



HAL
open science

Reduction of rain-induced errors for wind speed estimation on SAR observations using convolutional neural networks

Aurélien Colin, Pierre Tandeo, Charles Peureux, Romain Husson, Ronan Fablet

► **To cite this version:**

Aurélien Colin, Pierre Tandeo, Charles Peureux, Romain Husson, Ronan Fablet. Reduction of rain-induced errors for wind speed estimation on SAR observations using convolutional neural networks. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2023, pp.1 - 13. 10.1109/jstars.2023.3291236 . hal-04149533

HAL Id: hal-04149533

<https://imt-atlantique.hal.science/hal-04149533>

Submitted on 3 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Reduction of rain-induced errors for wind speed estimation on SAR observations using convolutional neural networks

Aurélien Colin^{1,2}, Pierre Tandeo^{1,3}, Charles Peureux², Romain Husson², Ronan Fablet^{1,3},

Abstract—Synthetic Aperture Radar is known to be able to provide high-resolution estimates of surface wind speed. These estimates usually rely on a Geophysical Model Function (GMF) that has difficulties accounting for non-wind processes such as rain events. Convolutional neural network, on the other hand, have the capacity to use contextual information and have demonstrated their ability to delimit rainfall areas. By carefully building a large dataset of SAR observations from the Copernicus Sentinel-1 mission, collocated with both GMF and atmospheric model wind speeds as well as rainfall estimates, we were able to train a wind speed estimator with reduced errors under rain. Collocations with in-situ wind speed measurements from buoys show a root mean square error that is reduced by 27% (resp. 45%) under rainfall estimated at more than 1 mm/h (resp. 3 mm/h). These results demonstrate the capacity of deep learning models to correct rain-related errors in SAR products.

Index Terms—Synthetic Aperture Radar, Deep Learning, Oceanography, Wind.

I. INTRODUCTION

SYNTHETIC Aperture Radar (SAR) is a powerful tool for studying the ocean surface. C-Band SAR are sensitive to variations in sea surface roughness, and have been used to detect various meteorological and ocean processes, referred to as metocean, such as atmospheric or ocean fronts [1], icebergs [2], oil surfactants from pollution [3] or generated by plankton [4], and some species of seaweed [5]. They are particularly useful for

studying waves [6] and extreme events like cyclones [7], [8]. There has been particular attention given to estimating wind speed using these sensors.

As the number of satellite missions with C-SAR sensors increases and archives of these data accumulate, it is becoming easier to build large SAR datasets. This paper focuses on the Sentinel-1 mission from the Copernicus program, which consists of two satellites, Sentinel-1A (launched in 2014) and Sentinel-1B (launched in 2016, which has been out of operation since December 2021). Sentinel-1C is planned to be launched in 2023. Ground Range Detected Higher Resolution Interferometric Wide-swath (GRDH IW) observations have a range of 250 km, an azimuth of about 200 km, and a resolution of 10 m/px. These observations are mainly routinely acquired over coastal areas. Systematic processes are used to produce geophysical products from these observations, including wind speed estimates. Several Geophysical Model Functions (GMFs) have been developed for this purpose, including CMOD3 [9], CMOD4 [10], CMOD5 [11], CMOD5.N [12], CMOD6 [13], CMOD7 [14] and C_SARMOD2 [15]. These GMFs use the vertical-vertical polarization, which is Sentinel-1's default co-polarization in coastal areas. However, the co-polarization channel saturates at high wind speed [16]. Therefore, the cross-polarization has also been used to estimate the wind speed [7]. In particular, H14E [17] was found to provide accurate wind speed measurements even in extreme events such as category 5 hurricanes [18]. The cross-polarization has the advantage of lower incidence and wind direction dependency [19]. GMFs have also been developed

¹ IMT Atlantique, Lab-STICC, UMR CNRS 6285, F-29238, France.

² Collecte Localisation Satellites, Brest, France.

³ Odyssey, Inria/IMT, France

for horizontal-horizontal polarization [20]. Bayesian nonparametric wind speed estimation has also been proposed, offering the advantage of not relying on wind direction priors that are sensitive to meteorological processes with rapid temporal and spatial evolution [21].

However, these GMFs are sensitive to contamination from non-wind processes. In particular, rainfall can either increase or decrease sea surface roughness [22], making it difficult to correct for its effects.

Deep learning models, particularly Convolutional Neural Networks (CNNs), have demonstrated their ability to detect rain signatures in SAR observations [23]. These models are known to be able to tackle denoising [24] and inpainting [25] tasks because they use contextual information to estimate the original signal. This paper is dedicated to estimating wind speed in rainy areas using a model that does not require an explicit rainfall prior and only uses the parameters available to GMFs.

In the first section, we present the SAR data used to train the model and the ancillary information available. The second section describes the methodology used to build the dataset, with special attention given to ensuring a balanced representation of rainfall observations. The final section presents the results on the training set and confirms them with in-situ measurements from buoys, demonstrating the model's ability to correct for rain overestimates.

II. DATASET

The SAR measurements used in this chapter come from 19,978 IW observations acquired globally between March 03rd, 2018 and the February 23rd, 2022, inclusive. Each of these observations covers approximately 44 000 km² and has a resolution of 100 m/px, downscaled from the GRDH products available at 10 m/px. The radiometric information used as input is the Sea Surface Roughness (SSR) defined in [26] as the normalized radar cross section σ_0 divided by the σ_0 of a wind of 10 m/s and a direction of 45° relative to the antenna look angle. The σ_0 of this neutral wind is given by the GMF CMOD5.N.

Obtaining global information on rain that can be used in conjunction with SAR observations can be difficult. A previous study conducted using a global Sentinel-1 dataset found only 2,304 partial collocations with the satellite-based radar GPM/DPR [27] out of 182,153 IW. "Partial collocations" refers to instances where at least 20x20 km of a swath is observed by the spaceborne weather radar 20 minutes before or after the SAR observation. Coastal ground-based radars like NEXRAD [28] could provide rainfall estimates, but they are affected by topography and may not capture all wind regimes. Therefore, SAR-based rain estimation is preferred to maximize the number of available observations and simplify the collocation process. We used a recent SAR rainfall estimator [23] that emulates NEXRAD's reflectivity and proposes three rainfall thresholds that roughly correspond to 1 mm/h, 3 mm/h, and 10 mm/h.

Ancillary information, such as incidence angle and satellite heading, is retrieved from Sentinel-1 Level-2 products. It also includes collocations with atmospheric models from the European Centre for Medium-Range Weather Forecasts, which provide modelled wind speed and direction, as well as the surface wind speed computed by the GMF. The atmospheric models have a spatial resolution of 0.25x0.25 degrees and a temporal resolution of 3 hours [29], while the GMF is computed at a spatial resolution of 1 km/px and corresponds to the observation itself. The GMF used in Level-2 product is IFR2 [30] until July 2019 and CMOD5.N [12] afterwards. The ancillary information contained in the Level-2 products is available at a spatial resolution of 1 km/px. However, it is interpolated to 100 m/px to match the grid of the radiometric channels.

III. METHODOLOGY

This section presents the methodology for building the rain-invariant wind speed estimator. We first describe the deep learning architecture of the model, then we discuss the creation of the dataset, which is biased to have a large number of rain examples. The final section describes the evaluation procedure.

A. Deep Learning Model

The architecture used in this chapter is the UNet architecture [31] depicted in Fig. 1. UNet is an autoencoder architecture with the advantage of being fully convolutional, meaning it has translation equivariance properties (translations of the input result in translations of the output). In addition, skip connections between the encoder and the decoder facilitate training, especially by reducing the vanishing gradient issue [32]). Introduced in 2015, UNet has been used in various domains and has demonstrated its importance for segmentation of SAR observations [33], [23], [34].

The output of the model always contains a single convolution kernel, activated by the ReLU function to ensure that the prediction is in the interval $[0, +\infty]$. All convolution kernels in the hidden layers are also activated by ReLU functions. The model is set to take input of 256×256 pixels during training, but since the weights only describe convolution kernels, it is possible to use the model for inference on images of any shape as long as the input resolution remains at 100 m/px. Variants of the model are trained with different numbers of input channels. The architecture is modified by changing the size of the first convolution kernel, which is defined as a kernel of size $(3, 3, c, 32)$, where c is the number of input channels.

B. Dataset balancing procedure

In this section, we describe the process for building a balanced dataset. Our goals are to (I) ensure that the wind distribution of each dataset is close to the real-world distribution, (II) prevent information leak between the training, validation, and test subsets, (III) ensure that the groundtruth wind speed, obtained from an atmospheric model, accurately represents the real-world wind speed, and (IV) include enough rain samples to allow the model to learn from them.

a) Rain and rainless patches selection: As discussed earlier, rainfall estimation is provided by a deep learning model at a resolution of 100 m/px, on the same grid as the SAR observation. Therefore, it is possible to separate the observations into two

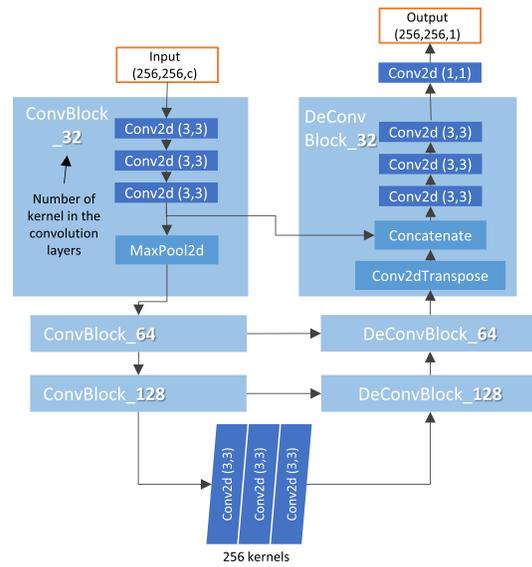


Fig. 1: Architecture of the UNet model used for estimating the wind speed.

areas, \mathcal{A}^+ and \mathcal{A}^- , based on the 3 mm/h threshold from the rainfall estimation.

$$\mathcal{A}^+ = \{x : Rainfall(x) \geq 3mm/h\} \quad (\text{Eq. 1})$$

$$\mathcal{A}^- = \{x : Rainfall(x) < 3mm/h\} \quad (\text{Eq. 2})$$

However, most SAR observations do not contain rain signatures. Collocations with GPM's dual polarization radar, a satellite-based weather radar, indicated that the probability of rain rates higher than 3 mm/h was 0.5%. Thus, by dividing the SAR observations into tiles of 256 by 256 pixels, we call "rain patches" the tiles with more than 5% of their surface predicted to have rain rates higher than 3 mm/h, and "rainless patches" those without rain signatures. We denote n_+ as the number of rain patches and n_- as the number of rainless patches. To ensure that the model will learn regardless of the rain-situation, we set $n_+ = n_-$.

Fig. 2 indicates collocations between the reanalysis from ERA5 and the satellite-based weather radar GPM/DPR. It shows the wind speed distribution for rainfall higher than 3 mm/h (a), 30 mm/h (b), and the overall distribution (orange curve). Stronger

rainfall is associated with a lower probability of strong wind speeds. However, the impact is mostly marginal at moderate rain rates.

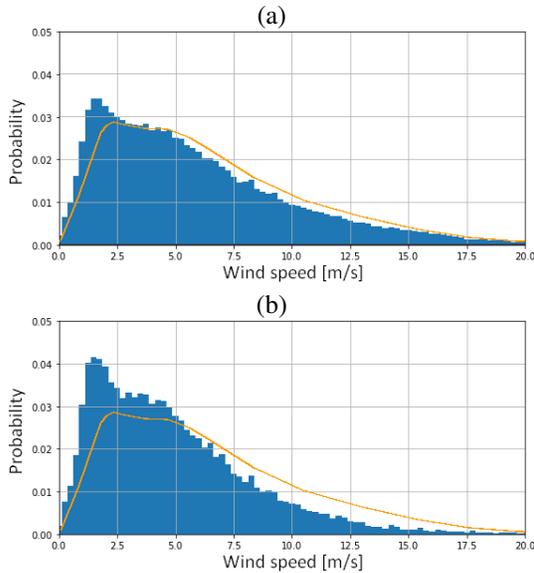


Fig. 2: Distribution of ERA5 wind speed collocated with surface rain rate from GPM/DPR for rainfall higher than 3 mm/h (a) and 30 mm/h (b). They amount for 0.55% and 0.02% of the collocations, respectively. The orange curve in the figure shows the wind distribution regardless of the rainfall.

b) Restriction to a priori Accurate Model Wind Speeds: Atmospheric models have been known to have coarse resolution (0.25x0.25 degrees spatially, 3 hours temporally) and to be unable to accurately depict fine-scale wind fields. However, they are computed globally and independently of the SAR observations. On the other hand, SAR-based wind fields from the GMFs are known to be accurate on rainless patches, but are diversely affected by rainfall. At low and moderate wind speeds, impinging droplets lead to an overestimation of the wind speed. Under high wind speed conditions, dominant attenuation results in an underestimation of the wind speed. Because of the difficulty of collocating a sufficient quantity of high wind speed and rainfall events, this study mainly aims to correct the wind speed overestimation. We

calculate $\Delta_{\mathcal{A}^-}$ as the discrepancy between the GMF and the atmospheric model on rain-free pixels.

$$\Delta_{\mathcal{A}^-} = MSE|_{\mathcal{A}^-}(Atm, GMF) \quad (\text{Eq. 3})$$

In our experiments, the threshold was set at $\Delta_{\mathcal{A}^-} < 1$ m/s. All patches containing a higher discrepancy between the two wind speed sources were discarded. By ensuring agreement between the atmospheric model and the SAR-based GMF on rainless patches, we assume that the atmospheric model is close to the real wind speed and use it as a target for optimizing the deep learning model.

c) Balancing to the real-world wind distribution: It should be noted that this condition ensures accurate modeled wind speeds and rain distribution, especially because the rainfall estimator is known to overestimate rainfall at high wind speeds.

We denote:

- P^+ as the wind speed distribution on n_+ .
- P^- as the wind speed distribution on n_- .
- P as the wind speed distribution on $n_- \cup n_+$.

Balancing the dataset to the real-world wind distribution translates to the following condition:

$$\forall x, P(x) = \frac{n_+P^+(x) + n_-P^-(x)}{n_+ + n_-} \quad (\text{Eq. 4})$$

As we choose to keep all rain patches and to set $n_+ = n_-$, Eq. 4 leads to:

$$\forall x, P^-(x) = \frac{1}{2}(P^+(x) - P(x)) \quad (\text{Eq. 5})$$

For some wind speeds x , $P^+(x)$ is higher than twice $P(x)$. In these cases, we relax the condition from Eq. 5 in order to avoid removing rain patches. Fig. 3 depicts the wind speed distribution for rain and rainless patches. The mean squared error between P and $\frac{1}{2}(P^+ + P^-)$ reach 8.8%.

The dataset can be further balanced to ensure that, for each wind speed, the number of rain and rainless patches is equal. However, this leads to remove of 84% of the data. Appendix 1 compares the performance of this second dataset. As did not provide improvements, this dataset is left out of the main document.

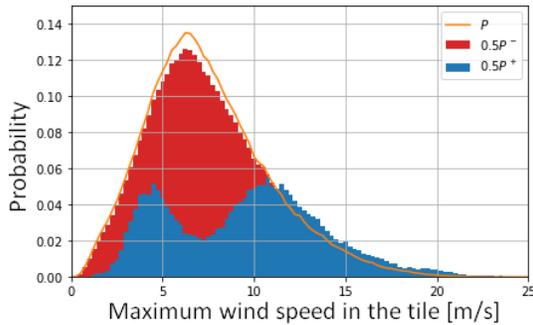


Fig. 3: Wind speed distributions for rain (blue) and rainless (red) patches. Orange curve correspond to the real world wind distribution.

1) Training, Validation and Test Set Division:

After extracting the patches following the distributions P^+ and P^- , they are split into training, validation, and test sets. Each subset preserves the same distributions. Furthermore, to avoid information leakage, if a patch from one IW is in a subset, every patch from the same IW belongs to the same subset. The stochastic brute forcing method described in Algorithm 1 draws random IWs and computes the distribution of the validation and test subsets, compares them to the overall distributions, and returns the solution that minimizes the difference. In this algorithm, \bar{P}_e indicates the wind speed distribution multiplied by the number of patches in e and divided by the total number of patches. It ensures that the validation and test subsets each contain approximately 10% of all the patches.

The initial 19,978 IW observations account for 105,164 rain patches and 2,094,370 rainless patches. At the end of the process, the distribution in the subsets is as follow:

- 168,349 patches from 14,169 IWs in the training set;
- 20,944 patches from 1,763 IWs in the test set;
- 21,010 patches from 1,763 IWs in the validation set.

Before training the model, we compute the mean and standard deviation of each channel on the training set and use them to normalize the inputs during training, validation, and inference. The

Algorithm 1: Stochastic brute forcing

input : The list L of all swath i and their patch distribution P_i .
output: L_{val} and L_{test} , the list of the swaths contained in the validation and test sets.

```

1  $\mathcal{L} = \text{MAE};$ 
2  $e_{min} = +\infty;$ 
3  $n_1 = \text{len}(L)/10;$ 
4  $n_2 = \text{len}(L)/3;$ 
5 for  $i \leftarrow 0$  to 1000000 by 1 do
6    $n = \text{random.integer}(\text{min}=n_1, \text{max}=n_2);$ 
7    $candidate = \text{random.choice}(L, \text{size}=n,$ 
8      $\text{replace}=\text{False});$ 
9    $c_{val} = candidate[: n/2];$ 
10   $c_{test} = candidate[n/2 :];$ 
11   $e_{val} = \mathcal{L}(0.1P^+, \bar{P}_{c_{val}}^-) + \mathcal{L}(0.1P^-, \bar{P}_{c_{val}}^-);$ 
12   $e_{test} =$ 
13     $\mathcal{L}(0.1P^+, \bar{P}_{c_{test}}^+) + \mathcal{L}(P^-, 0.\bar{P}P_{c_{test}}^-);$ 
14   $e = \frac{2 \cdot e_{val} \cdot e_{test}}{e_{val} + e_{test}};$ 
15  if  $e < e_{min}$  then
16     $L_{val} = c_{val};$ 
17     $L_{test} = c_{test};$ 
18     $e_{min} = e;$ 
19  end
20 end

```

output, however, is not normalized. We train the model for 100,000 weight updates (i.e. steps of the stochastic gradient descent) to minimize the Mean Square Error (MSE) between the prediction and the wind speed from the atmospheric model, with a batch size of 16 and a learning rate of 10^{-5} using the Adam optimizer.

C. Evaluation procedure

To evaluate the impact of each input channel, we train various variants of the model:

- I uses only the VV channel.
- II uses the VV channel, the incidence angle and the *a priori* wind direction.
- III uses both the VV and the VH channel, the incidence angle and the *a priori* wind direction.
- IV uses both the VV and the VH channel, the incidence angle, the *a priori* wind direction and the wind speed prior.
- V uses only the wind speed prior.

All channels are interpolated to 100 m/px and concatenated on the same grid.

Incidence angles and wind directions from ECMWF atmospheric model are obtained from the Level-2 products and notably used for the computation of the GMF. Therefore, architecture II contains the same inputs as the GMF though interpolated at a finer resolution.

Each architecture is trained five times to reduce the impact of random initialization on the evaluation results. The results is presented as the mean and standard deviation over these five independent trainings.

We compare the results using the Root Mean Square Error (RMSE) and the Pearson correlation coefficient (PCC). The PCC is formulated in Eq. 6.

$$PCC_{Y,\hat{Y}} = \frac{\mathbb{E}[(Y - \mu_Y)(\hat{Y} - \mu_{\hat{Y}})]}{\sigma_Y \sigma_{\hat{Y}}} \quad (\text{Eq. 6})$$

The results are computed against both the groundtruths from the atmospheric model, which provides a large test set, and against collocations with buoys, which have good temporal resolution and are in-situ measurements.

IV. RESULTS

A. Benchmarking experiments

The performance of the models compared to ECMWF are calculated on the test subset for each input variant and the baseline GMF. The results of this analysis can be found in Table I. It appears that the most important input is the GMF itself, as both IV and V have better results than the other variants. I, II and III are unable to achieve better results than the GMF, except under strong rainfall, even though II and III have access to all the channels used by the GMF.

Examples of rain-induced wind speed overestimation are presented in Fig. 4. Fig. 4.a shows the SSR on a logarithmic scale, with bright patterns appearing in the observation and segmented as rainfall in Fig. 4.c. These patterns correlate with high wind speeds estimated by the GMF in Fig. 4.b. The deep learning method in Fig. 4.d ignores this spike in the SSR, but it also seems to underestimate

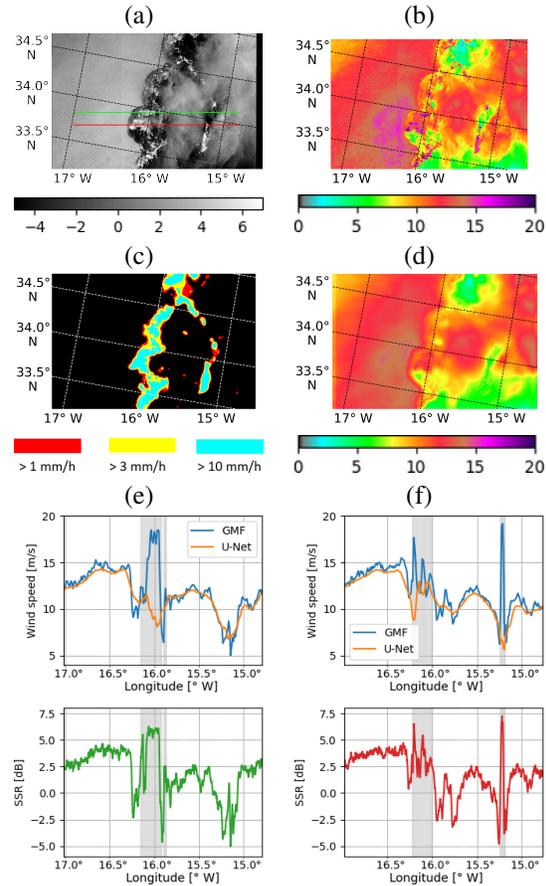


Fig. 4: Observation from the November 11th 2018 at 04:56:47. Sea Surface Roughness (SSR) of the VV channel in dB (a), geophysical model function estimated wind speed in m/s (b), deep learning estimated wind speed in m/s (d), prediction of the rainfall estimator (c) and transects of the green (e) and red (f) lines, with rainfall predicted higher than 10 mm/h indicated as grey areas.

the wind speed in the bottom-left quarter of the observation. Both of these behaviours are especially visible in the transect shown in Fig. 4.f.

MODEL AND CHANNELS	RAIN RATE	RMSE		PCC	
		Balanced Dataset	Neutral Dataset	Balanced Dataset	Neutral Dataset
I [VV]	[0, 1[mm/h	1.38 [0.016]	2.40 [0.019]	89.9% [0.19%]	74.2% [0.42%]
	[1, 3[mm/h	1.64 [0.046]	2.78 [0.038]	92.7% [0.39%]	79.5% [0.58%]
	[3, 10[mm/h	1.59 [0.038]	3.18 [0.070]	92.7% [0.38%]	78.9% [0.62%]
	≥ 10mm/h	2.12 [0.052]	★3.27 [0.135]	81.7% [1.03%]	74.3% [0.90%]
II [VV, INC, WDIR]	[0, 1[mm/h	0.87 [0.009]	2.26 [0.019]	96.2% [0.06%]	77.7% [0.24%]
	[1, 3[mm/h	1.04 [0.086]	2.87 [0.061]	97.2% [0.47%]	77.6% [0.24%]
	[3, 10[mm/h	1.19 [0.053]	3.39 [0.193]	96.2% [0.37%]	75.0% [3.04%]
	≥ 10mm/h	2.27 [0.151]	3.86 [0.700]	81.7% [1.89%]	63.2% [9.19%]
III [VV, VH, INC, WDIR]	[0, 1[mm/h	0.83 [0.002]	2.18 [0.006]	96.5% [0.02%]	79.3% [0.07%]
	[1, 3[mm/h	0.93 [0.020]	2.73 [0.022]	97.7% [0.07%]	80.1% [0.29%]
	[3, 10[mm/h	1.09 [0.022]	3.25 [0.052]	96.7% [0.07%]	78.0% [0.44%]
	≥ 10mm/h	2.13 [0.050]	3.68 [0.310]	83.9% [0.66%]	70.8% [2.95%]
IV [VV, VH, INC, WDIR, GMF]	[0, 1[mm/h	★0.64 [0.007]	★1.90 [0.028]	★97.9% [0.04%]	★85.0% [0.33%]
	[1, 3[mm/h	★0.63 [0.015]	★2.29 [0.075]	★98.9% [0.03%]	★87.1% [0.56%]
	[3, 10[mm/h	★0.78 [0.040]	★2.55 [0.132]	★98.4% [0.08%]	★87.1% [1.12%]
	≥ 10mm/h	★1.63 [0.162]	3.37 [0.113]	★90.9% [1.17%]	73.4% [2.31%]
V [GMF]	[0, 1[mm/h	0.67 [0.003]	3.16 [2.650]	97.7% [0.02%]	80.7% [17.55%]
	[1, 3[mm/h	0.68 [0.005]	4.17 [3.582]	98.8% [0.01%]	82.5% [12.67%]
	[3, 10[mm/h	0.88 [0.021]	4.53 [3.739]	98.0% [0.03%]	83.4% [8.19%]
	≥ 10mm/h	1.94 [0.087]	4.79 [2.917]	87.9% [0.66%]	74.0% [8.19%]
GMF	[0, 1[mm/h	0.77	2.41	97.0%	81.2%
	[1, 3[mm/h	0.84	3.16	98.1%	80.0%
	[3, 10[mm/h	1.25	3.42	96.5%	81.9%
	≥ 10mm/h	4.65	3.70	52.5%	★ 75.4%

Table I: Comparison of the five variants of the model and the two datasets. RMSE and PCC are computed on the respective test set and for five training with random initialization. Results are given as mean and standard deviation in brackets. The best result for each metric is indicated by ★. Results better than the GMF are in bold.

B. Application to SAR observation with groundtruthed in-situ data

However, ECMWF wind speeds are model data and not in-situ, which can be obtained using anemometers on buoys. Using the dataset created in [35], 4732 collocation points between Sentinel-1 and NDBC buoys are identified. The rain prediction model estimates that 4643 of these points are rainless, 75 record rainfall of more than 1 mm/h, and 14 record rainfall of more than 3 mm/h. On a side note, the height at which in-situ measurements were taken varies, with most being between 3.8 m and 4.1 m above sea level. As mentioned in [35], the SAR inversion and deep learning prediction are both normalized to the altitude of the corresponding in-situ measurement. Denoting w the wind speed at 10 m estimated from the SAR, the wind speed at elevation h is given by the exponential law [36]:

$$w_h = w \cdot \left(\frac{h}{10}\right)^{0.11} \quad (\text{Eq. 7})$$

Table II indicates that the performances of the deep learning are higher than the GMF for both the Root Mean Square Error and the Pearson Correlation Coefficient for all rain ranges. The RMSE decreases by 0.04, 0.37 and 1.33 m/s for rainless, light rain, and moderate rain-situations, respectively. The bias decreases by 0.26 and 1.01 m/s under light and moderate rain, but increases by 0.02 m/s for rainless situations. Table II also demonstrates the importance of the dataset building scheme as a dataset composed of collocations without the aforementioned sample selection, referred to as the "neutral dataset," consistently has lower performances than the balanced dataset.

Fig. 5 presents a comparison of the performances against NDBC buoys for various wind speeds. Unfortunately, collocations are scarce at very low and high wind speeds, which are the intervals where both the GMF and deep learning models have higher errors.

	Balanced Dataset	Neutral Dataset	GMF	
Bias	< 1mm/h	0.73 [0.04]	1.32 [0.04]	★ 0.71
	[1, 3]mm/h	★ 1.38 [0.04]	1.47 [0.04]	1.64
	> 3mm/h	★ 0.92 [0.07]	1.96 [0.29]	2.38
RMSE	< 1mm/h	★ 1.44 [0.03]	1.76 [0.12]	1.48
	[1, 3]mm/h	★ 1.81 [0.04]	1.95 [0.18]	2.18
	> 3mm/h	★ 1.60 [0.10]	2.42 [0.21]	2.93
PCC	< 1mm/h	★ 93.6% [0.16%]	92.9% [0.20%]	93.4%
	[1, 3]mm/h	★ 96.3% [0.22%]	93.4% [0.19%]	95.3%
	> 3mm/h	★ 95.9% [0.35%]	93.4% [2.04%]	91.3%

Table II: Bias, Root Mean Square Error, and Pearson Correlation Coefficient of model IV, the GMF, for each rainfall level. The best result for each metric and rainfall level is indicated by ★. Results better than the baseline are in bold. Results are given as mean and standard deviation in brackets.

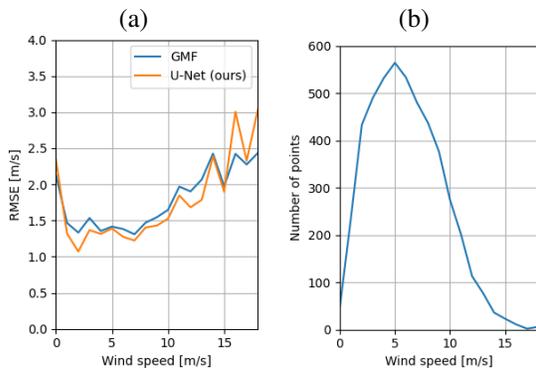


Fig. 5: Evaluation of the Root Mean Square Error across the wind speed (a) and number of collocation with NDBC buoys in rainless areas (b).

In the following, we observe two cases where rainfall was detected on the buoy position at the time of observation.

1) *2017-01-08 01:58:19 at NDBC 46054:* The observation from 2017-01-08 01:58:19 covers the north of the Californian Channel Islands (Fig. 6.a). Several meteorological buoys are dispersed over the channel, including NDBC 46054 and NDBC 46053, which are indicated as red dots. The wind speed over the area is mostly around 6 m/s, but a squall line appears at the position of NDBC 46054 and spans over a dozen kilometers. A surge in the air column reflectivity is recorded by weather radar from the NEXRAD network. The ground station is located at 119 km. The GMF indicates very

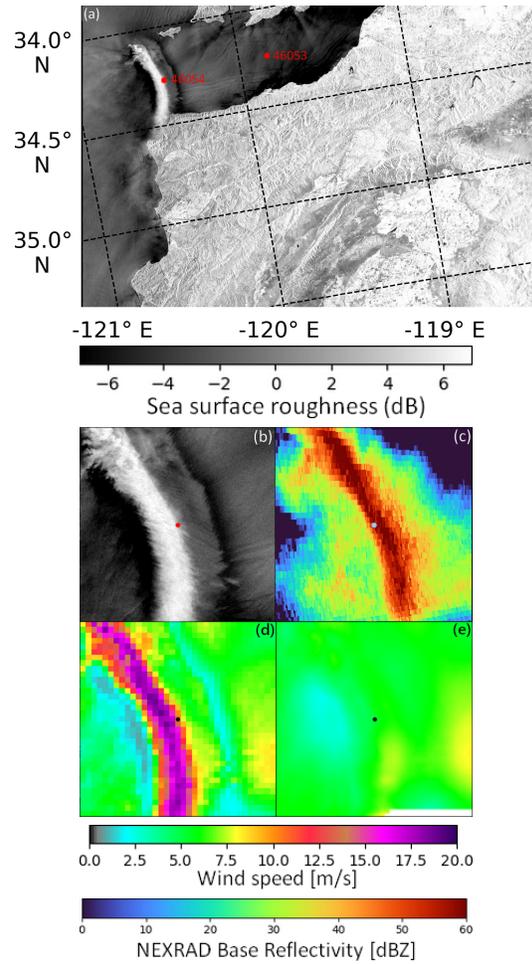


Fig. 6: Sea surface roughness from the January 08th 2017 at 01:58:19 in VV channel (a), zoom on an area of 35x35 km centred on the buoy NDBC 45054 (b), NEXRAD base reflectivity acquired at 01:57:58 (c), wind speed given by the GMF (d) and by the deep learning model (e).

high wind speeds, higher than 20 m/s (Fig. 6.d). The deep learning model attenuates these values to between 6 m/s and 8 m/s (Fig. 6.e).

For NDBC 46054, only one measurement of wind speed and direction per hour is available. It recorded a wind speed of 6.3 m/s eight minutes before the SAR observation. The GMF and the deep learning

model estimated wind speeds of 15.1 m/s and 5.9 m/s, respectively. While the temporal resolution of NDBC 46054 is one measurement per hour, NDBC 46053 records data every ten minutes. Furthermore, the gust front appears to be moving toward the right part of the observation. This can be seen in the time series in Fig. 6 as a large variation in wind direction between 02:40:00 and 03:00:00. The variation in wind speed seems to precede the variation in direction, first increasing then decreasing to a lower wind regime. On NDBC 46053, the GMF and the deep learning model agree on a wind speed of 4.5 m/s, which is slightly lower than the in-situ data of 5.4 m/s. Since the distance between NDBC 46054 and NDBC 46053 is approximately 60 km, the progression of the gust front can be estimated to be around 90 km/h. With a width of around 5 or 6 km, the whole system would pass the buoys in three minutes. This means that even NDBC 46053 may not have been able to accurately estimate the wind speed due to its low temporal resolution. However, it is worth noting that even the gust speed at NDBC 46054, defined as the maximum wind speed over a given number of seconds, does not record a speed higher than 9 m/s.

2) SAR-20191006T232853 NDBC-41009:

The observation from 2019-10-06 23:28:53 was recorded on the east coast of Florida. While most of the swath covers the marshes around Orlando and Cap Canaveral rather than the ocean, convective precipitation can be observed in the right part of the image (Fig. 8.a). The cells are moving downward (north-north-east), as indicated by the stronger gradient of the convective front. Since the wind from the convection is opposing the underlying wind regime, an area of lower wind speed appears as an area of lower backscatter. Rainfall was detected by a weather radar from the NEXRAD network located at 64 km (Fig. 8.c). The GMF is impacted by these rain signatures and estimates a very high local wind speed (Fig. 8.d). The deep learning model is less affected by the rain signatures, but also appears to blur the low wind speed area (Fig. 8.e).

The time serie from NDBC 41009 in Fig. 9 indicates that the lower backscattering was indeed

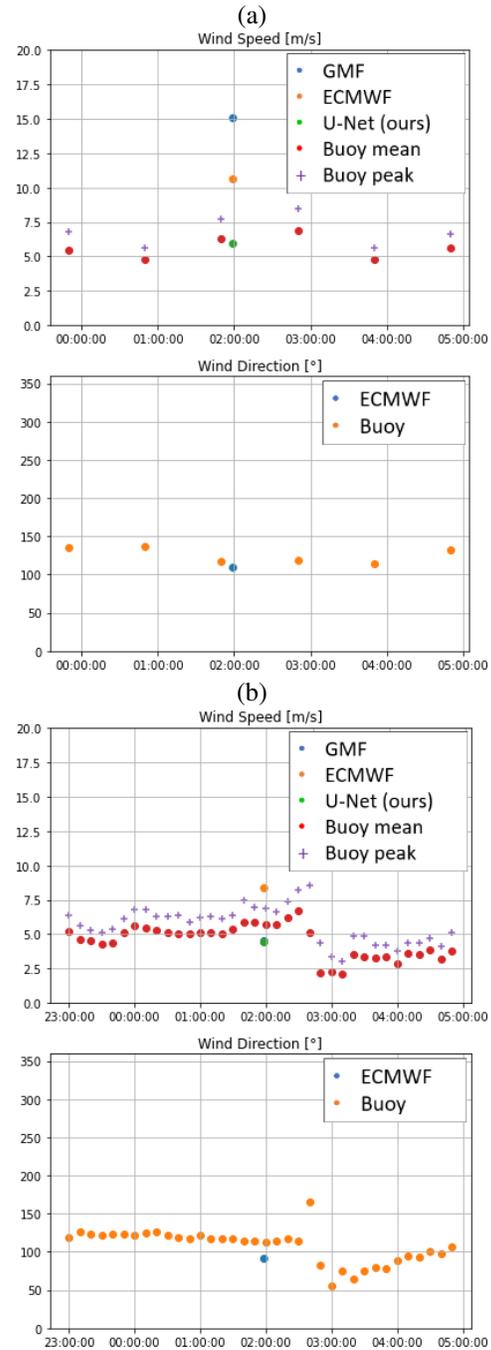


Fig. 7: Time series of the NDBC buoy wind measurements around January 08th 2017 01:58:19 for NDBC 46054 (a) and NDBC 46053 (b), and the estimation from the GMF, the deep learning model and the atmospheric model.

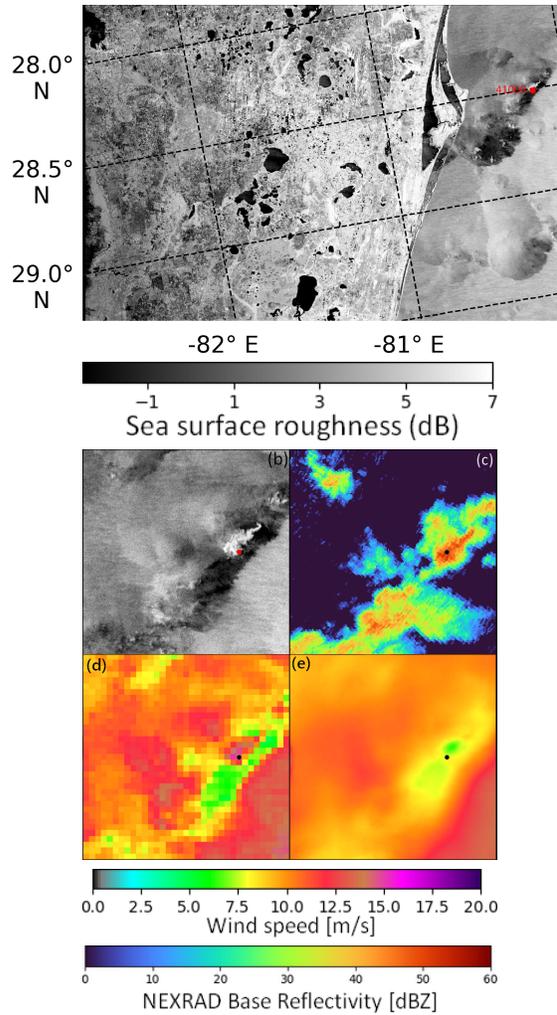


Fig. 8: Sea surface roughness from October 06th 2019 at 23:28:53 in VV channel in dB (a), zoom on an area of 35x35 km around the buoy NDBC 41009 (b), NEXRAD observation acquired at 23:27:33 (c), wind speed given by the GMF (d) and by the deep learning model (e).

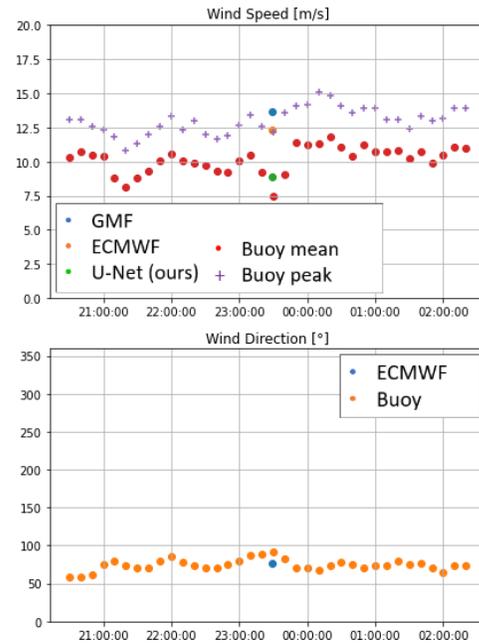


Fig. 9: Time series of the NDBC buoy wind measurements around October 06th 2019 at 23:28:53 for NDBC 41009, and the estimation from the GMF, the deep learning model and the atmospherical model.

caused by a drop in wind speed rather than a change in direction, as the latter does not significantly change during the passage of the convective cell (possibly because the underlying wind regime is strong). It does record a sudden drop in wind speed to 7.5 m/s one minute after the SAR observation, while the GMF and the deep learning model estimated wind speeds of 13.7 m/s and 8.9 m/s, respectively.

V. CONCLUSION

Previous studies have shown that high-resolution rain signatures can be automatically extracted from SAR observations. Using this SAR rainfall segmenter, we built a wind estimation dataset where 50% of the patches contain rainfall examples. Samples were chosen so that a SAR-based and a SAR-independent wind speed model agree on non-rain pixels, ensuring that their estimates are close to the

true wind speed. A UNet architecture was trained on this dataset to estimate wind speeds based on the SAR-independent atmospheric model. We tested several input combinations and found that the most important parameter was the wind speed prior from the geophysical model function, which the deep learning model had difficulty emulating.

Collocations with buoy in-situ measurements show that the model outperforms the current Geophysical Model Function (GMF) on rain areas, reducing the RootMean Square Error (RMSE) by 27% (resp. 45%) for rain rates higher than 1 mm/h (resp. 3 mm/h). On rainless areas, performances are similar with a small reduction of the RMSE by 2.7%. However, since the buoys have a time resolution of ten minutes, some quick sub-mesoscale processes, such as gust fronts, are difficult to register. The limited spatial range of the buoys also makes it challenging to observe rare phenomena. Future work should address these concerns.

APPENDIX

A secondary dataset was created differing with the main dataset by Eq. 4. Here, the balancing policy is defined as:

$$\forall x, P(x) = P^+(x) = P^-(x) \quad (\text{Eq. 8})$$

The distributions P^+ and P^- are presented in Fig. 10. The balancing is performed for every wind speed, which lead to remove 84% of the rain patches -especially at 5 and 10 m/s- since the number of rain patches at 8 m/s is limited.

Models were optimized under the same training process as in subsection III-B1, in particular with the same number of weight updates. Comparison with the first balancing scheme display lower performances despite the more accurate balancing, as indicated Table III.

REFERENCES

[1] C. Wang, P. Tando, A. Mouche, J. E. Stopa, V. Gressani, N. Longepe, D. Vandemark, R. C. Foster, and B. Chapron, "Classification of the global sentinel-1 SAR vignettes for ocean surface process studies," *Remote Sensing of Environment*, vol. 234, p. 111457, Dec. 2019. [Online]. Available: <https://doi.org/10.1016/j.rse.2019.111457>

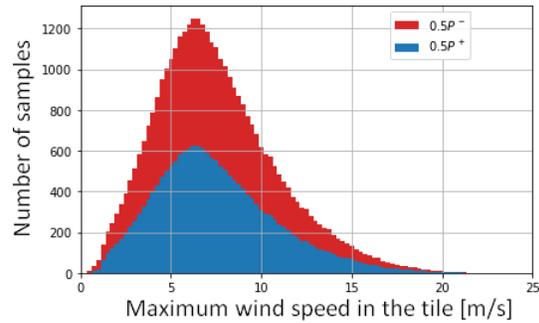


Fig. 10: Wind speed distribution for rain (blue) and rainless (red) patches.

	Dataset I	Dataset II	GMF
Bias	< 1mm/h	0.73 [0.04]	★ 0.71
	[1, 3]mm/h	★1.38 [0.04]	2.18
	> 3mm/h	0.92 [0.07]	★0.83 [0.14]
RMSE	< 1mm/h	★1.44 [0.03]	1.46 [0.06]
	[1, 3]mm/h	★1.81 [0.04]	1.85 [0.10]
	> 3mm/h	★1.60 [0.10]	1.61 [0.10]
PCC	< 1mm/h	★93.6% [0.16%]	93.3% [0.15%]
	[1, 3]mm/h	★96.3% [0.22%]	96.1% [0.23%]
	> 3mm/h	★95.9% [0.35%]	95.0% [0.46%]

Table III: Bias, Root Mean Square Error of model IV, the GMF, for each rainfall level. The best result for each rainfall level is indicated by ★. Results better than the baseline are in bold. Results of deep learning models are given as mean and standard deviation in brackets.

[2] M. M. Barbat, C. Wesche, A. V. Werhli, and M. M. Mata, "An adaptive machine learning approach to improve automatic iceberg detection from SAR images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 156, pp. 247–259, Oct. 2019. [Online]. Available: <https://doi.org/10.1016/j.isprsjprs.2019.08.015>

[3] F. Ronci, C. Avolio, M. di Donna, M. Zavagli, V. Piccialli, and M. Costantini, "Oil spill detection from SAR images by deep learning," in *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, Sep. 2020. [Online]. Available: <https://doi.org/10.1109/igarss39084.2020.9323590>

[4] J. Svejksky and J. Shandley, "Detection of offshore plankton blooms with AVHRR and SAR imagery," *International Journal of Remote Sensing*, vol. 22, no. 2-3, pp. 471–485, Jan. 2001. [Online]. Available: <https://doi.org/10.1080/014311601450040>

[5] H. Shen, W. Perrie, Q. Liu, and Y. He, "Detection of macroalgae blooms by complex SAR imagery," *Marine Pollution Bulletin*, vol. 78, no. 1-2, pp. 190–195, Jan. 2014. [Online]. Available: <https://doi.org/10.1016/j.marpolbul.2013.10.044>

[6] F. Arduin, B. Chapron, and F. Collard, "Observation of

- swell dissipation across oceans,” *Geophysical Research Letters*, vol. 36, no. 6, Mar. 2009. [Online]. Available: <https://doi.org/10.1029/2008gl037030>
- [7] B. Zhang and W. Perrie, “Cross-polarized synthetic aperture radar: A new potential measurement technique for hurricanes,” *Bulletin of the American Meteorological Society*, vol. 93, no. 4, pp. 531–541, Apr. 2012. [Online]. Available: <https://doi.org/10.1175/bams-d-11-00001.1>
- [8] A. A. Mouche, B. Chapron, B. Zhang, and R. Husson, “Combined co- and cross-polarized SAR measurements under extreme wind conditions,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 12, pp. 6746–6755, Dec. 2017. [Online]. Available: <https://doi.org/10.1109/tgrs.2017.2732508>
- [9] A. Long, “C-band v-polarized radar sea-echo model from ERS-1 haltenbanken campaign,” *Journal of Electromagnetic Waves and Applications*, vol. 9, no. 3, pp. 373–391, Jan. 1995. [Online]. Available: <https://doi.org/10.1163/156939395x00532>
- [10] A. Stoffelen and D. Anderson, “Scatterometer data interpretation: Estimation and validation of the transfer function CMOD4,” *Journal of Geophysical Research: Oceans*, vol. 102, no. C3, pp. 5767–5780, Mar. 1997. [Online]. Available: <https://doi.org/10.1029/96jc02860>
- [11] H. Hersbach, “Cmod5 an improved geophysical model function for ers c-band scatterometry,” 2003. [Online]. Available: <https://www.ecmwf.int/node/9861>
- [12] —, “Cmod5.n: A c-band geophysical model function for equivalent neutral wind.” 2008. [Online]. Available: <https://www.ecmwf.int/node/9873>
- [13] A. Elyouncha, X. Neyt, A. Stoffelen, and J. Verspeek, “Assessment of the corrected CMOD6 GMF using scatterometer data,” in *SPIE Proceedings*, C. R. Bostater, S. P. Mertikas, and X. Neyt, Eds. SPIE, Oct. 2015. [Online]. Available: <https://doi.org/10.1117/12.2195727>
- [14] A. Stoffelen, J. A. Verspeek, J. Vogelzang, and A. Verhoef, “The CMOD7 geophysical model function for ASCAT and ERS wind retrievals,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 5, pp. 2123–2134, May 2017. [Online]. Available: <https://doi.org/10.1109/jstars.2017.2681806>
- [15] Y. Lu, B. Zhang, W. Perrie, A. Mouche, X. Li, and H. Wang, “A c-band geophysical model function for determining coastal wind speed using synthetic aperture radar,” in *2018 Progress in Electromagnetics Research Symposium (PIERS-Toyama)*. IEEE, Aug. 2018. [Online]. Available: <https://doi.org/10.23919/piers.2018.8598163>
- [16] H. Shen, Y. He, and W. Perrie, “Speed ambiguity in hurricane wind retrieval from SAR imagery,” *International Journal of Remote Sensing*, vol. 30, no. 11, pp. 2827–2836, Jun. 2009. [Online]. Available: <https://doi.org/10.1080/01431160802555879>
- [17] P. A. Hwang, A. Stoffelen, G.-J. van Zadelhoff, W. Perrie, B. Zhang, H. Li, and H. Shen, “Cross-polarization geophysical model function for c-band radar backscattering from the ocean surface and wind speed retrieval,” *Journal of Geophysical Research: Oceans*, vol. 120, no. 2, pp. 893–909, Feb. 2015. [Online]. Available: <https://doi.org/10.1002/2014jc010439>
- [18] A. Mouche, B. Chapron, J. Knaff, Y. Zhao, B. Zhang, and C. Combet, “Copolarized and cross-polarized SAR measurements for high-resolution description of major hurricane wind structures: Application to irma category 5 hurricane,” *Journal of Geophysical Research: Oceans*, vol. 124, no. 6, pp. 3905–3922, Jun. 2019. [Online]. Available: <https://doi.org/10.1029/2019jc015056>
- [19] B. Zhang, W. Perrie, J. A. Zhang, E. W. Uhlhorn, and Y. He, “High-resolution hurricane vector winds from c-band dual-polarization SAR observations,” *Journal of Atmospheric and Oceanic Technology*, vol. 31, no. 2, pp. 272–286, Feb. 2014. [Online]. Available: <https://doi.org/10.1175/jtech-d-13-00006.1>
- [20] Y. Lu, B. Zhang, W. Perrie, A. Mouche, and G. Zhang, “CMODH validation for c-band synthetic aperture radar HH polarization wind retrieval over the ocean,” *IEEE Geoscience and Remote Sensing Letters*, vol. 18, no. 1, pp. 102–106, Jan. 2021. [Online]. Available: <https://doi.org/10.1109/tgrs.2020.2967811>
- [21] S. Wang, K.-V. Yuen, X. Yang, and B. Zhang, “A nonparametric tropical cyclone wind speed estimation model based on dual-polarization SAR observations,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022. [Online]. Available: <https://doi.org/10.1109/tgrs.2022.3188328>
- [22] W. Alpers, B. Zhang, A. Mouche, K. Zeng, and P. W. Chan, “Rain footprints on c-band synthetic aperture radar images of the ocean - revisited,” *Remote Sensing of Environment*, vol. 187, pp. 169–185, Dec. 2016. [Online]. Available: <https://doi.org/10.1016/j.rse.2016.10.015>
- [23] A. Colin, P. Tandeo, C. Peureux, R. Husson, N. Longépé, and R. Fablet, “Rainfall estimation with sar using nexrad collocations with convolutional neural networks,” 2022. [Online]. Available: <https://arxiv.org/abs/2207.07333>
- [24] G. Fracastoro, E. Magli, G. Poggi, G. Scarpa, D. Valsesia, and L. Verdoliva, “Deep learning methods for synthetic aperture radar image despeckling: An overview of trends and perspectives,” *IEEE Geoscience and Remote Sensing Magazine*, vol. 9, no. 2, pp. 29–51, Jun. 2021. [Online]. Available: <https://doi.org/10.1109/mgrs.2021.3070956>
- [25] S. Wei, H. Zhang, X. Zeng, Z. Zhou, J. Shi, and X. Zhang, “CARNet: An effective method for SAR image interference suppression,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 114, p. 103019, Nov. 2022. [Online]. Available: <https://doi.org/10.1016/j.jag.2022.103019>
- [26] C. Wang, A. Mouche, P. Tandeo, J. E. Stopa, N. Longépé, G. Erhard, R. C. Foster, D. Vandemark, and B. Chapron, “A labelled ocean SAR imagery dataset of ten geophysical phenomena from sentinel-1 wave mode,” *Geoscience Data Journal*, vol. 6, no. 2, pp. 105–115, Jul. 2019. [Online]. Available: <https://doi.org/10.1002/gdj3.73>
- [27] Precipitation Processing System (PPS) At NASA GSFC, “Gpm dpr precipitation profile 12a 1.5 hours 5 km v07,” 2021. [Online]. Available: <https://disc.gsfc.nasa.gov/datacollection/>
- [28] R. O. C. NOAA National Weather Service, “Noaa next generation radar (nexrad) level ii base data,” 1991. [Online]. Available: <https://www.ncei.noaa.gov/metadata/geoportal>
- [29] M. Bourbigot, P. Vincent, H. Johnsen, and R. Piantanida, “Sentinel-1 IPF Auxiliary Product Specification - Sentinel Online — sentinel.esa.int.”

[30] Y. Quilfen, B. Chapron, T. Elfouhaily, K. Katzaros, and J. Tournadre, "Observation of tropical cyclones by high-resolution scatterometry," *Journal of Geophysical Research: Oceans*, vol. 103, no. C4, pp. 7767–7786, Apr. 1998. [Online]. Available: <https://doi.org/10.1029/97jc01911>

[31] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *CoRR*, vol. abs/1505.04597, 2015. [Online]. Available: <http://arxiv.org/abs/1505.04597>

[32] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," 2012. [Online]. Available: <https://arxiv.org/abs/1211.5063>

[33] A. Colin, R. Fablet, P. Tandeo, R. Husson, C. Peureux, N. Longépé, and A. Mouche, "Semantic segmentation of meteorologic processes using SAR observations and deep learning," *Remote Sensing*, vol. 14, no. 4, p. 851, feb 2022. [Online]. Available: <https://doi.org/10.3390/rs14040851>

[34] I. de Gelis, A. Colin, and N. Longepe, "Prediction of categorized sea ice concentration from sentinel-1 SAR images based on a fully convolutional network," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 5831–5841, 2021. [Online]. Available: <https://doi.org/10.1109/jstars.2021.3074068>

[35] L. de Montera, H. Berger, R. Husson, P. Appelghem, L. Guerlou, and M. Fragoso, "High-resolution offshore wind resource assessment at turbine hub height with sentinel-1 synthetic aperture radar (SAR) data and machine learning," *Wind Energy Science*, vol. 7, no. 4, pp. 1441–1453, Jul. 2022. [Online]. Available: <https://doi.org/10.5194/wes-7-1441-2022>

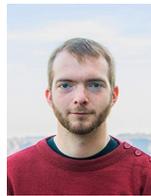
[36] E. W. Peterson and J. P. Hennessey, "On the use of power laws for estimates of wind power potential," *Journal of Applied Meteorology*, vol. 17, no. 3, pp. 390–394, Mar. 1978. [Online]. Available: [https://doi.org/10.1175/1520-0450\(1978\)017<0390:otuopl>2.0.co;2](https://doi.org/10.1175/1520-0450(1978)017<0390:otuopl>2.0.co;2)



Aurélien COLIN After graduating from the IMT Atlantique as an engineer, he began a Ph.D and is currently studying the segmentation of various meteorological and ocean phenomena on Synthetic Aperture Radar using deep learning models.



Pierre TANDEO was born in France in 1983. He received the M.S. degree in applied statistics from Agrocampus Ouest, Rennes, France, and the Ph.D. degree from the Oceanography from Space Laboratory at IFREMER, Brest, France, in 2010. Then, he spent two years as a Postdoctoral Researcher with the Atmospheric Science Research Group, University of Corrientes, Argentina, and three years at Télécom Bretagne, Brest, France. Since 2015, he is an associate professor at IMT Atlantique, Brest, France, and a researcher at Lab-STICC, CNRS, France. Since 2019, he is an associate researcher at the Data Assimilation Research Team, RIKEN Center for Computational Science, Kobe, Japan. His main research interests are focused on IA, data assimilation, and inverse problems for geophysics.



Charles PEUREUX is a research engineer at CLS since 2020 where he is mainly involved with SAR oceanography. He previously worked at IFREMER where he defended his thesis on the observation and modelling of the directional properties of short gravity waves. Thereafter, he worked on the SKIM satellite mission, especially for the production of high resolution sea state numerical simulations.



Romain HUSSON is a scientist and project engineer at CLS, in the Environmental Monitoring and Climate Business Unit. After working as intern at NASA JPL and as YGT at ESRIN on SAR wind-wave-current activities, he completed in 2012 his PhD in Physical Oceanography at CLS Brest developing methods to estimate synthetic swell field from SAR wave mode observations. Since then, he has been involved in several applicative and R&D projects such as wind/wave products for marine renewable energy and contributed to several ESA and European projects such as Cal/Val activities for ENVISAT and S-1 mission performance center as Level-2 expert.



Ronan FABLET is a Professor at IMT Atlantique and a research scientist at Lab-STICC in the field of Data Science and Computational Imaging. He is quite involved in interdisciplinary research at the interface between data science and ocean science, especially space oceanography and marine ecology. His current research interests include deep learning for dynamical systems and applications to the understanding, analysis, simulation and reconstruction of ocean dynamics, especially using satellite ocean remote sensing data.