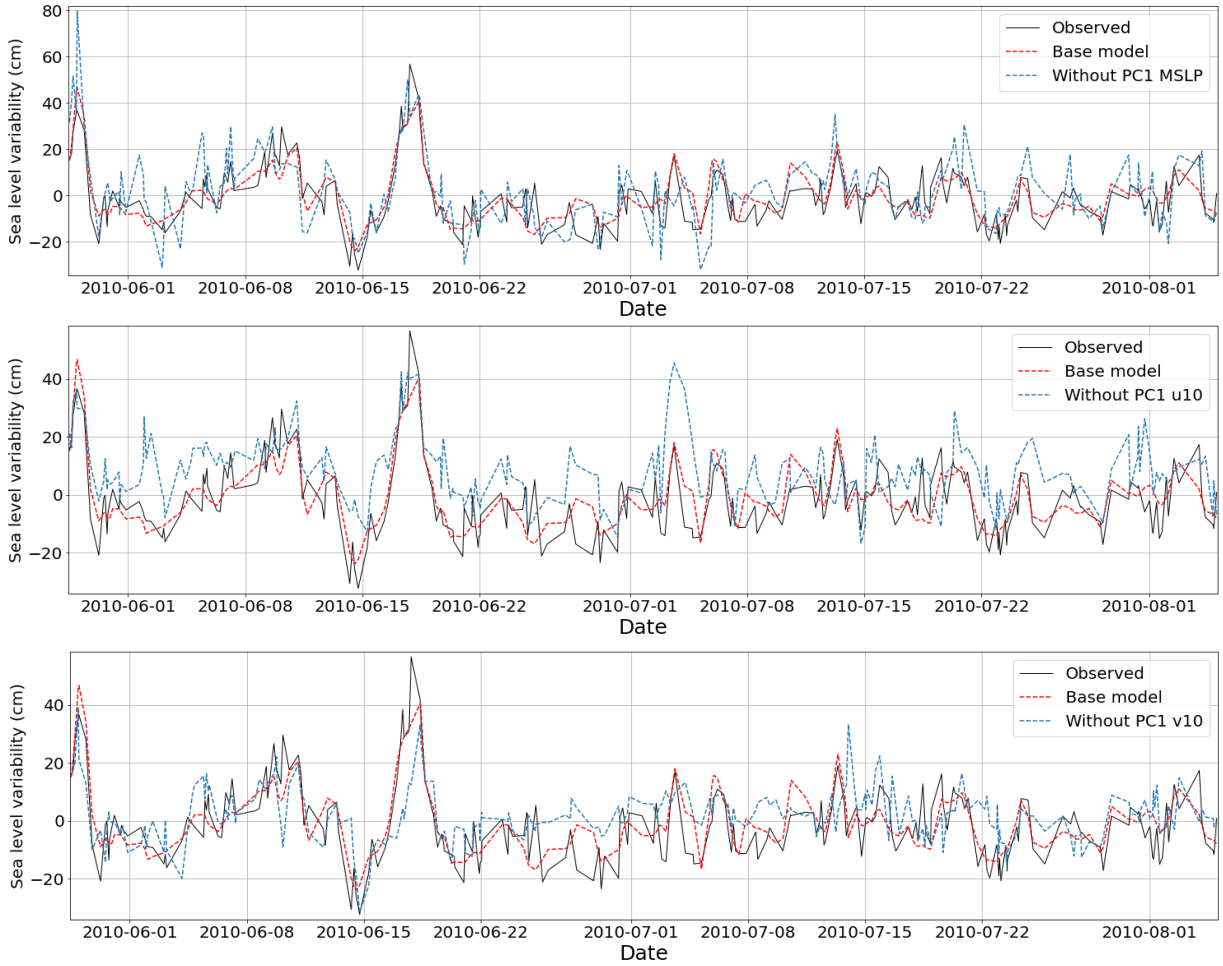

Figure 4.6 Model predictions with the most important features permuted, where at the top the PC1 of MSLP is permuted, in the middle the PC1 of u10 is permuted and in the bottom the PC of v10 is permuted. For comparison, I display the base prediction with all features included and the observed sea level variability at West-Terschelling. Here an arbitrary time period is chosen for easier visualization.

On the other hand, when PC1 of u10 is permuted (**Figure 4.6**, middle), the network clearly overestimates the sea levels, which means that the lowering influence of this feature is important for the model, at least for this time. This relationship was not found in the EOF analysis, where the correlation between the PC and sea level was insignificant. This might be due to a delay in response of wind to sea level, which can be captured in the model, where the previous 47 hours are considered to enable the analysis of a long-term relationship. Or, there is a more complex relationship that can be better captured by the network. Actually, machine learning algorithms have shown to perform particularly good upon complex tasks, sometimes even better than more classic mathematical models (Hieronymus et al., 2019; Géron, 2017). However, when trying to connect the model dependencies to real world relationships one needs to be cautious, as one disadvantage with machine learning is that you get a result but without an explanation or description of why or how the network did. You never really know how the network learned, as it operates in a hidden, so-called, black box (Géron, 2017), which is why one should be careful not to draw too strong conclusions related to the real world out of the results without investigating the results further.

The last model performance to be examined is the prediction having PC1 of v10 permuted. This variability difference to the base prediction is rather random, although there is a tendency of a delayed response in the prediction where this feature is permuted, e.g. the highest peaks in June and July are predicted later than they actually happened. This would attribute this feature as being an important factor in preventing the network from delays in the prediction, which in a physical sense could indicate that this driver, at these peaking events, preceded the other drivers.

4 Conclusion & Outlook

With the aim of using machine learning to predict high-frequency sea level variability at West-Terschelling, as well as determine what drivers are the most important, I used hourly sea level observation as target in the model, and ERA5 data of potential drivers as predictors to train the neural network upon. The use of LSTM networks has shown its potential in this task as it enables sequential data as input and thus can consider temporally dependencies. The use of hourly data enables the ability of capturing high-frequency variability forcing, which are related to extreme sea levels and flooding. My main findings are:

- The LSTM network performs well on the task of predicting local sea level variability using all potential drivers as input, with an accuracy of 7.21 cm and 7.36 cm for the time periods 2010-2011 and 2016-2017, with a correlation of 0.91.
- The most important driver of SLV at the location of West-Terschelling is attributed to the dominant MSLP pattern over the North Sea region, resembling the NAO pattern. However, the MSLP pattern is strongly linked to the second PC of zonal wind, which indicates that the pressure system influences the sea level indirectly by driving the wind system which in turn forces the water masses. This is consistent with previous work (Sterlini et al., 2015; Dangendorf et al., 2013; Plag & Tsimplis, 1999).
- The network found a strong relationship to the dominant zonal wind-pattern which could not be detected by classical correlation, which indicates that LSTM networks may perform better on finding complex relationships not captured by conventional mathematical methods.
- During a positive NAO phase, the sea level is found to be higher during winter months (DJF), presumably due to enhanced westerlies forcing water masses towards the coast of Terschelling.
- This approach can easily be done upon other targets, producing locally tailored models, which facilitates decision makers wanting to adapt with proper strategies and mitigate threats from flooding and a rising ocean.
- These results can be used for future predictions of SLV in the studied location.

The experimental dimension of machine learning allows for countless possibilities for future studies. The same is true for the exploration of climatological factors forcing sea level. Here I present some of the thoughts I have upon future development and research as well as some

comments upon my own methods. To start with, it would be interesting to also look at local drivers of sea level, by using in-situ observed climate data in an LSTM model, to compare the results with those I found when using PCs of a larger region. Local data could also serve as a complement, since some of the drivers causing flooding are of more local character (e.g. river-runoff). One should also consider different possible drivers depending on the location of interest. Based on previous findings I did not use much ocean data, but it would be interesting to explore that by training models with such data. Furthermore, there is of course room for improvement in my model and my method. For example, I would have liked to find a reasonable method that follows the machine learning practice of conducting data cleaning on test and training data separately. Overall, the data processing methods are always a subject for discussion as they will affect the model results and reliability in the end. The tidal analysis also has potential for improvement where I for example could have filtered out more signals using Fourier analysis. Although it is very time-consuming to prepare the data, tune and train the model, once it is done, the method is quite straight-forward, which gives the use of machine learning great potential. I believe in a more extended collaboration between scientific researchers and data scientists, where both parties' knowledge are needed to develop the best possible models. Or why not encourage people to learn both.

Appendix

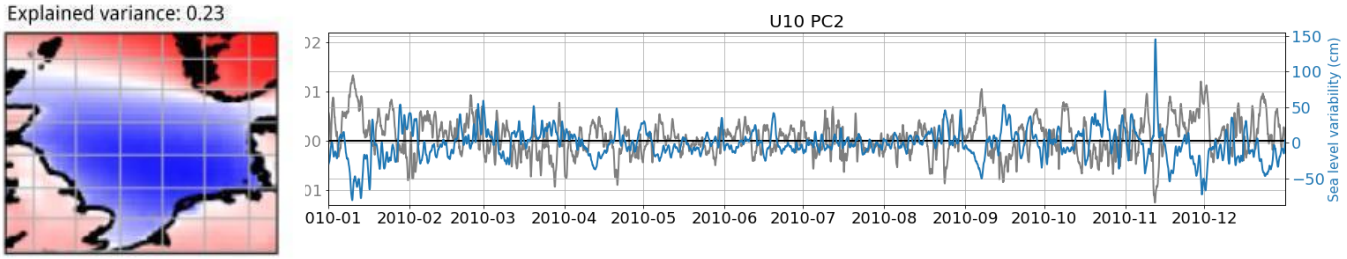

Figure 1 The EOF and PC time series of the second PC of zonal wind, compared to the sea level variability in 2010. This PC was excluded prior to the modeling due to its high correlation of 0.80 with PC1 of MSLP.

5 References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... & Zheng, X. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems, software available from tensorflow. org (2015). *URL https://www. tensorflow. org*.

Arias, P.A., N. Bellouin, E. Coppola, R.G. Jones, G. Krinner, J. Marotzke, V. Naik, M.D. Palmer, G.-K. Plattner, J. Rogelj, M. Rojas, J. Sillmann, T. Storelvmo, P.W. Thorne, B. Trewin, K. Achuta Rao, B. Adhikary, R.P. Allan, K. Armour, G. Bala, R. Barimalala, S. Berger, J.G. Canadell, C. Cassou, A. Cherchi, W. Collins, W.D. Collins, S.L. Connors, S. Corti, F. Cruz, F.J. Dentener, C. Dereczynski, A. Di Luca, A. Diongue Niang, F.J. Doblas-Reyes, A. Dosio, H. Douville, F. Engelbrecht, V. Eyring, E. Fischer, P. Forster, B. Fox-Kemper, J.S. Fuglestvedt, J.C. Fyfe, N.P. Gillett, L. Goldfarb, I. Gorodetskaya, J.M. Gutierrez, R. Hamdi, E. Hawkins, H.T. Hewitt, P. Hope, A.S. Islam, C. Jones, D.S. Kaufman, R.E. Kopp, Y. Kosaka, J. Kossin, S. Krakovska, J.-Y. Lee, J. Li, T. Mauritsen, T.K. Maycock, M. Meinshausen, S.-K. Min, P.M.S. Monteiro, T. Ngo-Duc, F. Otto, I. Pinto, A. Pirani, K. Raghavan, R. Ranasinghe, A.C. Ruane, L. Ruiz, J.-B. Sallée, B.H. Samset, S. Sathyendranath, S.I. Seneviratne, A.A. Sörensson, S. Szopa, I. Takayabu, A.-M. Tréguier, B. van den Hurk, R. Vautard, K. von Schuckmann, S. Zaehle, X. Zhang, and K. Zickfeld, 2021: Technical Summary. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 33−144. doi:10.1017/9781009157896.002.

Balogun, A. L., & Adebisi, N. (2021). Sea level prediction using ARIMA, SVR and LSTM neural network: assessing the impact of ensemble Ocean-Atmospheric processes on models' accuracy. *Geomatics, Natural Hazards and Risk*, *12*(1), 653-674.

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*, 1877-1901.

Chen, X., Dangendorf, S., Narayan, N., O'Driscoll, K., Tsimplis, M. N., Su, J., ... & Pohlmann, T. (2014). On sea level change in the North Sea influenced by the North Atlantic Oscillation: local and remote steric effects. *Estuarine, coastal and shelf science*, *151*, 186-195.

Chen, X., Zhang, X., Church, J. A., Watson, C. S., King, M. A., Monselesan, D., ... & Harig, C. (2017). The increasing rate of global mean sea-level rise during 1993–2014. *Nature Climate Change*, *7*(7), 492-495.

Chollet, F., & others. (2015). Keras. GitHub. Retrieved from https://github.com/fchollet/keras

Codiga, D.L., 2011. Unified Tidal Analysis and Prediction Using the UTide Matlab Functions. Technical Report 2011-01. Graduate School of Oceanography, University of Rhode Island, Narragansett, RI. 59pp. ftp://www.po.gso.uri.edu/pub/downloads/codiga/pubs/2011Codiga-UTide-Report.pdf

Dangendorf, S., Calafat, F. M., Arns, A., Wahl, T., Haigh, I. D., & Jensen, J. (2014). Mean sea level variability in the North Sea: Processes and implications. *Journal of Geophysical Research: Oceans*, *119*(10).

Dangendorf, S., Hay, C., Calafat, F. M., Marcos, M., Piecuch, C. G., Berk, K., & Jensen, J. (2019). Persistent acceleration in global sea-level rise since the 1960s. *Nature Climate Change*, *9*(9), 705-710.

Dangendorf, S., Mudersbach, C., Wahl, T., & Jensen, J. (2013). Characteristics of intra-, inter-annual and decadal sea-level variability and the role of meteorological forcing: the long record of Cuxhaven. *Ocean Dynamics*, *63*, 209-224.

Dieng, H. B., Cazenave, A., Meyssignac, B., & Ablain, M. (2017). New estimate of the current rate of sea level rise from a sea level budget approach. *Geophysical Research Letters*, *44*(8), 3744-3751.

Dommenget, D., & Latif, M. (2002). A cautionary note on the interpretation of EOFs. *Journal of climate*, *15*(2), 216-225.

Fahad, A., Burls, N.J. & Strasberg, Z. How will southern hemisphere subtropical anticyclones respond to global warming? Mechanisms and seasonality in CMIP5 and CMIP6 model projections. *Clim Dyn* 55, 703–718 (2020). https://doi.org/10.1007/s00382-020-05290-7

Fox-Kemper, B., H.T. Hewitt, C. Xiao, G. Aðalgeirsdóttir, S.S. Drijfhout, T.L. Edwards, N.R. Golledge, M. Hemer, R.E. Kopp, G. Krinner, A. Mix, D. Notz, S. Nowicki, I.S. Nurhati, L. Ruiz, J.-B. Sallée, A.B.A. Slangen, and Y. Yu, 2021: Ocean, Cryosphere and Sea Level Change. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1211–1362, doi:10.1017/9781009157896.011.

Géron, A. (2017). *Hands-On Machine Learning with Scikit-Learn & TensorFlow.* O´Reilly Media, Inc.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Gupta, A. S., & McNeil, B. (2012). Variability and change in the ocean. *The future of the world's climate*, 141-165.

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

Hall, A. (2013). The north sea flood of 1953. *Arcadia*, *5*, 148. Rachel Carson Center for Environment and Society. https://doi.org/10.5282/rcc/5181

Hieronymus, M., Hieronymus, J., & Hieronymus, F. (2019). On the application of machine learning techniques to regression problems in sea level studies. *Journal of Atmospheric and Oceanic Technology*, *36*(9), 1889-1902.

Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, *18*(7), 1527-1554.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thépaut, J-N. (2023): ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), DOI: 10.24381/cds.adbb2d47

Hochreiter, S., & Schmidhuber, J"urgen. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Hurrell, J. W. (1995). Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science*, *269*(5224), 676-679.

Ibebuchi, C. C. (2022). Patterns of atmospheric circulation in Western Europe linked to heavy rainfall in Germany: preliminary analysis into the 2021 heavy rainfall episode. *Theoretical and Applied Climatology*, *148*(1-2), 269-283.

Ibebuchi, C. C. (2022). Patterns of atmospheric circulation in Western Europe linked to heavy rainfall in Germany: preliminary analysis into the 2021 heavy rainfall episode. *Theoretical and Applied Climatology*, *148*(1-2), 269-283.

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.

IPCC, 2021: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*[Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press,

Cambridge, United Kingdom and New York, NY, USA, In press, doi:10.1017/9781009157896.

Ishida, K., Tsujimoto, G., Ercan, A., Tu, T., Kiyama, M., & Amagasaki, M. (2020). Hourly-scale coastal sea level modeling in a changing climate using long short-term memory neural network. *Science of the Total Environment*, *720*, 137613.

Jänicke, L., Ebener, A., Dangendorf, S., Arns, A., Schindelegger, M., Niehüser, S., ... & Jensen, J. (2021). Assessment of tidal range changes in the North Sea from 1958 to 2014. *Journal of Geophysical Research: Oceans*, *126*(1), e2020JC016456.

Karpathy, A., Johnson, J., & Fei-Fei, L. (2015). Visualizing and understanding recurrent networks. *arXiv preprint arXiv:1506.02078*.

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

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B. E., Bussonnier, M., Frederic, J., ... & Willing, C. (2016). *Jupyter Notebooks-a publishing format for reproducible computational workflows* (Vol. 2016, pp. 87-90).

Kwadijk, J. C., Haasnoot, M., Mulder, J. P., Hoogvliet, M. M., Jeuken, A. B., van der Krogt, R. A., ... & de Wit, M. J. (2010). Using adaptation tipping points to prepare for climate change and sea level rise: a case study in the Netherlands. *Wiley interdisciplinary reviews: climate change*, *1*(5), 729-740.

Larson, M., Donnelly, C., Jiménez, J. A., & Hanson, H. (2009, September). Analytical model of beach erosion and overwash during storms. In *Proceedings of the Institution of Civil Engineers-Maritime Engineering* (Vol. 162, No. 3, pp. 115-125). Thomas Telford Ltd.

LeCun, Y., & Bengio, Y. (1995). Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, *3361*(10), 1995.

Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., & Talwalkar, A. (2017). Hyperband: A novel bandit-based approach to hyperparameter optimization. *The Journal of Machine Learning Research*, *18*(1), 6765-6816.

Liyew, C. M., & Melese, H. A. (2021). Machine learning techniques to predict daily rainfall amount. *Journal of Big Data*, *8*, 1-11.

Mitchell, T. M. (1995). *Machine learning*. New York: McGraw-hill.

Nerem, R. S., Beckley, B. D., Fasullo, J. T., Hamlington, B. D., Masters, D., & Mitchum, G. T. (2018). Climate-change–driven accelerated sea-level rise detected in the altimeter era. *Proceedings of the national academy of sciences*, *115*(9), 2022-2025.

Nieves, V., Radin, C., & Camps-Valls, G. (2021). Predicting regional coastal sea level changes with machine learning. *Scientific Reports*, *11*(1), 1-6.

NOAA. North Atlantic Oscillation (NAO). National Ocean Service website, https://www.cpc.ncep.noaa.gov/products/precip/CWlink/pna/nao.shtml, accessed on 8/09/22.

O'Malley, T., Bursztein, E., Long, J., Chollet, F., Jin, H., & Invernizzi, L. (2019). Keras tuner. *Retrieved May*, *21*, 2020.

Oppenheimer, M., B.C. Glavovic , J. Hinkel, R. van de Wal, A.K. Magnan, A. Abd-Elgawad, R. Cai, M. Cifuentes-Jara, R.M. DeConto, T. Ghosh, J. Hay, F. Isla, B. Marzeion, B. Meyssignac, and Z. Sebesvari, 2019: Sea Level Rise and Implications for Low-Lying Islands, Coasts and Communities. In: *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate* [H.-O. Pörtner, D.C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama, N.M. Weyer (eds.)]. Cambridge University Press, Cambridge, UK and New York, NY, USA, pp. 321-445. https://doi.org/10.1017/9781009157964.006.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, 2825-2830.

Plag, H. P., & Tsimplis, M. N. (1999). Temporal variability of the seasonal sea-level cycle in the North Sea and Baltic Sea in relation to climate variability. *Global and Planetary Change*, *20*(2-3), 173-203.

Pörtner, H. O., Roberts, D. C., Adams, H., Adler, C., Aldunce, P., Ali, E., ... & Ibrahim, Z. Z. (2022). *Climate change 2022: Impacts, adaptation and vulnerability* (p. 3056). Geneva, Switzerland:: IPCC.

Reddi, S. J., Kale, S., & Kumar, S. (2019). On the convergence of adam and beyond. *arXiv preprint arXiv:1904.09237*.

Roshni, T., Samui, P., & Drisya, J. (2019). Operational use of machine learning models for sea-level modeling.

Slangen, A. B. A., Carson, M., Katsman, C. A., Van de Wal, R. S. W., Köhl, A., Vermeersen, L. L. A., & Stammer, D. (2014). Projecting twenty-first century regional sea-level changes. *Climatic Change*, *124*, 317-332.

Sterlini, P., de Vries, H., & Katsman, C. (2016). Sea surface height variability in the North East Atlantic from satellite altimetry. *Climate Dynamics*, *47*, 1285-1302.

Sterlini, P., Le Bars, D., de Vries, H., & Ridder, N. (2017). Understanding the spatial variation of sea level rise in the North Sea using satellite altimetry. *Journal of Geophysical Research: Oceans*, *122*(8), 6498-6511.

Sušelj, K., Sood, A., & Heinemann, D. (2010). North Sea near-surface wind climate and its relation to the large-scale circulation patterns. *Theoretical and applied climatology*, *99*, 403-419.

Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.

VanKoningsveld, M., Mulder, J. P., Stive, M. J., VanDerValk, L., & VanDerWeck, A. W. (2008). Living with sea-level rise and climate change: a case study of the Netherlands. *Journal of coastal research*, *24*(2), 367-379.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... & Van Mulbregt, P. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods*, *17*(3), 261-272.

Vogt, B. (2019). The Afsluitdijk as a Complex System. *Insight*, *22*(1), 32-36.

Vousdoukas, M. I., Mentaschi, L., Voukouvalas, E., Verlaan, M., & Feyen, L. (2017). Extreme sea levels on the rise along Europe's coasts. *Earth's Future*, *5*(3), 304-323.

Winona, A. Y., & Adytia, D. (2020, August). Short Term Forecasting of Sea Level by Using LSTM with Limited Historical Data. In *2020 International Conference on Data Science and Its Applications (ICoDSA)* (pp. 1-5). IEEE.