

# Scale aware deep learning for wide swath altimetry calibration

Quentin Febvre, *Student, IEEE*, Ronan Fablet, Clément Ubelmann, Julien Le Sommer

**Abstract**—Sea surface height (SSH) maps are used for observing and studying mesoscale surface ocean dynamics. For the past few decades, one dimensional satellite observations have been used to produce such maps. The Surface Water and Ocean Topography (SWOT) mission will deploy a new sensor that will acquire two dimensional profiles of the SSH. One aim of this mission is to observe the ocean at previously unresolved spatio-temporal scales. One challenge that needs to be addressed before exploiting the SWOT data is to separation of the SSH signal from other instrument or geophysical errors. In this paper, we propose a novel learning based approach to sensor calibration that relies on a scale decomposition step informed by the structure of the different signals. In a supervised setting, our method reaches/improves state of the art residual error while proposing a correction on the full spectrum on the observation and using weaker constraints on the modeled error signal.

**Index Terms**—Deep Learning, Altimetry, Calibration, SWOT.

## I. INTRODUCTION

For the last decades the sea surface height (SSH) observations from NADIR satellites have been mapped to study ocean surface dynamics. However due to the sparse and irregular sampling of the observations, the spatial scales below 100km are not resolved. The SWOT mission with its radar interferometer will take high resolution two dimensional measures along a wide swath. It should make it possible to resolve the scales down to 15km. However these new observations will come with unknown errors that will need to be calibrated. The calibration step will be challenging because the error amplitude can be an order of magnitude higher than the SSH's. The approaches considered so far for the cross-calibration of KaRIN data consist in parameterizing the different components of the error signals and then using the information from calibrated sensors to estimate the parameters. [1][2]. In this work, we propose a novel learning-based approach for the cross-calibration problem. On one hand we use the fact that the error signals will have a specific energy distribution in the spectral domain thanks to the mission specifications of the error budget at different spatial scales. On the other hand, the estimation error made from interpolating the NADIR tracks also has a characteristic spectral distribution. We will show how we can use the specificity of those two error signals to produce an estimate of the SSH observed by the SWOT mission. Our method rely on leveraging the different spectral structures of the interpolation error and the observation error through a scale decomposition and renormalization step. We then train a convolutional neural network to separate the SSH from the errors using the different scale informations.

Thanks to IDRIS, Ifremer, CNES, CLS

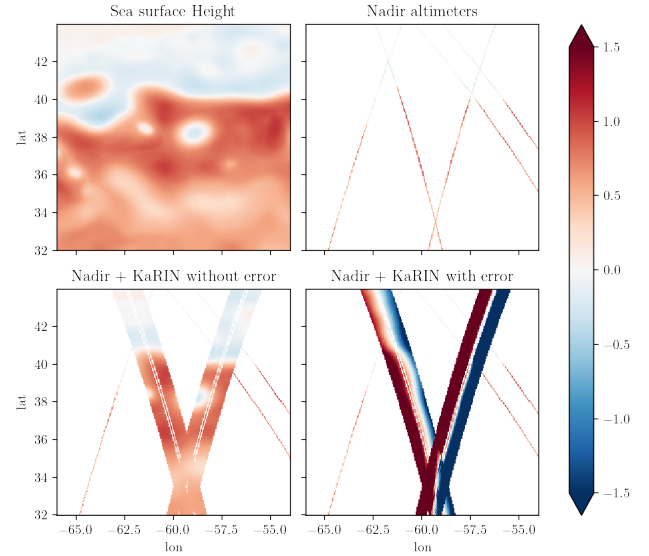


Fig. 1. SSH, Nadirs, SWOT KaRIN with and without errors

Contrary to existing approaches, our method doesn't directly rely on explicit formulation of the error signals, but requires simulated SSH and error data to train the model. Therefore in order for our method to be considered for calibrating KaRIN data, it remain to be shown that our calibration operator trained on simulated data can generalize to real signals. On the other hand since our method rely on both the error signal structure and the SSH structure, we could hope that our method could manage to capture unplanned error signals where existing approaches can't since they rely on explicit formulations of the estimated error. Finally one advantage of our method is that it can provide a correction on the whole spatial spectrum whereas other approaches only estimate the errors above a certain scale

Our contribution are the following:

- We formulate a calibration learning problem
- We propose a scale decomposition scheme that allows for better separability of the SSH and the errors
- We present a convolutional neural network architecture that exploits a gridded product to reach state of the art calibration performance

## II. PRELIMINARIES

### A. Sensors

In this paper we propose a cross calibration method, meaning that we rely on external calibrated sensor data.

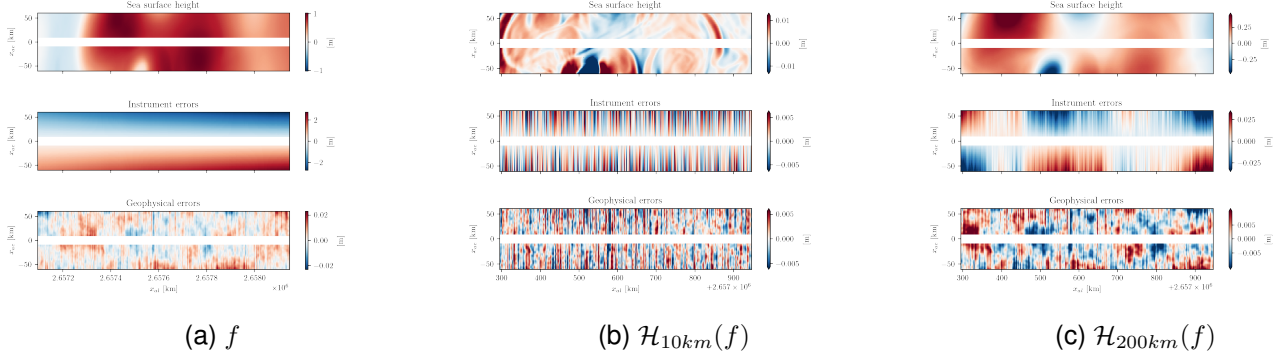


Fig. 2. Uncalibrated KaRIN observation components unfiltered and through high pass filter

We consider a NADIR satellite constellation as source of calibrated data, and we aim to calibrate the data from the Ka-band radar interferometer (KaRIN) sensor from the SWOT mission. The NADIR altimeters acquire one dimensional profiles of the SSH vertically under the satellite. In addition to a nadir altimeter, The SWOT mission will deploy a novel instrument: the KaRIN. Contrary to the nadir altimeters, the KaRIN instrument will observe a two dimensional swath approximately 120 kilometers wide.

In Figure 1, you can visualize a simulated SSH field over the Gulfstream domain, as well as the gridded SSH observed by a constellation of 4 nadirs altimeters with and without the SWOT mission. In the bottom right panel, we can see how the error signals completely obfuscate the observed SSH in the KaRIN observation.

### B. Observation errors

In figure 2 left most panel we display in the swath geometry the composition of the uncalibrated KaRIN observations considered, it's composed of three parts: the target SSH quantity, instrument errors and geophysical errors.

We can see that the overall uncalibrated signal is largely dominated by the instrument error. Numerically we can compute that over 75% of the energy of the uncalibrated observation is due to the instrument error. We show in section IV-C that if we use the raw uncalibrated observation as input, traditional neural architecture fail to learn to separate the SSH from the errors. However the instrument errors seem to vary slowly in the along track dimension compared to the other components. This is due to the spectral error specification of the SWOT mission that restricts the error energy in certain spectral domain but don't specify a constraint above a certain scale. To further illustrate this point, we introduce an highpass filter  $\mathcal{H}_\sigma$  of wavelength  $\sigma$  detailed in equation 1. This operator takes the residual of the signal  $f$  after application of a gaussian filter of standard deviation  $\sigma$  in the along track direction  $x_{ac}$ . The output of this filter for sigmas of 10km and 200km are

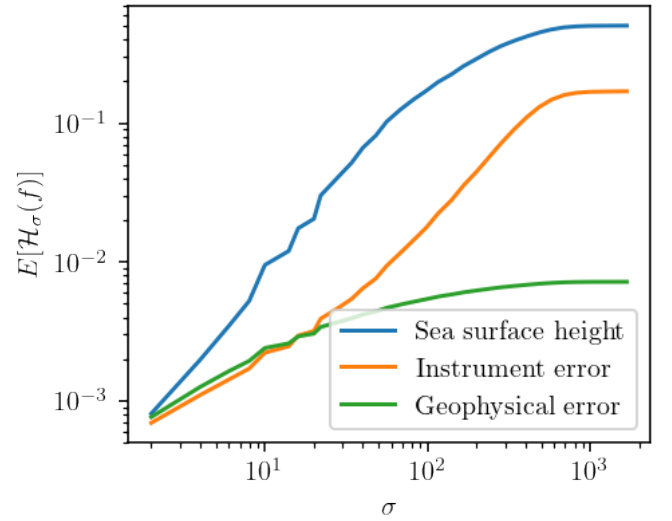


Fig. 3. Energy of filtered KaRIN observations components in function of  $\sigma$

shown in figure 2 center and right panels.

$$\mathcal{H}_\sigma(f(x_{ac}, x_{al})) = f(x_{ac}, x_{al}) - \sum_{k \in [-N, N]} K e^{-\frac{(k\delta x)^2}{\sigma^2}} f(x_{ac}, x_{al} + k\delta x) \quad (1)$$

If we look at energy of the observation components for different  $\sigma$  as in 3, we can see that even for high values of sigmas, the target SSH signal is prevalent compared to the instrument and geophysical errors. Meaning that at these scales, we actually filter out significantly more instrument error than SSH signal. Note that the flat behaviour we observe for sigmas 500 km, is due to the side effects induced by fact that we are reaching a filtering scale close to the whole domain considered.

### C. Sea surface height interpolation

Additionally instead of directly considering the NADIR observations as reference for the cross calibration task, we use interpolated maps that are produced with methods described in this section as well.

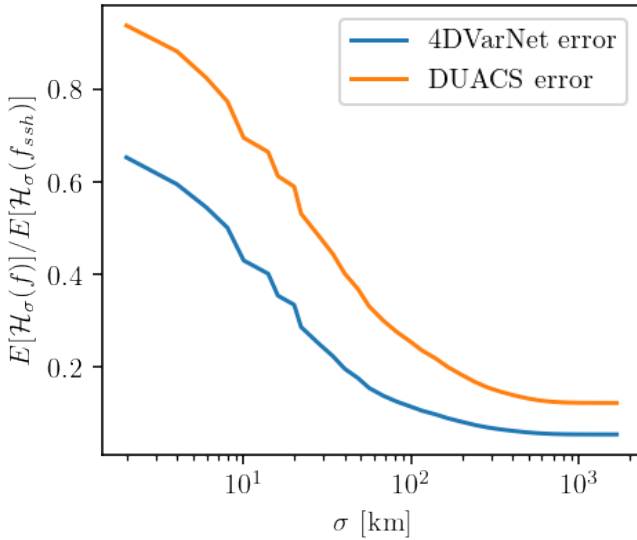


Fig. 4. Relative interpolation error energies to the SSH energy at different scales

The sea surface height quantity is determined by complex phenomena that behave differently at different spatial and temporal scales. There is a correlation between the spatial scales, temporal scales and energy of the structures we find in the SSH field. The smaller and more short-lived structures will contain less energy than the bigger and more durable phenomena. We have an illustration of this in the figure 3 when we observe the exponential evolution of the SSH energy in function of the filtering parameter  $\sigma$ : The bigger the scale the more energy it represent. This is relevant when considering the reconstruction of the interpolated field from partial observations. Given a domain of which we observe a few percents each day as in figure 1 top right panel, we will only be able to track and reconstruct the larger structures and phenomena (100km). These reconstructions will capture most of the energy of the total signal but will miss the dynamics that occur at finer scales.

In figure 4 we display the relative energy of the interpolation error on the SWOT swath at different filtering scales for two state of the art interpolation methods for altimetry data: the DUACS [3] operational product applied to a constellation of 4 NADIR altimeter satellites and the learning-based 4DVarNet which additionally uses the SWOT mission's nadir altimeter [4][5]. We can see the same trend on both curves: we consider the high scales, most of the energy of the SSH signal is recovered but the finer the scale we look at, the less SSH energy we have in the remaining signal. As explained above, this is to be expected independently of the interpolation method due to the NADIR sampling and the nature of the SSH dynamics.

#### D. Convolutional neural networks

Convolutional neural network have been widely considered the standard architecture for image processing in the deep learning field for a variety of task such as image classification[6],

[7], image inpainting[8], object detection[9] and more. More recently, convolutional architecture have also applied to remote sensing problems such as SAR image segmentation[10], [11], altimetry data interpolation [12] and even sensor calibration [13]. Related to the work presented in this paper, a learning based approach for the joint problem of SWOT calibration and mapping using NADIR altimeters has been proposed in [14]. Eventhough this approach opened some interesting perspectives on jointly treating the calibration and mapping problem for error signals with low amplitude, it failed to handle more realistic error amplitude like the one we deal with in this work. We believe that this is in part due to the fact that it treats the calibration problem in a reference grid geometry in which it fails to exploit the along track spectral features of the error signals.

### III. PROPOSED METHODOLOGY

We consider the cross calibration problem of separating the error signals from the SSH in the KaRIN observation using calibration NADIR observations. In order to compare the NADIR observations with the KaRIN, we first produce an interpolated map that we reinterpolate on the KaRIN observation domain. We train a neural network to compute the KaRIN SSH signal using the uncalibrated KaRIN observation and the interpolated nadirs as inputs.

The data, mapping algorithm and neural network architecture are described in this section.

#### A. Observing-systems simulation experiments (OSSE)

We rely on simulated data in this study.

The sea surface height field is taken from the NATL60 [15] simulation of the NEMO ocean model. We then sample nadirs and wide-swath profile corresponding to real orbits for the nadirs and projected 21-day cycle orbit for the SWOT satellite. We focus on calibrating the spatially correlated error signals, we use the simulated error signals from the SWOT simulator [16]. The instrument errors considered are: the roll, timing, baseline dilatation, phase. The geophysical error signal that we use is the wet troposphere residual error. The detail behind the generation of each of those signal can be found here [16], [1]

#### B. Network architecture

The overall architecture considered is shown in figure 5. The model takes as inputs the uncalibrated KaRIN observations and a gridded product made from the calibrated NADIRS using a mapping algorithm as described in section III-D. The gridded product is then is reinterpolated in the KaRIN geometry. The inputs go through a scale decomposition a renormalization step as described in section III-E that produces a unified input with channels corresponding to different scale components. An input convolution layer is used to change the number of channels. The data is then processed by a series a blocks composed of a convolution layer, a ReLU non-linearity, a skip connection and a mixing layer as described in III-C. Finally a last convolution layer is used to compute a point wise

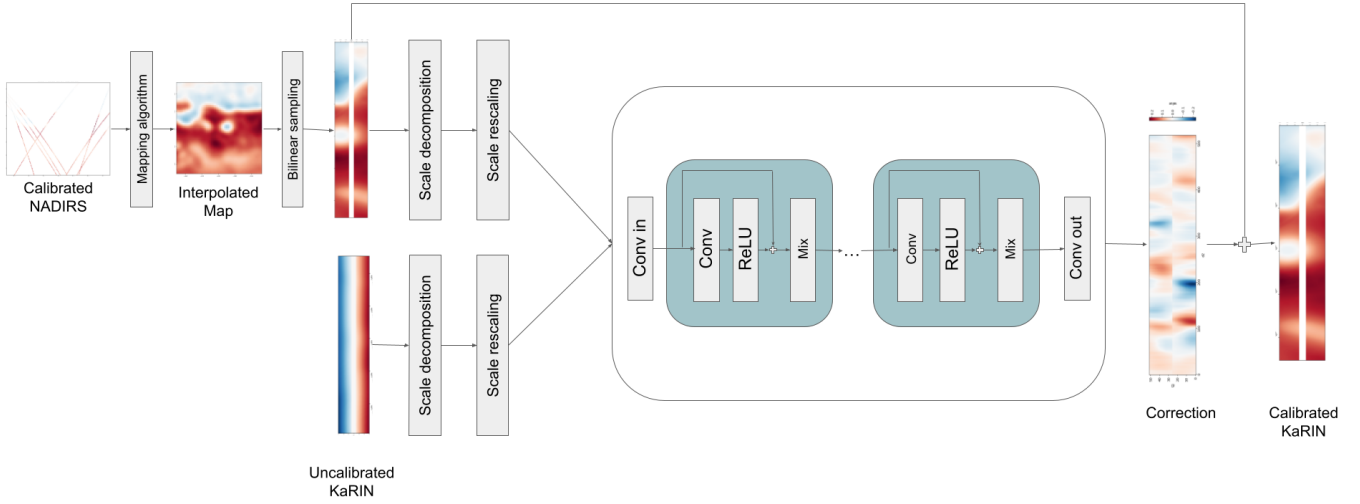


Fig. 5. Overview of the proposed architecture

correction of the gridded product from the different channels. The network is therefore trained to extract the difference between the true SSH observed by SWOT and the gridded product. Each convolution layer input is padding so that the height and width of the input remain unchanged throughout the network.

As one can see in figure 2, the swath observed from the KaRIN sensor is not contiguous in the across-track dimension. In order to avoid the for the convolution kernel to be spread over the two sides of the swath which could add some unwanted side effect, each side is treated separately by each convolution layer.

### C. Mixing Layer

As shown in recent work [17], [18] mixing layers, can help with the expressivity of neural networks. We implement our own layer based on this idea using a convolution layer that we apply using the across-track dimension as channels. This layer behaves like a fully connected layer in the across-track dimension, and is applied for each channel and for each along-track position .

If we note  $v_{in}(c, x_{al}) \in \mathbb{R}^{N_{ac}}$  the across track section of the input of a mixing layer at channel  $c$  and along-track coordinate  $x_{al}$  of dimension  $N_{ac}$  the number of across track coordinates, and  $\mathcal{M}(v_{in}, W, b)$  the mixing layer with weight  $W \in \mathbb{R}^{N_{ac} \times N_{ac}}$  and bias  $b \in \mathbb{R}^{N_{ac}}$ , the computation performed by the mixing layer is described in 2.

$$\mathcal{M}(v_{in}(c, x_{al}), W, b) = v_{in}(c, x_{al}) \times W + b \quad (2)$$

with  $\times$  being the matrix multiplication. We analyse in section IV-C the contribution of this layer.

### D. Mapping Algorithm

The main mapping algorithm considered produce the interpolated field from the NADIR observation is the

4DVarNet algorithm[12]. This method is based on a trainable adaptation of the 4DVar[19] variational data assimilation method. The interpolated state  $\hat{x}$  is reconstructed using the NADIR observations  $y$  by minimizing a variational cost  $\mathcal{U}(x, y)$

$$\hat{x} = \underset{x}{\operatorname{argmin}} [\mathcal{U}(x, y)] \quad (3)$$

$$= \underset{x}{\operatorname{argmin}} [\mathcal{U}_{obs}(x, y) + \mathcal{U}_{prior}(x)] \quad (4)$$

with  $\mathcal{U}_{obs}(x, y)$  a data fidelity term and  $\mathcal{U}_{prior}(x)$  a regularization term trained on simulated data to minimize the reconstruction loss. The minimization algorithm is a LSTM [20] based gradient descent as described in [21]

This method outperforms concurrent approaches on the GULFSTREAM domain in an OSSE setup [22]. Therefore, it provides the best first guess for the calibrated KaRIN SSH and therefore should allow to produce the best calibration. However the method we propose is not restricted to this mapping algorithm. Indeed as mentionned in section II-C, the spectral structure of the interpolation error that enables the approach proposed here is common across different interpolation method. We show how our method depends on the mapping algorithm in section IV-D

### E. Scale decomposition

The task of extracting relevant SSH information from the uncalibrated observation is challenging because the fine scales signals that are not reconstructed by the interpolation of the NADIRS contain very low energy compared to the uncalibrated KaRIN. Indeed the interpolation correction would represent around one percent of the raw KaRIN observation. But as seen in section 1, if we look at the signal at different scales, this ratio can change significantly. This is illustrated in Figure 9 if we look at the relative energy of the observation and interpolation errors, we can see that



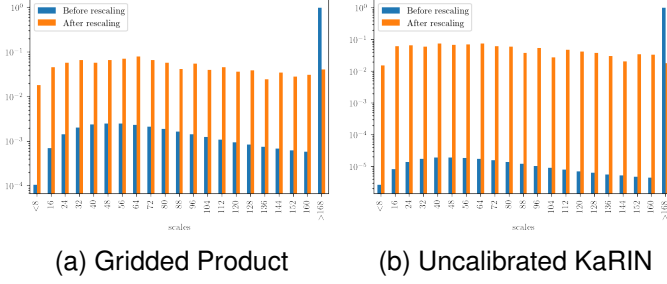


Fig. 6. Percentage of explained variance of scale decomposition components before and after learnt rescaling

there is a spatial scale domain between 10km and 100km where the observation errors are less energetic than the interpolation error. More over learning a scale separation similar to what we introduce in equation 1 with a convolutional neural network would either require very large convolutional filters or a deeper architecture that we would like. Therefore we introduce the following explicit scale decomposition scheme: given a set of increasing scales  $[\sigma_0, \sigma_1, \dots, \sigma_S]$ , we decompose the input signal  $f$  into the  $S + 1$  components  $[\mathcal{H}_{\sigma_0}(f), \mathcal{H}_{\sigma_1}(f) - \mathcal{H}_{\sigma_0}(f), \dots, \mathcal{H}_{\sigma_S}(f) - \mathcal{H}_{\sigma_{S-1}}(f), f - \mathcal{H}_{\sigma_S}(f)]$ . The different components are then considered as channels for the convolutional networks. As of now we consider the decomposition based on a set of 20 scales evenly spaced by 8km. We discuss in section IV-E how sensitive the proposed method is to the choice of the decomposition. In figure 6 in which we display the explained variance of each component. We observe that without some kind of normalization the variance of the input is concentrated on the last component of the decomposition. The variance of the other components can be up to 4 order of magnitude lower. In order to rescale the components, we introduce a batch normalization layer [23] that will track the statistics of the different channels and learn a different output variance and means for each component. The explained variance of each component after the batch normalization rescaling is also depicted in figure 6

The intuition behind this decomposition is that we expect the KaRIN SSH field to be closer to the interpolated NADIRS for some components and closer to the uncalibrated observation for others, and that the CNN will be able to learn to reconstruct the best estimate from those. In order to confirm this idea, we show the ratio of the NADIR interpolation error and the observation error at different across track coordinate  $x_{ac}$  for each component in figure 7. We indeed see that there is a shift in the ratio at different scales depending on  $x_{ac}$ . The overall tendency is that the interpolation error is bigger than the observation error at lower scales the ratio shifts between 50km and 100km. We can note that we observe the biggest ratios for the first and last component, and for both the observation error is more than 10 times larger than the interpolation error.

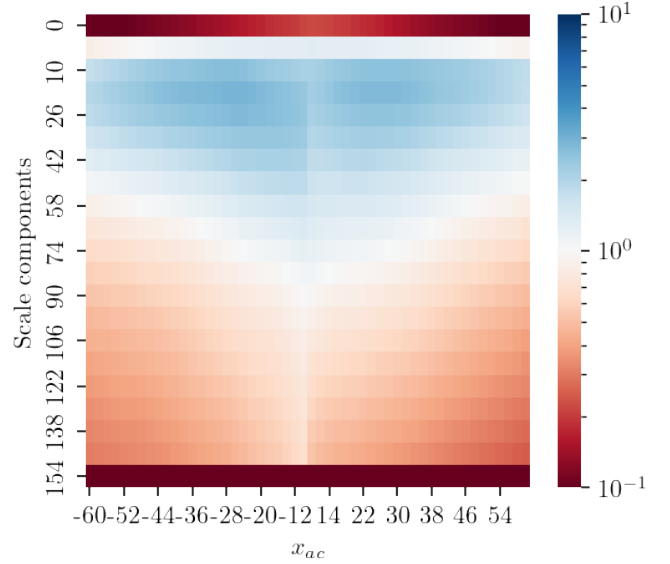


Fig. 7. Ratio of the interpolation error and the observation error  $\frac{E[f_{map} - f_{ssh}]}{E[f_{obs} - f_{ssh}]}$

## IV. EXPERIMENTAL RESULTS

### A. Setup

The results of this section have been computed using the one year ocean simulation NATL60, over the gulfstream domain. Forty days are used for model evaluation and the rest is used for the training of the mapping and calibration models. The experiment setup used is the same as [22]. The base configuration uses three convolutional blocks with 128 channels as presented in figure 5. The model is trained on the reconstruction of the SSH, the gradients and the laplacian. The default scale decomposition used is made of twenty 8 kilometers band. The calibration model is trained for 250 epochs with an annealing triangular cyclical learning rate [24].

### B. Calibration

The main result of this paper is shown in table I. We compare our approach CalCNN with a standard UNet [25] architecture that is trained to output the calibrated observation from the gridded product and the uncalibrated KaRIN stacked as two channels. We also put as baseline the metrics of estimating the KaRIN SSH from the NADIR observations only, using the 4DVarNet method. We compare those three estimations on two metrics, the root mean squared error (RMSE) of the SSH field, and the RMSE of the amplitude of the gradients of the SSH field. As we can see where the UNet fails to produce a better estimate than the mapping algorithm our CalCNN improves the estimation of the SSH and its gradient by over 35% and brings the residual error below 1.5 centimeters.

In figure 9, we show the evolution for different  $\sigma$  of the relative energy for the different error fields filtered by  $\mathcal{H}_\sigma$  as described in 1. We see that for all the scales considered the calibration output manages to reach a lower error than both the observation and the interpolation. We see how at larger scales

|               | RMSE (cm) | RMSE $\ \nabla_{ssh}\ $ |
|---------------|-----------|-------------------------|
| CalCNN        | 1.39e-02  | 6.46e-03                |
| UNet          | 2.34e-02  | 1.07e-02                |
| 4DVarNet-5nad | 2.17e-02  | 9.57e-03                |

TABLE I  
RESIDUAL ERROR THE PROPOSED CALIBRATION FRAMEWORK CALCNN

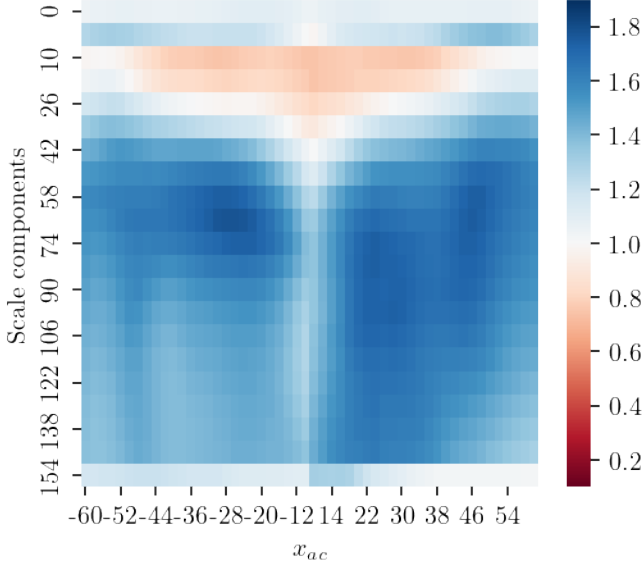


Fig. 8. Scale wise improvement of the calibration output w.r.t. the inputs  $\frac{\min(E[f_{obs}-f_{ssh}], E[f_{map}-f_{ssh}])}{E[f_{cal}-f_{ssh}]}$

the error get closer to the interpolation due to the high errors of the observation. Interestingly, at scales lower than 10km, the algorithm manages to improve on the interpolation even though the observation error is quite high. Between 10-100km we see that the observation errors are lower than the interpolation errors, and our method manages to take advantage of that to improve on the interpolation estimate. Moreover, if we look at the learnt rescaling by the batch normalization layer in figure 10, we see that the scale components with the most weight correspond to this specific scale domain. In order to understand where this improvement comes from we apply our scale decomposition step to the output of our network and compare for each component at each across track position. The results are visible in figure 8. The main takeaway from this figure is that for most of the heatmap we observe values well above one meaning that the calibration network manages to combine information from both inputs produce a better scale estimate than either one. This observation holds for all scale components but for some lower components around 10-30 kilometers, we see that it depends on the across-track position. This can be attributed by the fact that this domain corresponds to the scales and swath positions where the geophysical error is dominant compared to the instrument error. And by design, our approach is better at calibrating the instrument error that has structural features easy to differentiate from an SSH field than the geophysical error for which it is more challenging.

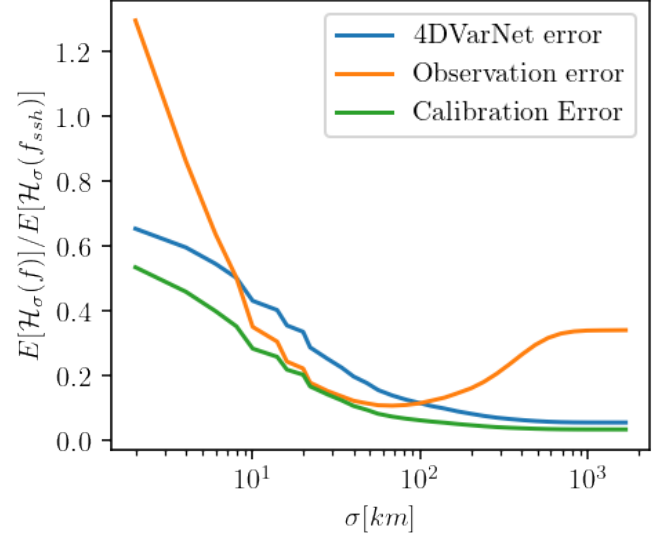


Fig. 9. Relative error energies to the SSH energy at different scales

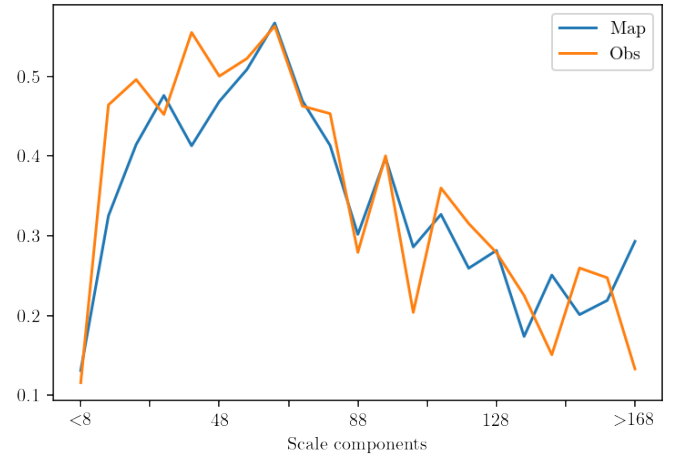


Fig. 10. Standard deviation after rescaling

### C. Ablation Study

In this section we analyse the impact of the different architectural choices made. In the table IV-C we look at the performance metrics of three variations neural network chosen: A model without mixing layers, one without skip connections and one without gridded product as input. As we can see the version without skip connection fail to improve on the 4DVarNet estimation. The version without mixing layer reaches a 10% improvement on the rmse compared to the gridded product which confirms the relevance of those mixing layers. We can also note that without the use of a gridded product as input, the proposed methods does not reach acceptable metrics.

In the following table we show the sensitivity to the size of the network for the same training routine, we compare the base architecture 3x128 with a bigger network 5X512, a smaller network 1x32 as well as a linear version. We can first note that the linear version fail to extract geophysical information

|           | RMSE (cm) | RMSE $\ \nabla_{ssh}\ $ |
|-----------|-----------|-------------------------|
| xp        |           |                         |
| Ref       | 1.39e-02  | 6.46e-03                |
| No skip   | 2.17e-02  | 9.57e-03                |
| No map    | 1.70e-01  | 2.47e-02                |
| No scales | 2.17e-02  | 9.58e-03                |
| No mixing | 1.94e-02  | 9.60e-03                |

|             | RMSE (cm) | RMSE $\ \nabla_{ssh}\ $ |
|-------------|-----------|-------------------------|
| xp          |           |                         |
| 128x3 (Ref) | 1.39e-02  | 6.46e-03                |
| Linear      | 2.13e-02  | 1.02e-02                |
| 32x1        | 1.44e-02  | 6.22e-03                |
| 512x5       | 1.49e-02  | 7.19e-03                |

from the uncalibrated information. Then the smaller and bigger architecture performs similarly with a slight increase in the residual error but the smaller model show a slight improvement in the gradient reconstruction and spatial resolution. Overall, this shows that our training is robust to different sizing of the network and the conclusions we advanced in section IV-B are not sensitive to the hyperparameters of our network architecture.

#### D. Gridded product sensibility

In this section we analyse the impact of different gridded product used as input. In figure 11, we display the improvement in RMSE of the SSH on the swath and of the gradients of the SSH obtained by our CalCNN for different gridded product. We compare three gridded product, the DUACS operational maps produced from a constellation of four NADIR satellites (without the SWOT's mission NADIR altimeter), the 4DVarNet applied to the same four NADIR satellites but using the SWOT NADIR altimeter. And finally the 4DVarNet-SST that uses the 5 NADIR altimeters and the sea surface temperature (SST) field which is integrated in the  $\mathcal{U}_{lf}$  of equation 4 as described in ??.

First result is that for all three interpolated product, we observe the same evolution i.e an improvement of both the RMSE and the RMSE of the gradients.

The fact that DUACS doesn't use the SWOT NADIR altimeter seems to have three consequences. First it is penalized when estimating the SSH on the swath because it is lacking some close by observations. Secondly, the improvement of the calibrated output on the gridded metrics is relatively bigger that for the other two, this show an overlap between the SSH information of the uncalibrated KaRIN and SWOT's NADIR. Finally the calibrated outputs is better when considering other gridded products, which would mean that the gap in the gridded product performance could not be compensated in the calibration step. When comparing the two 4DVarNets, the results are more nuanced, the 4DVarNet-SST has better metrics than the 4DVarNet-5nad, however the difference of RMSE is greatly reduced after calibration whereas the gap in RMSE of the gradients is conserved. This could be interpreted as the gain of RMSE we get from using the SST can be obtained from the uncalibrated

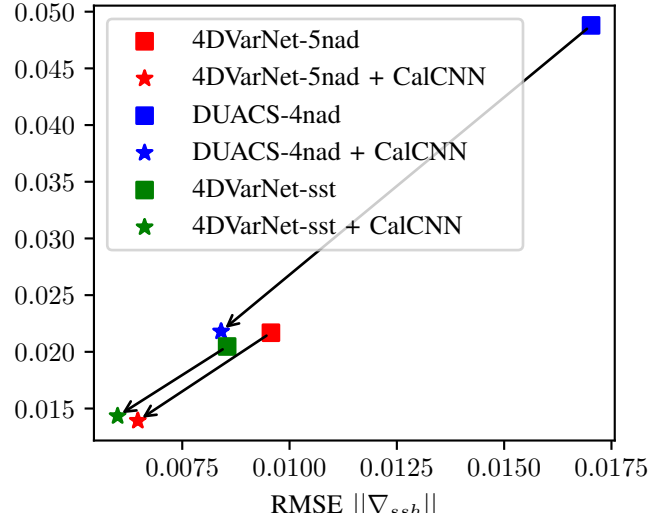


Fig. 11. Evolution of the RMSE and the RMSE  $\|\nabla_{ssh}\|$  of the calibrated observation for different gridded products.

| $N_{band}$ | $\delta_{band}$ | RMSE (cm) | RMSE $\ \nabla_{ssh}\ $ |
|------------|-----------------|-----------|-------------------------|
| 20         | 8               | 1.39e-02  | 6.46e-03                |
| 40         | 8               | 1.54e-02  | 6.88e-03                |
| 40         | 4               | 1.44e-02  | 6.64e-03                |
| 40         | 8               | 1.54e-02  | 6.88e-03                |
| 10         | 16              | 1.48e-02  | 6.75e-03                |
| 10         | 8               | 1.56e-02  | 6.81e-03                |
| 5          | 32              | 1.41e-02  | 6.65e-03                |

TABLE II  
CALIBRATION METRICS IN FUNCTION OF THE SCALE DECOMPOSITION

KaRIN, however some of the gradients we reconstruct through the SST are not easily extractable from the observations. Overall this shows interesting relations between the redundant informations in the uncalibrated KaRIN and the interpolated products product.

#### E. Scale decomposition sensitivity

In table II, we display the calibration metrics for different scale decompositions. We vary the number of scales considered and the spacing between two consecutive ones. We can see that the overall metrics are very similar for quite different scales amplitudes and spacing. The point of this section is to show that our results are not prohibitively sensitive to the choice of a certain scale decomposition.

### V. CONCLUSION

We have proposed in this work a new calibration approach based on a scale decomposition scheme and a convolutional neural network. This approach robustly achieves less than 1.5cm residual error. Additionnally we have shown some interesting interactions between the quality of the gridded product and the calibrated output. This motivates some future work combining the approach described in this paper with some trainable mapping method such as the 4DVarNet.

## REFERENCES

- [1] C. Ubelmann, et al., “A Cross-Spectral Approach to Measure the Error Budget of the SWOT Altimetry Mission over the Ocean,” *Journal of Atmospheric and Oceanic Technology*, vol. 35, no. 4, pp. 845 – 857, 2018, Place: Boston MA, USA Publisher: American Meteorological Society.
- [2] G. Dibarboure, et al., “Empirical Cross-Calibration of Coherent SWOT Errors Using External References and the Altimetry Constellation,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 6, pp. 2325–2344, June 2012, Conference Name: IEEE Transactions on Geoscience and Remote Sensing.
- [3] G. Taburet, et al., “DUACS DT2018: 25 years of reprocessed sea level altimetry products,” *Ocean Science*, vol. 15, no. 5, pp. 1207–1224, 2019.
- [4] R. Fablet, et al., “Learning Variational Data Assimilation Models and Solvers,” *Journal of Advances in Modeling Earth Systems*, vol. n/a, no. n/a, pp. e2021MS002572, \_eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2021MS002572>.
- [5] R. Fablet, et al., “END-TO-END PHYSICS-INFORMED REPRESENTATION LEARNING FOR SATELLITE OCEAN REMOTE SENSING DATA: APPLICATIONS TO SATELLITE ALTIMETRY AND SEA SURFACE CURRENTS,” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. V-3-2021, pp. 295–302, 2021.
- [6] Y. Lecun, et al., “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov 1998.
- [7] K. He, et al., “Deep residual learning for image recognition,” 2016, p. 770–778.
- [8] G. Liu, et al., “Image inpainting for irregular holes using partial convolutions,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 85–100.
- [9] J. Redmon, et al., “You only look once: Unified, real-time object detection,” 2016, p. 779–788.
- [10] A. Colin, et al., “Segmentation of sentinel-1 sar images over the ocean, preliminary methods and assessments,” in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, Jul 2021, p. 4067–4070.
- [11] A. Colin, et al., “Semantic segmentation of metoceanic processes using sar observations and deep learning,” *Remote Sensing*, vol. 14, no. 44, pp. 851, Jan 2022.
- [12] R. Fablet, et al., “Joint Interpolation and Representation Learning for Irregularly Sampled Satellite-Derived Geophysical Fields,” *Frontiers in Applied Mathematics and Statistics*, vol. 7, 2021.
- [13] X. Li, et al., “A Convolutional Neural Network-Based Relative Radiometric Calibration Method,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2022, Conference Name: IEEE Transactions on Geoscience and Remote Sensing.
- [14] Q. Febvre, et al., “Joint calibration and mapping of satellite altimetry data using trainable variational models,” in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2022, p. 1536–1540.
- [15] A. Ajayi, et al., “Spatial and Temporal Variability of the North Atlantic Eddy Field From Two Kilometric-Resolution Ocean Models,” *Journal of Geophysical Research: Oceans*, vol. 125, no. 5, pp. e2019JC015827, 2020, \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1029/2019JC015827>.
- [16] C. Ubelmann, et al., “SWOT Simulator documentation,” .
- [17] A. Vaswani, et al., “Attention is all you need,” in *Advances in Neural Information Processing Systems*, I. Guyon, et al., Eds. 2017, vol. 30, Curran Associates, Inc.
- [18] I. O. Tolstikhin, et al., “Mlp-mixer: An all-mlp architecture for vision,” in *Advances in Neural Information Processing Systems*. 2021, vol. 34, p. 24261–24272, Curran Associates, Inc.
- [19] A. Carrassi, et al., “Data assimilation in the geosciences: An overview of methods, issues, and perspectives,” *Wiley Interdisciplinary Reviews: Climate Change*, vol. 9, no. 5, pp. e535, Sept. 2018, Publisher: Wiley.
- [20] S. Hochreiter et al., “Long short-term memory,” *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [21] M. Andrychowicz, et al., “Learning to learn by gradient descent by gradient descent,” in *Advances in Neural Information Processing Systems*. 2016, vol. 29, Curran Associates, Inc.
- [22] ,” Mar 2022.
- [23] S. Ioffe et al., “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in *Proceedings of the 32nd International Conference on Machine Learning*. Jun 2015, p. 448–456, PMLR.
- [24] L. N. Smith, “Cyclical learning rates for training neural networks,” in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Mar 2017, p. 464–472.
- [25] O. Ronneberger, et al., “U-net: Convolutional networks for biomedical image segmentation,” , no. arXiv:1505.04597, May 2015, arXiv:1505.04597 [cs].

**Quentin Febvre** Quentin Febvre is a 2nd year PhD student at IMT Atlantique Brest