



Φ-lab



## SuperIce project

### Deliverable 2.1

### Report on NextSIM simulation and validation

Version 2.0

24.06.2024

ESA Contract No. 4000142335/231/I-DT • Physically-based synthetic data for Earth observation  
(SD4EO)



History of Changes		
Version	Publication date	Changes
1.0	15/02/2024	Initial Version
2.0	24/06/2024	Updated dataset
Contributors		
Authors	Anton Korosov, Fabio Mangini	
Contributors		
Reviewers	Julien Brajard	

## 1. Summary of main results

### 1.1 neXtSIM simulation

The next Generation Sea Ice Model (neXtSIM, Olason et al., 2022) was run for 10 winters starting from October 2013 until April 2023. The model was run at 3 km spatial resolution and was forced by the following combinations of ocean and atmosphere forcings:

2013 – 2017 : TOPAZ reanalysis

2018 – 2023 : TOPAZ forecasts

2013 – 2022 : ERA 5 reanalysis

2023 – 2023 : ECMWF forecast

Every winter the model fields were initialized from sea ice concentration and sea ice thickness from the CS2SMOS v2.6 product. Ice thickness from CS2SMOS product was assimilated into neXtSIM everyday using data insertion method with weighted average of observations and model fields, with a weight of the observations equal to 0.05 and weight of the model equal to 0.95. The model runs from mid-October to mid-April, when the CS2SMOS data is available.

The model outputs fields of sea ice thickness, concentration, drift and snow depth on grids in polar stereographic projection, at resolution of 3 km, four times a day, averaged over 6-hours. Figure 1 shows examples of simulated sea ice concentration, thickness, and U- V- components of the sea ice drift for 1 January 2023.

The impact of a different forcing would be marginal. We could have slightly lower or slightly higher deformation rates, and the position and orientation of linear kinematic features could be slightly different. However, downscaling to the satellite resolution would smear away all these effects.

The simulated ice thickness was compared to satellite derived sea ice thickness, and it was found that due to initialization and assimilation the bias of the modeled thickness is practically zero during the

entire season. The RMSD of the modeled thickness is quite high, however, this is due to very high resolution of simulations compared to observations.

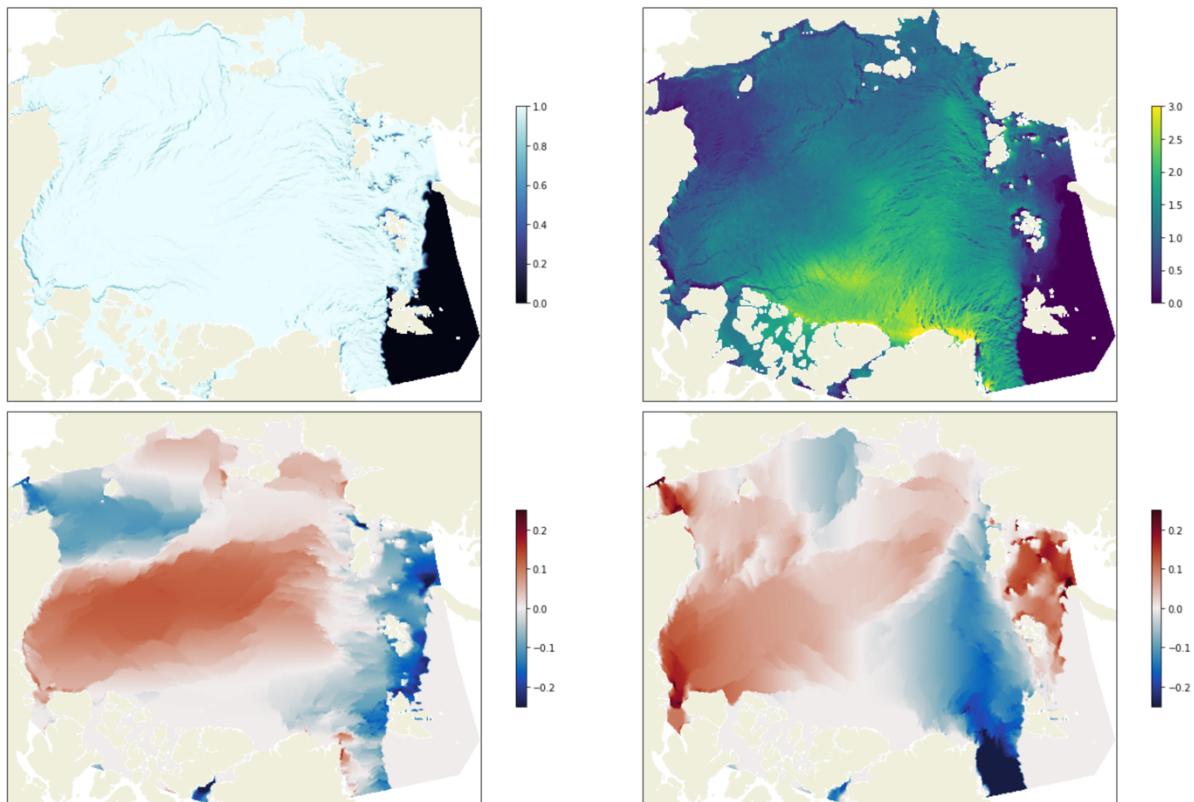


Figure 1. Examples of simulated sea ice concentration, thickness, and U- V- components of the sea ice drift for 1 January 2023.

## References

Ólason, E., Boutin, G., Korosov, A., Rampal, P., Williams, T., Kimmritz, M., Dansereau, V., and Samaké, A.: A new brittle rheology and numerical framework for large-scale sea-ice models, EGU General Assembly 2023, Vienna, Austria, 24–28 Apr 2023, EGU23-11613, <https://doi.org/10.5194/egusphere-egu23-11613>, 2023.

## 1.2 Generating a synthetic low-resolution sea-ice dataset from the neXtSIM output

SuperIce aims to develop an AI-based diffusion model to reconstruct small-scale features of sea ice from low-resolution satellite observations. The first step in this process involves creating a synthetic



low-resolution dataset using neXtSIM to train the AI-based diffusion model. This section provides a brief overview of this dataset.

We presented a first version of the low-resolution dataset in February 2024 (version 1.0 of this deliverable). This was based on neXtSIM's sea-ice concentration (SIC), sea-ice thickness (SIT), divergence, and shear, with the latter two variables having been derived from the U- and V-components of sea-ice drift. The synthetic dataset was generated by applying a 2D Gaussian filter to the neXtSIM dataset, with the standard deviation of the Gaussian kernel (sigma) adjusted to make the model output resemble observations from January 1<sup>st</sup>, 2021.

Three datasets were used to evaluate the quality of the synthetic dataset:

1. "AMSR-2 sea ice concentration product of the EUMETSAT Ocean and Sea Ice Satellite Application Facility (OSI SAF, <https://osi-saf.eumetsat.int>)". The sea-ice concentration (variable name: ice\_conc) ranges between 0 and 100% and has a temporal resolution of 1 day. Additional information can be found at: <https://osisaf-hl.met.no/osi-408-desc>.
2. "CryoSat-SMOS Merged Sea Ice Thickness" version 2.6. The sea-ice thickness (variable name: analysis\_sea\_thickness\_variable) is provided in meters and has a temporal resolution of 1 day. Additional information can be found at: <https://spaces.awi.de/display/CS2SMOS/CryoSat-SMOS+Merged+Sea+Ice+Thickness>.
3. "Low resolution sea ice drift product of the EUMETSAT Ocean and Sea Ice Satellite Application Facility (OSI SAF, <https://osi-saf.eumetsat.int>)". The x- and y-displacements of sea ice (variable names: 'dX' and 'dY') are provided in km and are computed over a period of 2 days. Additional information can be found at: <https://osi-saf.eumetsat.int/products/osi-405-c>.

A few reprocessing steps were needed for the observations and the model output to be comparable. At first, neXtSIM's SIC was adjusted by a factor of 100, for it to range from 0 to 100, instead of 0 to 1. Secondly, the x- and y- sea-ice displacements provided by the observations were used to compute the x- and y- components of sea-ice drift and, in turn, the sea-ice divergence and shear. Importantly, the x- and y- components of sea-ice drift were previously converted from km/2-days into m/s. Finally, the observations were interpolated on the same grid of neXtSIM. The reprocessing applied to the observations is summarized in Table 1.

Dataset Name	Variables	Reprocessing
AMSR-2 sea ice concentration product of the EUMETSAT Ocean and Sea Ice Satellite Application Facility	SIC	Interpolation on: <ul style="list-style-type: none"><li>• NeXtSIM's grid</li></ul>
CryoSat-SMOS Merged Sea Ice Thickness	SIT	Interpolation on: <ul style="list-style-type: none"><li>• NeXtSIM's grid</li></ul>
Low resolution sea ice drift product of the EUMETSAT Ocean and Sea Ice Satellite Application Facility	Sea-ice drift: <ul style="list-style-type: none"><li>• x-component</li><li>• y-component</li></ul> (units: km/2-days)	1. Conversion: <ul style="list-style-type: none"><li>• from km/2-day</li><li>• to m/s</li></ul> through multiplication factor:

		<ul style="list-style-type: none"> <li><math>1000 / (2 * 24 * 3600)</math></li> </ul> <p>2. Computation of:</p> <ul style="list-style-type: none"> <li>• divergence</li> <li>• shear</li> </ul> <p>3. Interpolation on:</p> <ul style="list-style-type: none"> <li>• NeXtSIM's grid</li> </ul>
--	--	--

Table 1: Summary of the observations datasets and the variables used in the analysis, and the reprocessing procedures applied.

The quality of the synthetic dataset was based on the visual inspections of the 2D fields and on analysis of the corresponding power spectra. The results are shown in Figure 2.

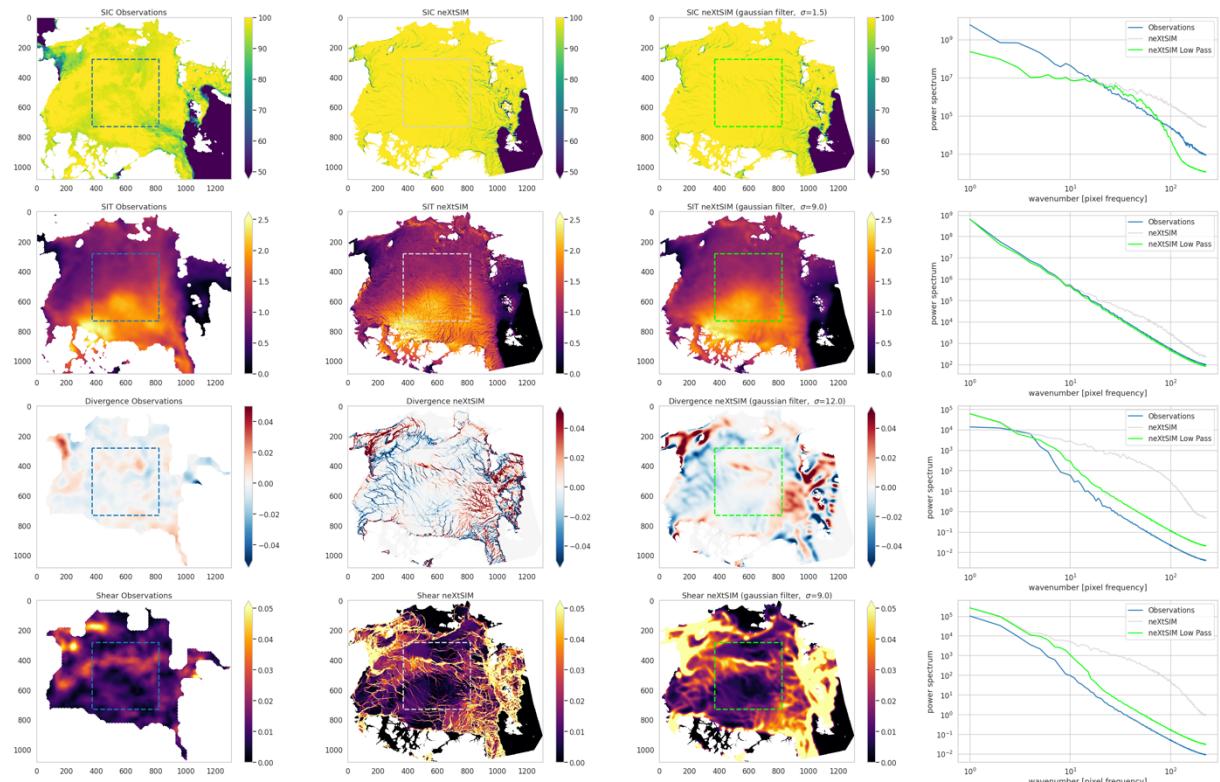


Figure 2. Results from the tuning. The first row presents the results for SIC, the second for SIT, the third for divergence, and the fourth for shear. Each column represents different data sources or processing steps: the first column displays the fields derived from remote sensing observations, the second column presents the output from neXtSIM, and the third column shows the low pass filtered output from neXtSIM. The fourth column provides the power spectra for each variable, based on the three fields within the selected box. The tuning process is based on observations and model output from January 1<sup>st</sup>, 2021.



Figure 2 indicates that the low pass filtering procedure helps increase the similarity between the observations and neXtSIM. The tuning returns a good agreement for SIT, both in terms of spatial variability and amplitude. It also partially improves the similarity between the modelled and the observed divergence and shear, with the modelled fields showing a spatial variability comparable with the observations, but larger amplitude. The least agreement is found for SIC. We hypothesize this to result from the presence of atmospheric noise in the observations.

Overall, the results in Figure 2 suggest that further reprocessing is needed for the synthetic dataset to better resemble the observations. Therefore, we conducted additional analysis and further reprocessing on the neXtSIM's dataset to increase its resemblance to the observations. At first, we tried to determine a sigma value for each variable that is representative of the entire winter season, rather than just a single day. Then, we adjusted the variance of neXtSIM's divergence and shear for them to more closely mirror the observations and, for the same reason, we introduced a synthetic noise to SIC from neXtSIM. These steps and their outcomes are summarized in Table 2 and are now briefly described.

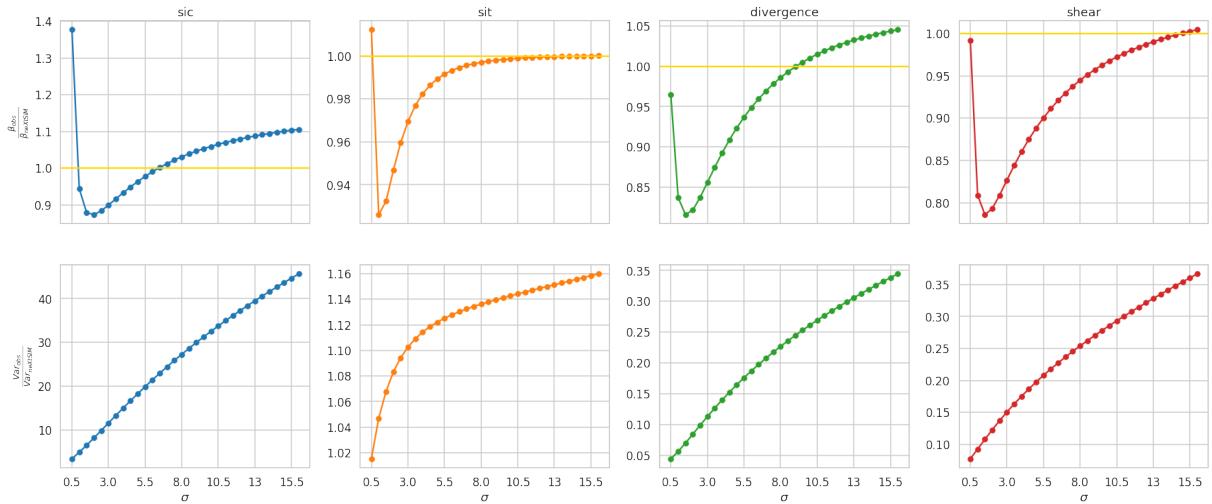
Variable name	Standard deviation of Gaussian kernel (units: n° of pixels)	Further reprocessing
SIC	4.0	Noise added (see main text for more information)
SIT	11.0	None
Divergence	9.0	Multiplication factor of 0.49
Shear	15.0	Multiplication factor of 0.59

Table 2: For each modelled variable (Column 1), list of the standard deviation of the Gaussian kernel used for the low-pass filtering (column 2), and of additional reprocessing procedures (Column 3).

To determine the values of sigma that are more representative for the entire winter season, we considered the winter days between November 20<sup>th</sup>, 2020, and April 12<sup>th</sup>, 2021. For each variable, we first computed the power spectrum of the fields as provided by observations and by neXtSIM. The model fields were previously low pass filtered with varying values of sigma ranging from 0.5 to 16 grid points in increments of 0.5. Then, we winter-averaged the power spectra of the observed and of the low pass filtered model fields, and we computed the corresponding linear trends. The optimal values of sigma were determined by selecting the sigmas that resulted in a ratio of the trends from the observations and the low pass filtered model closest to one. These correspond to 11 grid points for SIT, 9 for divergence, and 15 for shear (upper panel in Figure 3). This analysis also returns a value of 6.5 grid points for SIC; however, this value was discarded for the reasons described later in this section.

For the amplitudes of the synthetic dataset to more closely resemble those from observations, we calculated the daily variance of each variable as provided by observations and by neXtSIM. Notably, the variances were calculated within the box highlighted in each subplot in Figure 2, and the neXtSIM's data had been previously low pass filtered with the same varying values of sigma as for the linear trends. Then, for each value of sigma, we computed the ratio of the variances from the observations and the model, and we winter-averaged the result. Finally, we selected the divergence

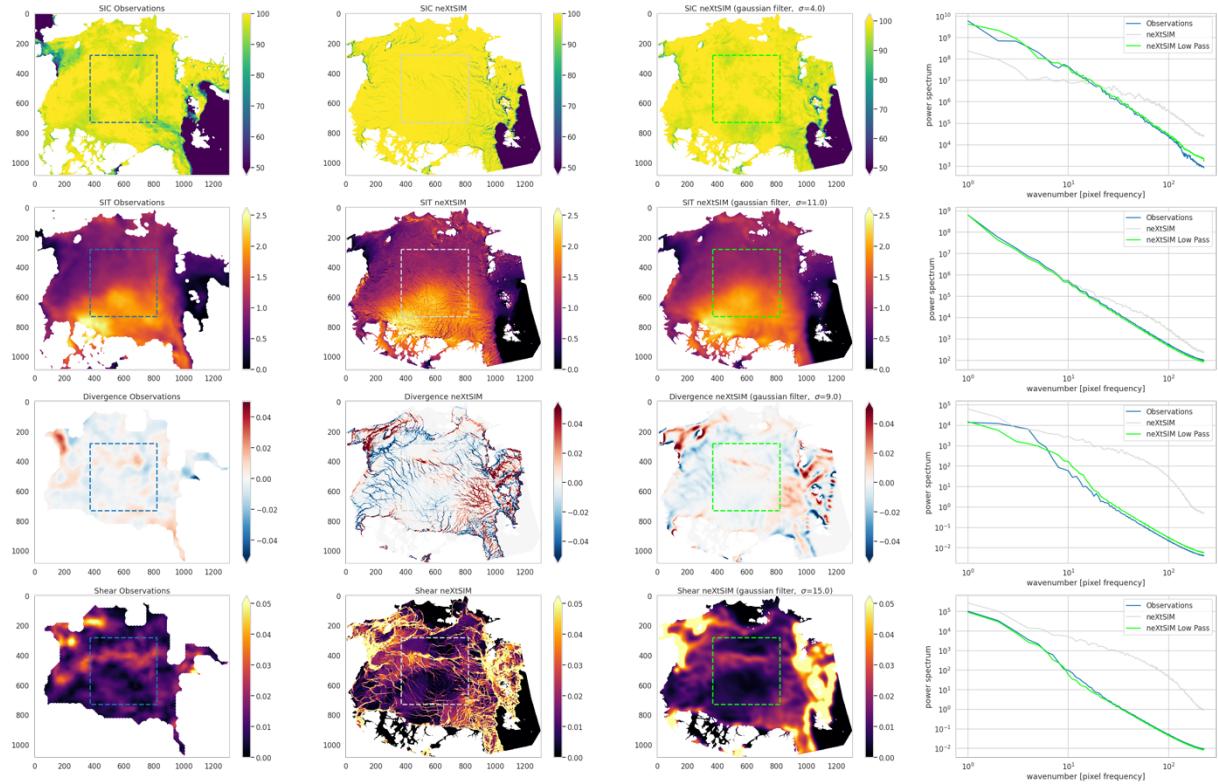
and shear that were low pass filtered with the values of sigma listed in the previous paragraph, and we multiplied their amplitudes by the corresponding square root of the ratios of variances (the dependence of the ratios on sigma is shown in the lower panel in Figure 3). The multiplying factors correspond to 0.49 for divergence and 0.59 for shear.



**Figure 3.** (Upper row) For each variable, dependence on sigma of the ratio between the trend of the winter-averaged power spectrum from observations and from neXtSIM. (Lower row) For each variable, dependence on sigma of the ratio between the winter-averaged variance of the field derived from observations and neXtSIM. Both the power spectrum and the variance for each variable are based on the field contained within the boxes shown in each subplot of Figure 2.

To improve the agreement between SIC from observations and neXtSIM, we added noise to the latter. This was created by adding a 2D random field with a predefined spectrum to the low pass filtered SIC provided by neXtSIM, where the predefined spectrum consisted in the difference between the winter-averaged spectrum from the observed and the modelled SIC. This procedure was tested for the January 01<sup>st</sup>, 2021, and was performed using the same varying values of sigma as for the linear trends. The best agreement was found for sigma equal to 4 grid points. Higher values were found to smooth the modelled SIC excessively. As a result, the added noise completely hid the signal, resulting in its loss. For this reason, the optimal value of sigma was set to 4, not 6.5, which was the value obtained using the linear-trend approach described above.

These additional reprocessing procedures, applied to modelled variables for January 01<sup>st</sup>, 2021, appears to have increased the similarity between neXtSIM and the observations (Figure 4). The better agreement results both from visual inspection and from the analysis of the power spectra. We note that the low pass filtered SIC from neXtSIM has now a power spectrum comparable to that from observations. This is a result of the added noise, which produces large-scale variations of SIC in the modelled field which partly mimic those in observations. Moreover, we note that the amplitudes of the low pass filtered divergence and shear from neXtSIM resemble those from observations. This is an interesting result because the variance adjustment has not been tailored to January 01<sup>st</sup>, 2021, but to the 2020/2021 winter season.



*Figure 4. It follows the same structure as Figure 2 but, compared to Figure 2, the third and fourth columns have been updated to show the outcomes of the additional reprocessing procedures.*

We also propose an alternative approach to evaluate the outcome of the reprocessing procedure applied to neXtSIM's SIC, SIT, divergence, and shear. For each variable, we produce a set of nine histograms, displayed as a 3x3 matrix (Figs. 5, 6, 7, 8). In each figure, the first row shows the results for the observations, the second for high-resolution neXtSIM, and the third for the synthetic dataset. The columns show the values for the first seven days of November 2020, January 2021, and March 2021, respectively. These three different periods were chosen to assess whether the similarity between the observations and the synthetic dataset varies over time.

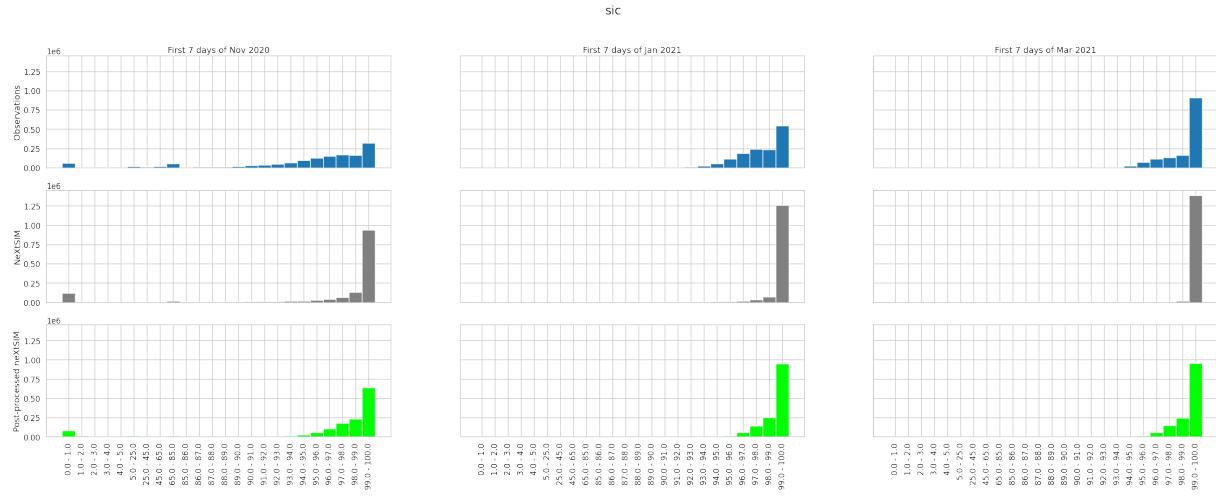
The histograms only display the values of each dataset within the region where the power spectra were computed (highlighted in Figs. 2 and 4).

Overall, the histograms indicate that the reprocessing procedure enhances the similarity between the model output and the observations. For example, the distribution of the synthetic SIC (Fig. 5) widens near 100% when compared to the SIC from neXtSIM, resulting in a better agreement with the observations.

Interestingly, for the shear (Fig. 8), the reprocessing procedure increases the similarity between the shapes of the distribution of the model output and of the observations. However, the two

distributions are characterized by different means during the first seven days of November. This discrepancy likely arises because the observations do not provide the sea-ice displacement in proximity of the sea-ice edge, thus lacking information on the shear in the region. In contrast, neXtSIM provides information on sea-ice deformation at the sea-ice edge, where values appear significantly larger than those farther away from the boundaries. This discrepancy is visible only during the first seven days of November because the sea-ice edge crosses the region over which the histograms are computed.

For the divergence (Fig. 7), the low-pass filtering applied to the neXtSIM dataset leads to a widening of the corresponding distribution, therefore increasing the similarity with the observations. This is particularly evident during the first seven days of November and January, and to a lesser extent in March.



*Figure 5: Histograms showing the distribution of SIC as provided by the observations, neXtSIM, and the synthetic dataset over three different periods. The first row shows the results for the observations, the second for neXtSIM, and the third for synthetic dataset. The first column shows the values for the first seven days of November 2020, the second for the first seven days of January 2021, and the third for the first seven days of March 2021. The histograms show the values within the region where the power spectra were computed.*

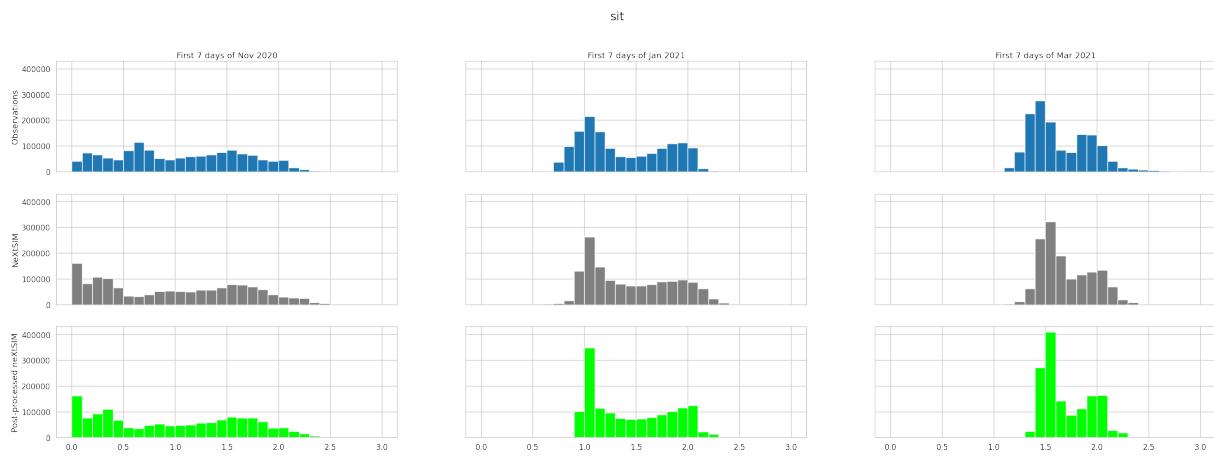


Figure 6: It follows the same structure as Figure 5, but it shows the results for SIT.

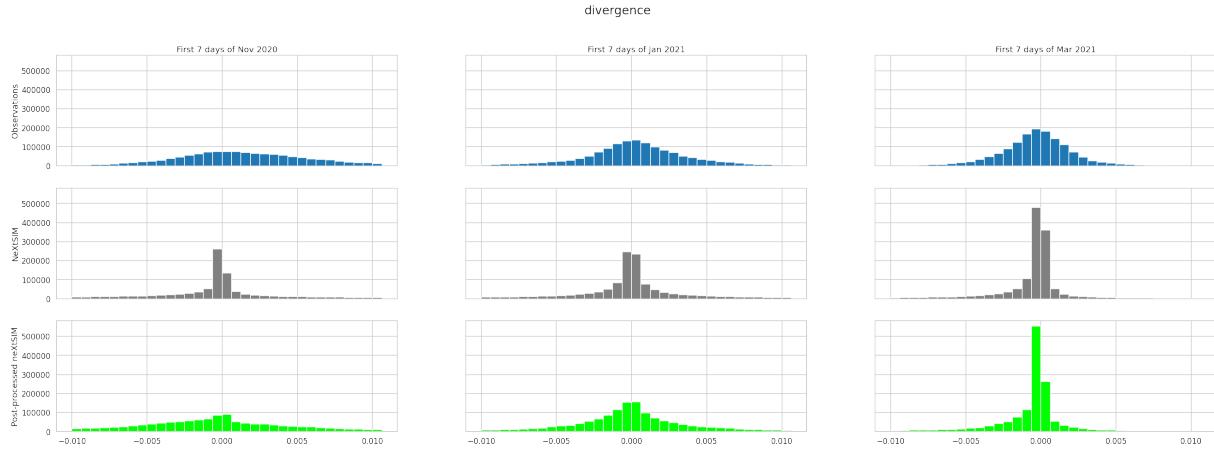


Figure 7: It follows the same structure as Figure 5, but it shows the results for the divergence.

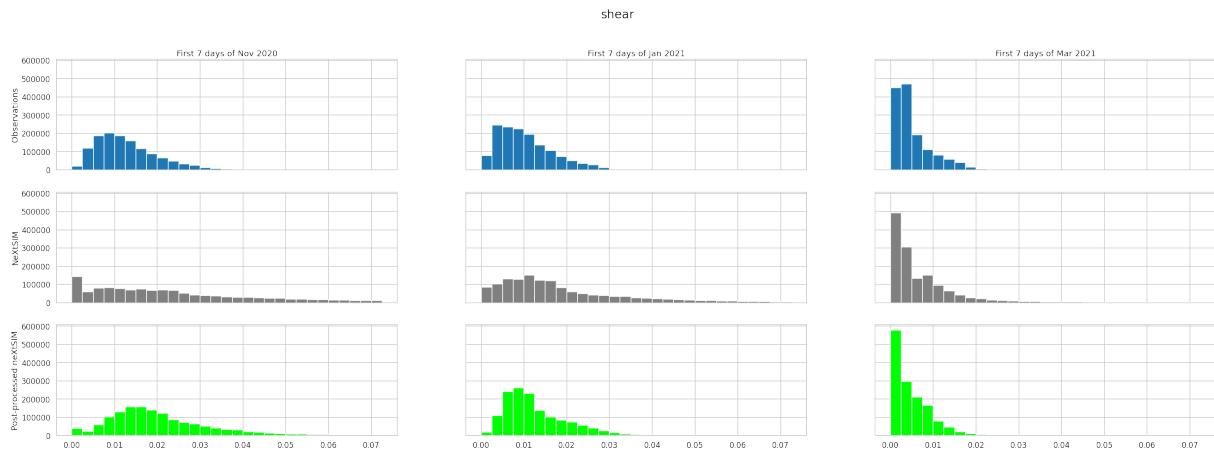


Figure 8: It follows the same structure as Figure 5, but it shows the results for the shear.