CAMBRIDGE
UNIVERSITY PRESS

APPLICATION PAPER

Multimodal learning–based reconstruction of high-resolution spatial wind speed fields

Matteo Zambra[1,2] , Nicolas Farrugia[1,2], Dorian Cazau[3,2], Alexandre Gensse[4] and Ronan Fablet[1,2]

[1]IMT Atlantique, Brest, France
[2]Lab-STICC, Brest, France
[3]ENSTA Bretagne, Brest, France
[4]Naval Group, Toulon, France
Corresponding author: Matteo Zambra; Email: matteo.zambra1@gmail.com

Abstract

Wind speed at the sea surface is a key quantity for a variety of scientific applications and human activities. For its importance, many observation techniques exist, ranging from in situ to satellite observations. However, none of such techniques can capture the spatiotemporal variability of the phenomenon at the same time. Reanalysis products, obtained from data assimilation methods, represent the state-of-the-art for sea-surface wind speed monitoring but may be biased by model errors and their spatial resolution is not competitive with satellite products. In this work, we propose a scheme based on both data assimilation and deep learning concepts to process spatiotemporally heterogeneous input sources to reconstruct high-resolution time series of spatial wind speed fields. This method allows to us make the most of the complementary information conveyed by the different sea-surface information typically available in operational settings. We use synthetic wind speed data to emulate satellite images, in situ time series and reanalyzed wind fields. Starting from these pseudo-observations, we run extensive numerical simulations to assess the impact of each input source on the model reconstruction performance. We show that our proposed framework outperforms a deep learning–based inversion scheme and can successfully exploit the spatiotemporal complementary information of the different input sources. We also show that the model can learn the possible bias in reanalysis products and attenuate it in the output reconstructions.

Impact Statement

This application article presents a novel approach to sea-surface wind speed monitoring. It is addressed to a multidisciplinary audience involving geophysicists, signal analysts, and machine learning practitioners in the fields of geosciences. Because of its multidisciplinary scope, this work invests as much in the scientific as in the operational aspects.

1. Introduction

The characterization, quantification, and study of wind speed at the sea surface are primary interests for a broad ensemble of scientific and operational purposes. Because of the spatiotemporal-scale heterogeneity, different sensors and observation techniques have been developed and implemented through the years to

target some given features of the flow field. In situ measures (Gould et al., 2013) allow us to observe the phenomenon near-continuously in time with good measurement accuracy but limited spatial coverage. On the other hand, remote-sensing products (Amani et al., 2022), especially synthetic aperture radar (SAR) imagery (Monaldo et al., 2013), can observe a spatially wide region at high spatial resolution, but the revisit time of the remote sensor of a given region may not suffice to capture the temporal evolution of surface wind patterns. Numerical weather prediction (NWP) approaches, based on mathematical modeling, are another source of wind speed information. Environmental variables *reanalyses* (Valmassoi et al., 2023) are obtained by data assimilation methods. Despite being state-of-the-art for wind speed forecasting and reconstruction, they may be biased by timing and/or intensity errors (Storto et al., 2019) and their spatial resolution may not suffice to resolve the smaller scales. In last years, machine learning modeling gained popularity among the geophysical communities thanks to its efficacy in modeling high-dimensional dynamical systems and to the ever-growing volumes of oceanic observations and data bases (Karpatne et al., 2017, Karpatne et al., 2018; Yu and Ma, 2021; Bergen et al., 2019). The problem of wind speed forecasting and estimation has been the object of many studies (Arabi et al., 2023; Saxena et al., 2021; Arnold and Asgarimehr, 2021). In this work, we shift the focus of the analysis on the problem of reconstructing complete, high-resolution (HR) wind speed fields time series from its partial observations. In other words, the objective is to exploit the complementary information heterogeneous input sources to retrieve the HR spatiotemporal information related to the sea-surface wind speed. This kind of *inverse problems* (Snieder and Trampert, 1999) is typically solved with data assimilation methods (Bannister, 2017; Carrassi et al., 2018) and deep learning–based inversion techniques (Barth et al., 2020; Manucharyan et al., 2021) that are used to spatiotemporally interpolate sparse geophysical fields. Recent work proposed to bridge data assimilation and deep learning techniques (Arcucci et al., 2021; Farchi et al., 2022). Here, we elaborate on the work of Fablet et al., (Fablet et al., 2021a; b). They propose a framework, the 4DVarNet, which is inspired by the 4DVar variational data assimilation scheme (see Talagrand, 2015) and based on deep learning modeling. The 4DVar inversion scheme is kept explicitly and is parameterized in some of its parts by trainable neural network operators, giving and end-to-end trainable and differentiable architecture. The appeal of this method is that the time-dependent processes stated by the 4DVar state-space formulation better inform the learning-based part of the inversion.

Given the scale heterogeneity of the input observations mentioned above, the wind speed retrieval from its partial observations is an inherently *multimodal* problem (Hong et al., 2020). We hypothesize that the HR wind speed time series reconstruction can profit from the complementary information in the input data. We design an observing system simulation experiment (Hoffman and Atlas, 2016) based on synthetic wind speed data to test the 4DVarNet framework for the reconstruction task and assess the impact of each input source. HR satellite products, in situ time series, and low-resolution (LR) NWP products are obtained by the original synthetic wind speed fields. We prove that the 4DVarNet inversion outperforms a vanilla deep learning–based inversion scheme. We also show how the 4DVarNet can jointly exploit the HR spatial information of the satellite pseudo-observation and the temporal information of the local in situ time series. To conclude, we show that the 4DVarNet, suitably parameterized to ingest a multimodal dataset, can attenuate the bias imputable to errors in LR NWP products.

The rest of this article is structured as follows. Section 2 gives an overview of the dataset used. Section 3 discusses the methodological aspects of the 4DVar inversion, the models used, and the experimental settings. Section 4 presents and discusses the results, and Section 5 closes the paper and restates the main highlights of our analyses.

2. Data

For this work, we prioritize the use of a simulation dataset not to have co-location and grid compatibility issues between satellite and reanalysis fields and time series. We use the output of the RUWRF mesoscale model (Optis et al., 2020), based on the version 4.1.2 of the WRF model (Powers et al., 2017). This model
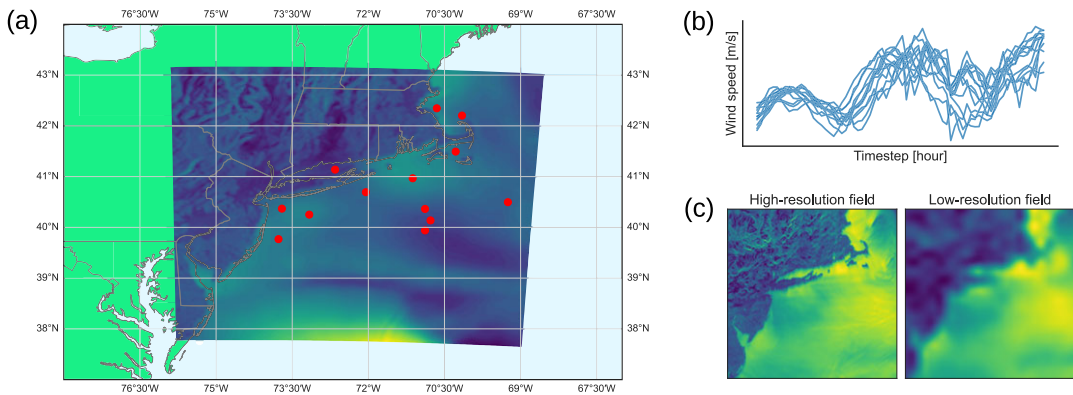
Figure 1. *Qualitative description of the dataset. Panel (a): Geographical region and buoys positions. Panel (b): In situ pseudo-observations. Panel (c). Original HR fields and the emulated LR fields.*

runs a parent nest with a resolution of 9 km for a time interval of 120 hours and a child nest with a resolution of 3 km ($\sim 0.03\,°$) out of 48 hours. The resolution of the child nest is the reference HR for our analyses. The region selected (a portion of the North American East Coast) has dimensions of 644×645 km and the complete time series selected ranges from January 1, 2019, 00:00 to January 1, 2021, 23:00. Panel (a) of Figure 1 shows the region selected. The dataset is formatted as a collection of 24-hour time series, from 00:00 to 23:00 for each day. We process the wind speed components into the vector norm to work with the wind speed modulus.

LR data. We obtain LR fields to emulate data assimilation products. Reanalyses are well suited for mesoscale phenomena, characterized by spatial scales ranging from tens to hundreds of kilometers and temporal scales ranging from hours to some days. We manufacture the LR fields by downsampling the original data and reinterpolating the result on the HR grid, see Panel (c) of Figure 1. The resulting data are prepared to have a spatial and temporal resolution of 30 km and 6 hours, at hours 00:00, 06:00, 12:00, 18:00, and 23:00. The spatial resolution chosen matches the resolution of the ECMWF ERA-5 data base products (Hersbach et al., 2020).

HR data. We use the original fields to emulate noise-free HR satellite pseudo-observations. Given the typical temporal sampling frequency of SAR products, we choose to prepare our dataset to have the HR fields with a temporal sampling frequency of 12 hours. In particular, we set the HR observations at hours 06:00 and 18:00 of each input time series.

In situ *time series.* We obtain local in situ time series by the pixel values of the original HR fields on the positions of the NOAA National Data Buoy Center network (reference site: https://www.ndbc.noaa.gov/). The buoys selected are those that were active in the time window and spatial extent chosen. In situ pseudo-observations have hourly temporal resolution.

2.1. Experimental configurations

We can identify four experimental configurations for our experiments as combinations of the input sources described above. We can state these configurations as follows. (i) SR: only LR fields, (ii) C1: LR fields and HR fields with frequency of 12 hours, (iii) C2: LR fields and in situ observations with hourly sampling frequency, (iv) C3: LR fields, HR fields with frequency of 12 hours and hourly in situ time series. The LR fields have sampling frequency of 6 hours in all the configurations. The name SR for the first configurations refers to the *superresolution* task. The model is required to reconstruct the ground truth HR fields time series from partial time series of LR fields. These configurations are used to assess the value of each input source on the overall model reconstruction performance. Note that these configurations relate only with the combinations on input sources and do not refer to model architectures.

3. Methods

The classical 4DVar data assimilation scheme is based on the following state-space formulation

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathcal{M}(\mathbf{x}(t)) + \boldsymbol{\eta}(t) \\ \mathbf{y}(t) = \mathcal{H}(\mathbf{x}(t)) + \boldsymbol{\epsilon}(t) \end{cases} \tag{1}$$

where $\mathbf{x} \in X$ is the objective state variable, $\mathbf{y} \in Y$ is the partial observation of $\mathbf{x}$ provided by the observation operator $\mathcal{H} : X \to Y$. The spaces $X \subset \mathbb{R}^m$ and $Y \subset \mathbb{R}^d$ are high-dimensional state and observation spaces. One typically has that $d < m$. The operator $\mathcal{M} : X \to X$ encodes the time evolution of the state variable and $\boldsymbol{\eta} \in \mathbb{R}^m$ and $\boldsymbol{\epsilon} \in \mathbb{R}^d$ are independent identically distributed noise processes. The objective is to *invert* the forward processes $\mathcal{M}$ and $\mathcal{H}$ from the observations to retrieve the state variable. In this case, the observations contain information about HR and LR fields and in situ time series. The state variable contains the information about the complete HR wind speed fields. The state-space formulation is discretized and the continuous dynamical prior $\mathcal{M}$ is replaced by the one-step-ahead predictor $\Phi :$ $\mathbb{R} \times X \to X$ defined as

$$\Phi(t + \Delta t, \mathbf{x}) = \mathbf{x}(t) + \int_t^{t + \Delta t} \mathcal{M}(\mathbf{x}(t)) \, dt \tag{2}$$

The 4DVar scheme states the problem as the minimization of the following variational cost[1]

$$\begin{aligned} U_\Phi(\mathbf{x}, \mathbf{y}; \Omega) &= \lambda_1 \|\mathcal{H}(\mathbf{x}) - \mathbf{y}\|^2 + \lambda_2 \|\mathbf{x} - \Phi(\mathbf{x})\|^2 \\ &= \lambda_1 \|\mathbf{x} - \mathbf{y}\|_\Omega^2 + \lambda_2 \|\mathbf{x} - \Phi(\mathbf{x})\|^2 \end{aligned} \tag{3}$$

As usual in inverse problems, the variational cost involves a data proximity term and a prior knowledge constraint. In this case, this constraint refers to prior physical knowledge. The symbol $\|\cdot\|$ is an L_2 norm. The parameters $\lambda_{1,2}$ are tunable weights, and Ω represents the spatiotemporal sampling domain. The observation operator $\mathcal{H}$, in our case, enforces the temporal sampling frequencies of each input source, as prescribed by the configurations introduced above. For this reason, the observation term can be stated compactly as the distance between observations and the state of the domain Ω.

3.1. Direct learning–based inversion

To fix a baseline, we choose to use a vanilla learning–based inversion scheme. This kind of trainable scheme has proven effective in imaging problems (Ongie et al., 2020). This inversion can be stated as follows:

$$\mathbf{x} = f_\theta(\mathbf{y}) \tag{4}$$

where f_θ is a neural network parameterized by the parameters $\boldsymbol{\theta}$. This neural network is trained to reconstruct the full-time series of HR wind speeds from the time series of partial observations, with no need for the variational cost formulation. This kind of *end-to-end* learning scheme directly relates the observations to the state variable. Some examples of learning-based direct inversion to perform spatio-temporal interpolation of satellite products are provided by the work of Barth et al. (2020), for sea-surface temperature, and Manucharyan et al. (2021) for sea-surface height.

3.2. Trainable data assimilation scheme

In the classical 4DVar, the operator Φ is implemented by Euler or Runge–Kutta ODE integration schemes (Cash, 2003). In the 4DVarNet case, Φ is parameterized by a neural network operator (Fablet et al., 2021a). By this design choice, the flow operator is parameterized and trained to learn the state-space

---

[1] The expression of the variational cost uses the compact matrix notation. The explicit summation over time steps is omitted.

dynamics of the state variable. This makes the variational cost itself a learnable function. The minimization of the variational cost is performed by a second neural network operator, the gradient solver Γ. The iterative rule to update the state variable is the following:

$$\begin{cases} \boldsymbol{\delta}^k = \Gamma\left(\nabla_{\mathbf{x}} U_{\Phi}\left(\mathbf{x}^k, \mathbf{y}, \Omega\right)\right) \\ \mathbf{x}^{k+1} = \mathbf{x}^k - \alpha \boldsymbol{\delta}^k \end{cases} \tag{5}$$

where k refers to the iteration number. For our experiments, we set the total number of iterations to 5. The 4DVarNet scheme leverages the *automatic differentiation* capabilities of modern deep learning frameworks to evaluate the gradients of the variational cost (Baydin et al., 2018). Figure 2 represents the 4DVarNet scheme graphically, the organization of the input observations and the output reconstructions and the architectures of the trainable networks used. The 4DVarNet can be seen as a *bi-level optimization* framework to jointly learn the variational model and solver (Fablet and Drumetz, 2020). As the state variable is available from the variational cost minimization, it is used to evaluate a training loss function. This loss function is in turn optimized to fit the trainable networks parameters. These two steps are repeated for the number of training epochs selected by the user.

The 4DVarNet scheme has been successfully applied to sea-surface height (Beauchamp et al., 2023), sediment (Vient et al., 2022), and turbidity (Dorffer et al., 2023). The open-source 4DVarNet code base is available on the following repositories, the started version: https://github.com/CIA-Oceanix/4dvarnet-starter and the core version: https://github.com/CIA-Oceanix/4dvarnet-core.

Multimodal version. We mentioned that the problem is inherently multimodal. Both the direct inversion and the 4DVarNet configuration discussed do not process the heterogeneous HR, LR, and local components of the observations **y** separately. We propose the following modification to account explicitly for this input source heterogeneity. The observation term in the variational cost has a further term that evaluates the distance between the feature maps of the observations and the state variable. These feature maps are not constrained by the spatiotemporal characteristics of the observation components and can better relate. The new variational cost reads
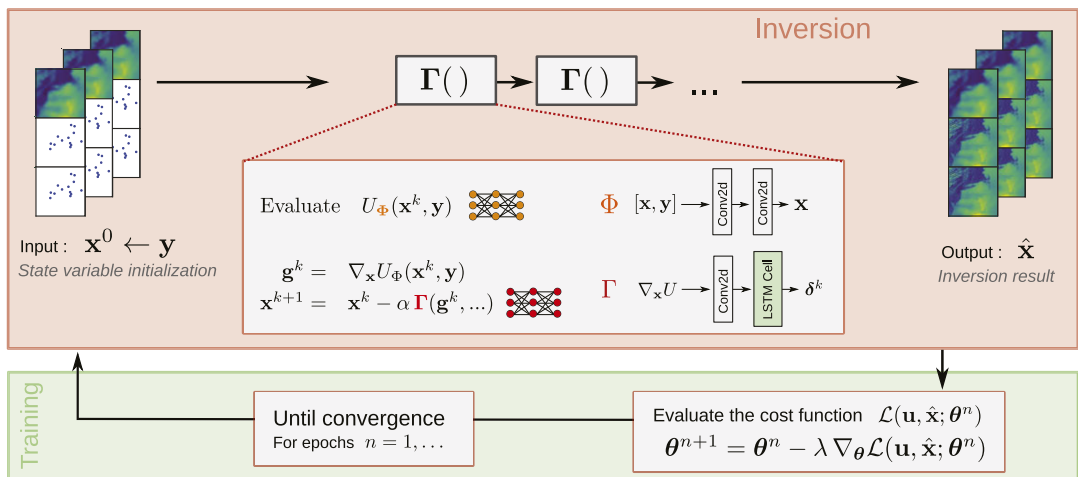
Figure 2. *Schematic illustration of the 4DVarNet framework. The observations and the state variable are prepared as concatenations of the input sources available. The symbol θ represents the networks Φ and Γ parameters and n denotes the training epoch index. The symbol **u** represents the ground truths. The statement $\mathbf{x}^0 \leftarrow \mathbf{y}$ means that the initial guess of the state variable is initialized with the observations. The red box contains the inversion part to solve for the state variable **x**, being the parameters θ fixed as of the previous training iteration. The green box highlights the parameters training part to solve for θ, with the state variable **x** fixed as returned by the inversion part.*

$$U_\Phi(\mathbf{x}, \mathbf{y}; \Omega) = \lambda_1 \|\mathbf{x} - \mathbf{y}\|_\Omega^2 + \lambda_1 \|\psi_\mathbf{x}(\mathbf{x}) - \psi_\mathbf{y}(\mathbf{y})\|^2 + \lambda_2 \|\mathbf{x} - \Phi(\mathbf{x})\|^2 \qquad (6)$$

where $\psi_{\mathbf{x},\mathbf{y}}$ is the trainable neural networks. For simplicity, we call *single-modal* the 4DVarNet based on the plain variational cost 3 and we call *multimodal* the 4DVarNet based on the variational cost 6. In both the single and multimodal versions of the 4DVarNet, we choose to train the weights $\lambda_{1,2}$ as well.

3.3. Numerical implementation

The observations and state variable objects are obtained by concatenating the LR and anomaly fields, see Figure 2. The anomaly is the difference between the HR and LR fields. The LR represents the average field information, while the anomaly is the deviation from the LR of the HR field. This choice prevents the model to process redundantly the LR information. We choose a 2D convolutional neural network to parameterize both for the direct inversion and the dynamical prior Φ of the 4DVarNet. The gradient solver Γ is parameterized by a 2D convolutional LSTM (Hochreiter and Schmidhuber, 1997). The networks $\psi_{\mathbf{x},\mathbf{y}}$ are composed by 2D or 1D convolutional layers depending on the experimental configuration. Recall that by configuration we mean the input data combinations. 2D layers process spatial image fields and 1D layers are well suited for multivariate time series.

3.4. Learning scheme

The models are trained with the following supervised mean-squared-error cost function:

$$\mathcal{L}(\mathbf{u}, \widehat{\mathbf{x}}) = \frac{1}{M} \sum_{i=0}^{M} \sum_{t=0}^{T} \lambda_l \|\mathbf{u}_{it}^{lr} - \widehat{\mathbf{x}}_{it}^{lr}\|^2 + \lambda_h \|\mathbf{u}_{it}^{hr} - \widehat{\mathbf{x}}_{it}^{hr}\|^2 + $$
$$ + \lambda_g \|\nabla \mathbf{u}_{it}^{hr} - \nabla \widehat{\mathbf{x}}_{it}^{hr}\|^2 + \lambda_p \|\widehat{\mathbf{x}}_{it}^{hr} - \Phi\left(\widehat{\mathbf{x}}_{it}^{hr}\right)\|^2 \qquad (7)$$

Let $\mathbf{u}$ and $\widehat{\mathbf{x}}$ be the ground truths and the reconstructions, respectively. The symbols M and T represent the number of elements in each data batch and the time window length, that is, 24 hours. To evaluate the loss function, we obtain the HR fields summing the LR and anomaly components of the reconstructed state variable. The spatial gradients term enforces the spatial coherence of the output fields. These gradients, evaluated with the finite difference method available in the Pytorch library, are computed with respect to the coordinates that define the fields spatial grid. Note that these spatial finite difference-based gradients are evaluated numerically. This has nothing in common with the gradients of the variational cost, as in the first line of Equation 5, which are evaluated by automatic differentiation. The last training loss term acts as a regularization and enforces the parameters to be learned in compliance with the dynamics specified by the first equation of the state-space formulation 1. This term should not be dominant, so the weight λ_p is set to a small value. In this case, $\lambda_p = 10^{-3}$. We train the model for 50 epochs with the Adam algorithm (Kingma and Ba, 2014). Training is regularized by the L_2 method and by early stopping based on the validation loss.

The learning rates for Φ, Γ, and the parameters $\lambda_{1,2}$ are, respectively, 10^{-4}, 10^{-3}, and 10^{-4}. The weight decay coefficients are, in the same order, 10^{-5}, 10^{-5}, and 10^{-8}. The coefficients λ_l, λ_h, and λ_g are, respectively, 2.5, 15, and 10^{-2}. In the multimodal 4DVarNet case, the models $\psi_{\mathbf{x},\mathbf{y}}$ have learning rate and weight decay coefficient set to 10^{-4} and 10^{-7}. We used a batch size 8 for all the simulations[2].

4. Results

In the presentation of the results, we adopt the following notation. Let LDI be the direct inversion baseline. Let 4DVN-SM denote the single-modal 4DVarNet version and 4DVN-MM the multimodal version. Let

---

[2] The simulations are run on a machine mounting 8 Nvidia A100-SWM4 graphical processor units. These units have 80 GB memory and GA100 graphics processors with 1593 MHz memory clock speed. Training times are reported in Section 4.

the symbols {LDI, 4DVN-SM, 4DVN-MM}-{SR, C[1,2,3]} indicate the model-data configuration combinations. These abbreviations are, for readers' convenience, restated systematically in part A of Table 1. The target of each model are the ground truth HR original fields.

4.1. Evaluation framework

To set up the evaluation scheme, we chose a proper reference metric for the sea-surface wind speed reconstruction. In the literature, reference values for the root-mean-squared error (RMSE) can be found for SAR-derived wind speed fields (Monaldo et al., 2013; Ahsbahs et al., 2020) and reanalyzed products (Brune et al., 2021; Gualtieri, 2021; Potisomporn et al., 2023). Part B of Table 1 reports the reference values provided by the cited work. The RMSE score is evaluated for SAR and reanalyses using in situ observations and assuming that these observations match the true wind values. In our case, we use synthetic simulated data and we have no in situ observations to be used as ground truth.[3] To fix a reference performance value, we do the following. We interpolate the LR fields in time on the 24-hour window. Recall that our experimental configurations have LR fields with observation frequency of 6 hours. We then evaluate the RMSE between these interpolated time series and the ground truth HR fields. Let B-RMSE represent this metric, which attains the value of 1.1234 m s^{-1}. We assume that our HR ground truth fields match the real physical phenomenon. In this way, we can directly compare our B-RMSE with the RMSE range found in the literature for reanalyzed fields, that is, $1.35–1.9$ m s^{-1}. Our B-RMSE attains a lower value than the lower bound of this interval. This is due to the fact that our interpolated LR fields, simulating real reanalyses, are not affected by model errors, as happens for real reanalyses. In this way, we can fix our B-RMSE as a quantitative reference for this case study. That is, if only LR NWP products were available. Next, we evaluate how two classes of trainable models (direct inversion and 4DVarNet) give better reconstruction performance leveraging (i) learnable modules and/or (ii) heterogeneous input observations. We may emphasize that we choose as baseline the superresolution direct inversion model LDI-SR because it is the simplest method that aims to reconstruct the HR ground truth fields, while the LR interpolation approach does not. Therefore, we keep the mentioned B-RMSE as quantitative reference, but it would not be a fair baseline to assess the 4DVarNet performance improvement.

The reconstructions of our models are evaluated with the RMSE score related to an ensemble of 10 model runs. We assess the RMSE in two ways. First, we collect the reconstruction RMSEs between each model's output and the ground truths. We evaluate the average RMSE and the standard deviation as measures of centrality and spread. In formulas

$$\mathrm{Avg\,RMSE}(\widehat{\mathbf{x}}, \mathbf{u}) = \frac{1}{R}\sum\nolimits_{r=1}^{R}\mathrm{RMSE}(\widehat{\mathbf{x}}_r, \mathbf{u}) = \frac{1}{R}\sum\nolimits_{r=1}^{R}\sqrt{\frac{1}{M}\sum\nolimits_{i=1}^{M}(\widehat{\mathbf{x}}_r - \mathbf{u})^2}$$

$$\mathrm{Std\,RMSE}(\widehat{\mathbf{x}}, \mathbf{u}) = \sqrt{\frac{1}{R}\sum\nolimits_{r=1}^{R}(\mathrm{Avg\,RMSE} - \mathrm{RMSE}(\widehat{\mathbf{x}}, \mathbf{u}))^2}$$

(8)

where $\mathbf{u}$ is the ground truth, $\widehat{\mathbf{x}}_r$ is the model r reconstruction, R is the number of model runs, equivalent to the ensemble members, and M is the number of samples. Second, the RMSE is evaluated between the ground truths and the median of an ensemble of 10 model runs (Rincy and Gupta, 2020). In formulas

$$\mathrm{Median\ RMSE}(\widehat{\mathbf{x}}, \mathbf{u}) = \mathrm{RMSE}(\mathbf{u}, \mathrm{Median}(\{\widehat{\mathbf{x}}_r; r=1, \ldots, R\}))$$

(9)

We define a relative gain to compare the model improvement w.r.t. the chosen baseline as $(1 - p_m/p_b) \times 100$, where p_m and p_b are the performance of the model and baseline, respectively.

4.2. Model benchmark

The 4DVarNet is compared against the LDI-SR model. This baseline model is trained to perform a superresolution task to retrieve the finer-scaled information from the LR input fields. The neural network

---

[3] Do not confuse our in-situ *pseudo*-observations as true in-situ values.

Table 1. *Global overview on the benchmark experiments*

Table A—*Table of the abbreviations used in the results presentation. The symbol ✖ states that a given input data source is missing. 1 h, 6 h, and 12 h stand, respectively, for observation sampling frequency of 1, 6, and 12 h.*

Model	Explanation	Data configuration	LR	HR	In situ
LR-int	LR fields interpolation	SR	6 h	✖	✖
LDI	Direct inversion	C1	6 h	12 h	✖
DVN-SM	Single-modal 4DVarNet	C2	6 h	✖	1 h
DVN-MM	Multimodal 4DVarNet	C3	6 h	12 h	1 h

Table B—*Benchmark test results. To contextualize our results, the first two rows report the typical reconstruction errors expected when using SAR Sentinel-1A imagery and the wind speed reanalyses of the ECMWF ERA-5 catalog. In the second part of this Table, we report the RMSE scores for our simulations, expressed by Equations 8 and 9. Relative gains are expressed in percentage. The gains are referred to the LDI-SR baseline, marked in* *orange**. The black* **boldface** *highlights the best result. We follow the names conventions stated in part A of this Table.*

Method	RMSE [m s^{-1}]			References	
SAR Sentinel–1A	1.4–1.6			Monaldo et al. (2013); Ahsbahs et al. (2020)	
ECMWF ERA5	1.35–1.9			Potisomporn et al. (2023); Brune et al. (2021)	
Model	**RMSE** [m s^{-1}]		**Gain [%]**	**Δ Gain** (C3 vs C1) [%]	
	Avg ± Std	**Median**			
LR-int		1.1234			
LDI	SR	0.9968 ± 0.0003	0.9960		
	C1	0.9617 ± 0.0004	0.9605	3.56	3.91–3.56 = 0.35
	C2	0.9964 ± 0.0002	0.9957	0.03	
	C3	0.9586 ± 0.0004	0.9571	3.91	
4DVN-SM	C1	0.9073 ± 0.0018	0.9000	9.64	9.65–9.64 = 0.01
	C2	0.9673 ± 0.0008	0.9619	3.42	
	C3	0.9066 ± 0.0016	0.8999	9.65	
4DVN-MM	C1	0.8983 ± 0.0016	0.8802	11.63	13.48–11.63 = **1.85**
	C2	0.9319 ± 0.0015	0.9197	7.66	
	C3	**0.8876 ± 0.0035**	**0.8617**	**13.48**	

used for the direct learning–based inversion shares the same architecture used to parameterize the dynamical prior Φ of the 4DVarNet framework. Part B of Table 1 lists systematically the simulations results. We can see that the 4DVarNet, in both the single and multimodal versions, outperforms the direct inversion baseline. However, the 4DVN-SM model does not benefit fully from the in situ observations, and indeed, the baseline does. The rightmost column of the table evaluates the difference between the performance associated to the configuration C3 w.r.t. the configuration C1. That is, the added value of in situ time series. In the case of 4DVN-MM, thanks to the features maps term in the variational cost 6, the framework can effectively benefit from the in situ time series. Figure 3 illustrates the spatial distribution of the time-average gain of the C3 configuration w.r.t. the C1 configuration, for both the 4DVN-SM and 4DVN-MM models. The single-modal 4DVN-SM, despite outperforming the direct inversion baseline, does not fully use the HR information of the satellite pseudo-observations. 4DVN-MM, on the other hand, can effectively extract the spatiotemporal HR information from both satellite pseudoproducts and in situ
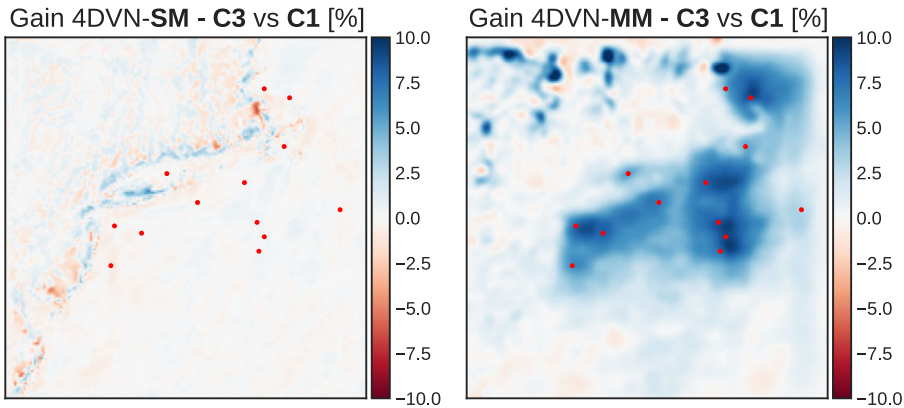
Figure 3. *Average gains maps. Left panel: plain 4DVarNet, average gain of 4DVN-SM-C3 vs 4DVN-SM-C1. Right panel: 4DVarNet with the additive trainable observation term in the variational cost, 4DVN-MM-C3 vs 4DVN-MM-C1.*

Table 2. *Computational effort associated to the models used. Training times refer to the time required to train the ensemble of 10 models. The trainable parameters are the number of parameters of one single model. Memory size refers to the space required to save the 10-members model ensemble*

Model	Training time	Parameters	**Size** [MB]
LDI	20 min	115 K	39
DVN-SM	3 h	742 K	89
DVN-MM	4 h 40 min	1.3 M	157

time series. The effect of this multimodal skill is reflected in the area affected by the reconstruction improvement, on tens-of-kilometers neighborhoods of the in situ buoys.

To complete the systematic benchmark of our models, we show in Table 2 that the computational effort associated with each model is used. We report the training time for the ensemble of 10 model runs, the trainable parameters of one single model of the ensemble, and the memory size required to store the 10-member model ensemble. Not surprisingly, the multimodal 4DVarNet 4DVN-MM is the most expensive model in terms of resources.

4.3. Biased LR data

Reanalyzed NWP products can be biased by model errors. To emulate this scenario, we train our models on an artificially biased dataset where the LR fields are modified by a random delay of $[-4, +4]$ hours or a random phase amplitude of $[0.5, 1.5]$. Formally,

$$\mathbf{y}^{lr}(t) = \begin{cases} \mathbf{y}^{lr}(t+\Delta t) \,\text{with}\, \Delta t \sim \mathcal{U}(-4, +4) \\ \alpha \mathbf{y}^{lr}(t) \,\text{with}\, \alpha \sim \mathcal{U}(0.5, 1.5) \end{cases} \tag{10}$$

where $\mathcal{U}$ represents the uniform distribution. Interestingly, this at-train-time modification can be seen as a dynamic data augmentation process (Xu et al., 2021). At test time, the model is evaluated using a test set that is systematically modified by each one of the modification factors. In this way, we can visualize a set of performance curves that depict that reconstruction error as a function of the modification factor, see Figure 4. In the case of random delay, the reconstruction curves for 4DVN-MM-C3 are lower at the extremes. This means that the model trained on the biased dataset learns the error in LR fields.
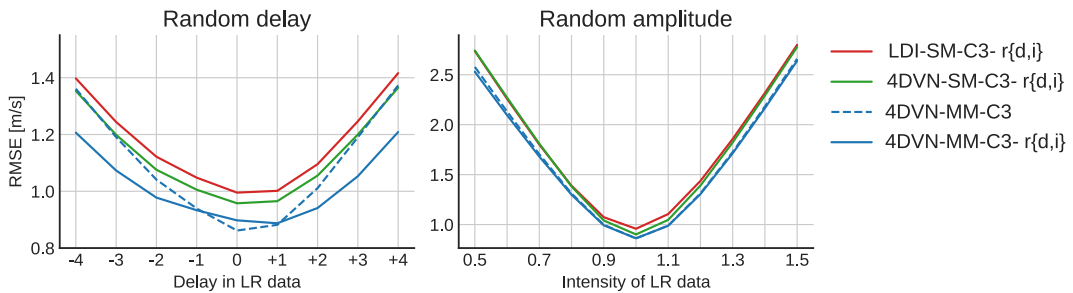
Figure 4. *Biased LR field tests. Left panel: random delay. Right panel: random intensity. The suffixes "-rd" and "-ri" identify the models trained in the case of random delay and intensity, respectively.*

Intriguingly, the curve associated with 4DVN-MM-C3 is systematically lower than the curve of 4DVN-SM-C3. This proves that the modeling choice of the additional term in the variational cost 6 gives the model the capability to learn and generalize even in the case of biased input data. In the amplitude case, the effect is less pronounced. An explanation of this phenomenon could be the following. In the delay case, there is a clear mismatch between the HR field and the swapped LR field. The LR field remodulation by the amplitude factors we chose may not suffice to make the field drastically different from the original. This implies a lesser model capability to detect and attenuate the LR data bias.

5. Conclusions

Our analysis presents the application of a hybrid data assimilation and deep learning framework used to retrieve HR wind speed fields from the partial observations of sea-surface wind. We showed that this framework can outperform a deep learning–based direct inversion scheme. We see that the 4DVarNet scheme improves the baseline in two independent ways. The first is imputable to the underlying 4DVar formulation that accounts explicitly for time processes. This informs the reconstruction of the temporal features of the phenomenon. On the other hand, the 4DVN-MM model better exploits the different input sources. This multimodal skill covers a primary importance in oceanography and geosciences since the large volumes of Earth observations are characterized by different spatiotemporal features. Our results show that proper modeling choices allow us to profit from the complementary information conveyed by these diverse observation sources. We also proved that these modeling choices, thanks to the flexibility of deep learning modeling, can endow the 4DVarNet framework with improved generalization capabilities that help to attenuate the errors in NWP products.

References

Ahsbahs T, **Maclaurin G**, **Draxl C**, **Jackson CR**, **Monaldo F and Badger M** (2020) Us east coast synthetic aperture radar wind atlas for offshore wind energy. *Wind Energy Science 5*(3), 1191–1210.

Amani M, **Mohseni F**, **Layegh NF**, **Nazari ME**, **Fatolazadeh F**, **Salehi A**, **Ahmadi SA**, **Ebrahimy H**, **Ghorbanian A**, **Jin S**, et al. (2022) Remote sensing systems for ocean: A review (part 2: Active systems). *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 15*, 1421–1453.

Arabi S, **Asgarimehr M**, **Kada M and Wickert J** (2023) Hybrid cnn-lstm deep learning for track-wise gnss-r ocean wind speed retrieval. *Remote Sensing 15*(17),4169.

Arcucci R, **Zhu J**, **Hu S and Guo Y-K** (2021) Deep data assimilation: integrating deep learning with data assimilation. *Applied Sciences 11*(3), 1114.

Arnold C and Asgarimehr M (2021) Global ocean wind speed estimation with cygnssnet. In *NeurIPS 2021 Workshop on Tackling Climate Change with Machine Learning*, https://www.climatechange.ai/papers/neurips2021/12.

Bannister RN (2017) A review of operational methods of variational and ensemble-variational data assimilation. *Quarterly Journal of the Royal Meteorological Society 143*(703), 607–633.

Barth A, **Alvera-Azcárate A**, **Licer M and Beckers J-M** (2020) Dincae 1.0: a convolutional neural network with error estimates to reconstruct sea surface temperature satellite observations. *Geoscientific Model Development 13*(3), 1609–1622.

Baydin AG, **Pearlmutter BA**, **Radul AA and Siskind JM** (2018) Automatic differentiation in machine learning: a survey. *Journal of Machine Learning Research 18*(153), 1–43.

Beauchamp M, **Febvre Q**, **Georgenthum H and Fablet R** (2023) 4dvarnet-ssh: end-to-end learning of variational interpolation schemes for nadir and wide-swath satellite altimetry. *Geoscientific Model Development Discussions 16*, 2119–2147. https://gmd.copernicus.org/articles/16/2119/2023/.

Bergen KJ, **Johnson PA**, **de Hoop MV and Beroza GC** (2019) Machine learning for data-driven discovery in solid earth geoscience. *Science 363*(6433).

Brune S, **Keller JD and Wahl S** (2021) Evaluation of wind speed estimates in reanalyses for wind energy applications. *Advances in Science and Research 18*, 115–126.

Carrassi A, **Bocquet M**, **Bertino L and Evensen G** (2018) Data assimilation in the geosciences: an overview of methods, issues, and perspectives. *WIREs Climate Change 9*(5), e535. https://doi.org/10.1002/wcc.535. https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcc.535.

Cash JR (2003) Review paper. efficient numerical methods for the solution of stiff initial-value problems and differential algebraic equations. *Proceedings: Mathematical, Physical and Engineering Sciences 459*(2032), 797–815. ISSN 13645021.

Dorffer C, **Jourdin F**, **Mouillot D**, **Devillers R and Fablet R** (2023) Observation-only learning of 4dvarnet neural schemes for the reconstruction of sea surface turbidity dynamics from gappy satellite images. In *EGU General Assembly Conference Abstracts*. pp. EGU–15977. Vienna, Austria: European Geophysical Union.

Fablet R, **Beauchamp M**, **Drumetz L and Rousseau F** (2021b) Joint interpolation and representation learning for irregularly sampled satellite-derived geophysical fields. *Frontiers in Applied Mathematics and Statistics 7*. ISSN 2297–4687. https://doi.org/10.3389/fams.2021.655224. https://www.frontiersin.org/articles/10.3389/fams.2021.655224.

Fablet R, **Chapron B**, **Drumetz L**, **Mémin E**, **Pannekoucke O and Rousseau F** (2021a) Learning variational data assimilation models and solvers. *Journal of Advances in Modeling Earth Systems 13*(10). https://doi.org/10.1029/2021MS002572.

Fablet R, **Drumetz L and Rousseau F** (2020) Joint learning of variational representations and solvers for inverse problems with partially-observed data. *CoRR*, abs/2006.03653.

Farchi A, **Bocquet M**, **Laloyaux P**, **Bonavita M**, **Chrust M and Malartic Q** (2022) Model error correction with data assimilation and machine learning. In *EGU General Assembly Conference Abstracts*. pp. EGU22-5692.

Gould J, **Sloyan B and Visbeck M** (2013) In situ ocean observations: a brief history, present status, and future directions. *International Geophysics 103*, 59–81.

Gualtieri G (2021) Reliability of era5 reanalysis data for wind resource assessment: a comparison against tall towers. *Energies 14*(14), 4169.

Hersbach H, **Bell B**, **Berrisford P**, **Hirahara S**, **Horányi A**, **Muñoz-Sabater J**, **Nicolas J**, **Peubey C**, **Radu R**, **Schepers D**, **Simmons A**, **Soci C**, **Abdalla S**, **Abellan X**, **Balsamo G**, **Bechtold P**, **Biavati G**, **Bidlot J**, **Bonavita M**, **De Chiara G**, **Dahlgren P**, **Dee D**, **Diamantakis M**, **Dragani R**, **Flemming J**, **Forbes R**, **Fuentes M**, **Geer A**, **Haimberger L**, **Healy S**, **Hogan RJ**, **Hólm E**, **Janisková M**, **Keeley S**, **Laloyaux P**, **Lopez P**, **Lupu C**, **Radnoti G**, **Rosnay Pde**, **Rozum I**, **Vamborg F**, **Villaume S and Thépaut J-N** (2020) The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society 146*(730), 1999–2049. https://doi.org/10.1002/qj.3803. https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3803.

Hochreiter S and Schmidhuber J (1997) Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. ISSN 0899–7667.

Hoffman RN and Atlas R (2016) Future observing system simulation experiments. *Bulletin of the American Meteorological Society 97*(9), 1601–1616.

Hong D, **Gao L**, **Yokoya N**, **Yao J**, **Chanussot J**, **Du Q and Zhang B** (2020) More diverse means better: multimodal deep learning meets remote-sensing imagery classification. *IEEE Transactions on Geoscience and Remote Sensing 59*(5), 4340–4354.

Karpatne A, **Atluri G**, **Faghmous JH**, **Steinbach M**, **Banerjee A**, **Ganguly A**, **Shekhar S**, **Samatova N and Kumar V** (2017) Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on Knowledge and Data Engineering 29*(10), 2318–2331.

Karpatne A, **Ebert-Uphoff I**, **Ravela S**, **Babaie HA and Kumar V** (2018) Machine learning for the geosciences: Challenges and opportunities. *IEEE Transactions on Knowledge and Data Engineering 31*(8), 1544–1554.

Kingma D and Ba J (2014) Adam: a method for stochastic optimization. *International Conference on Learning Representations 12*.

Manucharyan GE, **Siegelman L and Klein P** (2021) A deep learning approach to spatiotemporal sea surface height interpolation and estimation of deep currents in geostrophic ocean turbulence. *Journal of Advances in Modeling Earth Systems 13*(1), e2019MS001965.

Monaldo FM, **Jackson CR and Pichel WG** (2013) Seasat to radarsat-2: Research to operations. *Oceanography 26*(2), 34–45.

Ongie G, **Jalal A**, **Metzler CA**, **Baraniuk RG**, **Dimakis AG and Willett R** (2020) Deep learning techniques for inverse problems in imaging. *IEEE Journal on Selected Areas in Information Theory 1*(1), 39–56.

Optis M, **Kumler A**, **Scott GN**, **Debnath MC and Moriarty PJ** (2020) Validation of RU-WRF, the custom atmospheric mesoscale model of the rutgers center for ocean observing leadership. Technical Report, National Renewable Energy Lab.(NREL), Golden, CO.

Potisomporn P, **Adcock TAA and Vogel CR** (2023) Evaluating era5 reanalysis predictions of low wind speed events around the UK. *Energy Reports 10*, 4781–4790.

Powers JG, **Klemp JB**, **Skamarock WC**, **Davis CA**, **Dudhia J**, **Gill DO**, **Coen JL**, **Gochis DJ**, **Ahmadov R**, **Peckham SE**, et al. (2017) The weather research and forecasting model: Overview, system efforts, and future directions. *Bulletin of the American Meteorological Society 98*(8), 1717–1737.

Rincy TN and Gupta R (2020) Ensemble learning techniques and its efficiency in machine learning: a survey. In *2nd International Conference on Data, Engineering and Applications (IDEA)*. pp. 1–6. IEEE.

Saxena BK, **Mishra S and Rao KVS** (2021) Offshore wind speed forecasting at different heights by using ensemble empirical mode decomposition and deep learning models. *Applied Ocean Research 117*, 102937.

Snieder R and Trampert J (1999) Inverse problems in geophysics. In *Wavefield Inversion*. Springer, pp. 119–190.

Storto A, **Alvera-Azcárate A**, **Balmaseda MA**, **Barth A**, **Chevallier M**, **Counillon F**, **Domingues CM**, **Drevillon M**, **Drillet Y**, **Forget G**, **Garric G**, **Haines K**, **Hernandez F**, **Iovino D**, **Jackson LC**, **Lellouche J-M**, **Masina S**, **Mayer M**, **Oke PR**, **Penny SG**, **Peterson KA**, **Yang C and Zuo H** (2019) Ocean reanalyses: recent advances and unsolved challenges. *Frontiers in Marine Science 6*. ISSN 2296–7745. https://doi.org/10.3389/fmars.2019.00418. https://www.frontiersin.org/articles/10.3389/fmars.2019.00418.

Talagrand O (2015) *4D-Var: Four-Dimensional Data Assimilation*. Oxford University Press, pp. 3–30.

Valmassoi A, **Keller JD**, **Kleist DT**, **English S**, **Ahrens B**, **Ďurán IB**, **Bauernschubert E**, **Bosilovich MG**, **Fujiwara M**, **Hersbach H**, et al. (2023) Current challenges and future directions in data assimilation and reanalysis. *Bulletin of the American Meteorological Society 104*(4), E756–E767.

Vient J-M, **Fablet R**, **Jourdin F and Delacourt C** (2022) End-to-end neural interpolation of satellite-derived sea surface suspended sediment concentrations. *Remote Sensing 14*(16), 4024.

Xu D, **Lee ML and Hsu W** (2021) Classification with dynamic data augmentation. In *2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*. pp. 1434–1441. https://doi.org/10.1109/ICTAI52525.2021.00228.

Yu S and Ma J (2021) Deep learning for geophysics: current and future trends. *Reviews of Geophysics 59*(3), e2021RG000742.