Local Ocean Wave Field Estimation Using A Deep Generative Model of Wave Buoys

Peihua Han, *Member, IEEE,* Hans Petter Hildre, Houxiang Zhang, *Senior Member, IEEE*

***Abstract*—Estimating oceanic wave fields from sparse observations has been a long-standing challenge in oceanography and an important environmental metric desired for maritime operations. The requirement for frequent real-time updates of the wave field within the local area poses difficulties for data assimilation approaches, as they can be computationally complex and rely on external atmospheric forcing. The relationship between the wave field and local sparse observations is embedded in reanalysis or hindcast data. We propose a data-driven deep learning model capable of estimating the local wave field using sparsely distributed floating wave buoys. This novel model simultaneously produces wave height, period, and direction, along with their respective uncertainties. In a year-long test period within a local fjord region characterized by complex wave patterns influenced by intricate geography, the proposed model demonstrates remarkable accuracy and efficiency in estimating wave fields. This study demonstrates the promising potential of data-driven deep learning models as an alternative to rapidly estimating the wave field.**

***Index Terms*—Ocean wave field estimation, wave buoys, neural networks.**

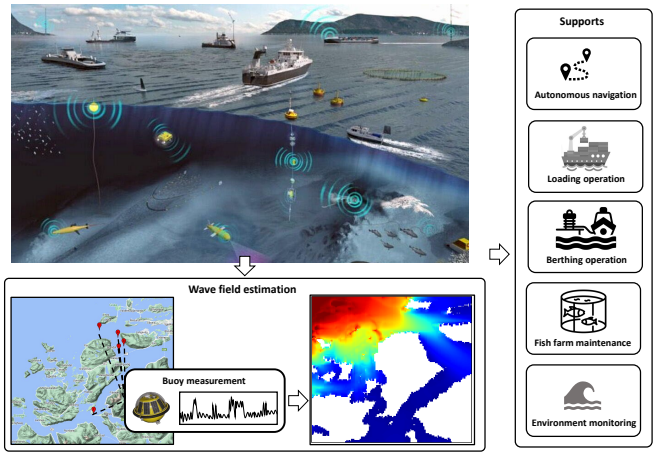


Fig. 1. An illustration of possible future digitalization transformation of the ocean in local areas. Various sensors will be placed in the local area to closely monitor the ocean. Ocean waves are one of the important parameters in this context, and frequent information on the entire ocean wave field will benefit various applications.

I. Introduction

Ocean waves are a critical factor that impacts all sea-related activities, such as daily planning, energy management, and transportation. Understanding the behavior of waves is critical to maximize economic opportunities, ensure safety at sea, and promote sustainable management of marine resources [1]. With the ongoing transformation toward digitization of ocean space in local areas (as illustrated in Fig. 1), an array of sophisticated sensors will be deployed for meticulous monitoring. Obtaining accurate ocean information is particularly important for guiding maritime activities, including fish farm maintenance, efficient docking of ships in the harbor, etc. To obtain precise ocean wave information in local areas, floating wave buoys have become the primary method [2], [3]. Although these buoys provide valuable wave statistics at their specific locations, their considerable distance from each other results in only point information for the entire area. Acquiring information about the current ocean wave field for the entire area can have numerous benefits over making point estimates only. For instance, the ship can use the ocean wave field information to plan routes and loading operations, or perform optimal berthing operations, and fish farms can schedule maintenance based on the ocean wave field information. This necessitates constructing ocean wave fields from limited in-situ observations obtained from wave buoys in the local area. The challenge is that localized areas may involve complex geographies, and wave statistics vary widely within about 200 m, requiring high-resolution and frequent updates of ocean wave fields.

In weather forecasting, estimating weather conditions is treated as a state estimation problem, where observations are combined with numerical models and data assimilation techniques to determine the current weather state and initiate forecasts. Similarly, for ocean waves, widely used numerical models like WaveWatch 3 [4] and SWAN [5] generate wave forecasts based on wind and boundary wave conditions. Data assimilation techniques refine these forecasts by incorporating observations. These refined forecasts are an estimate of the state of the ocean waves. The accuracy of numerical models depends on grid resolution, which poses a computational bottleneck that is challenging to overcome [6]. Standard data assimilation methods, such as 4D-VAR [7], require iterative solvers, making each iteration computationally intensive, similar to running a complete numerical model. Consequently, most operational wave models currently do not assimilate data or use sequential methods [8]. Even with the implementation of sequential data assimilation methods, real-time updates of the wave field pose challenges for local ocean wave field. Local areas are considered to be relatively small areas of around 25 km, and the local wave field is the wave information of this

P. Han, H. P. Hildre, and H. Zhang are with the Department of Ocean Operations and Civil Engineering, Norwegian University of Science and Technology, Ålesund 6009, Norway {peihua.han, hans.p.hildre, hozh}@ntnu.no

Manuscript received xx xx, 2023; revised xx xx, 2023.

area. The area can involve complex geographical features, such as rapid changes in water depth and potential obstruction to wind flow by high mountains. It therefore necessitates a high spatial resolution to accurately simulate intricate factors like wave refraction. Furthermore, the evolution of ocean waves is not solely an initial value problem. It also involves complex atmospheric forcing and boundary wave conditions obtained from global or regional simulations that may not be available. Therefore, these high spatial resolution and external input requirements pose challenges for local wavefield estimation using data assimilation.

On the other hand, reanalysis [9] or hindcast [10] is used to reconstruct historical ocean conditions. This process can be conducted offline with high grid resolution, where boundary wave conditions and atmospheric forces for the local area are obtained from global or regional reanalysis products. The availability of such high-resolution historical data enables the establishment of a relationship between wave buoy observations and the local wave field. In this paper, we propose a novel approach that employs deep neural networks to construct the ocean wave field based on sparse wave buoy observations. Neural network models learn the relationships between input observations and output variables directly from data. The objective is to learn a direct mapping from the wave buoys' observations to the local wave field. In this case, wave buoy observations serve as contextual information to generate a wave field. Once trained, the model provides an extremely fast way to generate wave fields that can be used for a range of tasks as shown in Fig. 1. Furthermore, this process can be inherently uncertain. Deep generative models offer a natural approach to modeling uncertainty. Such models can provide probabilistic outputs, which are valuable for decision-making purposes compared to deterministic outputs [11].

In this paper, a deep generative model is employed to construct the wave field. As demonstrated later, the model effectively reconstructed the ocean wave field within a challenging fjord environment, despite the limited availability of wave buoys. Waves are represented by three important parameters, providing detailed wave conditions rather than a single parameter in the literature. Importantly, it successfully captured the intricate relationship between wave height and periods. Additionally, the model demonstrated its capability to characterize and quantify the uncertainty associated with its outputs. Our findings highlight the remarkable potential of deep generative models in reconstructing fields of complex oceanic phenomena. The main contributions of this study are as follows:

- A deep generative model is developed to estimate the ocean wave field given sparse wave buoys' observations.
- The uncertainty of the estimation is well-represented by the model through a re-calibration procedure.
- The model provides rich wave information and captures the joint distribution of wave height and period.

II. Related works

A. Data assimilation

Data assimilation is commonly referred to as state estimation theory in geosciences [7], which is closely related to estimation theory and optimal control. Data assimilation takes a forecast (usually from a numerical evolution model) and applies a correction to the forecast based on observations. In this way, the best estimate of the state of the system is obtained. It is a standard practice in numerical weather prediction systems, especially for atmospheric forecasting systems [12]. Commonly used data assimilation methods can be categorized into variational approaches and sequential approaches. Variational methods [7], [12] are used to estimate the states that fit all the observations best within a prescribed observing window. It is formulated as an optimization problem to minimize a scalar error. The state estimate at an arbitrary time is therefore influenced by all observations within the prescribed window. The four-dimensional variational method (4D-Var) is one of the well-known methods that assumes a Gaussian background and observation error and a perfect model to establish the cost function. It is a standard practice today in atmospheric modeling such as in the European Centre for Medium-Range Weather Forecasts (ECMWF)[1]. The method has also been investigated to assimilate oceanographic data such as temperature, salinity, and currents [13], [14]. As for ocean waves, attempts have been made but developing an adjoint model is nontrivial due to the complex operators in the spectral space [15]. Furthermore, variational techniques require an iterative process to optimize the loss function, which can be computationally expensive. Therefore, the assimilation of ocean waves usually uses a sequential method [8]. The sequential methods alternate the forecast step with updated observations at each time step. By assuming Gaussian noise and a linear forecast model, the Kalman filter can be used. Smit et al. [8] used optimal interpolation (which can be viewed as a simplified Kalman filter by assuming a stationary covariance matrix) to perform data assimilation of the ocean wave. However, only significant wave height is assimilated. The same method is used to produce wave reanalysis products in the Black Sea[2], the Iberia-Biscay-Ireland area[3] and the Mediterranean Sea[4]. Similarly, Houghton et al. [16] applied the optimal interpolation framework to assimilate the Fourier coefficients of the wave spectrum at equidistant directions. In this way, the wave spectrum instead of a single bulk statistic is assimilated. In summary, data assimilation methods can be used to obtain (1) the best estimate of the current state to initialize forecasts; (2) the reanalysis of past data. The first term is closely related to our problem but it is usually performed at an interval of 6 hours, thus it is much less frequent than we desired. Besides, the wave assimilation problem is a mix of an initial value problem (swell) and a forced problem (wind) [16], where the atmospheric forcing field is needed. However, obtaining the atmospheric forcing field at a high frequency is also a challenge.

[1]ECMWF, https://www.ecmwf.int/

[2]Black Sea Waves Reanalysis, https://data.marine.copernicus.eu/product/BLKSEA_MULTIYEAR_WAV_007_006/description

[3]Atlantic -Iberian Biscay Irish- Ocean Wave Reanalysis, https://data.marine.copernicus.eu/product/IBI_MULTIYEAR_WAV_005_006/description

[4]Mediterranean Sea Waves Analysis and Forecast, https://data.marine.copernicus.eu/product/MEDSEA_ANALYSISFORECAST_WAV_006_017/description

B. Deep neural networks in geosciences

In recent years, geosciences have witnessed a major revolution from being a data-poor field to a data-rich field [17] and modern deep neural networks (DNN) have been intensively investigated. DNNs are capable of automatically learning complex patterns and relationships from large datasets, which can improve the accuracy of geoscientific models and predictions. Shoji et al. [18] used a convolutional neural network to classify volcanic ash particles according to their shape. The class probabilities returned by the network are used to determine the mixing ratio of ash particles with complex shapes. Lv et al. [19] developed a convolutional neural network to detect earthquakes from distributed acoustic sensing technology. It can achieve a high accuracy with a limited number of positive samples. Yao et al. [20] extended the visual transformer for multimodal learning and showed performance improvements on two remote sensing datasets. Deep neural network models have also been widely applied in hyperspectral image classification [21], [22], in which the features from different levels can be easily fused with neural network models. In addition to using DNNs for classification, the most relevant to our research is using them to estimate or predict weather variables. Ravuri et al. [11] performed high-resolution precipitation nowcasting using DNNs and showed that the model provides improved forecast quality, consistency, and value from the meteorologist evaluation. Zhao et al. [23] used a support vector machine for the hourly rainfall forecast. Espeholt et al. [24] developed a model that uses radar, assimilation, and geospatial features as input to produce twelve-hour precipitation forecasts at $1km \times 1km$ resolution. They show that the model outperforms state-of-the-art physics-based models currently operating in the Continental United States. Chen et al. developed a model that uses observations from the satellite and ground radar network as input to produce 12 hour precipitation forecasts at $1km \times 1km$ resolution. A similar study has been performed for ocean wave forecasting. Kim et al. [25] developed a DNNs-based ocean wave forecasting model to predict three integrated wave parameters up to 48 hours. The above models are all based on the principle that future predictions are generated by conditioning on past observations. For wave estimation, the wave field is generated by conditioning on the partial observations. Fablet et al. [26], [27] reconstructed the ocean wave field, e.g., sea surface temperature and sea surface height from satellite observations by learning a DNN model. It is shown that the model is more efficient and outperforms the optimal interpolation framework. Similarly, our aim is to reconstruct the ocean wave field as well, but we rely on sparser observations from wave buoys.

C. Deep generative models

Deep generative models are a type of model that can resemble data distribution and can be used to generate probabilistic outputs. These models learn a latent representation to represent the data distribution. By sampling from the latent representation, probabilistic outputs can be generated. There are several types of deep generative models, including variational autoencoders (VAEs) [28], generative adversarial networks (GANs) [29], and autoregressive models. These models have been used in [11], [24] to produce probabilistic precipitation forecasts. Probabilistic forecasting is expected to provide greater economic and decision-making value than deterministic ones. The same applies to wave field estimation. In addition, reconstructing wave fields from very sparse buoys' observations is a process that might involve large uncertainty, and making probabilistic estimation is necessary.

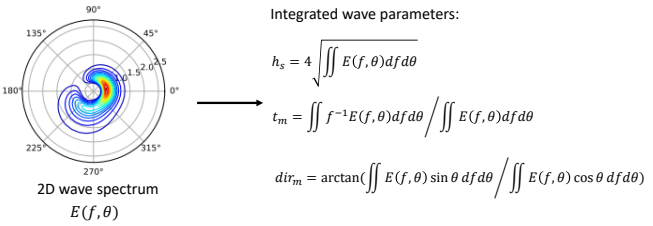


Fig. 2. The relationship of 2D wave spectrum and the three integrated wave parameters. The integrated parameters are used to represent the wave condition in this paper.

III. Deep generative model of wave buoys

A. Problem formulation

The proposed model is a conditional generative model that predicts the wave fields given past, or contextual, wave buoys' observations. The model includes latent random vectors z and can be described by:

$$p_\theta\left(X_t|x^{1:N}_{t-\delta t:t}\right) = \int p_\theta\left(X_t|z, x^{1:N}_{t-\delta t:t}\right) p_\theta\left(z|x^{1:N}_{t-\delta t:t}\right) dz \quad (1)$$

where X_t is the regional wave fields at time t. $x^{1:N}_{t-\delta t:t}$ are the N wave buoys' observations from time $t-\delta t$ to time t. θ is the parameters of the neural networks. z are the latent variables and the integration over z ensures that the model predictions are spatially dependent. For simplification, $x^{1:N}_{t-\delta t:t}$ will be denoted as c (context) and X_t will be denoted as X.

Learning of the parameter θ is framed within the framework of a conditional variational autoencoder with adversarial learning (CVAE-AL). Three consecutive wave buoys' observations (the past 20 min, wave buoys have an interval of 10 min) are used as context for the decoder that allows the sampling of multiple wave fields.

B. Ocean wave field representation

The ocean waves are represented by a 2D wave spectrum [3] that is discretized in direction and frequency. Integrated wave parameters are usually calculated from the 2D wave spectrum to represent the statistical properties of the wave. Here, three parameters are used to represent the wave condition at a single point [30]: significant wave height h_s, mean wave period t_m, and mean wave direction dir_m. Fig. 2 shows the relationship between the 2D wave spectrum and the integrated wave parameters.

For mean wave direction dir_m, $0°$ and $360°$ are essentially the same. We therefore encode dir_m into $(\cos dir_m, \sin dir_m)$. The dir_m can be easily reconstructed

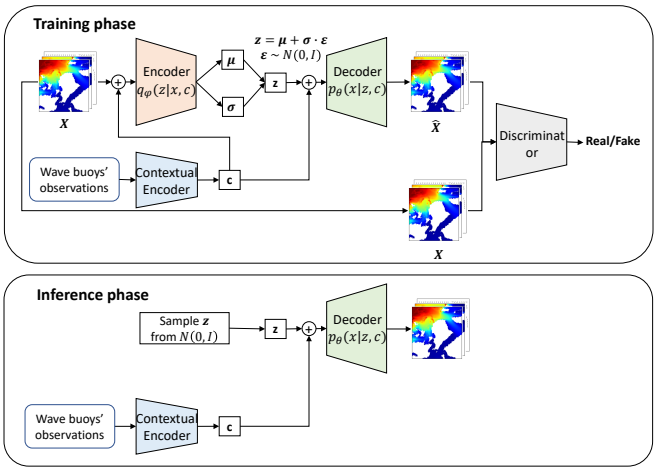


Fig. 3. System diagram depicting training procedure and inference procedure. The proposed model consists of an encoder, decoder, contextual encoder, and discriminator. Note that only the contextual encoder and decoder are used in the inference phase.

from $(\cos dir_m, \sin dir_m)$. In this way, the size of the original wave fields $(3, 128, 128)$ is transformed into $(4, 128, 128)$.

The $\cos dir_m$ and $\sin dir_m$ lie in the range of [-1, 1]. Similarly, the significant wave height h_s and the mean wave period t_m are normalized to the range of [-1, 1].

C. Conditional variational autoencoder with adversarial learning

The conditional variational autoencoder [31] is an extension of variational autoencoders (VAE) [28] that provides control over the VAE data generation process. CVAE consists of an encoder $q_\phi(z|X, c)$ and a decoder $p_\theta(X|z, c)$. The encoder $q_\phi(z|X, c)$ transforms the sampled points X and context c into the latent variables z. The decoder $p_\theta(X|z, c)$ reconstructs the sampled points x from latent variables z and context c. Note that the encoder and decoder are modeled in the structure of the neural network which is parametrized by ϕ and θ, respectively. The objective of CVAE is to maximize the variational lower bound, also called the evidence lower bound (ELBO), of the intractable marginal log-likelihood of data $\log p_\theta(X|c)$:

$$\log p_\theta(X|c) \geq \mathbb{E}\left[\log p_\theta(X|z, c)\right] - D_{KL}\left(q_\phi(z|X, c) \parallel p_\theta(z|c)\right) \tag{2}$$

where D_{KL} is the Kullback-Leibler (KL) divergence. Instead of maximizing the ELBO, the negative ELBO is usually used as the objective and the training is to minimize $-\log p_\theta(X|c)$. The negative ELBO can be viewed as the sum of reconstruction loss and KL divergence.

Reconstruction loss. For the negative log-likelihood $-\mathbb{E}\left[\log p_\theta(X|z, c)\right]$, we use mean square error (MSE):

$$l_{rec} = \parallel X - \mathbb{E}\left[p_\theta(X|q_\phi(z|X, c), c)\right] \parallel_2 \tag{3}$$

Even though MSE is often used as the reconstruction loss for VAE, it is problematic for maximum likelihood since it could be over-regularized during training [32]. However, we expect a large sample diversity and add a hyperparameter to the KL loss to balance regularization and reconstruction. Therefore, MSE is still used in this paper.

KL loss. The prior distribution $p_\theta(z|c)$ is modelled as a standard multi-variate Gaussian distribution $p_\theta(z|c) \sim \mathcal{N}(0, I)$. The encoder $q_\phi(z|X, c)$ is a deep neural network that outputs the mean and standard deviation of z, which is denoted as $\mu_\theta(X, c)$ and $\sigma^2_\theta(X, c)$, respectively. Sampling from $z \sim q_\phi(z|X, c)$ can be done using the reparameterization trick [28]. Therefore, the KL loss can be given by:

$$l_{kl} = \frac{1}{d}\sum \frac{1}{2}\left[\mu^2_\theta(X, c) + \sigma^2_\theta(X, c) - 1 - \log \sigma^2_\theta(X, c)\right] \tag{4}$$

where d is the the dimension of latent vector z.

a) Adversarial learning: To adopt adversarial training in the learning framework, a discriminator that distinguishes between the output generated by the decoder and the ground truth wavefield is added. The adversarial loss in training the generator and the discriminator are given by:

$$l_{adv-G} = l_{bce}\left(D_\varphi(\hat{X}), 1\right) \tag{5}$$

$$l_{adv-D} = \frac{1}{2} l_{bce}\left(D_\varphi(\hat{X}), 0\right) + \frac{1}{2} l_{bce}\left(D_\varphi(X), 1\right) \tag{6}$$

where D_φ is the discriminator parametrized by φ. $\hat{X}$ is the reconstructed wave fields from the decoder. l_{bce} is the binary cross-entropy loss that is defined as $l_{bce}(\hat{y}, y) = y \log \hat{y} + (1 - y) \log(1 - \hat{y})$.

b) Training and inference framework: Fig. 3 shows the training and inference procedures of our CVAE-AL framework. The model contains a contextual encoder that encodes the observations from the wave buoys into a contextual vector c. The encoder and decoder follow the CVAE framework that generates the latent vector and reconstructed wave fields conditioning on the context vector c. The discriminator is used for adversarial training.

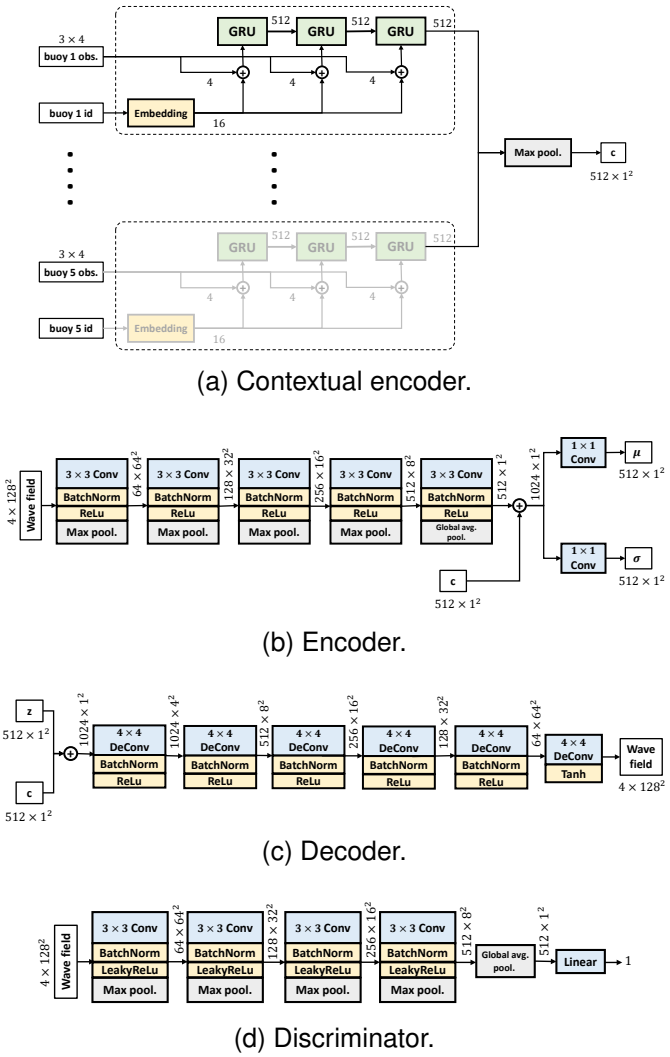
The four modules, context encoder, encoder, decoder, and discriminator, are parameterized by deep neural networks. The detailed network architectures for these four modules can be found in the supplementary materials. We consider 3 important wave parameters: significant wave height, mean wave period, and mean wave direction. In short, the estimated wave fields have 3 dimensions instead of 1.

Training. The training of the generator (including a contextual encoder, encoder, and decoder) and discriminator involves the minimization of the following loss function, respectively:

$$l_G = l_{rec} + \lambda_{kl} l_{kl} + \lambda_{adv} l_{adv-G} \tag{7}$$

$$l_D = l_{adv-D} \tag{8}$$

where λ_{kl} and λ_{adv} are two hyperparameters to balance the KL loss and adversarial loss, respectively. During training, we alternate between one gradient descent step on l_G, and then one step on l_D.



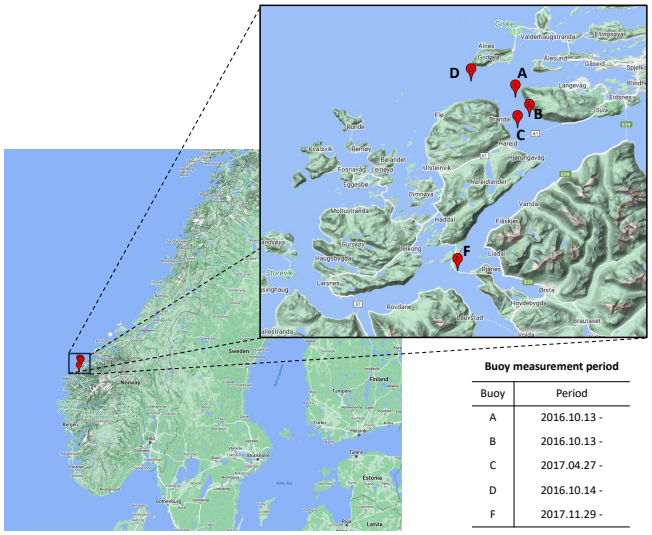
(a) Contextual encoder.

(b) Encoder.

(c) Decoder.

(d) Discriminator.

Fig. 4. Detailed network architectures.



Buoy measurement period

Buoy	Period
A	2016.10.13 -
B	2016.10.13 -
C	2017.04.27 -
D	2016.10.14 -
F	2017.11.29 -

Fig. 5. Spatial distribution of the wave buoys in the Sulafjord.

Inference. After the model is trained, only the contextual encoder and decoder will be used in the inference phase. Wave buoys' observations are fed into the contextual encoder to generate the context c. Then latent vector z is obtained by sampling from the standard multivariate Gaussian distribution $\mathcal{N}(0, I)$. The decoder outputs the wave field based on the combination of the context c and latent vector z. Now we can sample different wave fields by sampling the latent vector z.

D. Network architectures

This section introduces the network architecture. The model follows a VAE structure and therefore involves an encoder and a decoder for encoding the wave field into a latent space and reconstructing it. A contextual encoder is designed to encode the time series wave buoys' measurement into a conditional vector to guide the encoding and decoding process. A discriminator is designed to enable adversarial training to penalize high-level statistics to avoid blurry outputs.

a) Contextual encoder.: The primary objective of the contextual encoder is to encode the observations from the wave buoys into a compact vector c. The objective of the contextual encoder is to encode arbitrary number of wave buoys. As illustrated in Fig. 4a, we employ a shared gated recurrent unit (GRU) to encode the data from each wave buoy, resulting in a buoy embedding. To ensure distinctiveness, we encode the unique ID of each wave buoy into a positional embedding, which is then concatenated with the corresponding observation at each time step. All buoy embeddings are collectively processed through a max pooling layer to generate the context vector c. An essential advantage of this approach is that it utilizes a shared GRU for all wave buoys and a symmetric max pooling layer to aggregate information, thereby enabling the contextual encoder to effectively handle an arbitrary number of wave buoys.

b) Encoder.: The encoder adopts a deep convolutional neural network architecture, incorporating modules in the form of convolution-BatchNorm-ReLu, along with a max pooling layer. The 4×4 convolution is used to downsample the wave field. As depicted in Fig. 4b, the encoder receives the 4-dimensional wave field and context as input and produces the mean and variance of the latent vector as output.

c) Decoder.: The decoder utilizes both the contextual vector c and latent vector z to generate the wave field. It shares a similar architecture with the deep convolutional generative adversarial network [33]. In particular, the decoder employs the 4×4 transposed convolution operator to carry out the upsampling operations. For a comprehensive view of the decoder's architecture, refer to Fig. 4c.

d) Discriminator.: The discriminator also adopts a deep convolutional neural network architecture, but with LeakyReLu as the activation function, in contrast to ReLu. Its input is the wave field, and its task is to discriminate between real and decoder-generated wave fields. For a more detailed view of the discriminator's architecture, refer to Fig. 4d.

IV. Experimental setup

A. Dataset

The dataset provided has been collected from the Norwegian Meteorological Institute in Sulafjorden, Norway. Sulafjorden is known for its dynamic water depth, resulting in a complex wave pattern. The dataset encompasses a comprehensive collection of data that spans three years, from January 1, 2017, to February 29, 2020. The training dataset comprises the initial two years of data, while the remaining one year of data serves as the test dataset.

Observations from five wave buoys in this area are used. The location of each wave buoy is shown in Fig. 5 and the deployment date is also listed. The chosen area exemplifies a typical fjord region characterized by an abundance of islands, towering mountains, and deep waters. As a result, wind waves and swells exist simultaneously in this area, giving rise to a diverse and ever-changing wave pattern that exhibits considerable spatial and temporal variations. Note that buoys C and F are deployed after the starting date of the collected dataset. They are filled with NaN values and the contextual encoder is designed to be able to handle different numbers of wave buoys. Raw measurements from wave buoys are inevitably affected by various degradations, noise effects or variability. This effect has been discussed in the literature for many different data sources [34]. Here, the collected data has gone through quality control. In addition, the erroneous observations, e.g. extremely large values, are filtered out manually.

The wave field data are collected from a hindcast model simulation with the wave model SWAN in a 250m grid and has a time resolution of 1 hour. Even though it is hindcast data rather than reanalysis data, it shows a good correlation with actual observations. The original hindcast data covers the geospatial region from latitude 62.0° to 62.6° and longitude 5.3° to 6.8°. We extract only the region in the size of (128, 128) that are interesting to us.

Three parameters are selected to represent the wave conditions: significant wave height h_s, mean wave period t_m, and mean wave direction dir_m. Therefore, only these three parameters are extracted from wave buoys and hindcast data.

B. Evaluation metrics

Per-grid-cell metrics are computed over all target grid cells in all examples in the test dataset. These target grid cells are indexed by i. In the following, the model's prediction for target grid cell i will be denoted as P_i, while the corresponding ground truth will be denoted as O_i. n denotes the number of target grid cells in a single prediction.

a) Root mean square error (RMSE).: This metric gives a continuous measure of the accuracy of deterministic predictions:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(P_i - O_i\right)^2} \tag{9}$$

b) Continuous ranked probability score (CRPS).: CRPS is a measure of how good predictions are in matching observed outcomes, which is widely used in forecasting groups to evaluate ensemble weather forecasting. The per-grid-cell CRPS is defined as:

$$CRPS_i = \mathbb{E}\left|P - O_i\right| - \frac{1}{2}\mathbb{E}\left|P - P'\right| \tag{10}$$

where P and P' are drawn independently from the predictive distribution P_i. We compute the estimates of $CRPS_i$ with the $N = 10$ ensemble members as samples. These per-grid-cell CRPS are then averaged over all grid cells as $CRPS = 1/n\sum_{i=1}^{n} CRPS_i$.

Note that for significant wave height hs and mean wave period t_m, we use the above two metrics directly. To properly evaluate mean wave direction dir_m, we replace $|P_i - O_i|$ in the above two metrics with $\min(|P_i - O_i|, 360 - |P_i - O_i|)$ since 0° and 360° are essentially the same.

V. Experimental results

A. Accuracy during test period

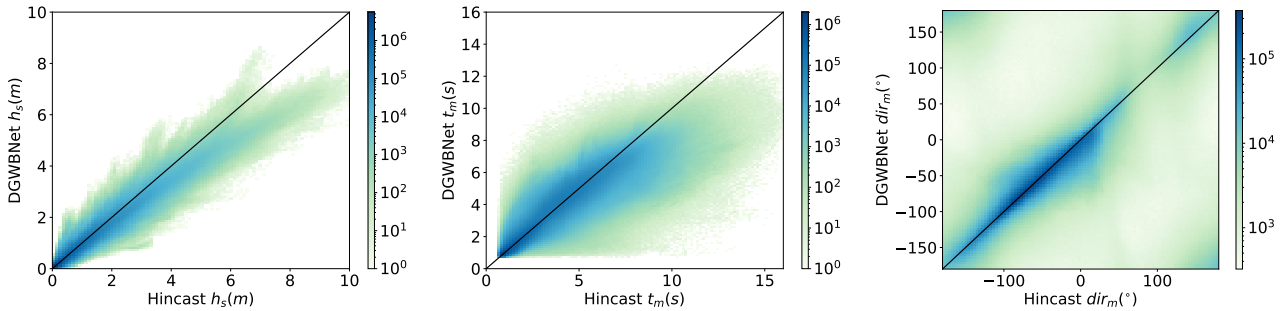
The predicted wave parameters from DGWBNet versus the hindcast wave parameters are calculated for all grid cells in the one-year test period to obtain the histogram plots as shown in Fig. 6. In general, the predicted wave parameters match the hindcast wave parameters, especially in the low wave height and wave period region. Conservative predictions are observed when the wave height and period are large.

The temporal and spatial distribution of the RMSE of the DGWBNet were calculated sample by sample from the estimation errors at all grids during the test period of 1 year and presented in Fig. 7 and Fig. 8, respectively. Analysis of the data, depicted in Fig. 7a and Fig. 7b, reveals that the RMSE of significant wave height and mean wave direction exhibits a slight increase during winter compared to summer, which might be due to the fact that rough sea conditions are more prevalent during the winter season. Conversely, during summer, the RMSE of the mean wave direction displays a slight elevation, potentially due to the inherent challenges in accurately observing wave direction in calm seas.

Referring to Fig. 8, it is evident that the RMSE for significant wave height is notably higher in the open sea area compared to other regions. This elevated RMSE can be attributed to the considerable variability present in the open sea, compounded by the significant distance between the wave buoys and this particular area. Conversely, the RMSE for the mean wave period exhibits a substantially higher value in the vicinity of the coastal region. This disparity can be attributed to various factors specific to the coastal dynamics. Furthermore, the fjord region demonstrates a higher RMSE for the mean wave direction, primarily owing to the predominant influence of wind patterns in this particular area.

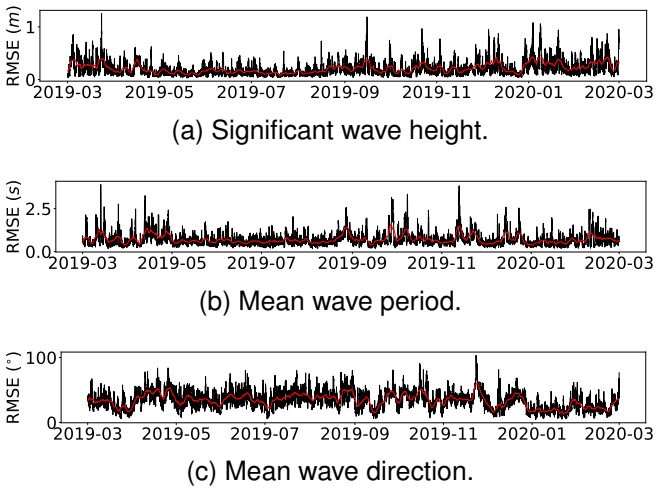
B. Joint distribution of wave height and period

To assess the model's ability to capture the wave pattern, scatter plots depicting the relationship between significant wave height and mean wave direction are presented for six



(a) Significant wave height. (b) Mean wave period. (c) Mean wave direction.

Fig. 6. Histogram plots of the hindcast data versus the outputs from the DGWBNet. Note that the histogram is in the log scale.



(a) Significant wave height.

(b) Mean wave period.

(c) Mean wave direction.

Fig. 7. Temporal distribution of RMSE during the test period. The red line represents the moving average of RMSE.

carefully chosen points within the region during the test period, as illustrated in Fig. 9. Specifically, three points in the open sea region and three points in the fjord region were selected for analysis. As shown in Fig. 9b, the proposed model successfully reconstructs the correlation between significant wave height and mean wave period. However, it is worth noting that the model's distribution appears to be slightly tighter than that of the observed data, suggesting that extreme cases may not be fully captured by the model. Furthermore, a noticeable distinction can be observed between the points located in the open sea area and those situated within the fjord region. This distinction becomes particularly evident when comparing points 1 and 2 with points 5 and 6. The differentiation can be attributed to the presence of swells and wind waves, with open sea areas primarily influenced by waves originating from the boundary, while fjord areas are predominantly impacted by local wind patterns.

C. Uncertainty calibration

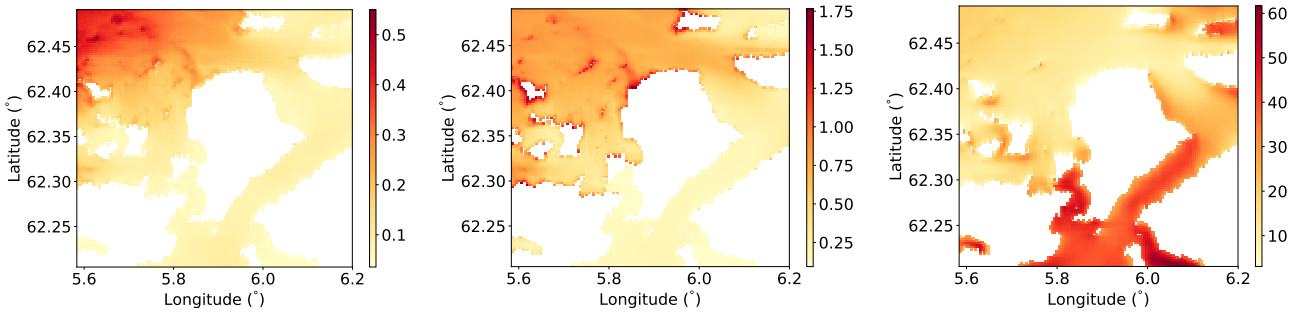
One of the key advantages of using a deep generative model is its ability to capture and represent the inherent uncertainty in wave estimation. To evaluate the prediction uncertainty of our proposed model, we employed a procedure involving the generation of 10 estimates for a given scenario. These estimates were then used to approximate the mean and standard deviation for each grid point. A held-out validation set was maintained, and this procedure was repeated for all scenarios within the validation set. Subsequently, we constructed an α-prediction, aiming to encompass observed values $\alpha\%$ of the time. By iterating different values of α, we assessed the proportion of validation data that fell within the prediction interval. The calibration plot depicted in Fig. 10a shows the predicted proportion of validation data expected to lie within the interval on the x-axis, while the observed proportion of validation data within the interval is represented on the y-axis. However, it is worth noting that our proposed model exhibits a slight tendency to be over-confident in its predictions. In order to enhance our model's ability to accurately represent uncertainty, we performed a re-calibration process, inspired by [11]. This involved scaling the variance of the latent vector during the inference phase, specifically by sampling from $\mathcal{N}(0, 2.5I)$ instead of the standard multivariate Gaussian distribution $\mathcal{N}(0, I)$. This modification resulted in a superior representation of uncertainty, as demonstrated in Fig. 10b, yielding more reliable results.

Fig. 8 presents the estimated significant wave heights obtained using the calibration model at four selected time stamps. Notably, the model not only generates estimates but also provides valuable information regarding the associated uncertainties. It is observed that the uncertainty in the estimates is predominantly concentrated in the upper left portion, corresponding to the open-sea area. Furthermore, as the actual wave height increases, the uncertainty in the estimates also tends to rise.

D. Comparison with other models

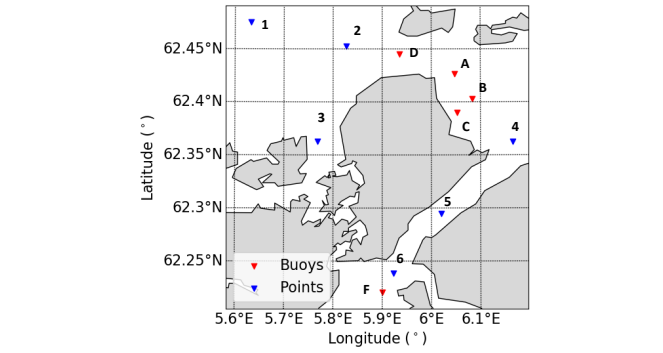
To assess the model's performance, we compute per-grid-cell metrics for point prediction and then aggregate them to derive full-grid metrics.

RMSE is employed to quantify the deviation between the actual wave field and the estimated wave field. Since our proposed model makes probabilistic predictions, we generate 10 samples from the probabilistic estimation and calculate their mean to evaluate the RMSE effectively. For probabilistic prediction assessment, we utilize CRPS, a proper scoring rule for univariate distributions. CRPS enables us to score the per-grid-cell marginals of the model's predictive distribution

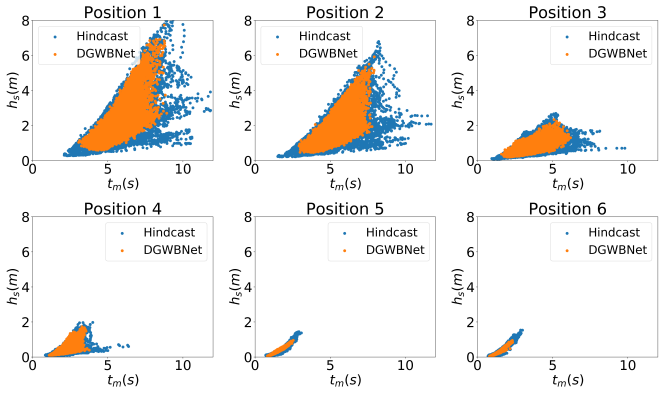


(a) Significant wave height. (b) Mean wave period. (c) Mean wave direction.

Fig. 8. Spatial distribution of RMSE during the test period.

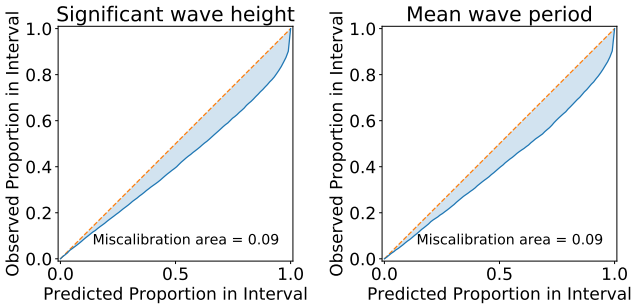


(a) 6 selected points in the map.

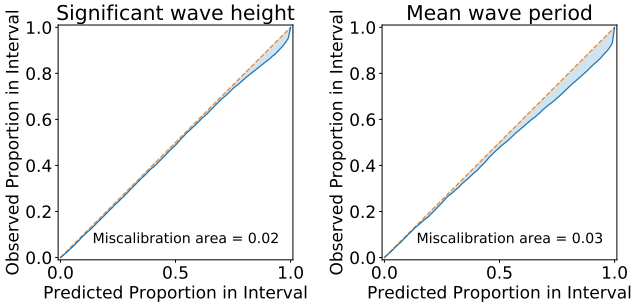


(b) joint distribution of the 6 points.

Fig. 9. Joint distribution of significant wave height and mean wave period for 6 selected points.



(a) before re-calibration.



(b) after re-calibration.

Fig. 10. Joint distribution of significant wave height and mean wave period for 6 selected points.

against the actual wave field, providing a robust evaluation. Table. I also presents the computational complexity of different neural network models, measured by the number of parameters and multiply-accumulate operations (MACs).

a) Baseline models: Four baselines are used in this paper:

- MyWaveWAM 800m: As mentioned above, data assimilation is usually not performed for ocean waves. One of the best practices for knowing the wave field in a particular area is to use a wave forecasting system. This suggests that the current wave field is actually represented by forecasting hours ago. MyWaveWAM 800m is the Norwegian coastal wave forecasting system from the Meteorological Institute of Norway. The operational numerical wave model is WAM and it is run on an 800-meter grid. The model is run with ECMWF and AROME atmospheric forcing. Note that this model has an 800-meter grid and therefore we perform a linear interpolation to downscale the result to our desired 200-meter grid.
- k-nearest neighbor (kNN): The second baseline is a k-nearest neighbor approach. We save the wave field and its corresponding wave buoys' observations in a database. Then in the inference phase, the distances between the input wave buoys' observations and those in the database are calculated. the top k corresponding wave fields are the estimated wave fields. We use k=10 for ensemble estimation and k=1 for deterministic estimation.
- GRU-Conv: The model is implemented in an encoder-decoder structure, where the GRU is used as the encoder to compress the time series wave buoys' observations and

TABLE I
QUANTITATIVE RESULTS OF ALL THE METHODS ON THE DATASET.

Model		MyWaveWAM 800m	kNN-1	kNN-10	GRU-Conv	Transformer-Conv	DGWBNet
RMSE	$h_s(m)$	0.192	0.217	0.175	0.204	0.234	**0.157**
	$t_m(s)$	3.017	0.688	0.538	0.651	0.667	**0.526**
	$dir_m(^\circ)$	83.46	27.20	36.81	30.66	27.06	**21.25**
CRPS	$h_s(m)$	-	-	0.133	0.177	0.199	**0.119**
	$t_m(s)$	-	-	0.401	0.550	0.549	**0.397**
	$dir_m(^\circ)$	-	-	**7.53**	19.08	17.31	11.19
Params(M)		-	-	-	**20.34**	22.70	28.75
MACs(G)		-	-	-	**2.43**	2.46	2.49

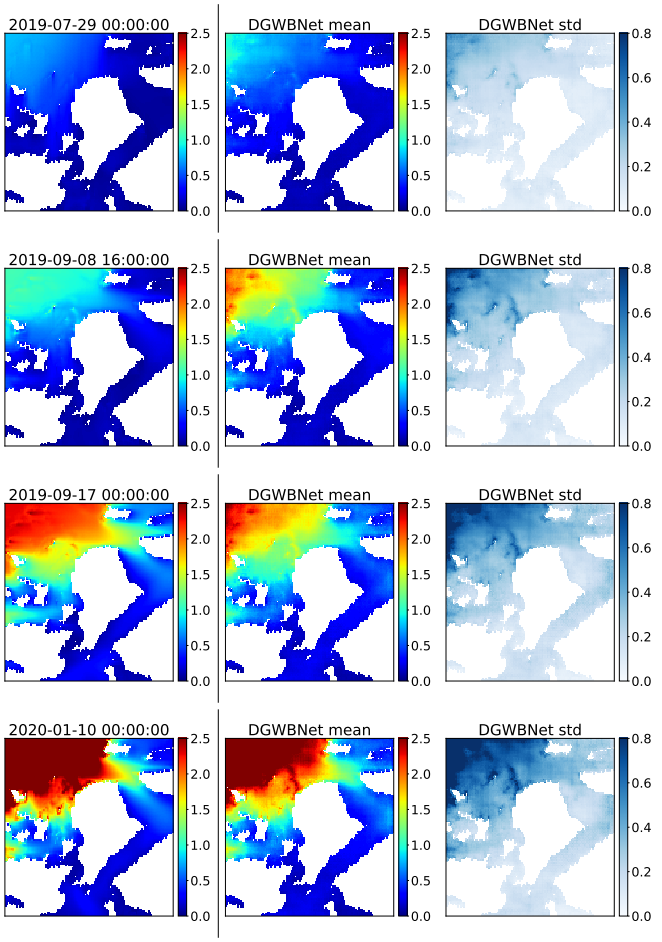


Fig. 11. Significant wave height estimation from four selected time stamps.

a convolutional neural network with transpose convolutional operation is used as the decoder to output the wave field. To produce probabilistic estimates, the output of the network on each grid is assumed to follow a Gaussian mixture model. The model is trained by minimizing the negative log-likelihood.

- Transformer-Conv: The model is similar to GRU-Conv. The transformer is used instead of GRU as the encoder to summarize all information in the buoys' observations. Similarly, to produce probabilistic estimates, a Gaussian mixture model is used for the output of the network on each grid.

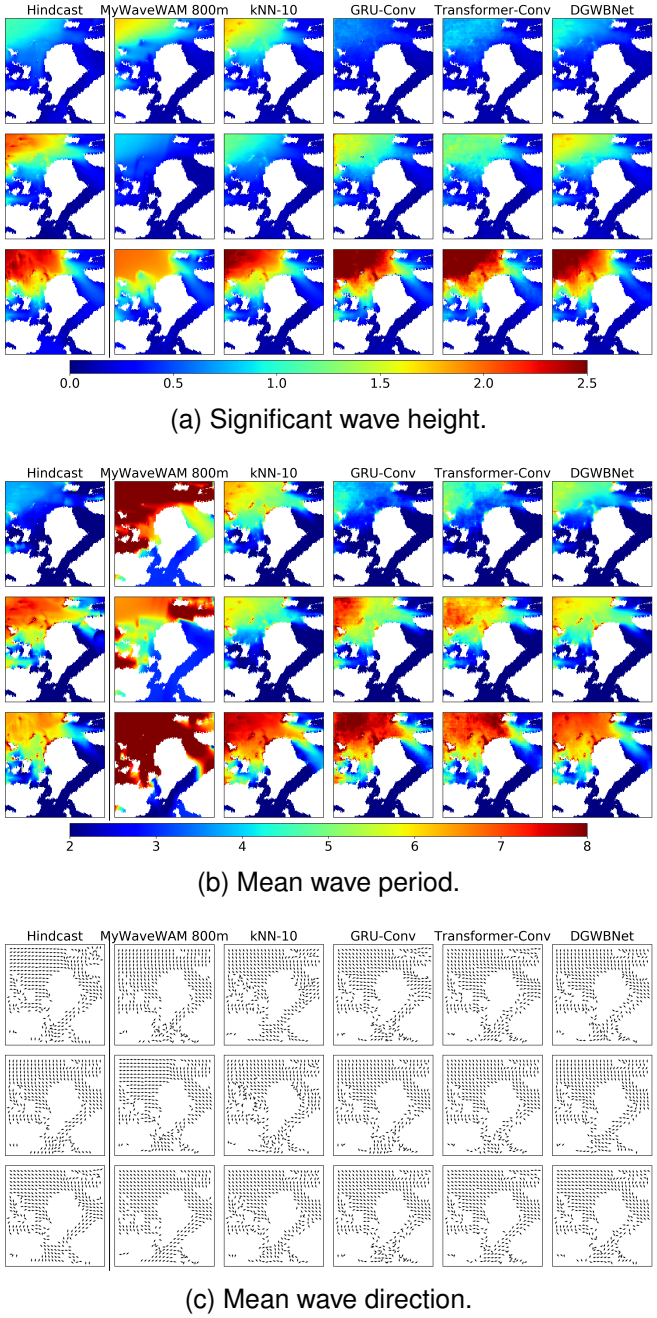


Fig. 12. Three randomly sampled wave fields and the estimates from different approaches

b) Results: Table. I shows the overall performance of our model compared with other baseline methods. The wave

TABLE II
ABLATION STUDY OF THE DISCRIMINATOR.

DGWBNet		w Discriminator	w/o Discriminator
RMSE	$h_s(m)$	**0.157**	0.172
	$t_m(s)$	**0.526**	0.557
	$dir_m(^\circ)$	**21.25**	23.35
CRPS	$h_s(m)$	**0.119**	0.135
	$t_m(s)$	**0.397**	0.409
	$dir_m(^\circ)$	11.19	**10.02**

forecast model MyWaveWAM 800m can be found to provide a reasonable error for significant wave height. However, it fails to characterize the wave period and direction in this region. All data-driven models are able to provide a low error in wave period and direction. The proposed model provides the least error in terms of RMSE and CRPS except for the CRPS for mean wave direction, where the k-nearest neighbor model with 10 neighbors gives the least error. The computational complexity of the proposed model is slightly higher than the two baseline models. However, performance is significantly improved.

Fig. 12 shows the estimated significant wave height, mean wave period, and mean wave direction of 3 random scenarios in the test set of different methods, respectively. It can be found that the MyWaveWAM 800m can reproduce the overall distribution of wave heights, but lacks detail due to its low resolution. In contrast, all learning-based methods are able to reconstruct detailed wavefields. This conclusion is the same for the mean wave period and the mean wave direction. The proposed DGWBNet is able to provide accurate and detailed sea state information.

E. Ablation study

To validate the design choice of the proposed model, an ablation study is performed. In particular, we conducted experiments by ablating the discriminator. Table II shows that including the discriminator module improves all the performance except the CRPS for mean wave direction.

VI. CONCLUSIONS

The ocean wave phenomenon plays a crucial role in influencing various activities taking place in marine environments. The rapid advancement of digitalization has highlighted the need for fast and accurate wave field estimation within local areas. Complex geographical features in local areas, such as rapid changes in water depth and frequent land-water interactions, demand high-resolution grids to accurately characterize the wave field. Moreover, the lack of external atmospheric forcing data adds to the challenges. In response, we propose DGWGNet, a novel approach for estimating wave fields in such intricate regions. The remarkable performance of DGWBNet in reconstructing the wave field using sparse wave buoy data, even under moderate and rough sea conditions, its ability to estimate uncertainty and capture variations, as well as its independence from physical simulations, underscore the feasibility, effectiveness, and strong potential of deep learning techniques in accurately estimating oceanic phenomena fields.

A deep learning model has the capability to generate highly accurate wave field estimates with minimal prior information, such as sparse wave buoy data, as demonstrated in our case. Furthermore, the model can be easily adapted and implemented in other areas of interest. The majority of the model's computational time is spent during the learning procedure to optimize the adjustable weights. Once trained, the deep learning model can efficiently generate forecasts without the need for iterative processes, resulting in rapid and real-time wave field predictions. In our study, it took only seconds to generate the wave field for the local area, highlighting the model's computational efficiency. By providing accurate wave fields quickly, a range of offshore applications can benefit, such as the aforementioned fish farms, ship operations, and garbage collection, which can all benefit from understanding wave fields. Future work will consider using the full wave spectrum as a representation rather than 3 integrated wave parameters. Validation of the model in different dataset and its integration with standard data assimilation methods, e.g. using model outputs as initial conditions, will be investigated.

REFERENCES

[1] B. G. Reguero, I. J. Losada, and F. J. Méndez, "A recent increase in global wave power as a consequence of oceanic warming," *Nature communications*, vol. 10, no. 1, p. 205, 2019.

[2] P. Han, G. Li, X. Cheng, S. Skjong, and H. Zhang, "An uncertainty-aware hybrid approach for sea state estimation using ship motion responses," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 2, pp. 891–900, 2021.

[3] P. Han, G. Li, S. Skjong, and H. Zhang, "Directional wave spectrum estimation with ship motion responses using adversarial networks," *Marine Structures*, vol. 83, p. 103159, 2022.

[4] H. L. Tolman, "A third-generation model for wind waves on slowly varying, unsteady, and inhomogeneous depths and currents," *Journal of Physical Oceanography*, vol. 21, no. 6, pp. 782–797, 1991.

[5] N. Booij, L. Holthuijsen, and R. Ris, "The "swan" wave model for shallow water," *Coastal Engineering 1996*, vol. 1, pp. 668–676, 1997.

[6] P. Bauer, A. Thorpe, and G. Brunet, "The quiet revolution of numerical weather prediction," *Nature*, vol. 525, no. 7567, pp. 47–55, 2015.

[7] A. Carrassi, M. Bocquet, L. Bertino, and G. Evensen, "Data assimilation in the geosciences: An overview of methods, issues, and perspectives," *Wiley Interdisciplinary Reviews: Climate Change*, vol. 9, no. 5, p. e535, 2018.

[8] P. Smit, I. Houghton, K. Jordanova, T. Portwood, E. Shapiro, D. Clark, M. Sosa, and T. Janssen, "Assimilation of significant wave height from distributed ocean wave sensors," *Ocean Modelling*, vol. 159, p. 101738, 2021.

[9] H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers *et al.*, "The era5 global reanalysis," *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 730, pp. 1999–2049, 2020.

[10] J. Perez, M. Menendez, and I. J. Losada, "Gow2: A global wave hindcast for coastal applications," *Coastal Engineering*, vol. 124, pp. 1–11, 2017.

[11] S. Ravuri, K. Lenc, M. Willson, D. Kangin, R. Lam, P. Mirowski, M. Fitzsimons, M. Athanassiadou, S. Kashem, S. Madge *et al.*, "Skilful precipitation nowcasting using deep generative models of radar," *Nature*, vol. 597, no. 7878, pp. 672–677, 2021.

[12] R. Bannister, "A review of operational methods of variational and ensemble-variational data assimilation," *Quarterly Journal of the Royal Meteorological Society*, vol. 143, no. 703, pp. 607–633, 2017.

[13] G. Broquet, C. Edwards, A. Moore, B. Powell, M. Veneziani, and J. Doyle, "Application of 4d-variational data assimilation to the california current system," *Dynamics of Atmospheres and Oceans*, vol. 48, no. 1-3, pp. 69–92, 2009.

[14] M. da Rocha Fragoso, G. V. de Carvalho, F. L. M. Soares, D. G. Faller, L. P. de Freitas Assad, R. Toste, L. M. B. Sancho, E. N. Passos, C. S. Böck, B. Reis *et al.*, "A 4d-variational ocean data assimilation application for santos basin, brazil," *Ocean Dynamics*, vol. 66, pp. 419–434, 2016.

[15] M. D. Orzech, J. Veeramony, and H. Ngodock, "A variational assimilation system for nearshore wave modeling," *Journal of Atmospheric and Oceanic Technology*, vol. 30, no. 5, pp. 953–970, 2013.
[16] I. A. Houghton, C. Hegermiller, C. Teicheira, and P. B. Smit, "Operational assimilation of spectral wave data from the sofar spotter network," *Geophysical Research Letters*, vol. 49, no. 15, p. e2022GL098973, 2022.
[17] A. Karpatne, I. Ebert-Uphoff, S. Ravela, H. A. Babaie, and V. Kumar, "Machine learning for the geosciences: Challenges and opportunities," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 8, pp. 1544–1554, 2018.
[18] D. Shoji, R. Noguchi, S. Otsuki, and H. Hino, "Classification of volcanic ash particles using a convolutional neural network and probability," *Scientific reports*, vol. 8, no. 1, pp. 1–12, 2018.
[19] H. Lv, X. Zeng, F. Bao, J. Xie, R. Lin, Z. Song, and G. Zhang, "Adenet: A deep neural network for das earthquake detection trained with a limited number of positive samples," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2022.
[20] J. Yao, B. Zhang, C. Li, D. Hong, and J. Chanussot, "Extended vision transformer (exvit) for land use and land cover classification: A multimodal deep learning framework," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.
[21] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 7, pp. 5966–5978, 2020.
[22] Z. Chen, D. Hong, and H. Gao, "Grid network: Feature extraction in anisotropic perspective for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, 2023.
[23] Q. Zhao, Y. Liu, W. Yao, and Y. Yao, "Hourly rainfall forecast model using supervised learning algorithm," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–9, 2021.
[24] L. Espeholt, S. Agrawal, C. Sønderby, M. Kumar, J. Heek, C. Bromberg, C. Gazen, R. Carver, M. Andrychowicz, J. Hickey *et al.*, "Deep learning for twelve hour precipitation forecasts," *Nature communications*, vol. 13, no. 1, p. 5145, 2022.
[25] J. Kim, T. Kim, J. Yoo, J.-G. Ryu, K. Do, and J. Kim, "Stg-oceanwavenet: Spatio-temporal geographic information guided ocean wave prediction network," *Ocean Engineering*, vol. 257, p. 111576, 2022.
[26] R. Fablet, M. Beauchamp, L. Drumetz, and F. Rousseau, "Joint interpolation and representation learning for irregularly sampled satellite-derived geophysical fields," *Frontiers in Applied Mathematics and Statistics*, p. 25, 2021.
[27] M. Beauchamp, Q. Febvre, H. Georgenthum, and R. Fablet, "4dvarnet-ssh: end-to-end learning of variational interpolation schemes for nadir and wide-swath satellite altimetry," *Geoscientific Model Development Discussions*, pp. 1–37, 2022.
[28] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.
[29] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.
[30] P. Han, G. Li, S. Skjong, B. Wu, and H. Zhang, "Data-driven sea state estimation for vessels using multi-domain features from motion responses," in *2021 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2021, pp. 2120–2126.
[31] K. Sohn, H. Lee, and X. Yan, "Learning structured output representation using deep conditional generative models," *Advances in neural information processing systems*, vol. 28, 2015.
[32] R. Yu, "A tutorial on vaes: From bayes' rule to lossless compression," *arXiv preprint arXiv:2006.10273*, 2020.
[33] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.
[34] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 1923–1938, 2018.