DELWAVE 1.0: Deep-learning surrogate model of surface wave climate in the Adriatic Basin

Peter Mlakar[1,6], Antonio Ricchi[2,3], Sandro Carniel[4], Davide Bonaldo[5,★], and Matjaž Ličer[1,7,★]

[1]Slovenian Environment Agency, Ljubljana, Slovenia
[2]University of L'Aquila, Department of Physical and Chemical Sciences (DSFC). L'Aquila, Italy
[3]CETEMPS (Center of Excellence in Telesensing of Environment and Model Prediction of Severe Events). L'Aquila, Italy
[4]NATO STO Centre for Maritime Research and Experimentation, La Spezia, Italy
[5]Institute of Marine Sciences of the National Research Council (CNR-ISMAR), Venice, Italy
[6]Faculty of Computer and Information Science, University of Ljubljana, Slovenia
[7]National Institute of Biology, Marine Biology location, Piran, Slovenia
[★]These authors contributed equally to this research.

Correspondence: Peter Mlakar (peter.mlakar@gov.si), Matjaž Ličer (matjaz.licer@gov.si)

Abstract. We propose a new point-prediction DEep Learning WAVe Emulating model (DELWAVE) which successfully emulates the behaviour of a numerical surface ocean wave model (SWAN) at a sparse set of locations, thus enabling numerically cheap large-ensemble prediction over synoptic to climate timescales. DELWAVE was trained on COSMO-CLM and SWAN input data during period 1971 - 1998, tested during 1998-2000 and cross-validated over the far-future climate timewindow of 2071-2100. It is constructed from a convolutional atmospheric encoder block, followed by a temporal collapse block and finally a regression block. DELWAVE reproduces SWAN model significant wave heights with a mean absolute error (MAE) between 5 and 10 cm, mean wave directions with a MAE of 10°-25° and mean wave period with a MAE of 0.2 s. DELWAVE is able to accurately emulate multi-modal mean wave direction distributions, related to dominant wind regimes in the basin. We use wave power analysis from linearized wave theory to explain prediction errors in the long-period limit during southeasterly conditions. We present a storm analysis of DELWAVE, employing threshold-based metrics of precision and recall to show that DELWAVE reaches a very high score (both metrics over 95%) of storm detection. SWAN and DELWAVE time series are compared against each other in the end-of-century scenario (2071-2100), and compared to the control conditions in the 1971-2000 period. Good agreement between DELWAVE and SWAN is found when considering climatological statistics, with a small ($\leq$5%), though systematic, underestimate of 99th percentile values. Compared to control climatology over all wind directions, the mismatch between DELWAVE and SWAN is generally small compared to the difference between scenario and control conditions, suggesting that the noise introduced by surrogate modeling is substantially weaker than the climate change signal.

1 Introduction

The multi-decadal characterisation of wave climate is a primary requirement for a number of applications. Coastal erosion, particularly in sandy, low-lying beaches, is largely dominated by wave-induced sediment transport at multiple time scales, with a short-term response at the seasonal, or even at the event scale, mainly given by cross-shore fluxes, and a long-term response at the annual to decadal scale resulting from the modulation of long-shore sediment fluxes and their spatial gradients (USACE, 2002). In transitional environments, wave climate can significantly affect morphodynamic processes both directly, by locally reworking morphological features such as shoals and salt marshes (Friedrichs, 2011), or indirectly, by controlling the potential sediment supply from the open coast (Di Silvio et al., 2010; Tognin et al., 2021). Wave climate is also an important factor controlling the safety and durability of human infrastructures, along the coast as well as offshore. Not least, in the framework of an ever-increasing demand for energy availability, particularly from renewable sources, information on wave climate and its variability is crucial for assessing the feasibility and improving the design of wave energy converter facilities (Astariz and Iglesias, 2015).

In recent decades, the progressively deeper understanding of the physical mechanisms underlying wave dynamics, together with an increasing availability of computational power, have contributed to making wave modelling the reference tool for a number of applications at different scales, from short-term forecasting to multi-decadal hindcasting and climate predictions (Cavaleri et al., 2007; Morim et al., 2020). Nonetheless, surface ocean wave modeling in particular is very numerically expensive. This is related to the fact that surface waves typically span deeply subgrid short spatial and temporal scales which are very far from being resolved in most ocean general circulation models. Modeling surface waves therefore typically translates into solving evolution equations of the directional wave-energy spectrum, requiring direction and frequency discretization at each model grid point, thus inflating computational demand. Furthermore, notwithstanding the continuous improvements, and particularly when dealing with long-term projections, numerical modelling maintains an intrinsic uncertainty at different levels. This impacts the very evolution of the global climate but also the propagation of the climate signal through different scales and systems, as well as the numerical description and parameterization of the processes involved. Part of this uncertainty can be addressed by means of an ensemble approach, in which multiple model descriptions are provided by considering different physical characterisations and different composition of the modelling chain (Parker, 2013). This approach comes at the cost of multiplying, usually by an order of magnitude, the requirements for computational power and data storage. This tends to limit the feasibility of extensive studies on future wave climate, particularly at the regional to local scale, and can require some heavy trade-off in terms of resolution, geographical and temporal coverage, or size of the model ensemble.

Deep learning has been shown to promise great potential to address these issues across multiple fields of science, including machine vision, natural language processing, and, more recently, in various subfields of meteorology (Janssens and Hulshoff, 2022; Beucler et al., 2021; Rasp et al., 2018) and oceanography (Rus et al., 2023; Sonnewald et al., 2021; Boehme and Rosso, 2021; Žust et al., 2021; Mallett et al., 2018). With particular reference to wave dynamics applications, James et al. (2018) proposed a machine learning system for predicting the steady-state response of the sea state in a coastal area to a given wind configuration, whereas Rodriguez-Delgado and Bergillos (2021) developed a framework for propagating onshore the open-sea information on incoming waves for renewable energy production purposes. In specific cases and for specific tasks deep learning methods have been shown to achieve state-of-the-art performance while keeping numerical costs low. This allows for performance gains which are often welcome, especially when considering computational requirements of classical geophysical numerical models on high spatial resolutions and on climate timescales.

In this paper we present a newly developed deep learning method, named DELWAVE, for emulating non-locationary modelled surface sea states, such as those produced by SWAN, albeit at a computational price smaller by several orders of magnitude, in response to given wind fields. The study site is the Adriatic Sea, a 200 km wide and 800 km long elongated epicontinental basin in the north-central Mediterranean. It is surrounded from all sides by the mountain ridges (Apennines in the west, Alps in the north and Dinaric Alps in the east) which topographically constrain winds over the basin (Figure 1). From a modelling point of view, this condition requires a high-resolution description of the atmospheric dynamics and a fine tuning of the physical parameterizations both at the air-sea and at the land-sea interfaces (Cavaleri et al., 2018). Dominant wind-wave forcings consist of the cold northeasterly Bora and warmer southeasterly Scirocco winds. Bora events are predominantly winter occurrences (November through March) of cross-basin continental air flow through the Dinaric orographic barriers over the Adriatic Sea. Scirocco is on the other hand a southerly wind transporting warm and moist air masses from northern Africa over the Adriatic, which can persist for several days and is channeled by the Apennines and Dinaric alps into an along-axis wind with a fetch much longer than in case of Bora. Wave dynamics in the Adriatic Sea are thus controlled by short-fetched wind seas and long-fetched swells, occasionally coexisting, propagating over a broad and shallow continental margin, and characterised by different multi-decadal trends (Pomaro et al., 2018) and possibly different responses to climate change (Bonaldo et al., 2020). As a typical example of some major challenges associated with wave modelling in semi-enclosed and coastal seas, the Adriatic Sea appears as a suitable testing site for wave model emulation within the DELWAVE framework.

DELWAVE is based on well established network architecture components, adapted to the field of wave forecasting, and it is benchmarked against SWAN performance, both models being forced by the COSMO-CLM atmospheric climate model of the far future climate (2071-2100) in the Adriatic basin (Bonaldo et al., 2020).

While DELWAVE model, presented in this manuscript, has been trained and tested on the outputs of COSMO-CLM and SWAN models, the model can be used with any regional atmospheric and wave modeling setup, or their ensembles, provided that available model results span a large enough time window to make DELWAVE training meaningful.

The paper is organized as follows. Classical geophysical models, COSMO-CLM for atmosphere and SWAN for surface wave modeling, are described in Section 2. DELWAVE deep network is thoroughly discussed in Section 3. Results and the far future climate simulations are presented in Section 5.

2 Numerical Models and Datasets

The wind and wave fields used as a reference for this application were retrieved from the numerical modelling chain described by Bonaldo et al. (2020) for the projection of future wave climate in a severe climate change scenario.

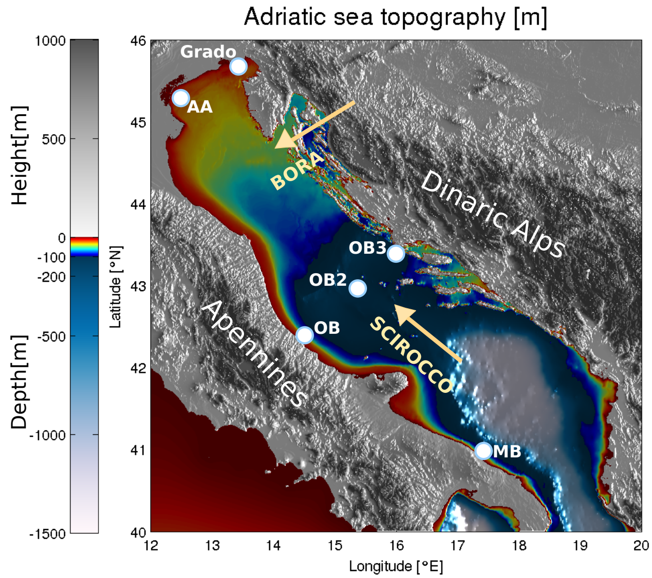

Figure 1. Topography and bathymetry of the Adriatic region. Abbreviations used on the map are as follows: AA - Acqua Alta tower, OB (2,3) - Ortona Buoy (2,3), MB - Monopoli Buoy. Directions of Bora and Scirocco are marked with beige arrows. The image was created by the authors based on EMODnet bathymetry data, available at https://portal.emodnet-bathymetry.eu/ (last access: 8 June 2022) and Copernicus European Digital Elevation Model, available at https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1-0-and-derived-products/eu-dem-v1.0 (last access: 8 June 2022).

2.1 Atmospheric Climate Model COSMO

The wind fields used for the present applications were retrieved from an implementation of the regional climate model (RCM) COSMO-CLM (Bucchignani et al., 2016), the climate version of the operational forecast model COSMO-LM (Steppeler et al., 2003), implemented over Italy and central Europe at a 0.0715° horizontal resolution (approximately 8 km, total 224 $\times$ 230 grid points), forced by the general circulation model (GCM) CMCC-CM (Scoccimarro et al., 2011). In that implementation the analysed period spanned from 1971 to 2100 reproducing first the CMIP5 historical experiment in the 1971-2005 period, and subsequently parting into two independent runs representing respectively the IPCC RCP4.5 (intermediate) and RCP8.5 (severe) scenarios. The evaluation of the model showed particularly good skills in reproducing the climatic features of air temperature and precipitation over Italy (Bucchignani et al., 2016; Zollo et al., 2016). A subsequent focus on the wind fields over the Adriatic sea (Bonaldo et al., 2017), whose reproduction is a challenging task also for hindcast and operational models due to the geometry of the basin and its complex coastal orography, highlighted outstanding skills for both intensity, although with some tendency to overestimate mean wind energy, and direction. Most interestingly for ocean modelling applications, COSMO-CLM proved capable of capturing the bimodality of Bora (north-easterly) and Sirocco (south-easterly) in the northernmost part of the basin, impossible to reproduce with previous climate models (Bellafiore et al., 2012). In a recent work (Benetazzo et al.,

2022) COSMO-CLM was also used to quantile-adjust near-surface wind speeds from ECMWF ERA5 reanalysis, thus merging the accuracy of the former with the higher temporal resolution and the synchronization with observed variability of the latter. For the wave modelling experiment described by Bonaldo et al. (2020) and for the present work the COSMO-CLM wind fields over the Adriatic Sea were retrieved for two 30-year periods in control conditions in the recent past (CTR, 1971-2000) and in the future in a severe RCP8.5 climate scenario (SCE, 2071-2100).

2.2 Wave Model SWAN

The modelling run described by Bonaldo et al. (2020) and providing wave data for this application was thus implemented in SWAN with reference to the Adriatic Sea in the CTR and SCE periods. SWAN provides a phase-averaged description of wind-generated sea states by solving a non-locationary wave action balance equation (Booij et al., 1999):

$$\frac{\partial N}{\partial t} + \frac{\partial c_x N}{\partial x} + \frac{\partial c_y N}{\partial y} + \frac{\partial c_\sigma N}{\partial \sigma} + \frac{\partial c_\theta N}{\partial \theta} = \frac{S_w}{\sigma} \tag{1}$$

N represents the action density, namely, the wave energy density divided by the relative frequency, and t is time. The propagation of N (second to fifth term in Eq. 1) is described in 2-D space (x and y, expressible in both Cartesian and spherical coordinates, with speed respectively c_x and c_y), and spectral space (radian frequency σ, relative to a frame moving with the ocean current, and angle θ normal to the wave crest, speed respectively c_σ and c_θ). S_w represents sources and sinks of wave energy density associated with generation, dissipation, and non-linear wave-wave interactions.

For the application presented here, the domain was discretised into an orthogonal curvilinear structured grid with horizontal resolution ranging from approximately 2 km in the northern region to 8-10 km in the southeasternmost part of the study area. Calm conditions were prescribed at the open boundary at the Otranto Strait, where waves generated within the basin were nonetheless permitted to radiate out of the domain. This assumption was made necessary by the lack of available wave fields consistent with the atmospheric forcing at the Mediterranean scale, but the validation confirmed that no major drawbacks in the results could be found beyond 100-200 km from the boundary. Wave spectra were discretised into 25 logarithmically-distributed frequencies ranging between 0.05 and 0.5 Hz and 36 directional sectors, whereas the timestep was set to 360 s. The bathymetry was reinterpolated from a 1-km resolution dataset used in previous applications (Benetazzo et al., 2014; Bonaldo et al., 2016) and obtained by merging recent surveys in the shallow northern basin and in the southern continental margin into previous basin-scale information. Sea level rise between the CTR and SCE period was taken into account by increasing the water depth in the latter scenario by 0.70 m, based on estimates by Antonioli et al. (2017), for the sake of simplicity uniformly distributed throughout the domain. As wind forcings from COSMO-CLM were provided with 6-hourly timestep, the same interval was maintained for the output, in which the main spectral parameters were saved for each grid point and the full spectra were saved for approximately 600 points along the Adriatic coast. The model validation was based on directional wave recordings from three observatories off the Italian coast along the main axis of the basin, namely the Acqua Alta oceanographic tower (AA,

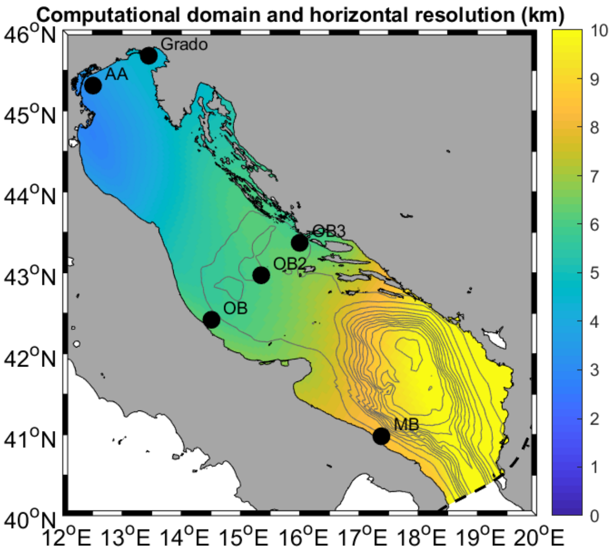

Figure 2. Geographical domain, validation locations and locations considered in SWAN and DELWAVE modeling.

12.51 °E, 45.31 °N, see Pomaro et al. 2018), and the Ortona and Monopoli buoys (respectively OB, 14.51°E, 42.42°N, and MB, 17.38°E, 40.98°N, Figure 2).

The comparison against observational data (carried out in statistical terms, as climate models are not synchronised with observed variability) was focused on significant wave height (H_S) and mean direction showed overall satisfactory performances for the SWAN implementation. The reported tendency of COSMO-CLM to overestimate mean wind energy had actually a moderate impact on wave modelling, being partially compensated by other factors such as the southern boundary conditions and some residual limitations in reproducing orographic jets, and its more marked effect was a partial overestimate of significant wave height statistics in the southernmost regions of the basin.

The end-of-century projections in a severe climate change condition outlined a composite scenario. While H_S in mean and stormy conditions appeared to decrease in most of the basin and for most directions, the effect of storms from the southern quadrant (southwest to southeast) on the northern Adriatic Sea expected to intensify. This result, interpreted as a consequence of a northbound shift of the storm tracks (Trenberth et al., 2003; Giorgi and Lionello, 2008) in the Mediterranean region, was shown to have significant implications for the coastal regions. Besides the obvious impact of stronger storms where this will happen, and besides the baseline sea level rise exacerbating the effect of storms even when their intensity are expected to decrease (Lionello et al., 2017), the spatial variability in the impact of climate change will result in a modification of the patterns of energy fluxes onto and along the Adriatic coast, thus modifying the sediment transport rates and gradients and ultimately coastal morphodynamic processes.

2.3 Training and evaluation datasets

Training and the application of DELWAVE was based on basin-wide wind fields from COSMO-CLM and pointwise wave time series at six locations exposed to different wave climates (Figure 2). AA (12.51 °E, 45.31 °N), OB (14.51 °E, 42.42 °N), and MB (17.38 °E, 40.98°N) which coincide with the observation points used for the SWAN model validation in Bonaldo et al. (2020) and are representative of nearshore conditions respectively along the northern, central and southern Italian coast. Grado (13.45 °E, 45.68 °N) lies at the edge of the gulf of Trieste in the northernmost end of the Adriatic Sea, facing south and is partially sheltered by the Istrian peninsula. OB2 (15.35 °E, 42.97 °N) and OB3 (16.00 °E, 43.37 °N) are located along an ideal transect off Ortona respectively in the middle of the basin and along the Croatian coast. Wind fields are provided as six-hourly meridional and zonal components, whereas wave data, also six-hourly, are given in terms of significant wave height (H_S), mean wave direction (d), and energy period ($T_{m-1,0}$). The model data from the control (CTR) period (1971-2000) were used for the network training, whereas future scenario (SCE, 2071-2100) data were used as a reference for assessing the network skills, particularly in terms of its capability to capture the features of the climate signal.

3 DELWAVE

In this section we present our DEep Learning WAVe Emulator (DELWAVE). DELWAVE is constructed from three logically separate parts. We proceed by providing an overview of the model input fields. Following that, we describe the DELWAVE architecture in detail and further argument specific model architecture decisions using ablation studies. Lastly, we provide a description of the training procedure.

3.1 Model input tensor

The data DELWAVE uses to conduct both training and inference is available in the form of a tensor, which contains three logically separate fields: spatial wind field, location encoding, and grid encoding. Each of this parts serves a specific purpose and we elaborate on each in the following subsections.

3.1.1 Wind field

Let's begin by first defining the input (wind) fields from which core information for surface wave prediction is extracted. Let $\mathbf{I}^t$ denote the spatial wind field over the Adriatic basin at time t. Then,

$$\mathbf{I}^t \in \mathbb{R}^{2\times n_x \times n_y}, \;\; \dim(\mathbf{I}^t) = [2, n_x, n_y], \tag{2}$$

where the first dimension corresponds to either u or v components of the wind vector, while the last two correspond to the zonal (n_x) and meridional (n_y) spatial dimensions of the modeled wind field, in our case $n_x = 90$, $n_y = 89$.

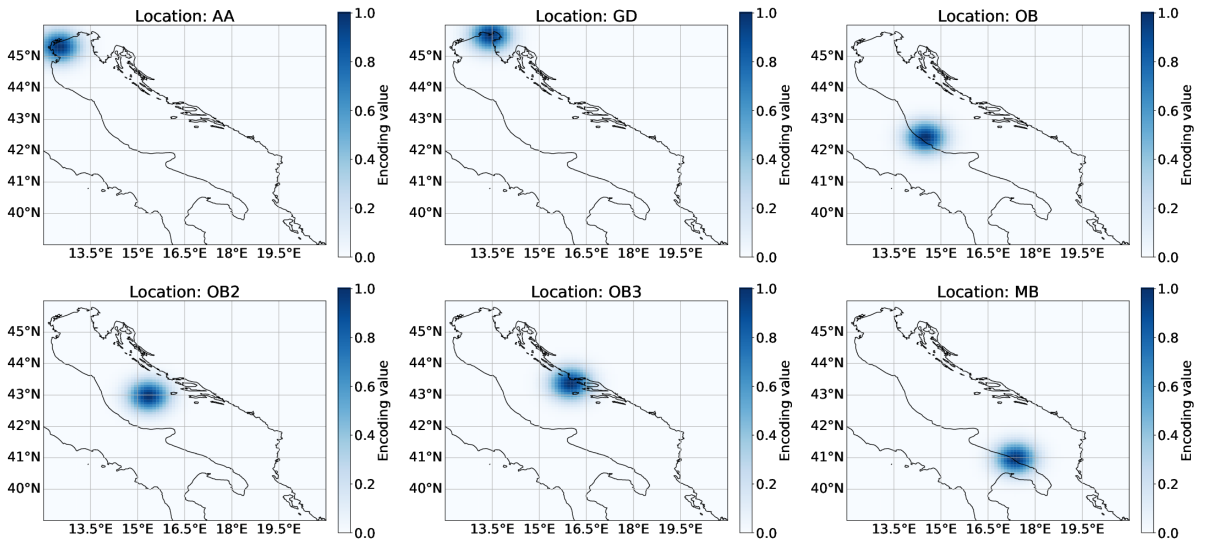

Figure 3. The visualization of the spatial encoding matrices for each location (the coastline is added for clarity). Each plot corresponds to one location encoding matrix which forms a part of the input sample tensor $\mathbf{I}_l$.

3.1.2 Location encoding and grid encoding

We further complement the wind field input tensor $\mathbf{I}^t$ by a spatial encoding matrix. The purpose of this matrix is to provide the network with information about the specific location for which we wish to predict surface wave attributes. This approach allows us to easily add new locations into the training procedure by simply defining new spatial encoding matrices, without the need for any other modifications to the algorithm or model architecture.

Let $\mathbf{L}_l$ denote the location encoding sparse matrix for location l. Then

$$\mathbf{L}_l \in \mathbb{R}^{n_x \times n_y}, \quad \dim(\mathbf{L}_l) = [n_x, n_y]. \tag{3}$$

We now denote each ith row ($i = 1, \ldots, n_x$) and jth column ($j = 1, \ldots, n_y$) entry of $\mathbf{L}_l$ as $\mathbf{L}_{l,(i,j)}$ and compute the matrix entries as

$$\mathbf{L}_{l,(i,j)} = \frac{1}{\sqrt{2\pi}\varsigma} \exp\left[-\frac{1}{2}\frac{(l_i - i + 1)^2 + (l_j - j + 1)^2}{\varsigma^2}\right], \tag{4}$$

where we set the spatial variance to $\varsigma^2 = 20$. This variance corresponds to a standard deviation of $\sqrt{20} \sim 4 - 5$ grid cells or $0.45°$ in longitude and latitude, as shown on Figure 3. We determined the value of the spatial variance empirically, by testing multiple value configurations where we finally selected the spatial variance value which produced the best results. The variables l_i and l_j denote the corresponding l location's position in the spatial field, expressed in terms of row and column indices. We illustrate examples of multiple encodings for different locations in Figure 3.

Finally, we normalize the matrix entries to the range $[0,1]$ by

$$\tilde{\mathbf{L}}_{l,(i,j)} = \frac{\mathbf{L}_{l,(i,j)} - \min(\mathbf{L}_l)}{\max(\mathbf{L}_l) - \min(\mathbf{L}_l)}. \tag{5}$$

We use this normalized location encoding matrix to augment the input wind field tensor to form the wind-location input tensor $\mathbf{I}_l^t$, where the tensor is now given for a specific location target and time. Here, the augmentation denotes the concatenation of the starting input tensor and the location encoding along the first dimension. This entails that $\dim(\mathbf{I}_l^t) = [3, n_x, n_y]$, where the increased size of the first dimension corresponds to this augmentation. To create training samples for all k locations, based on the same wind field, we use the following approach: we first randomly sample a wind field from the dataset. Then we augment the wind field with the k location matrices, where each individual augmentation produces its own $\mathbf{I}_l^t$ corresponding to a location l. This way, each training sample contains the wind field, together with a spatial encoding of a specific location. As we train the model, the training takes into account all different locations and all time steps during the same training process.

This input provides the necessary information for the model to distinguish between the different locations for which we require surface wave predictions. Without this encoding the model would most likely gravitate towards an average prediction at a specific time t for all locations, as it would not be able to distinguish between them. During DELWAVE training, we minimize the root mean squared differences loss function defined as

$$\mathcal{L} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{6}$$

where y_i denotes the SWAN values for sample i, and $\hat{y}_i$ DELWAVE's predictions. If we were to omit the location encoding from the input tensor for time t, then each location would share the same input tensor at time t (the location encoding is what differentiates input tensors for each target location), however the wave field attributes of each location are not the same. Therefore, the average prediction of all target locations is the minimizer in this case.

The final transformation of the input tensor is the concatenation of the grid encoding. A building block of DELWAVE's architecture is the convolution operation, which is, by design, translation invariant. This implies that same signal at different spatial locations produces the same output response. Since the location of specific wind patterns with relation to the target location of interest is important (wind fetch), we have to go against this inherent invariability of the convolution operation to translations. We do this using the grid encoding which assigns a unique value to each spatial location inside the input field. This enables the network to learn wind features in specific regions of the input. We denote the grid encoding matrix as $\mathbf{C} \in \mathbb{R}^{n_x \times n_y}$. Individual entries of the matrix are computed as

$$\mathbf{C}_{(i,j)} = \frac{(i-1)n_y + j - 1}{n_x n_y}, \tag{7}$$

where i is the index to the row and j to the column. We augment the wind-location tensor with the above defined grid matrix to produce the final wind-location tensor (we do not explicitly denote the grid encoding presence inside the tensor) in the same way as we did in the case of the location encoding. We end up with a tensor containing 4 input fields (zonal wind, meridional wind, location encoding, grid encoding) of dimension $[n_x, n_y]$:

$$\dim(\mathbf{I}_l^t) = [4, n_x, n_y]. \tag{8}$$

3.1.3 Temporal extent

The surface wave field at a specific location consists of the local wind sea and of the incoming swell, generated remotely in the hours preceding forecast time t. Consequently, predictions at time t require additional wind inputs from times preceding t. The number of preceding timesteps was estimated using a deep-water dispersion relation for gravity waves $\sigma^2(k) = gk$ and the corresponding gravity wave phase speed

$$c_f = \sigma/k = (g/k)^{1/2} = (g\lambda/2\pi)^{1/2}. \tag{9}$$

Using an estimate of surface wave wavelength $\lambda = 40$ meters, indicates that such waves traverse basin scale distances in about a day and a half. We consequently estimate that the wave field at a given location can be influenced by remotely generated swell over distances, traversed by swell waves in about 1 to 1.5 days, corresponding to about $10-14$ timesteps in three-hourly resolution input. We rounded this down to 10 temporally preceding timesteps.

We therefore take 10 preceding wind inputs from consecutive time instants leading up to t, namely $[\mathbf{I}_l^{t-10}, \mathbf{I}_l^{t-9}, \ldots, \mathbf{I}_l^{t-1}, \mathbf{I}_l^{t}]$. Repeating this for over 4 fields contained in the $\mathbf{I}_l^t$ tensor (zonal wind, meridional wind, location encoding, linear grid encoding), we end up with 11 timesteps of 4 fields over a spatial grid of $n_x \times n_y$ cells. Hence the dimensions of final concatenated input tensor are

$$\dim(\mathbf{I}_l) = [11, 4, n_x, n_y] = [11, 4, 90, 89]. \tag{10}$$

3.2 DELWAVE model architecture

DELWAVE is composed of three logically distinct components, each responsible for a specific processing task, as depicted in Figure 4. The atmospheric encoder is responsible for encoding the input fields for each timestep into high dimensional vectors. These vectors are then passed to the temporal collapse block where they are merged into a single vector, attenuated based on the temporal importance of the individual inputs, as explained in subsection 3.2.2 bellow. Finally, the regression block transforms this vector into the three required outputs. Individual blocks are described in more detail in the following subsections.

Additionally, let us define the notion of an *encoder*. An encoder is a sequence of transformations, a sub neural network, which maps a specific input to a, usually, high dimensional vector. This vector is said to be an *encoding* of the provided input, carrying information about it, albeit in a obtuse way. The encoder-decoder structure Cho et al. (2014), for example, is a common paradigm in machine learning leveraging this terminology.

3.2.1 Atmospheric encoder block

The atmospheric encoder block, displayed in Figure 5, is constructed from three sub-components: the per input atmospheric encoder, the joint atmospheric encoder, and the output atmospheric encoder.

Input atmospheric encoder: The input atmospheric encoder encodes each timestep individually before passing them to the joint atmospheric encoder. Each per-timestep input tensor has its own input atmospheric encoder block. This is to ensure that the initial processing of the wind field with the location encoding is unique to each timestep. The per-timestep encodings of

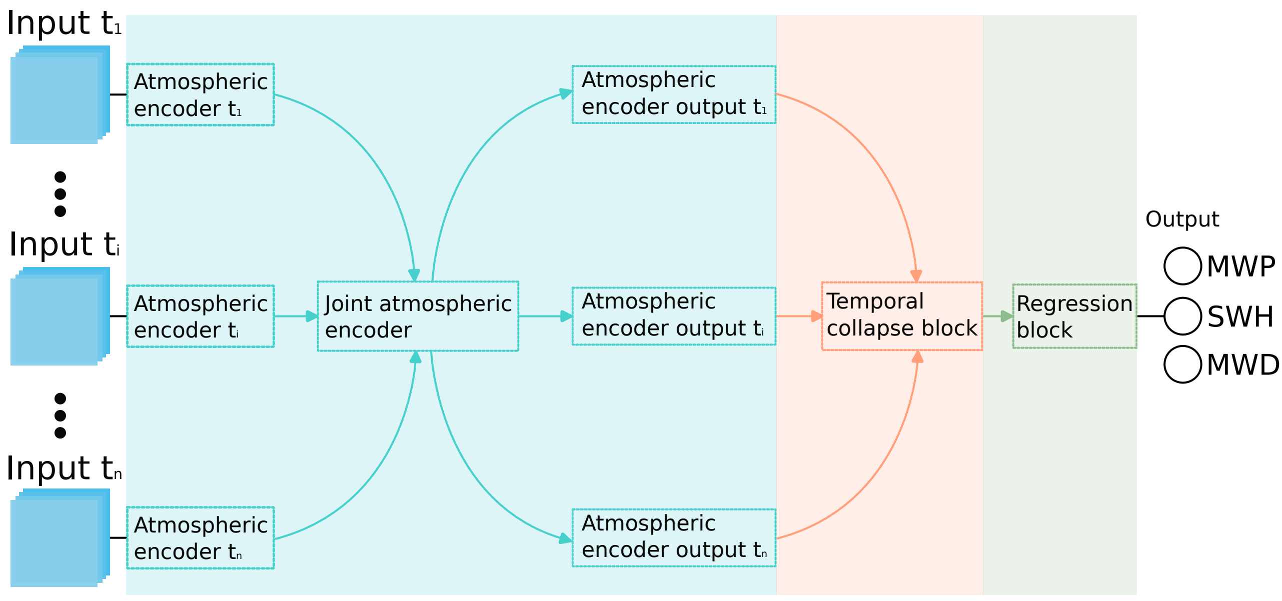

Figure 4. DELWAVE architecture overview. The network is comprised of three logically distinct sections. Each section is denoted using a different color. The information in the network flows from left to right. Inputs t_i denote the n timesteps passed to the network, while MWP, SWD, and MWD denote mean wave period, significant wave height, and mean wave direction, respectively.

spatial locations might be important for predicting wave characteristics at different timesteps, therefore per input encoders add to the flexibility of the model being able to adapt to such requirements. However, using completely separate encoders for each timestep would result in slow, hard to scale architectures with overfitting issues. Therefore, a shallow initial encoder structure for each timestep is a good compromise between the two approaches. Here, shallow denotes a architecture with only a few layers, as is denoted in Figure 5. Conversely, a deep neural network architecture constitutes of many tens of layers.

To be more formal about the atmospheric encoders, define as A_j the jth atmospheric encoder, in our case one of 11, each corresponding to one consecutive timestep. Then, DELWAVE proceeds by encoding each timestemp of the input tensor with its corresponding atmospheric encoder. this results in a set of atmospherically encoded input tensors

$$\{A_1(\mathbf{I}_{l,(1)}),\ldots,A_{11}(\mathbf{I}_{l,(11)})\}, \tag{11}$$

where $\mathbf{I}_{l,(j)}$ denotes the jth timestep from the input tensor $\mathbf{I}_l$ and $A_j(\mathbf{I}_{l,(j)})$ denotes the encoded tensor. This set is then passed to the next transformation, the joint atmospheric encoder.

Joint atmospheric encoder: The joint atmospheric encoder is the primary extractor of important wind field features as it is also the encoder with the most layers. It is shared between timesteps (we use the same joint atmospheric encoder to transform each per-timestep output of the previous block), since we care to locate important wind features independently of the time at which they occurred. For example, a specific wind pattern can occur at different timesteps. Therefore, we can use the same

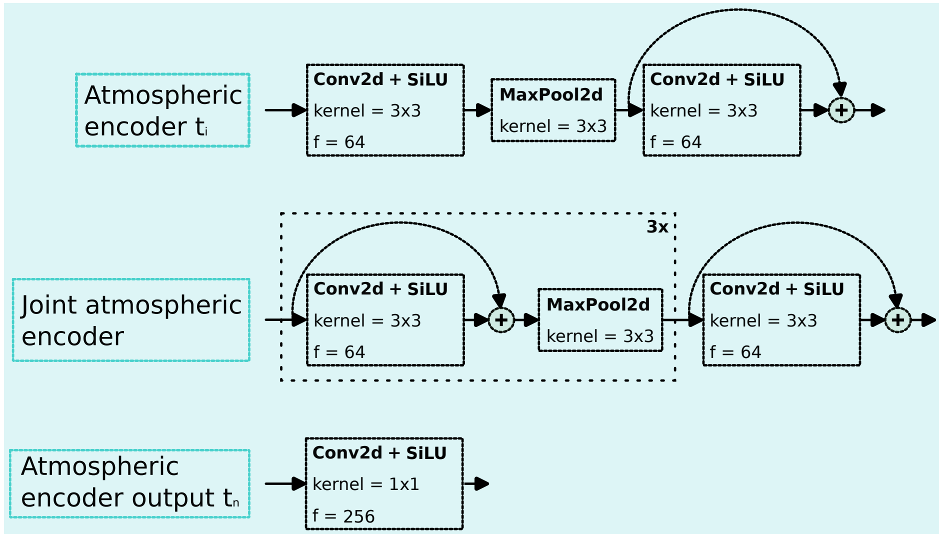

Figure 5. DELWAVE atmospheric encoder block sub-components. Each sub-component is shown in a separate row. The variable f denotes the number of output features of that operation and kernel denotes the kernel size of the convolution operation. Stride is always one for the convolution layers and two for the max pool layers. The activation function of choice is the Sigmoid Linear Unit (SiLU) (Hendrycks and Gimpel, 2016).

wind pattern detector to locate and recognize the pattern irrespective of the time of occurrence. The approach of weight sharing between timesteps also reduces the computational complexity, the number of required parameters, and acts as a regularizing method preventing overfitting. We denote this encoder as A_{joint} which is applied to all output of the individual atmospheric encoders

$$A_{\text{joint}}(A_1(\mathbf{I}_{l,(1)}),\ldots,A_{11}(\mathbf{I}_{l,(11)})), \tag{12}$$

and results in a single output tensor.

Output atmospheric encoder: The output atmospheric encoder selects the recognized wind features important for each timestep. We do this using a convolution operation with a kernel size of one, signifying a linear combination of the input features. The resulting per-timestep tensors of dimension $[b, 256, 4, 4]$, where b denotes the batch size, are summed across their last two dimensions, resulting in a 256 dimensional vector as the final output of this layer. These vectors serve as high dimensional weather descriptors for individual timesteps and contain wind information at each timestep.

3.2.2 Temporal collapse block

The temporal collapse block, displayed in Figure 6, collects the individual atmospheric vectors and encodes them into a single 256 dimensional vector. This is done by a sequence of two one-dimensional convolution operations Conv1d, where we set the the kernel size and output feature count to one for the later of the two. This essentially achieves a linear combination of the

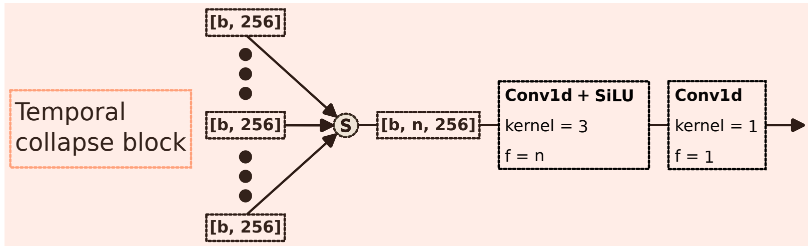

Figure 6. DELWAVE temporal collapse block. The variable n denotes the number of time steps used to train the model, b denotes the batch size. The n weather feature vectors of dimension $[b, 256]$ are stacked into a single tensor which is then reduced to a single 256 dimensional vector by passing through the convolutional operations. Stride is always one for the convolution layers. The activation function of choice is the Sigmoid Linear Unit (SiLU) (Hendrycks and Gimpel, 2016).

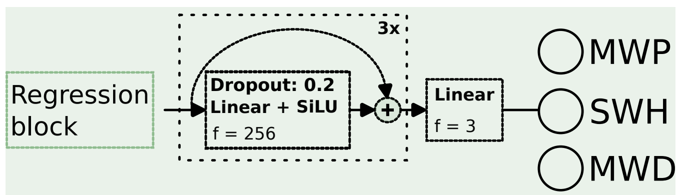

Figure 7. DELWAVE regression block reduces the output of the temporal collapse block into the final outputs: MWP, SWH, and MWD. The regression is conducted by a cascade of three dense skip connections followed by the final dense connection with three outputs. The activation function of choice is the Sigmoid Linear Unit (SiLU) (Hendrycks and Gimpel, 2016).

inputs across the timestep dimension. The first convolution produces a new set of interleaved temporal feature vectors. The second block reduces these temporal feature vectors into a single vector by means of a linear combination.

3.2.3 Regression block

Finally, the regression block, displayed in Figure 7, comprises of consecutive fully connected layers with skip connections. This block produces the final outputs: MWP, SWH, and MWD. To prevent overfitting and improve performance over unseen data in the cross-validation dataset, a dropout with a removal probability of 0.2 is applied between each fully connected layer, except the last two (the output layer and the penultimate layer).

3.3 Training protocol

The CTR period was used for training while the SCE period was used for testing the final, developed model. The CTR data was further split into two parts: the actual training dataset (CTR$_{trn}$) and the validation dataset (CTR$_{vld}$). CTR$_{trn}$ contains the first 80 percent of the training data, while CTR$_{vld}$ contains the remaining 20 percent. The data for each location and for each variable (significant wave height, mean wave period, mean wave direction) is separately standardized to exhibit zero mean and variance one. Prior to the standardization we log-transform significant wave heights by $\ln(\mathrm{SWH}+1)$. We give arguments for

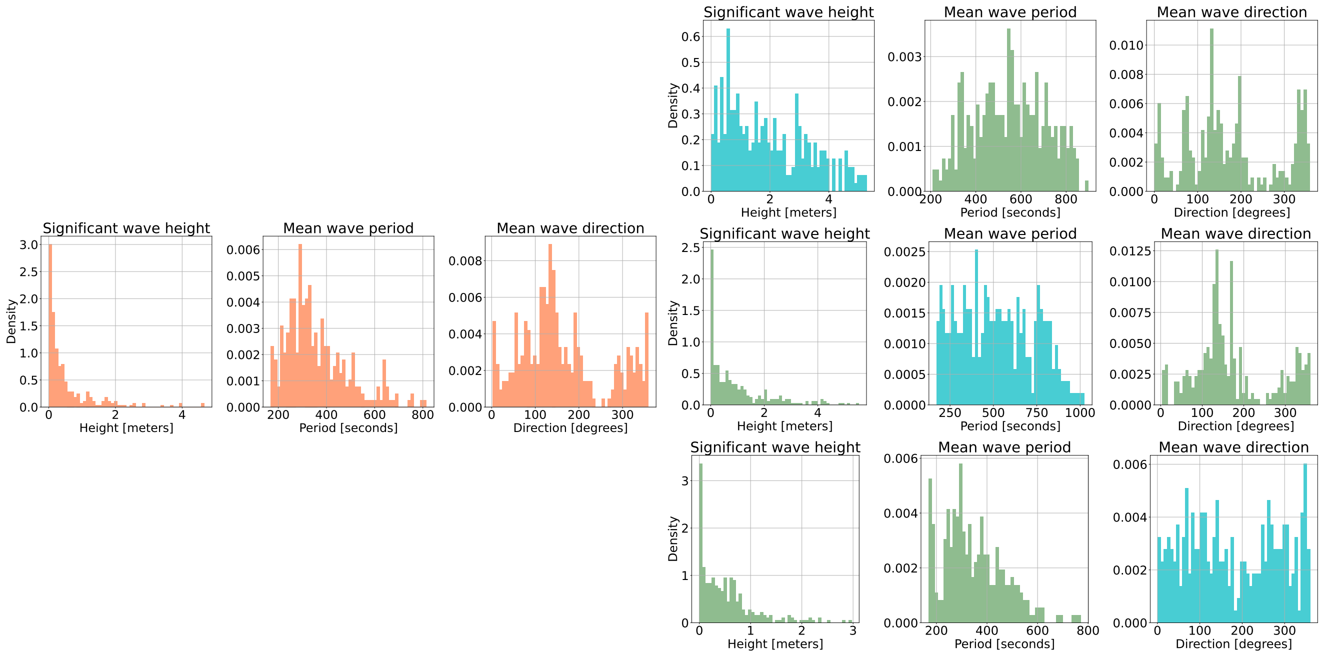

Figure 8. Random importance-sampling's effect on the batch constitution over training. The row on the left with orange columns represent the distributions of all three target variables in a random training batch without importance-sampling. The right three rows display importance-sampled batches, each row belonging to a specific variable that was importance-sampled. The histograms colored in blue contain those variables that were importance-sampled. Importance-sampling results in more uniform distributions for the sampled variables, which indicates a more equal sampling of the target variable realization space.

this transformation in the following paragraphs. Then at each training iteration a random batch of training samples is collected and the model loss function, the root mean squared difference, defined in Equation (6), is used to optimize DELWAVE's parameters, evaluated at these batches.

Neural networks often have difficulties predicting extreme events in the tails of the distributions, because these events are by definition rare and the network rarely encounters them in the training set. To better learn and model underrepresented values of the target variables we increased their presence in the training set by employing the so-called random importance-sampling. We illustrate random importance-sampling in Figure 8. If we observe the distributions of the three target variables in a randomly sampled batch (left panel in Figure 8) we can see that these are skewed. For example, the significant wave height is distributed similarly to a left slanted gamma-like distribution with a very long tail. Therefore, the model is not exposed to the tail of the distribution frequently which inhibits efficient training in that part of the distribution. This results in systematic errors, where the regression accuracy for significant wave height drops with increasing height. This is understandable as samples with wave heights above two meters constitute only a small fraction of the dataset, contributing less during training compared to samples with smaller wave heights.

Our implementation of importance-sampling is conducted on-the-fly at batch acquisition. The reasons we do this on-the-fly as opposed to conducting this statically (oversampling before training and saving the new samples on disk) is this: oversampling highly skewed distributions to a point of close uniformity would require a large amount of additional samples. Since we had limited disk space this was not an option. Therefore, we implemented single variable importance sampling that over-samples one of the target variables at random for a given training batch. However, when we over sample one of the variables, the remaining usually remain biased. We can observe this effect in the skewed green histograms in Figure 8, while the blue histograms are more uniform. To alleviate this issue, we alternated the sampling between the three target variables randomly to eliminate single variable importance sampling bias.

Furthermore, we alternated between regular sampling and importance-sampling, where every second batch was randomly importance sampled. This compromise offered the best performance out of the two approaches. We believe that this is due to the majority of data taking on only a small subset of values, thus these values influence the loss more than the rare events. This is especially true for significant wave heights where only five percent of all samples across all locations exhibit wave heights above two meters. Additionally, since fitting unbiased estimates of the tail of the distribution for significant wave heights was still challenging, we also penalized the network for miss-classifying significant wave height twice as much as for the remaining two variables.

We conducted our training procedure in two stages. Since we trained our model on the Vega cluster (Institute of Information Science, 2023) we were limited by the maximum time our training could take up. A single run could last up to two days maximum therefore, we first trained our model using the Adam (Kingma and Ba, 2014) solver with default Pytorch parameters, learning rate of 10^{-3}, and a weight decay of 10^{-6} for two days. Following this period, we extracted the model that best performed on the validation dataset, reinitialized the learning procedure with a reduced learning rate of 10^{-5}, and retrained for 600 more epochs. We again took the model that best performed on the validation dataset and used it to compute the test dataset results we present in the following sections.

4 Temporal ablation study of the input

In this section, we investigate the impact of the number of timesteps on the performance of the model. Adding multiple timesteps results in inputing more information into the model, therefore training performance might increase. However, due to overfitting this performance might not be reflected in the actual accuracy on unseen data. Therefore, we conducted a preliminary comparison of five DELWAVE variants, each trained with a different number of input timesteps. These variants are DELWAVE_{2}, DELWAVE_{4}, DELWAVE_{8}, DELWAVE_{11}, and DELWAVE_{16}, where the subscript denotes the number of used timesteps. Here, each of the five variants uses $n+1$ timestemps, where n denotes the number of previous timesteps (in case of DELWAVE_{8} this means seven previous timesteps with the addition of the current one). Results of this study are presented in Table 1 and their validation loss during training in Figure 9.

We can observe the diminishing returns nature of adding timesteps beyond the 11th timestep: the performance seems to be roughly identical between the DELWAVE_{16} and DELWAVE_{11}. Note also that DELWAVE_{16} contains more trainable parameters and is also slower to train comapred to DELWAVE_{11}. DELWAVE_{11} exhibits the best performance in four cases, equal to DELWAVE_{16}, followed by DELWAVE_{8} with two cases. Similarly, we can observe that after the threshold of 11 time samples is reached we enter the diminishing returns domain, where DELWAVE_{16} offers negligible or even worse performance in some cases compared to DELWAVE_{11}. Therefore, we concluded that DELWAVE_{11} is the most promising network variant for further training.

	AA	MB	OB
DELWAVE_2 $\text{RMS}_{\text{height}}$	0.145	0.134	0.072
DELWAVE_4 $\text{RMS}_{\text{height}}$	0.091	0.078	0.034
DELWAVE_8 $\text{RMS}_{\text{height}}$	**0.065**	0.082	0.033
DELWAVE_{11} $\text{RMS}_{\text{height}}$	0.067	0.083	**0.032**
DELWAVE_{16} $\text{RMS}_{\text{height}}$	0.073	**0.079**	**0.032**
DELWAVE_2 $\text{RMS}_{\text{period}}$	98.279	44.930	107.135
DELWAVE_4 $\text{RMS}_{\text{period}}$	82.057	30.432	76.555
DELWAVE_8 $\text{RMS}_{\text{period}}$	50.457	**24.402**	55.783
DELWAVE_{11} $\text{RMS}_{\text{period}}$	**43.546**	24.407	**55.614**
DELWAVE_{16} $\text{RMS}_{\text{period}}$	44.056	25.084	58.559
DELWAVE_2 $\text{RMS}_{\text{direction}}$	22.057	69.798	25.836
DELWAVE_4 $\text{RMS}_{\text{direction}}$	19.877	62.432	22.065
DELWAVE_8 $\text{RMS}_{\text{direction}}$	16.504	57.108	19.985
DELWAVE_{11} $\text{RMS}_{\text{direction}}$	16.775	**54.720**	19.626
DELWAVE_{16} $\text{RMS}_{\text{direction}}$	**16.270**	55.614	**18.961**

Table 1. Table containing the performance evaluations of DELWAVE which we constructed by varying the amount of timesteps used during training, for three training locations: AA, MB, and OB. RMS denotes the root mean squared error and the best performing (with the lowest RMS) variant is in bold.

5 Results

In order to assess the potential and the possible limitations of the DELWAVE network, the analysis of the results is divided into three phases. After an overview of the performance of the network in reproducing the main overall properties of the SWAN time series (Sec. 5.1), the analysis will focus on two aspects of particular relevance for practical purposes, namely the capability of reproducing storms (including, but not only, extreme events, Sec. 5.2) and their main properties, and the capability of capturing the main features of the climate change signal (Sec. 5.3).

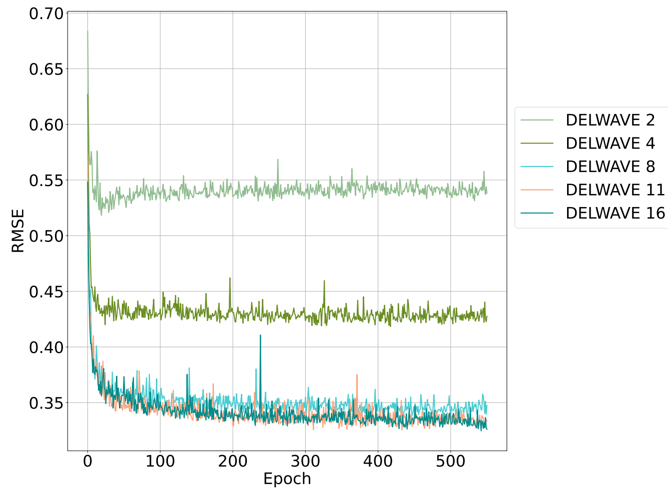

Figure 9. Root mean squared error on the validation dataset (averaged over all three variables) for all DELWAVE temporal ablation variants. The cut off number of epoch is the amount achieved by DELWAVE$_{16}$ in two days of training since it is the slowest of all the variants.

5.1 Deep Network vs SWAN under far future climate 2071-2100

In this section we present DELWAVE performance during the far future period of 2071-2100, as benchmarked against SWAN simulations. In other words: SWAN simulations represent the ground truth DELWAVE aims to model. Figure 10 depicts DELWAVE-SWAN heatmaps of H_S, d and $T_{m-1,0}$ at the locations of Acqua Alta oceanographic tower (AA) and the Ortona and Monopoli buoys (respectively OB and MB, see Figure 2 for locations). Results for other locations are provided in the Supplementary material.

We will proceed by analyzing DELWAVE performance using three related Figures. Figure 10 depicts DELWAVE predictions for H_S, d and $T_{m-1,0}$ compared to those from the SWAN model, obtained from the same wind fields, *i.e.* obtained for the same forecasting time window. Figure 11 shows the overlaps of histograms of H_S, d and $T_{m-1,0}$ from both DELWAVE and SWAN model. Note that close overlap of the distribution histograms from both models does not guarantee a good forecast since this overlap does not tell anything about the synchronicity of both forecasts - one therefore needs to view Figure 11 in conjunction with Figure 10. Additionally, Figure 11 illustrates how DELWAVE forecasting mean absolute errors change depending on which part of distribution we are modeling. Here mean errors imply error averaging over all the forecasting samples in a specific distribution bin. Consequently the error values are well defined only in the bins containing a large enough (e.g. over 100) number of samples. In what follows we will be basing our remarks on an interplay of messages from all three Figures.

Location AA in the northern Adriatic (off the Venetian shore, see Figure 2 for location) is marked by an excellent performance in H_S and d prediction, indicated by the near linear scatter plot displayed in Figure 10. The same aspect of DELWAVE performance is illustrated via histogram distribution for the same three parameters on Figure 11. d (Figure 11, top row, right column) exhibits two maximums related to two dominant Adriatic winds, northeasterly Bora at roughly $75°$ and southeasterly

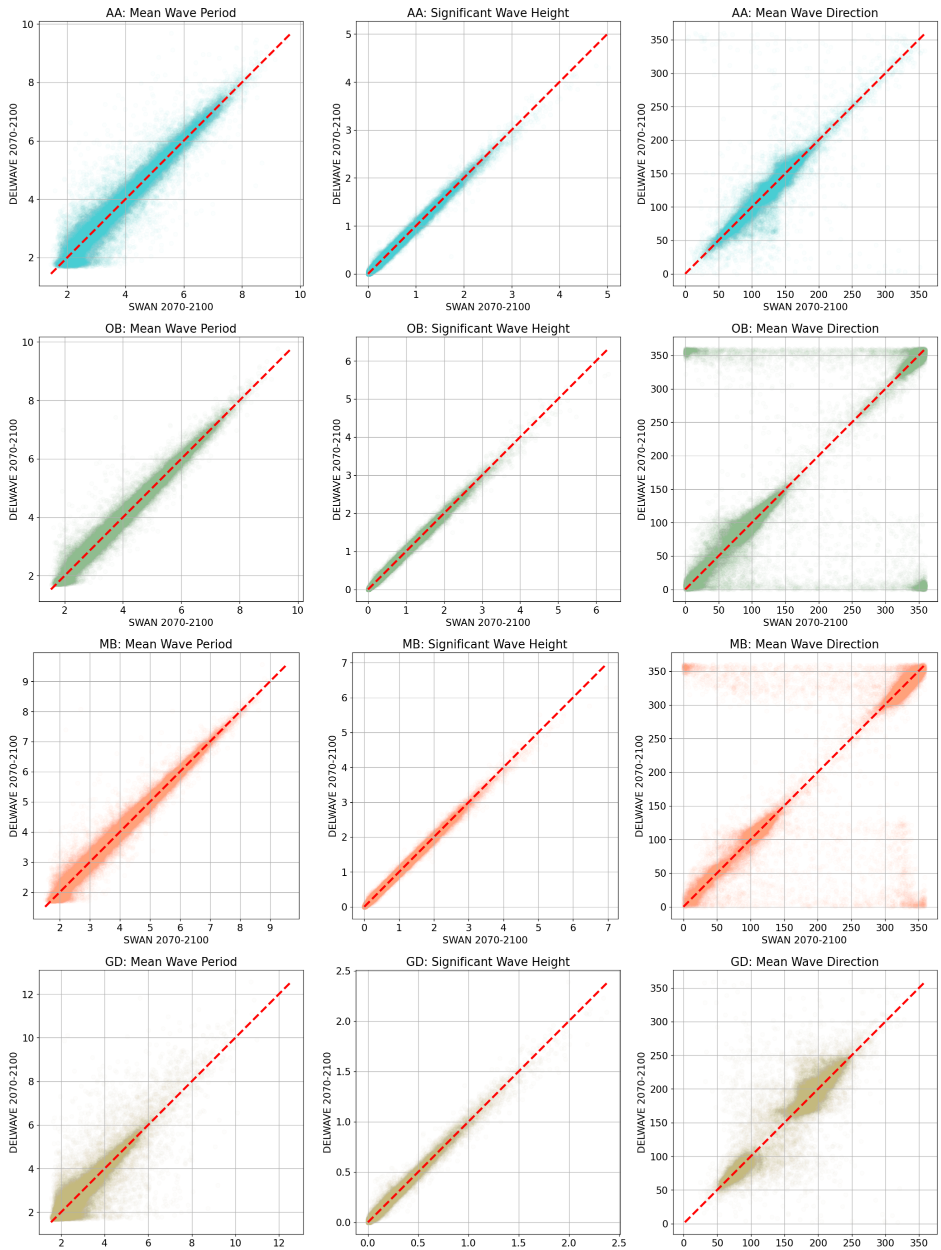

Figure 10. A scatter plot of DELWAVE forecasts (y-axis) compared to their SWAN targets (x-axis) for mean wave period [s] (**1st column**), significant wave height [m] (**2nd column**) and mean wave direction [°] (**3rd column**) at locations AA (**1st row**), OB (**2nd row**), MB (**3rd row**), and GD (**4th row**). Mean wave directions are listed in nautical notation (0° = North, 90° = East, etc.). Dashed diagonal line in each plot indicates a perfect forecast.

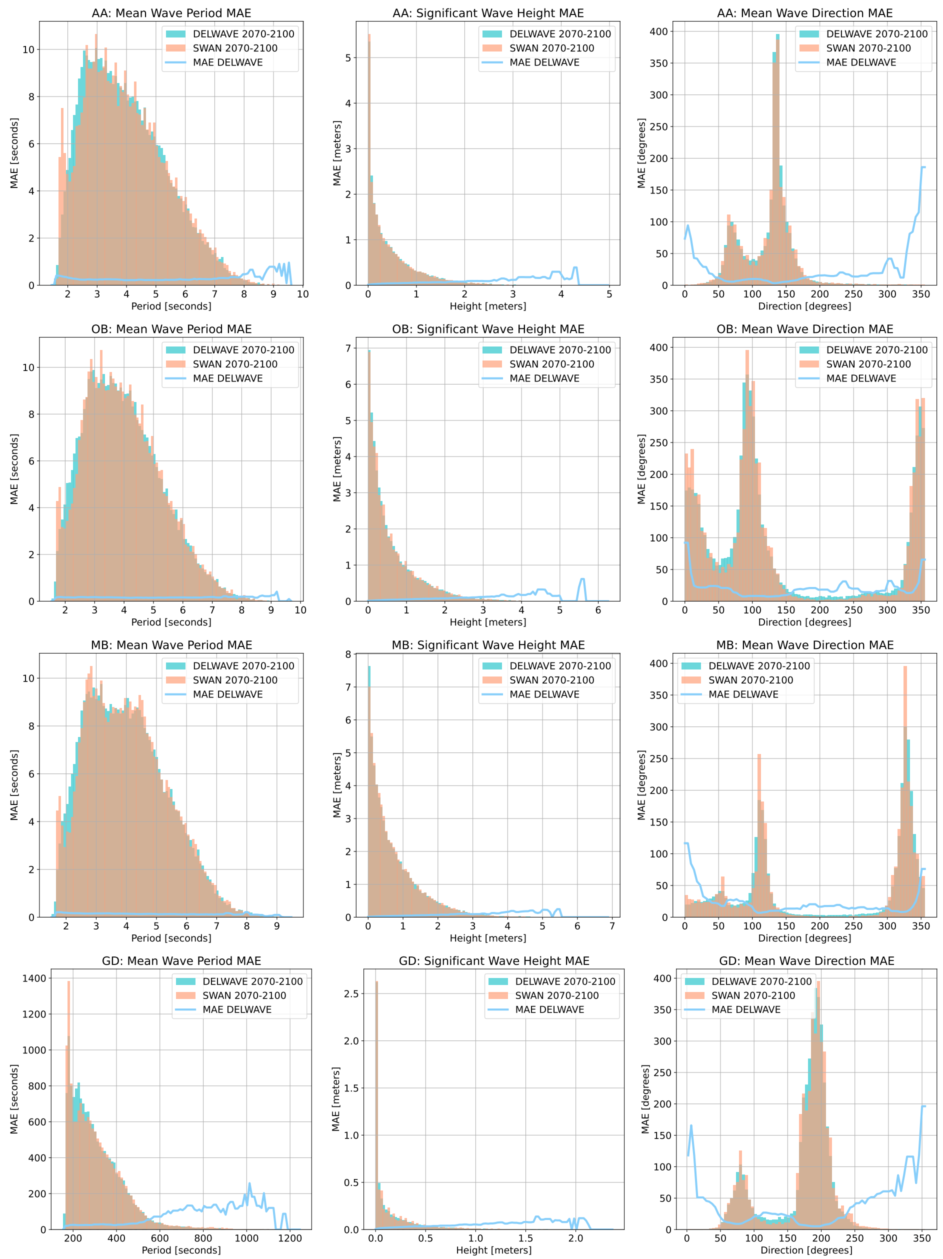

Figure 11. Histograms of DELWAVE-vs-SWAN distributions of H_S (left column), d (middle column) and $T_{m-1,0}$ (right column) from DELWAVE (turquoise bars) and SWAN model (brown bars) during the 2071-2100 timewindow at AA (1st row), OB (2nd row), MB (3rd row), and GD (4th row) location. Light blue lines are scaled on the y axis and depict MAE, averaged over number of samples in each bin.

Scirocco at roughly 135°. Short wave periods at AA location on the other hand seem to be the hardest to predict, as can be seen from in the left column in either of the Figures 10 or 11. This is to some extent expected: long wave periods correspond to longer waves and consequently windy atmospheric conditions. Short periods on the other hand correspond to calm conditions where the network is essentially modelling low amplitude, short wavelength, stochastic sea surface behaviour.

Similar observations can be made for OB and MB locations. SWAN H_S is modelled very reliably with DELWAVE. Multi-modal direction histograms at all locations are also reproduced to a high degree of accuracy, as can be seen from the middle column of Figure 11. On the other hand, the network seems to be struggling to reproduce northerly directions (rougly $0° \pm 10°$) at this location. This leads to horizontal strips of incorrect predictions displayed in the scatter plot of the right column, middle row in Figure 10 and to a bump in mean absolute error in the histogram displayed at the same location in Figure 11.

Figure 11 also hints at quantitative estimates of DELWAVE performance. When it comes to H_S predictions (middle column) errors at all locations grow with significant wave height from errors below 5 cm for H_S below 1 m to errors of order 10-15 cm for H_S over 3 m. DELWAVE predictions of mean wave direction d (right column) exhibit the smallest errors in the directional bins corresponding to prevalent wind patterns. In general directional errors are below $25°$, at AA even lower. High directional errors at $0°$ and $360°$ stem at least partly from the algorithm' s false distinction between $0°$ and $360°$ directions.

Wave period $T_{m-1,0}$ predictions are illustrated in the left column of Figure 11. At all locations, periods below 6 s are captured well by DELWAVE, with prediction errors below 0.25 s. Longer periods, likely corresponding to an incoming swell, however exhibit more diverse behaviour. MB location wave periods seem to be captured more accurately in the long period limit, with forecast error dropping below 0.1 s. At OB location the errors in the long period limit slightly rise, from 0.2 s to 0.3-0.5 s. AA location on the other hand indicates a sharp rise of $T_{m-1,0}$ prediction error which reaches 1 s for period above 8 s.

The error behaviour at the AA location is possibly explained by the differing roles played by the basin geometry, the local wind sea and swell. location AA is prevalently exposed to northeasterly Bora (blowing from roughly $75°$) and to southeasterly Scirocco (blowing from $135°$). In case of Bora the fetch is quite limited since Bora is a cross-basin wind. Therefore we do not expect swell to play a major role at AA location during Bora conditions: the wave field at AA location must be determined by local wind conditions. The case of Scirocco is very different. Scirocco is an along-axis wind with the largest fetch in the Adriatic basin. This means that during Scirocco, swell field at AA is determined to a large extent by non-local wind patterns in the south of the basin. Local wind conditions at AA location are furthermore often a poor proxy for winds in the south Adriatic. Bora in the north (promoting short fetch and shorter wave periods) coinciding with Scirocco in the south (promoting long period swell arriving at AA) is, for example, not unusual. These circumstances likely pose a challenge for the DELWAVE deep network, resulting in growing errors at longer wave periods (which most likely occur during Scirocco).

This explanation can be further substantiated by comparing wave period MAE to H_s, $T_{m-1,0}$ and wave power. The latter is computed from the linear theory to be

$$P = \frac{\rho g^2}{64\pi} H_s^2 T_{m-1,0}, \tag{13}$$

with ρ being the water density and g the acceleration due to gravity. This comparison is depicted in Figure 12, which corroborates this interpretation and constrains DELWAVE limitations in capturing the basin-scale dynamics.

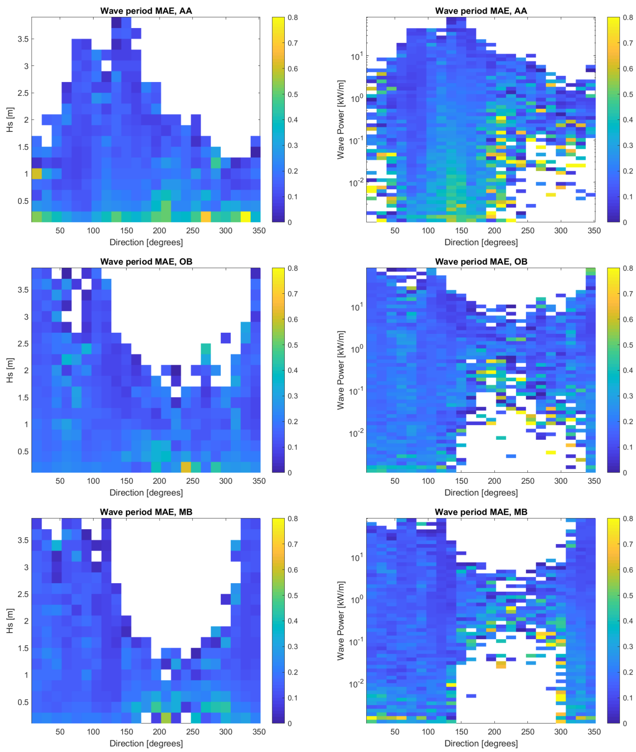

Figure 12. Comparison of wave period (indicated by color) relationship to significant wave height H_s (left column, y-axis) and to wave power (right column, y-axis) at all directions (x-axis) for AA (top row), OB (middle row) and MB (bottom row). The white parts in the plot refer to combinations of direction and variable for which no occurrence was found in the data.

Concentration of the highest values of MAE at low values of H_s and P (left and right columns respectively) confirms that largest errors tend to be associated with low-energy, nearly random sea states, even in the presence of relatively long waves (middle column) along the main basin axis (Scirocco at AA), thus with limited impacts on possible practical applications. It is further worth mentioning that a separate analysis, carried out by independently considering the rising and declining phases of the sea states (not shown), did not exhibit any preferential concentration of the higher values of MAE in either phase. Wave period error is therefore not systematically larger during either onset or calming of the storm, suggesting that it is not directly related to the sequential and monotonous temporal encoding of inputs within DELWAVE. Had this not been the case, we would have expected some error asymmetry with regard to the timing of the storm.

5.2 Storm analysis

The analysis of the storms was carried out by comparing the DELWAVE results against the SWAN time series during period 2071-2100. For each time series, the storms were identified following the method proposed by Boccotti (Boccotti, 2000), namely (i) finding the events with H_s larger than 1.5 times the mean value $\overline{H_s}$ of each respective series, (ii) merging the events parted by less than 10 hours and (iii) discard those overall shorter than 12 hours. Figure 13 compares SWAN and DELWAVE peak H_s and directions for each storm at AA, OB, and MB (the same is shown for the other locations in the supplementary material), considering separately the whole sets of storms occurring during the period and the annual maxima for each series. While the former provides a broader view on how DELWAVE reproduces the whole meteo-marine climate at each location, the latter aims at assessing its capability of addressing extreme events. The picture is flanked by a quantification of the DELWAVE precision (how many DELWAVE-predicted storms are actually present in the SWAN time series) and recall (how many SWAN-modelled storms are retrieved by DELWAVE). These two metrics are computed as

$$\text{Precision} = \frac{TP}{TP+FP}\,,\ \text{Recall} = \frac{TP}{TP+FN}, \tag{14}$$

where TP, FP and FN denote true positive (storm present in SWAN and predicted by DELWAVE), false positive (storm predicted by DELWAVE but not present in SWAN) and false negative (storm present in SWAN but not predicted by DELWAVE) classifications. Figure 14 shows an example of the application of Boccotti's method to SWAN and DELWAVE storms and the occurrence of false negatives and false positives.

All considered sets exhibit a satisfactory performance with very high scores (Precision, Recall $\geq$ 0.95) when all the storms are considered. When only annual maxima are taken into account, precision and recall are lower, though fairly high ($\geq$ 0.8), and without an evidently prevailing directional offset. Considering the whole storm sets, most of the false negatives and positives are generally clustered among the weakest events. This can be explained by considering that, for particularly weak or short events, small absolute errors can mean large relative errors. Therefore in small H_s limit, already a small error in the reproduction of H_s can significantly impact whether the criteria for the identification of storms are met or not (Figure 14).

This result seems to be in contradiction with the results for the yearly maxima sets, where prediction and recall scores decrease and the number of false negatives and positives increases. This contradiction is however only apparent and is related to the propagation of H_s prediction errors downstream into the identification of the yearly maxima. More precisely, in this case the mismatch does not seem related to the classification of an event as a storm, but rather to its classification as an yearly maximum: in fact, a slight error in predicting the peak height of storm events can introduce some noise in the ranking of the events, and in particular in the identification of the yearly maxima, leading to a mismatch between DELWAVE and SWAN. Nonetheless, as far as small errors in the prediction of the peak H_s are the cause for this mismatch, even if the events retrieved by DELWAVE are not exactly the ones resulting from the SWAN time series, their properties (or at least their peak H_s) should be quite close, which should be sufficient for most practical applications.

5.3 Climate change features

One of the main scopes of DELWAVE is to provide a computationally cheap model emulation system capable of providing large ensemble predictions for wave climate at a multi-decadal scale. This kind of applications is to some extent complementary to the event-scale analysis of single storms, and requires a specific assessment of the network capability of capturing the main statistical features of the climate signal. Figure 15 provides a twofold comparison of the climatological normals of the

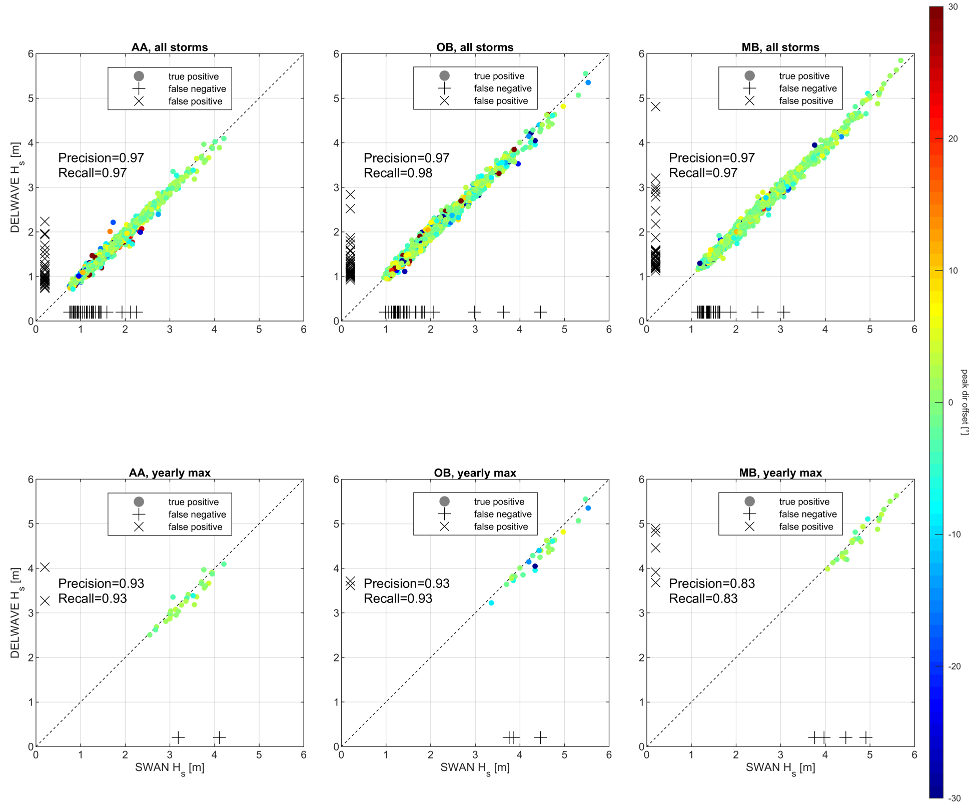

Figure 13. Comparison of SWAN and DELWAVE peak H_s value at AA, OB, and MB during all the storms (top row) and for the annual maxima (bottom row). Dashed diagonal line indicates a perfect match. The colormap represents the directional offset during the peak of each storm. Pluses and crosses along the plot axes represent false negatives ($+$) and false positives ($\times$).

monthly mean, median and 99th percentile of H_s at AA, OB, and MB (the same values for the other locations are provided as supplementary material) provided by SWAN and reproduced by DELWAVE. The statistics resulting from the SWAN and from the DELWAVE time series are compared against each other in the end-of-century scenario (2071-2100, SCE), and both are compared against the statistics from the control condition (CTR), available only for SWAN in the 1971-2000 period. The good agreement between DELWAVE and SWAN is confirmed also when considering climatological statistics, with a small ($\leq$5%), though systematic, underestimate of 99th percentile values, reflecting what was discussed in Section 5.1. Compared to the CTR climatologies, the mismatch between DELWAVE and SWAN is generally small compared to the difference between SCE and CTR conditions, suggesting that the noise possibly introduced by the model mimicking is weaker than the climate change signal in the considered locations. Not surprisingly, the only way in which the performance seems partially affected

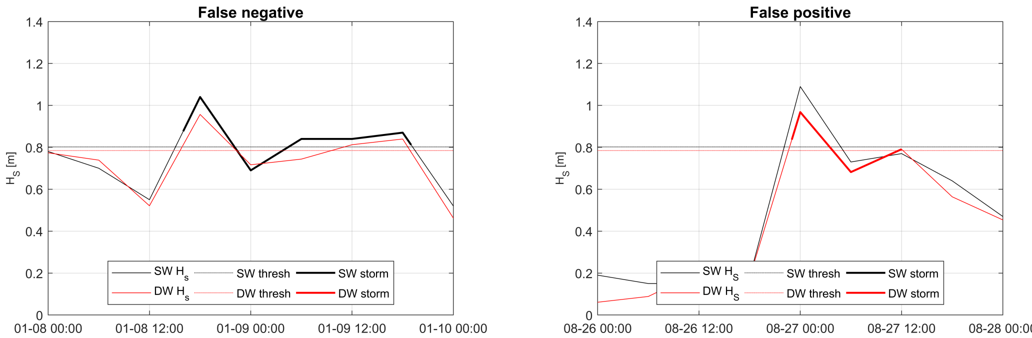

Figure 14. Examples of false negatives (left) and false positives (right) in the identification of storms (thick lines) following the method by Boccotti (2000) in the DELWAVE and SWAN H_S time series (thin lines). Dotted lines represent the reference threshold of $1.5\overline{H_S}$ for each time series.

by seasonality is through the modulation of significant wave height and the tendency of the network to underestimate higher (and therefore wintry) values. Nevertheless, Figure 15 shows that the potential modeling errors, introduced by the DELWAVE model, are substantially smaller than the difference between scenario (2070-2100) and control periods (1970-2100).

Following a similar approach for the directional wave climate, the linearized wave roses in Figure 16 show that the agreement between DELWAVE and SWAN allows to capture important impacts of climate change in the wave regime not only in absolute terms, but also in response to projected shifts in the wind regimes. This is for instance the case of the slight weakening of Bora (NE) storms associated with an intensification of Scirocco (SE) events in the northern Adriatic Sea in the broader framework of a tendency towards an overall decrease of the storminess in most of the basin, suggested by Bonaldo et al. (2020) and confirmed by the DELWAVE projections.

6 Conclusions

We have presented a new point-prediction deep learning method for surface gravity wave emulation in epicontinental Adriatic basin, which took about two and a half days to train and can process more than 100 wind fields per second, to be used for large-ensemble prediction over synoptic to climate timescales. DELWAVE input set consists of atmospheric winds during 1998-2000 and cross-validation period is the far-future climate timewindow of 2071-2100. We have thoroughly analyzed which architecture yields best results for wave emulation and these efforts led us to presented architecture of a convolution-based atmospheric encoder block, temporal collapse block and finally a regression block. We introduced random importance-sampling for improved modeling of underpopulated tails of variable data distributions. Detailed ablation studies were performed to determine optimal performance regarding input fields, temporal horizon of the training set and network architecture. We demonstrated that DELWAVE reproduces SWAN model significant wave heights with a mean absolute error (MAE) between 5 and 10 cm, mean wave directions with a MAE of $10°$-$25°$ and mean wave period with a MAE of 0.2 s. The network is able to accurately

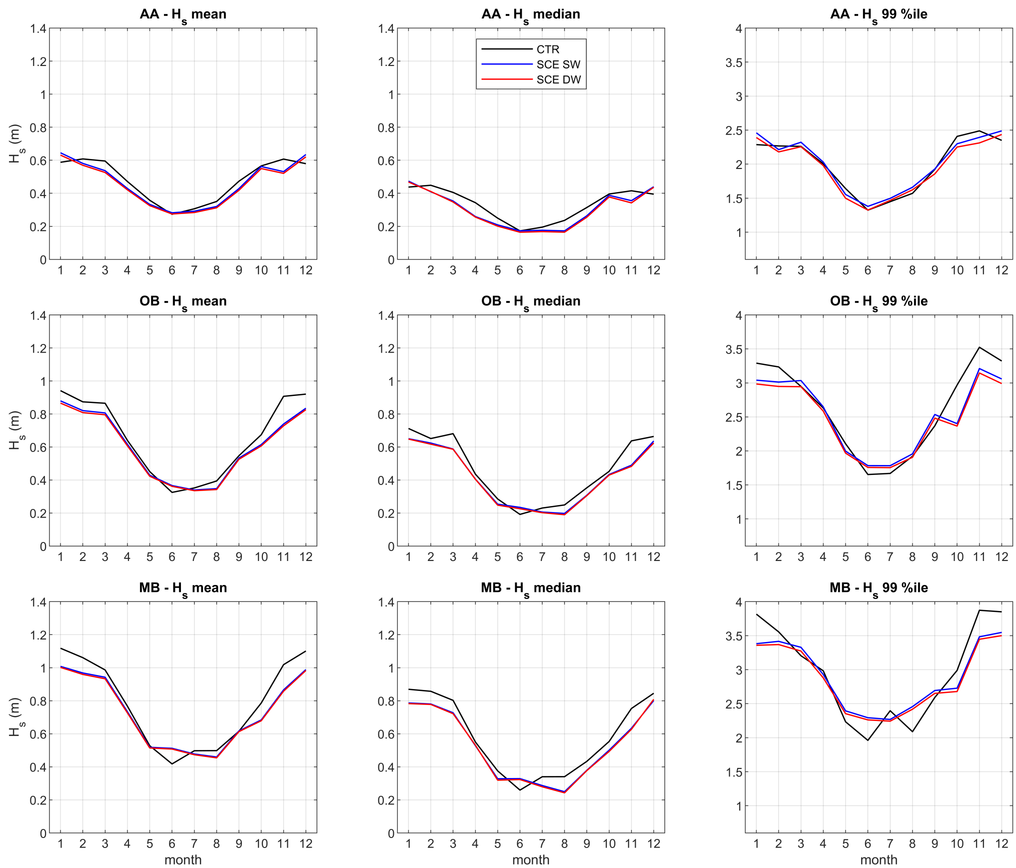

Figure 15. Comparison of SWAN (SW) and DELWAVE (DW) mean, median and 99th percentile H_s climatologies statistics in the future scenario (2071-2100, SCE) against the SWAN-modelled statistics referred to the control period (1971-2000, CTR), respectively at AA, OB, and MB.

emulate multi-modal distributions of mean wave directions, which are related to dominant wind regimes in the basin. An analysis of DELWAVE performance during storms was performed by employing threshold-based metrics of precision and recall. DELWAVE reached a very high score (both metrics over 95%) of storm detection.

SWAN and DELWAVE time series are further compared against each other in the end-of-century scenario (2071-2100), and both are compared to control period of 1971-2000. Compared to control climatology over all wind directions, the mismatch between DELWAVE and SWAN is generally small compared to the difference between scenario and control conditions, suggesting that the noise introduced by surrogate modeling is substantially weaker than the climate change signal. There is a number of things we would like to explore further: it is currently not clear how to leverage gaussian (or other) spatial encoding to generate, if possible, reliable predictions for locations which lie outside of the training set. This might open the door for

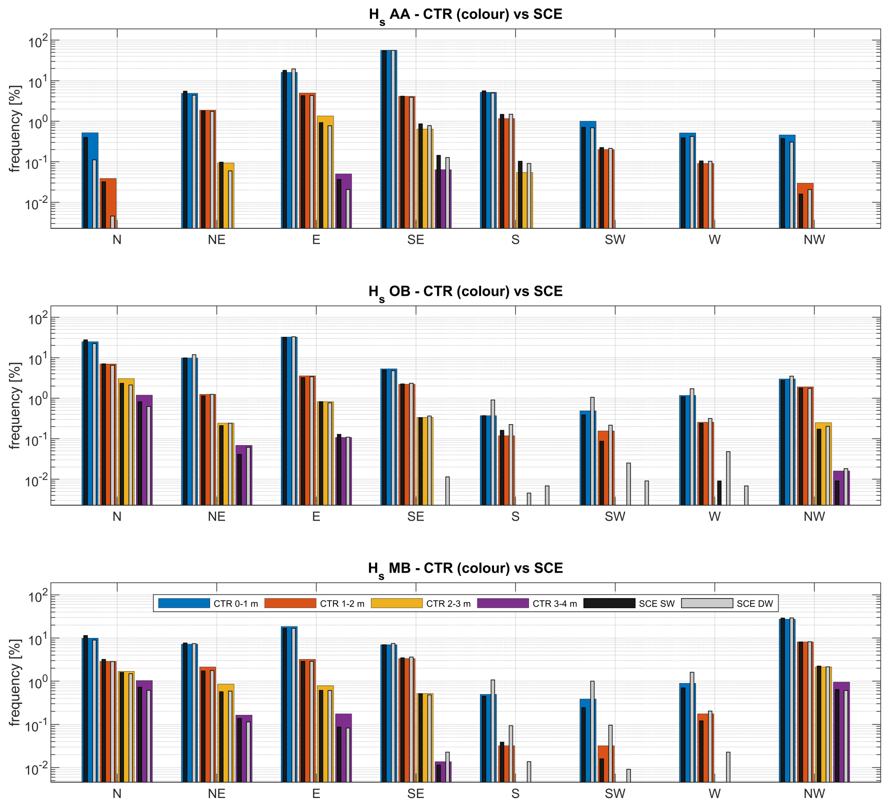

Figure 16. Comparison of SWAN (SW) and DELWAVE (DW) directional H_s statistics in the future scenario (2071-2100, SCE, black and grey bars respectively) against the same quantities modelled by SWAN with reference to the control period (1971-2000, CTR, coloured bars), respectively at AA, OB, and MB.

dense predictions of the wave field, at least in the vicinities of input data locations. It would furthermore be interesting to introduce temporal dependence of the Gaussian variances in the spatial encoding matrix to help the network focus on wider areas of input data as we feed it data from a more distant past.

Future research and potential applications may also focus on the larger scales, for example the entire Mediterranean Sea basin, using high-resolution wind and waves model to boost DELWAVE training. The objective would be to explore the behaviour of numerical and machine learning models in diverse wind and wave regimes, as well as wind and marine storms, which exhibit distinct physical characteristics in a basin with highly diverse morphological and dynamic features.

Last but not least, another promising venue is offered by recent developments in the field of physics-informed machine learning. Here, the solution subspace is further constrained by additional loss terms which nudge the learning process towards physically consistent solutions. Since the physical aspects of wind driven surface gravity waves are known in substantial detail, we expect there to be some immediate benefits to introducing dynamics laws into the training. Last but not least, it would be interesting to study how well the network generalizes to other domains and other models. All these will be topics of further research.

Code and data availability. DELWAVE model code is available publicly on GitHub: https://github.com/petermlakar/DELWAVE. Raw COSMO dataset can be found at the following repository, maintained by CMCC: https://doi.org/10.25424/cmcc-3hph-jy15. Preprocessed COSMO datasets, suitable for DELWAVE input, can be found on the following repository: https://doi.org/10.5281/zenodo.7816888.

Author contributions. PM designed, implemented and tested DELWAVE and all its ablations. ML wrote the initial version of the network. DB and ML contributed to geophysics-related aspects of DELWAVE. DB and AR provided SWAN simulations. PM, DB and ML performed the analyses and wrote the paper. All authors contributed the research plan and to the final version of the paper.

Competing interests. The authors declare no competing interests.

Acknowledgements. The authors acknowledge Dr. Edoardo Bucchignani (CIRA - Italian Aerospace Research Centre) and Dr. Paola Mercogliano (CMCC - Centro Euro-Mediterraneo sui Cambiamenti Climatici) for sharing thoughts and wind fields from COSMO-CLM. The manuscript was substantially improved by suggestions of the two reviewers.

Financial support. Antonio Ricchi acknowledges the financial support from the PON “Ricerca e Innovazione” 2014-2020, DM 1062 of 10/08/2021. Sandro Carniel gladly acknowledges the financial support from the NATO Chief Scientist. Matjaž Ličer acknowledges the financial support from the Slovenian Research Agency (research core funding No. P1-0237).

References

Antonioli, F., Anzidei, M., Amorosi, A., Lo Presti, V., Mastronuzzi, G., Deiana, G., De Falco, G., Fontana, A., Fontolan, G., Lisco, S., Marsico, A., Moretti, M., Orrù, P., Sannino, G., Serpelloni, E., and Vecchio, A.: Sea-level rise and potential drowning of the Italian coastal plains: Flooding risk scenarios for 2100, Quaternary Science Reviews, 158, 29–43, https://doi.org/10.1016/j.quascirev.2016.12.021, 2017.

Astariz, S. and Iglesias, G.: The economics of wave energy: A review, Renewable and Sustainable Energy Reviews, 45, 397–408, https://doi.org/https://doi.org/10.1016/j.rser.2015.01.061, 2015.

Bellafiore, D., Bucchignani, E., Gualdi, S., Carniel, S., Djurdjević, V., and Umgiesser, G.: Assessment of meteorological climate models as inputs for coastal studies, Ocean Dynamics, 62, 555–568, https://doi.org/10.1007/s10236-011-0508-2, 2012.

Benetazzo, A., Bergamasco, A., Bonaldo, D., Falcieri, F., Sclavo, M., Langone, L., and Carniel, S.: Response of the Adriatic Sea to an intense cold air outbreak: Dense water dynamics and wave-induced transport, Progress in Oceanography, 128, 115–138, https://doi.org/10.1016/j.pocean.2014.08.015, 2014.

Benetazzo, A., Davison, S., Barbariol, F., Mercogliano, P., Favaretto, C., and Sclavo, M.: Correction of ERA5 Wind for Regional Climate Projections of Sea Waves, Water, 14, https://doi.org/10.3390/w14101590, 2022.

Beucler, T., Ebert-Uphoff, I., Rasp, S., Pritchard, M., and Gentine, P.: Machine learning for clouds and climate (invited chapter for the agu geophysical monograph series “clouds and climate”), Earth and Space Science Open Archive, 27, 2021.

Boccotti, P.: Wave Mechanics for Ocean Engineering, Elsevier Science, Oxford, 496 pp., 2000.

Boehme, L. and Rosso, I.: Classifying oceanographic structures in the Amundsen Sea, Antarctica, Geophysical Research Letters, 48, e2020GL089 412, 2021.

Bonaldo, D., Benetazzo, A., Bergamasco, A., Campiani, E., Foglini, F., Sclavo, M., Trincardi, F., and Carniel, S.: Interactions among Adriatic continental margin morphology, deep circulation and bedform patterns, Marine Geology, 375, 82–98, https://doi.org/10.1016/j.margeo.2015.09.012, 2016.

Bonaldo, D., Bucchignani, E., Ricchi, A., and Carniel, S.: Wind storminess in the adriatic sea in a climate change scenario, Acta Adriatica, 58, 2017.

Bonaldo, D., Bucchignani, E., Pomaro, A., Ricchi, A., Sclavo, M., and Carniel, S.: Wind waves in the Adriatic Sea under a severe climate change scenario and implications for the coasts, International Journal of Climatology, 40, 5389–5406, https://doi.org/10.1002/joc.6524, 2020.

Booij, N., Ris, R. C., and Holthuijsen, L. H.: A third-generation wave model for coastal regions: 1. Model description and validation, Journal of Geophysical Research, 104, 7649, https://doi.org/10.1029/98JC02622, 1999.

Bucchignani, E., Montesarchio, M., Zollo, A. L., and Mercogliano, P.: High-resolution climate simulations with COSMO-CLM over Italy: Performance evaluation and climate projections for the 21st century, International Journal of Climatology, 36, 735–756, https://doi.org/10.1002/joc.4379, 2016.

Cavaleri, L., Alves, J. H., Ardhuin, F., Babanin, A., Banner, M., Belibassakis, K., Benoit, M., Donelan, M., Groeneweg, J., Herbers, T. H., Hwang, P., Janssen, P. A., Janssen, T., Lavrenov, I. V., Magne, R., Monbaliu, J., Onorato, M., Polnikov, V., Resio, D., Rogers, W. E., Sheremet, A., McKee Smith, J., Tolman, H. L., van Vledder, G., Wolf, J., and Young, I.: Wave modelling - The state of the art, Progress in Oceanography, 75, 603–674, https://doi.org/10.1016/j.pocean.2007.05.005, 2007.

Cavaleri, L., Abdalla, S., Benetazzo, A., Bertotti, L., Bidlot, J. R., Breivik, Carniel, S., Jensen, R. E., Portilla-Yandun, J., Rogers, W. E., Roland, A., Sanchez-Arcilla, A., Smith, J. M., Staneva, J., Toledo, Y., van Vledder, G. P., and van der Westhuysen, A. J.: Wave modelling in coastal and inner seas, Progress in Oceanography, 167, 164–233, https://doi.org/10.1016/j.pocean.2018.03.010, 2018.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y.: Learning phrase representations using RNN encoder-decoder for statistical machine translation, arXiv preprint arXiv:1406.1078, 2014.

Conv1d: PyTorch implementation, https://pytorch.org/docs/stable/generated/torch.nn.Conv1d.html.

Di Silvio, G., Dall'Angelo, C., Bonaldo, D., and Fasolato, G.: Long-term model of planimetric and bathymetric evolution of a tidal lagoon, Continental Shelf Research, 30, 894–903, https://doi.org/10.1016/j.csr.2009.09.010, 2010.

Friedrichs, C. T.: Tidal Flat Morphodynamics: A Synthesis, vol. 3, Elsevier Inc., https://doi.org/10.1016/B978-0-12-374711-2.00307-7, 2011.

Giorgi, F. and Lionello, P.: Climate change projections for the Mediterranean region, Global and Planetary Change, 63, 90–104, https://doi.org/10.1016/j.gloplacha.2007.09.005, 2008.

Hendrycks, D. and Gimpel, K.: Gaussian error linear units (gelus), arXiv preprint arXiv:1606.08415, 2016.

Institute of Information Science, G. o. t. R. o. S.: IZUM, https://izum.si/en/home, online; accessed 3 March 2023, 2023.

James, S. C., Zhang, Y., and O'Donncha, F.: A machine learning framework to forecast wave conditions, Coastal Engineering, 137, 1–10, https://doi.org/10.1016/j.coastaleng.2018.03.004, 2018.

Janssens, M. and Hulshoff, S. J.: Advancing Artificial Neural Network Parameterization for Atmospheric Turbulence Using a Variational Multiscale Model, Journal of Advances in Modeling Earth Systems, 14, e2021MS002 490, https://doi.org/https://doi.org/10.1029/2021MS002490, e2021MS002490 2021MS002490, 2022.

Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980, 2014.

Lionello, P., Conte, D., Marzo, L., and Scarascia, L.: The contrasting effect of increasing mean sea level and decreasing storminess on the maximum water level during storms along the coast of the Mediterranean Sea in the mid 21st century, Global and Planetary Change, 151, 80–91, https://doi.org/10.1016/j.gloplacha.2016.06.012, 2017.

Mallett, H. K. W., Boehme, L., Fedak, M., Heywood, K. J., Stevens, D. P., and Roquet, F.: Variation in the Distribution and Properties of Circumpolar Deep Water in the Eastern Amundsen Sea, on Seasonal Timescales, Using Seal-Borne Tags, Geophysical Research Letters, 45, 4982–4990, https://doi.org/https://doi.org/10.1029/2018GL077430, 2018.

Morim, J., Trenham, C., Hemer, M., Wang, X. L., Mori, N., Casas-Prat, M., Semedo, A., Shimura, T., Timmermans, B., Camus, P., Bricheno, L., Mentaschi, L., Dobrynin, M., Feng, Y., and Erikson, L.: A global ensemble of ocean wave climate projections from CMIP5-driven models, Scientific Data, 7, 1–10, https://doi.org/10.1038/s41597-020-0446-2, 2020.

Parker, W. S.: Ensemble modeling, uncertainty and robust predictions, WIREs Climate Change, 4, 213–223, https://doi.org/https://doi.org/10.1002/wcc.220, 2013.

Pomaro, A., Cavaleri, L., Papa, A., and Lionello, P.: Data Descriptor : 39 years of directional wave recorded data and relative problems , climatological implications and use, Scientific Data, 5, 1–12, https://doi.org/10.1038/sdata.2018.139, 2018.

Rasp, S., Pritchard, M. S., and Gentine, P.: Deep learning to represent subgrid processes in climate models, Proceedings of the National Academy of Sciences, 115, 9684–9689, https://doi.org/10.1073/pnas.1810286115, 2018.

Rodriguez-Delgado, C. and Bergillos, R. J.: Wave energy assessment under climate change through artificial intelligence, Science of The Total Environment, 760, 144 039, https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.144039, 2021.

Rus, M., Fettich, A., Kristan, M., and Ličer, M.: HIDRA2: deep-learning ensemble sea level and storm tide forecasting in the presence of seiches – the case of the northern Adriatic, Geoscientific Model Development, 16, 271–288, https://doi.org/10.5194/gmd-16-271-2023, 2023.

Scoccimarro, E., Gualdi, S., Bellucci, A., Sanna, A., Fogli, P. G., Manzini, E., Vichi, M., Oddo, P., and Navarra, A.: Effects of tropical cyclones on ocean heat transport in a high-resolution coupled general circulation model, Journal of Climate, 24, 4368–4384, https://doi.org/10.1175/2011JCLI4104.1, 2011.

Sonnewald, M., Lguensat, R., Jones, D. C., Dueben, P. D., Brajard, J., and Balaji, V.: Bridging observations, theory and numerical simulation of the ocean using machine learning, Environmental Research Letters, 16, 073 008, https://doi.org/10.1088/1748-9326/ac0eb0, 2021.

Steppeler, J., Doms, G., Schattler, U., Bitzer, H. W., Gassmann, A., Damrath, U., and Gregoric, G.: Meso-gamma scale forecasts using the nonhydrostatic model LM. Meteorology and Atmospheric Physics, Meteorology and Atmospheric Physics, 82, 75–96, 2003.

Tognin, D., D'Alpaos, A., Marani, M., and Carniello, L.: Marsh resilience to sea-level rise reduced by storm-surge barriers in the Venice Lagoon, Nature Geoscience, 14, 906–911, https://doi.org/10.1038/s41561-021-00853-7, 2021.

Trenberth, K. E., Dai, A., Rasmussen, R. M., and Parsons, D. B.: The Changing Character of Precipitation, Bulletin of the American Meteorological Society, 84, 1205–1217, https://journals-ametsoc-org.libezproxy.open.ac.uk/doi/pdf/10.1175/BAMS-84-9-1205, 2003.

USACE: Coastal Engineering Manual. Engineer Manual 1110-2-1100, U.S. Army Corps of Engineers, Washington, USA, 2002.

Žust, L., Fettich, A., Kristan, M., and Ličer, M.: HIDRA 1.0: deep-learning-based ensemble sea level forecasting in the northern Adriatic, Geoscientific Model Development, 14, 2057–2074, https://doi.org/10.5194/gmd-14-2057-2021, 2021.

Zollo, A. L., Rillo, V., Bucchignani, E., Montesarchio, M., and Mercogliano, P.: Extreme temperature and precipitation events over Italy: assessment of high-resolution simulations with COSMO-CLM and future scenarios, International Journal of Climatology, 36, 987–1004, https://doi.org/10.1002/joc.4401, 2016.