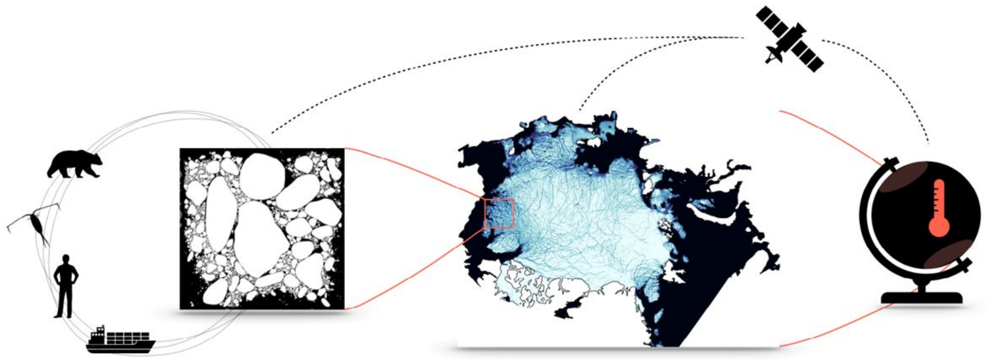
11th IICWG Data Assimilation Workshop
Tuesday March 21st 2023



Deep learning for surrogate modeling of neXtSIM

Charlotte Durand, Tobias Finn, Alban Farchi,
Marc Bocquet, Einar Ólason

Agenda

I- Introduction

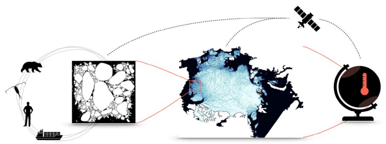
II- Workflow setup

III- Neural networks and training

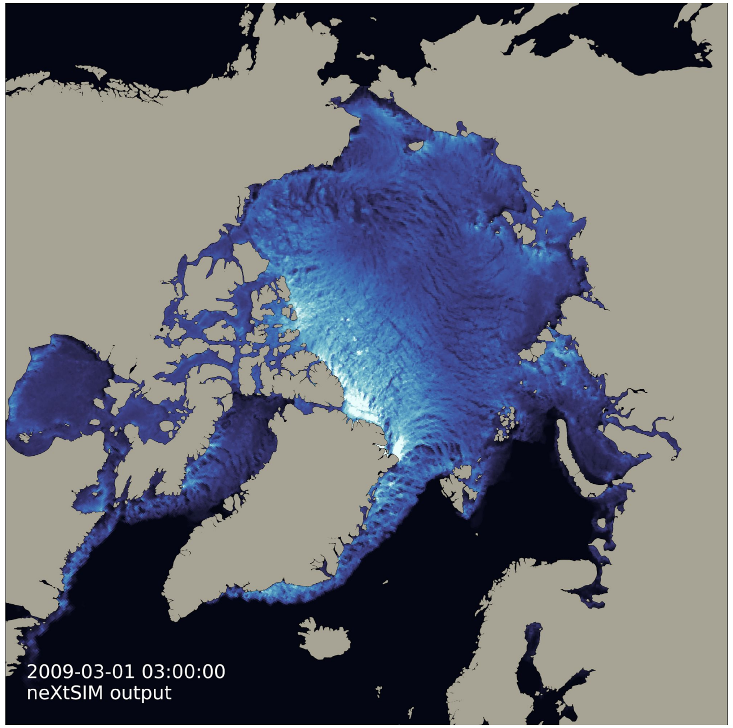
IV- Surrogate modeling, forecast skill

V- Diffusion analysis

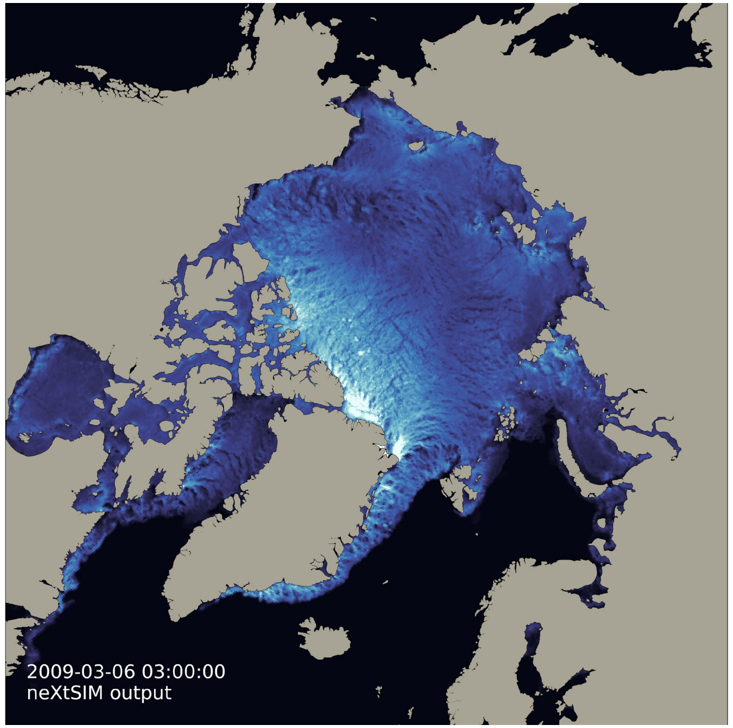
VI- Seasonal Forecast



Model Prediction using Deep Learning



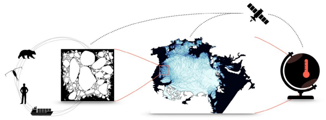
Neural Network



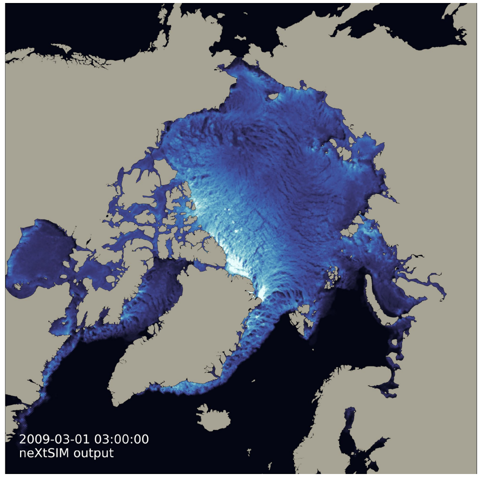
Problems :

- Multifractality
- Stochasticity
- High Resolution model
- Anisotropy

Can we find a fully neural network surrogate model that can predict the dynamics of sea-ice, with a cheaper computation cost that a physical model simulation ?

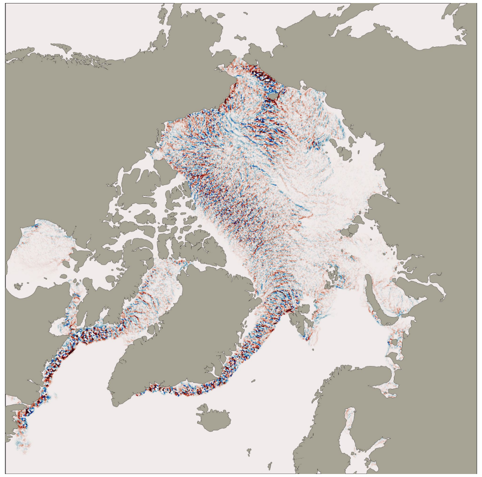


Model Prediction using Deep Learning



Neural network that can predict the dynamics of sea-ice thickness

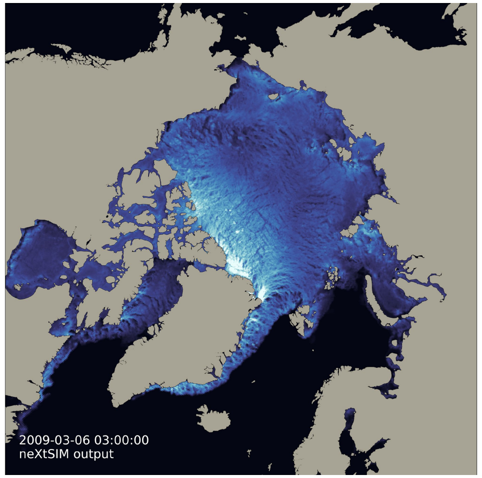
Trained to predict +12 hours dynamics

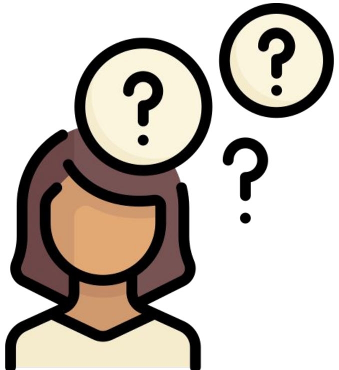


NN

NN

NN



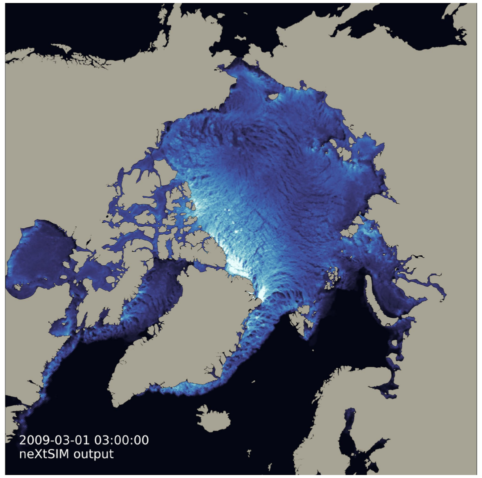


Is the neural network able to predict sea-ice dynamics for more than the 12 hours dynamic it has learned ?

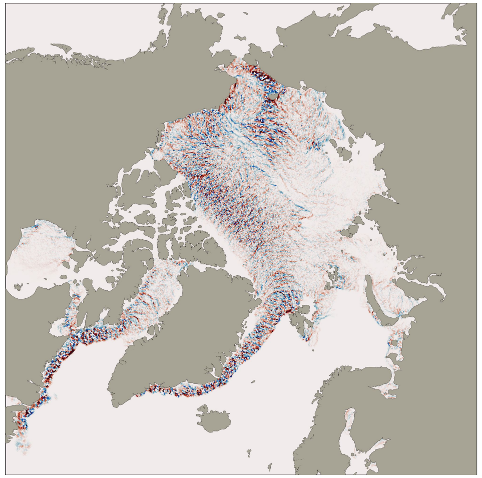
Dataset : Lagrangian neXtSIM

2009	Train Data
2010	Train Data
2011	Train Data
2012	Train Data
2013	Train Data
2014	Train Data
2015	Train Data
2016	Train Data
2017	Validation Data
2018	Test Data

Input : $x(t)$ = SIT



Output : $y = x(t+12 \text{ hours}) - x(t)$



Learning to predict the dynamic at t+12 hours

[1] G. Boutin *et al.*, « Arctic sea ice mass balance in a new coupled ice-ocean model using a brittle rheology framework », *The Cryosphere Discussions*, vol. 2022, p. 1-31, 2022, doi: 10.5194/tc-2022-142.

Forcings

Forcings from ERA5

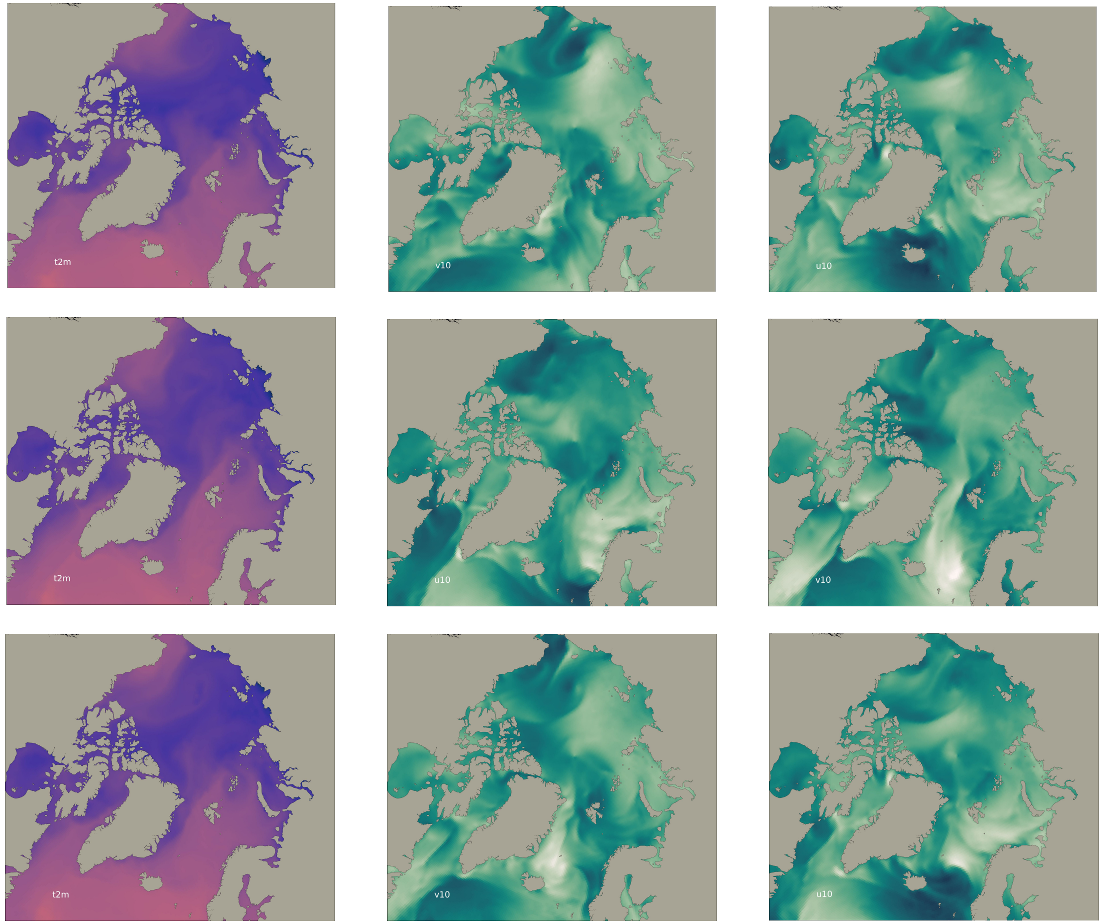
Interpolated from ERA5 grid to NeXtSim grid with nearest neighbour

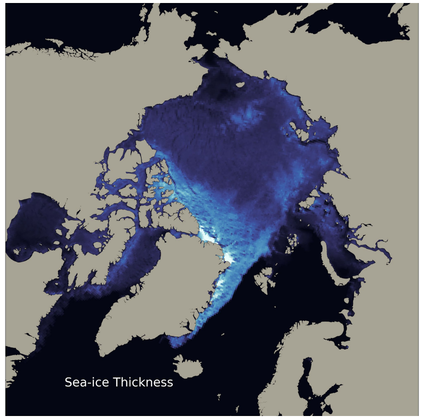
- **2m air temperature**
- **10m X velocity**
- **10m Y velocity**

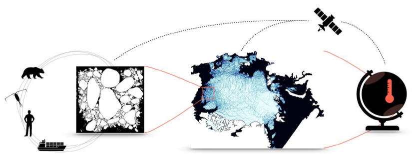
$t = t_0$

$t = t_0 + 6h$

$t = t_0 + 12h$

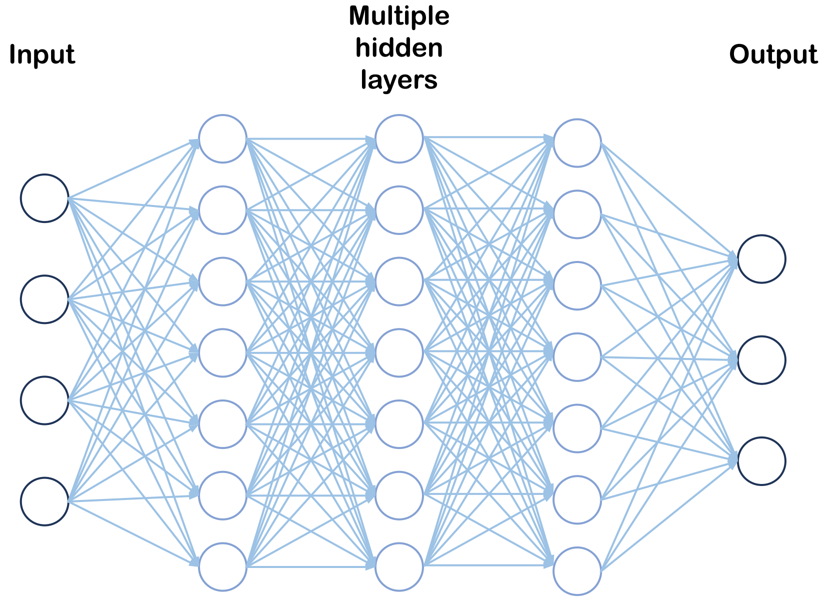




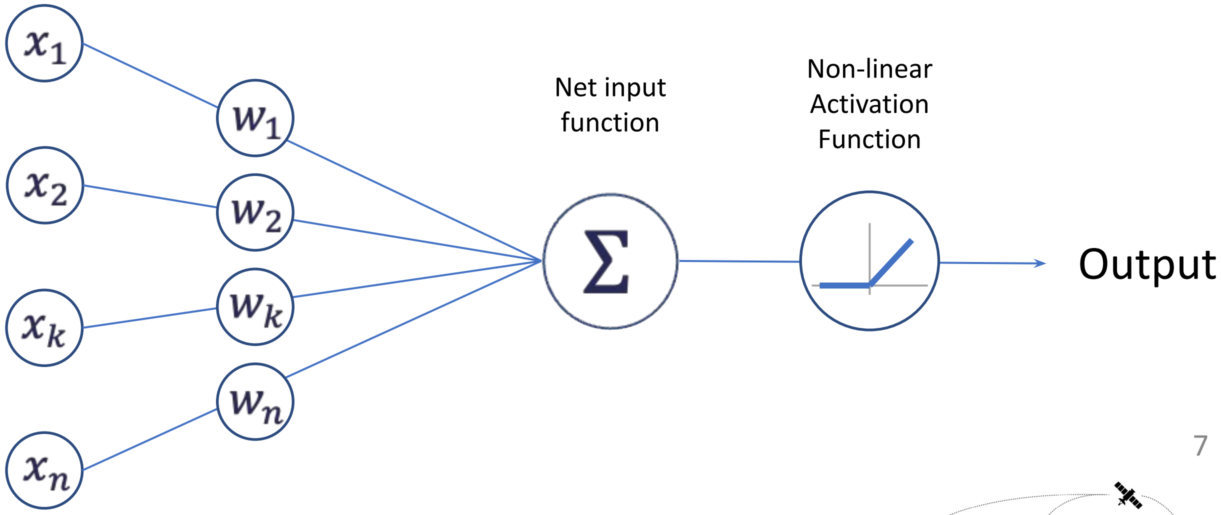


Deep Learning 101

Input X → Neural Network → Output Y

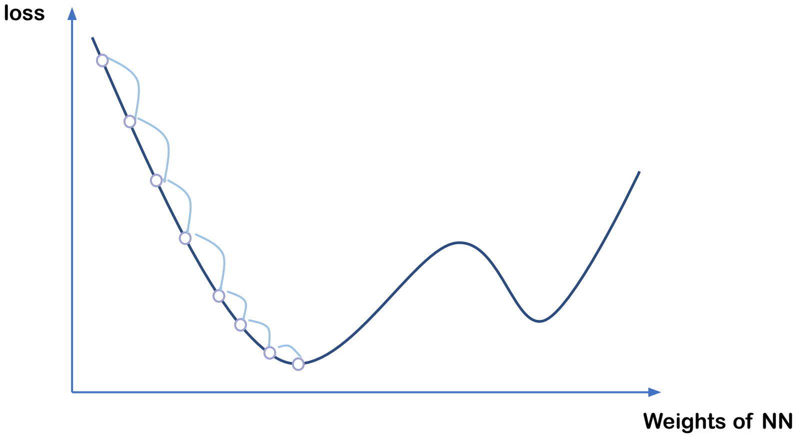


Each node



Deep Learning – Supervised Setting

Goal : Minimize a loss function
Let's consider m samples in our dataset



$$\mathcal{L}(\theta) = MSE(\theta, y, \hat{y}) = \frac{1}{m}\sum_{i=1}^{m}(\hat{y}(\theta) - y)^2$$

x : input of the dataset
y : « truth »
$\hat{y}$: prediction of the Neural Network

Optimizer : Determine how the network will be updated based on the loss function (Complex version of a stochastic Gradient Descent)

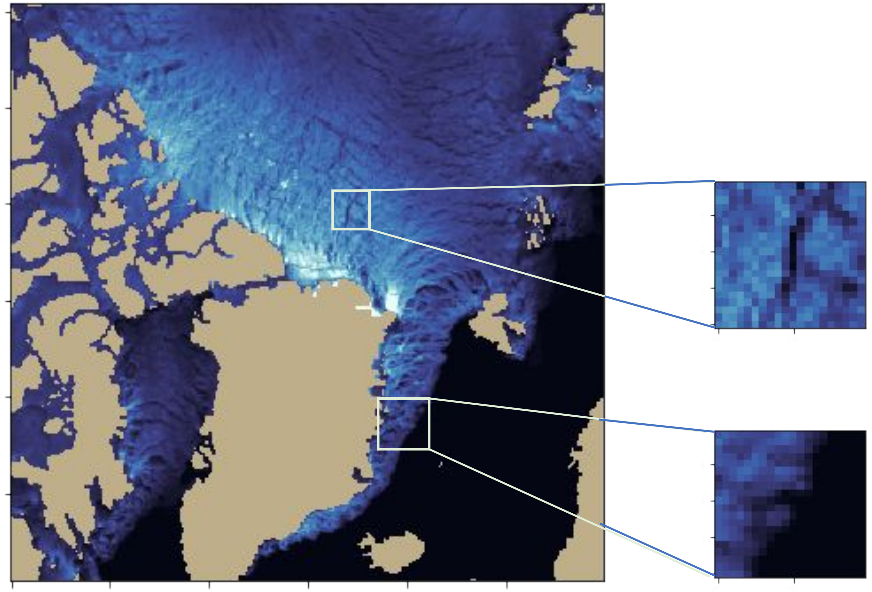
SGD update : $\theta_{t+1} = \theta_t - \eta \nabla \mathcal{L}(\theta_t)$

η : **learning rate**

Convolutional Neural Network (CNN)

Convolution layers learn patterns

- Learns cracks, ice accumulation…
- Learns apparition/disappearance of sea-ice (Marginal Ice Zones – MIZ)



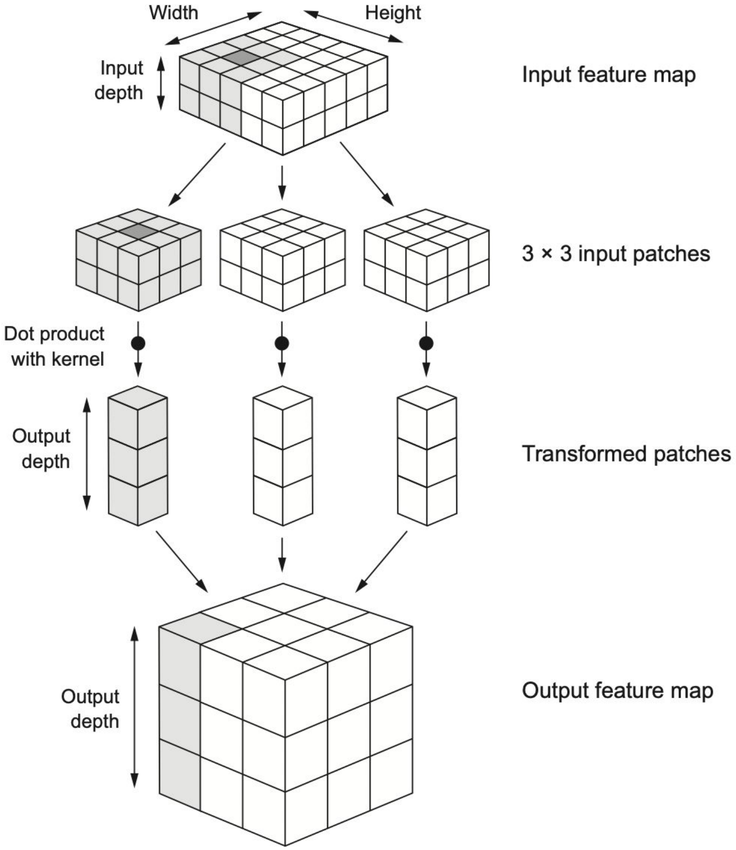
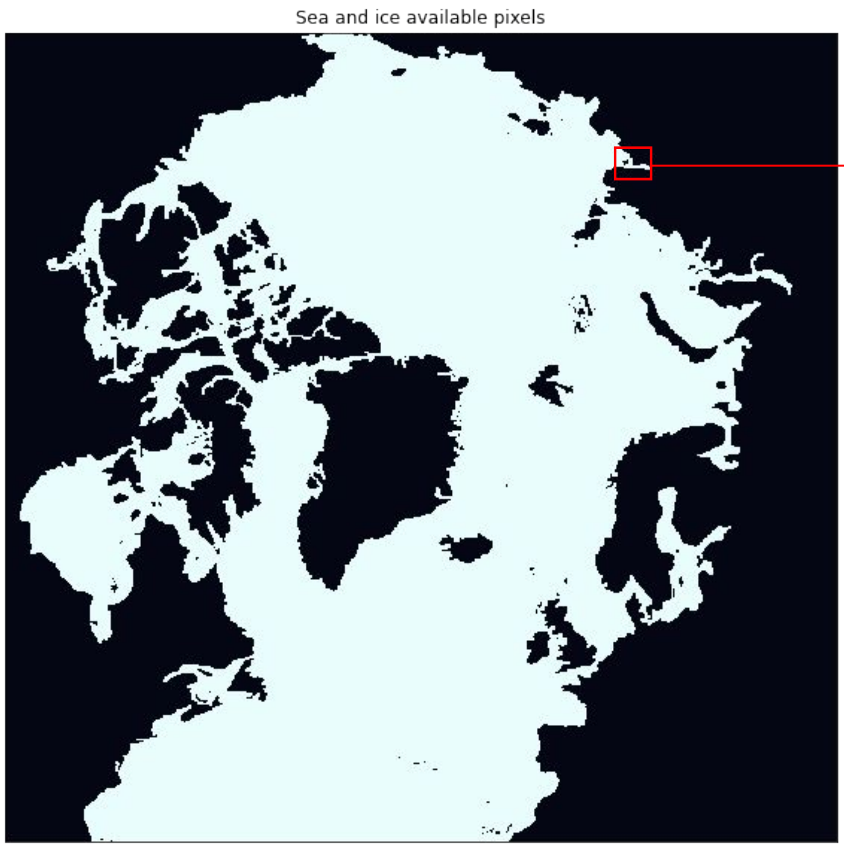
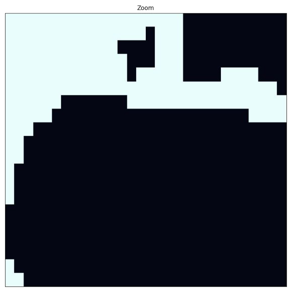


Image taken from Deep Learning with Python, François Chollet, ISBN 9781617294433

Partial Convolution





(light box)	Valid pixel : sea, ice
(black box)	Invalid pixel : earth

$$x' = W^T (X \odot M) \frac{sum(1)}{sum(M)} + b$$

$$\text{if } sum(M) > 0$$

Mask M

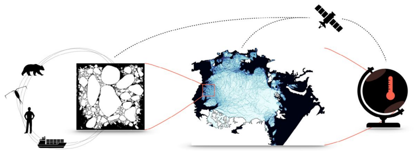
0	0	1	0
0	0	0	1
1	0	1	1
1	1	1	1

Input X

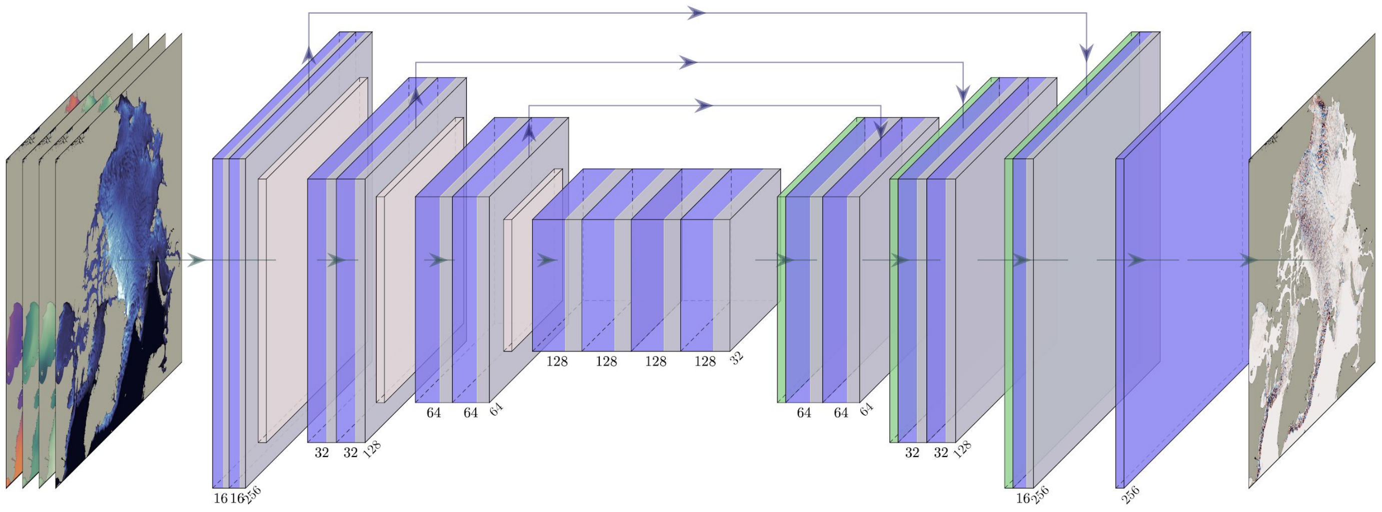
-2	-3	1	-1
-1	1	1	4
1	2	1	3
-1	4	2	2

$$W^T \times \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 4 \\ 0 & 1 & 3 \end{bmatrix} \times \frac{9}{4} + b$$

Liu, Guilin, Fitsum A. Reda, Kevin J. Shih, Ting-Chun Wang, Andrew Tao, et Bryan Catanzaro. « Image Inpainting for Irregular Holes Using Partial Convolutions ». *arXiv:1804.07723 [cs]*, 15 décembre 2018. http://arxiv.org/abs/1804.07723.



Network structures - UNet



Classical structure

- Multiscality
- Skip connection allows a quick training
- Added with Mask partial convolution

- Stacking several timesteps as input features

Convolution layer

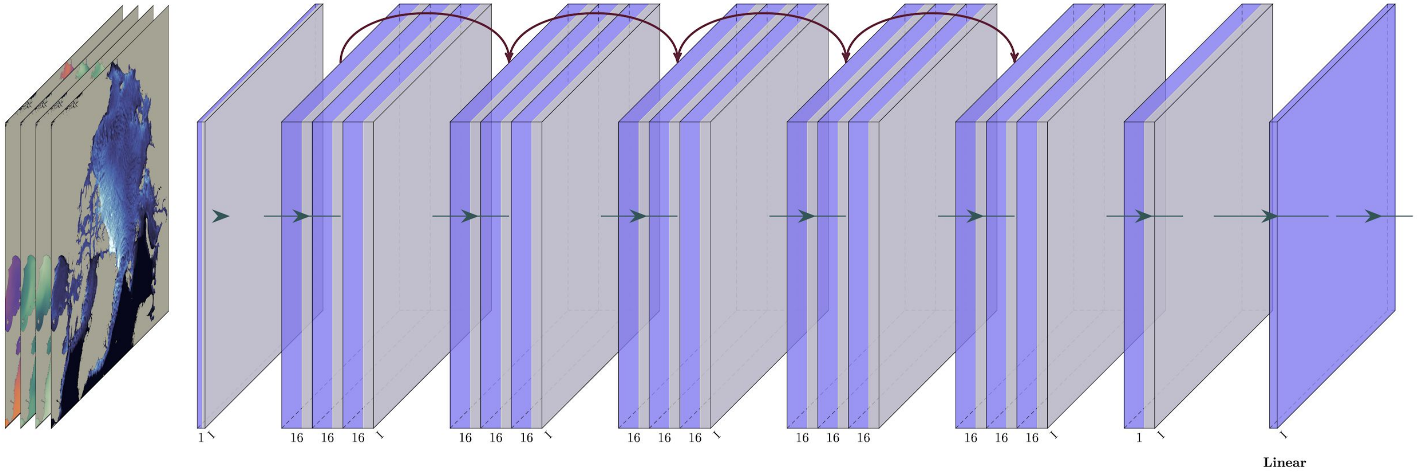
Max Pooling

mish activation

Up-sampling

Ronneberger, Olaf, Philipp Fischer, et Thomas Brox. « U-Net: Convolutional Networks for Biomedical Image Segmentation », 2015. https://doi.org/10.48550/ARXIV.1505.04597.

Network structures - ResNet



FixUp Initialization [2] method :
For each block : - last layer initialized at 0
- all previous layers initialized with He Normal [3] distribution

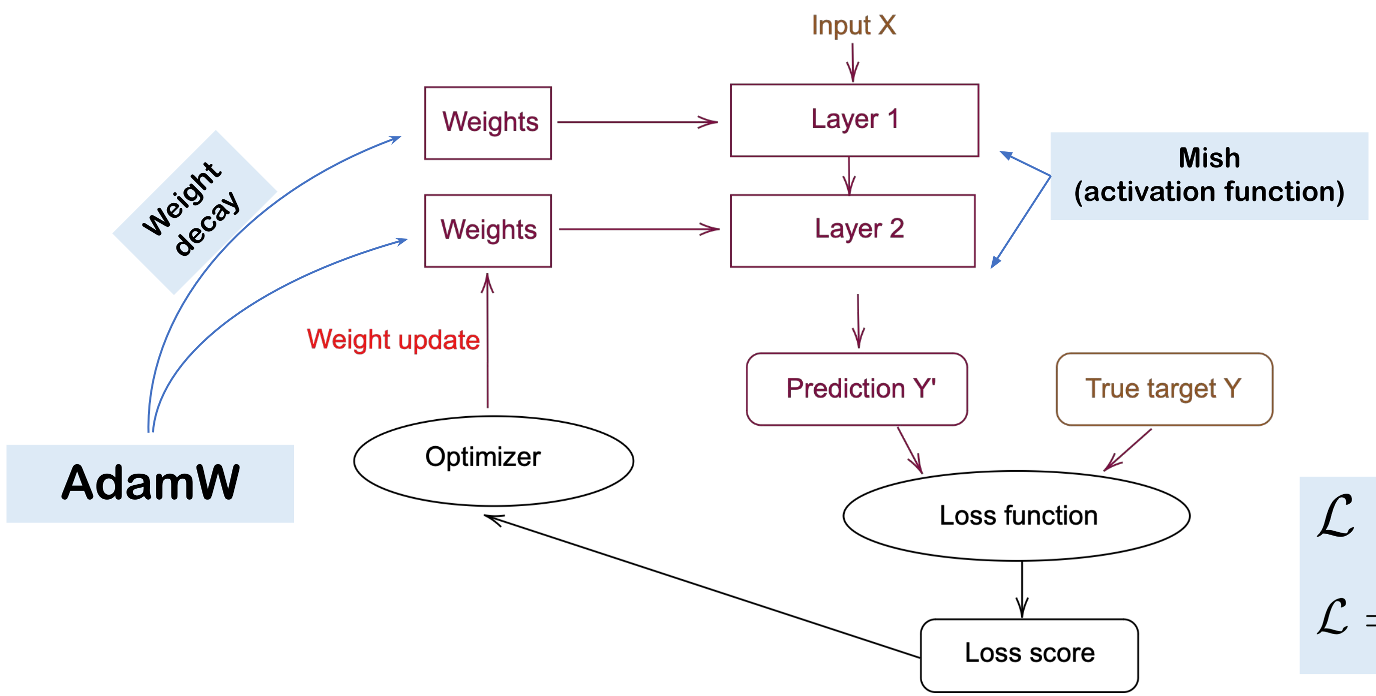
$$w_{init} \sim \mathcal{N}(0, \sqrt{2/n_l})$$

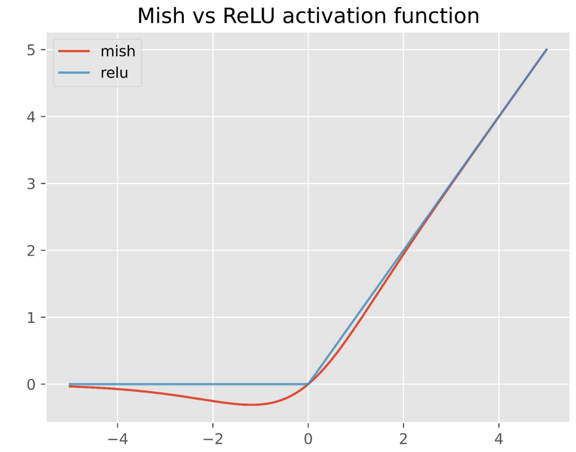
With n_l the number of layers

[2] Zhang, Hongyi, Yann N. Dauphin, et Tengyu Ma. « Fixup Initialization: Residual Learning Without Normalization », 2019. https://doi.org/10.48550/ARXIV.1901.09321.

[3] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, et Jian Sun. « Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification », 2015. https://doi.org/10.48550/ARXIV.1502.01852.

Neural Network training



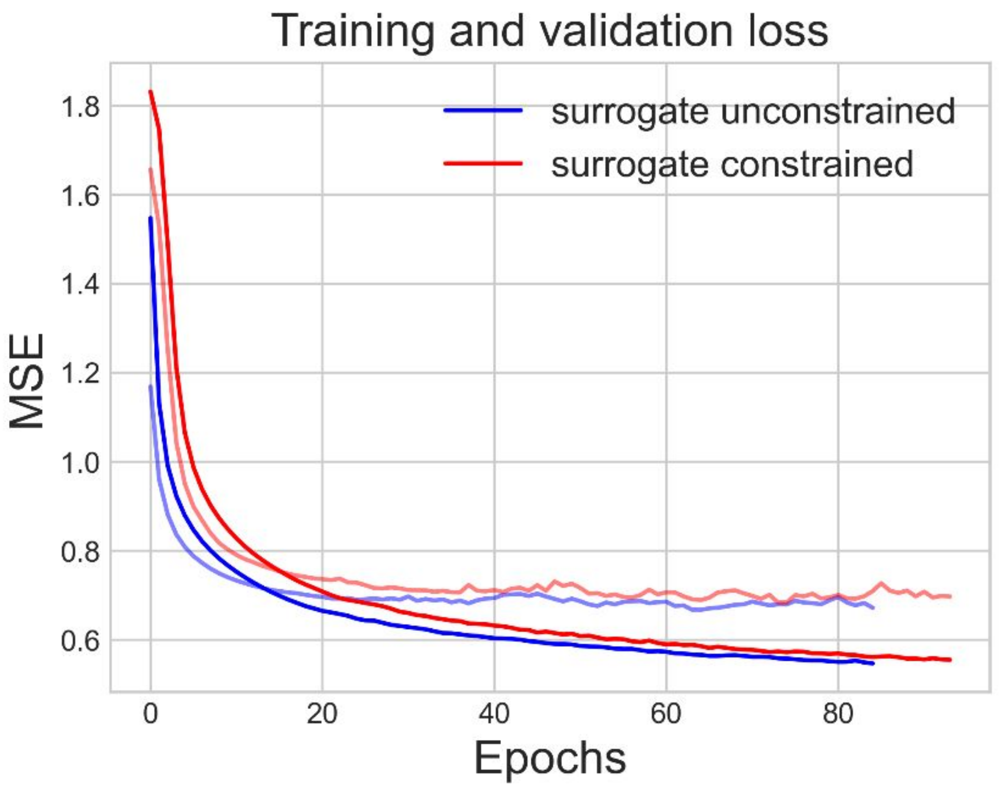


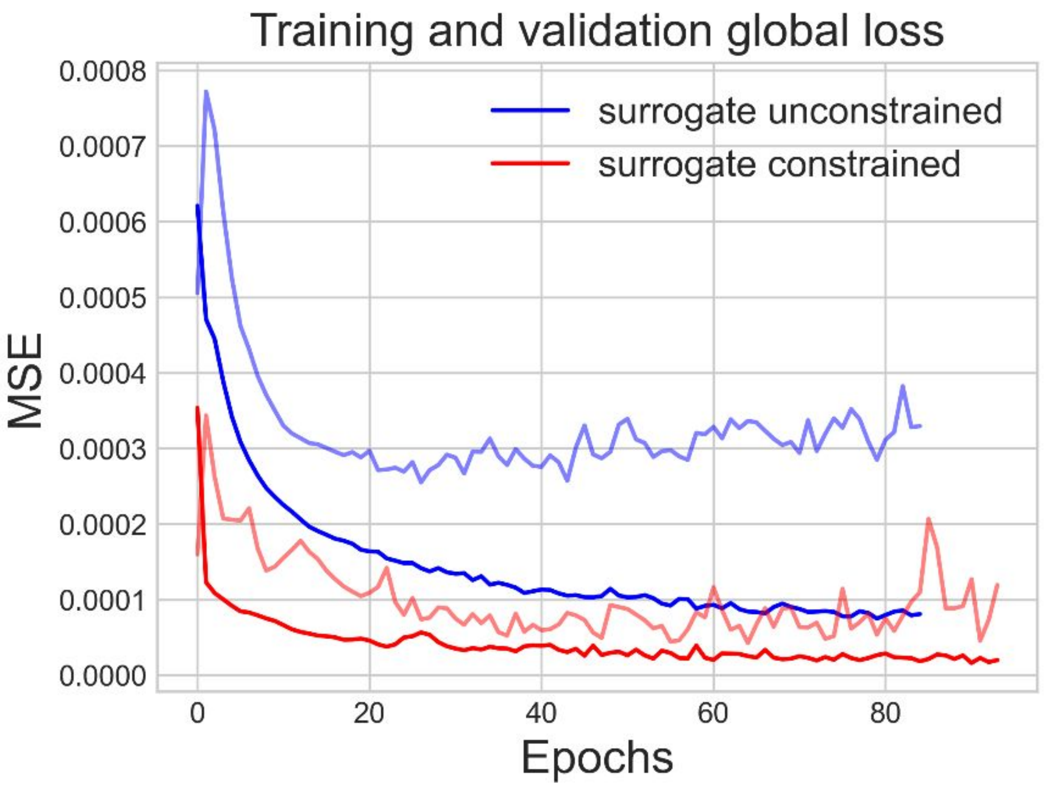
$$\mathcal{L} = \mathcal{L}_{\text{local}} + \mathcal{L}_{\text{global}}$$

$$\mathcal{L} = \mathbf{MSE}(\hat{\tilde{y}} \times M, \tilde{y} \times M) + \lambda\mathbf{MSE}(\overline{\hat{y} \times M}, \overline{y \times M})$$

[4] D. Misra, « Mish: A Self Regularized Non-Monotonic Activation Function », 2019, doi: 10.48550/ARXIV.1908.08681.
[5] I. Loshchilov et F. Hutter, « Decoupled Weight Decay Regularization ». arXiv, 4 janvier 2019. http://arxiv.org/abs/1711.05101
Deep Learning with Python, François Chollet

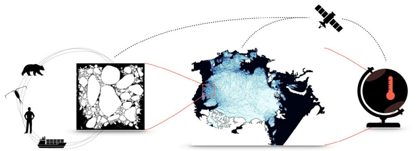
Results – UNet - High-resolution data





Constraining the loss allows a better evaluation of the global amount of sea-ice

Computation on IDRIS supercomputer Jean-Zay on A100 80GoRAM GPUs
Grant access to the HPC resources of IDRIS under the allocation 2021-AD011013069 made by GENCI



Surrogate model Prediction

How do we evaluate the ability of NN to predict sea-ice dynamic ?

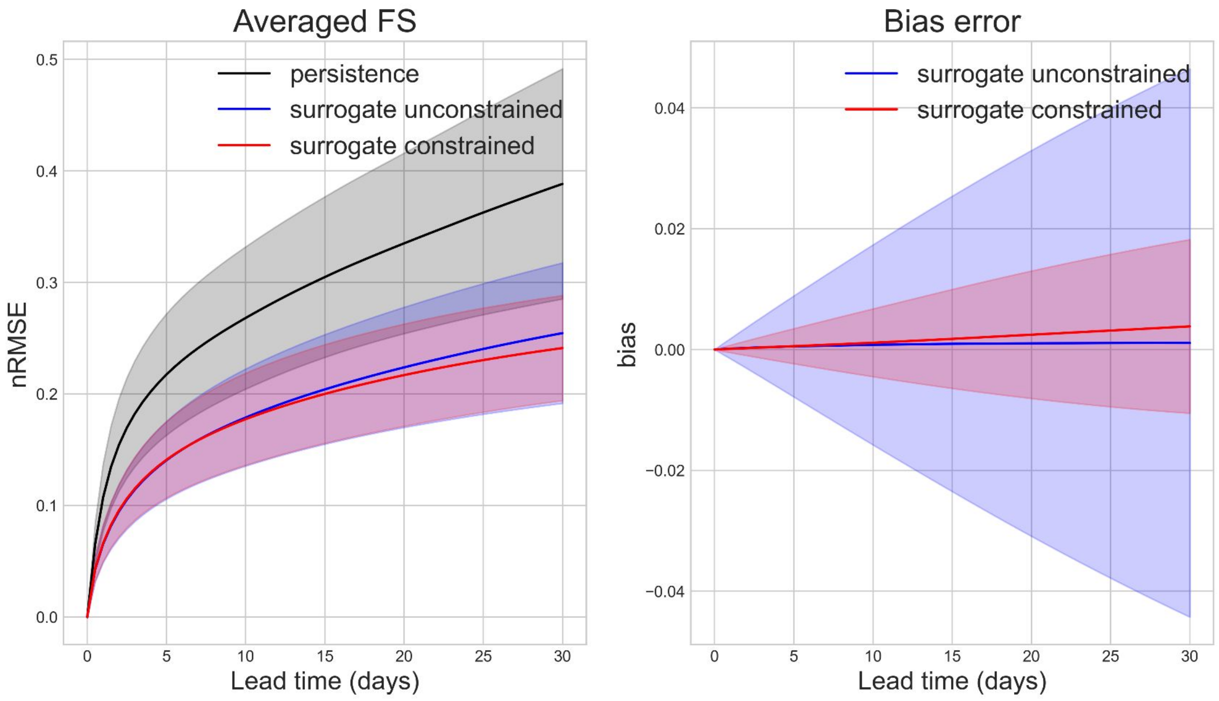
Let the model predict the dynamic over several timesteps

$x_{t \in \{0, T\}} \longrightarrow \tilde{x}_{t \in \{0, T\}}$ $y_{t \in \{1, T\}} = x_t - x_{t-1} \longrightarrow \tilde{y}_{t \in \{1, T\}}$

NN training | Surrogate modelling

$$\tilde{x}_t \longrightarrow \boxed{\text{Neural Network}} \longrightarrow \tilde{y}_{t+1} \text{ — } \boxed{\text{Renormalization: } \tilde{x}_{t+1} = \tilde{x}_t + \frac{\bar{y} + \sigma_y \hat{\tilde{y}}_{t+1}}{\sigma_x}} \text{ — } \tilde{x}_{t+1} \longrightarrow \boxed{\text{Neural Network}} \longrightarrow \tilde{y}_{t+2} \cdots$$

Results – Surrogate modelling



Metrics :

Averaged forecast skill :

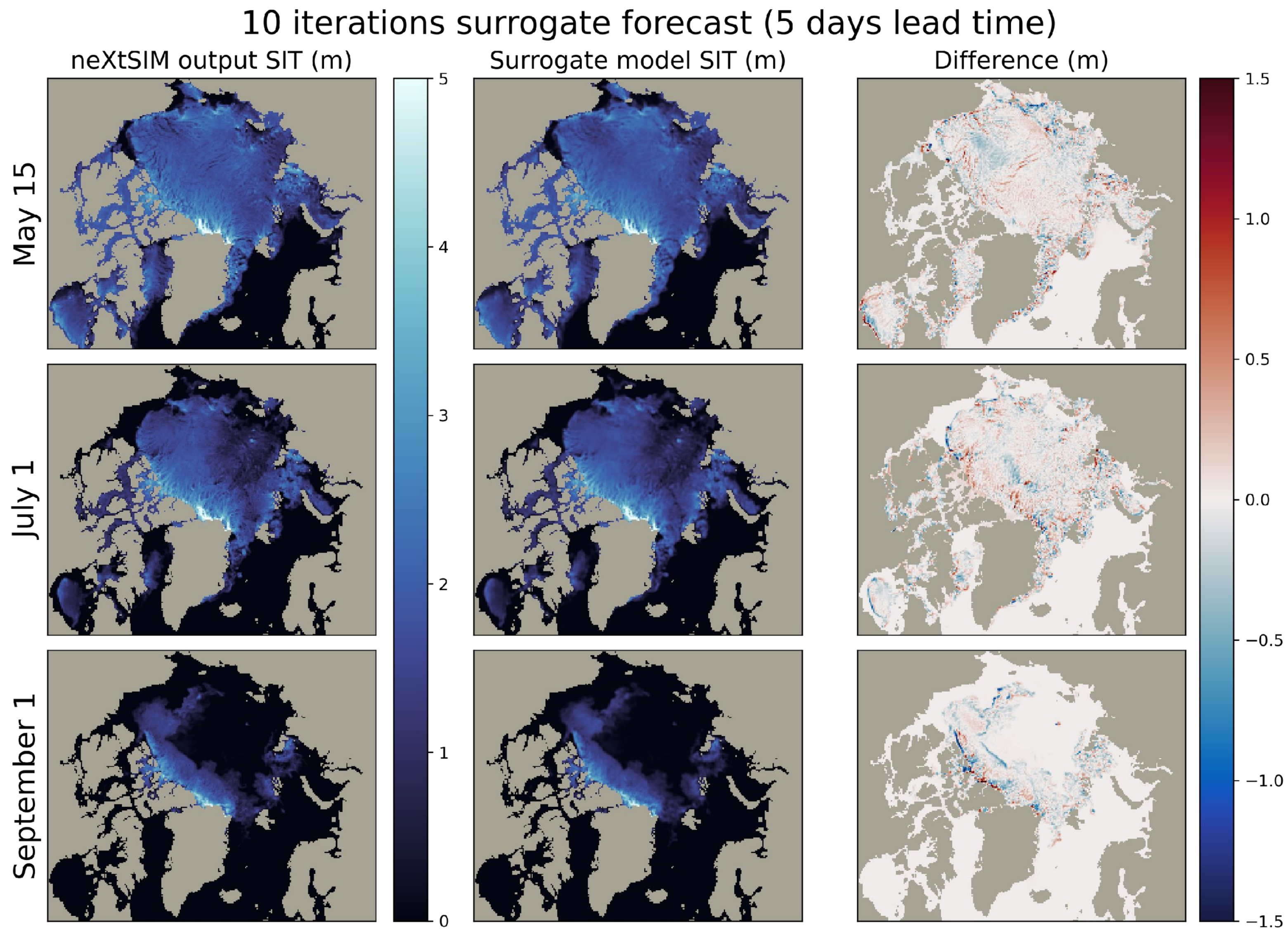
$$FS(k) = \frac{1}{N_i} \sum_{n=0}^{N-1} \sqrt{\frac{1}{N_x \times N_y} \sum_{i,j}^{N_x,N_y} (\mathcal{M}^{\mathrm{p}}_{i,j}(t_{n+k}) - \mathcal{M}^{\mathrm{s}}_{i,j}(t_{n+k}))^2}$$

With $N_i = 1370$ **and** $k = 6$**(30 lead days)**

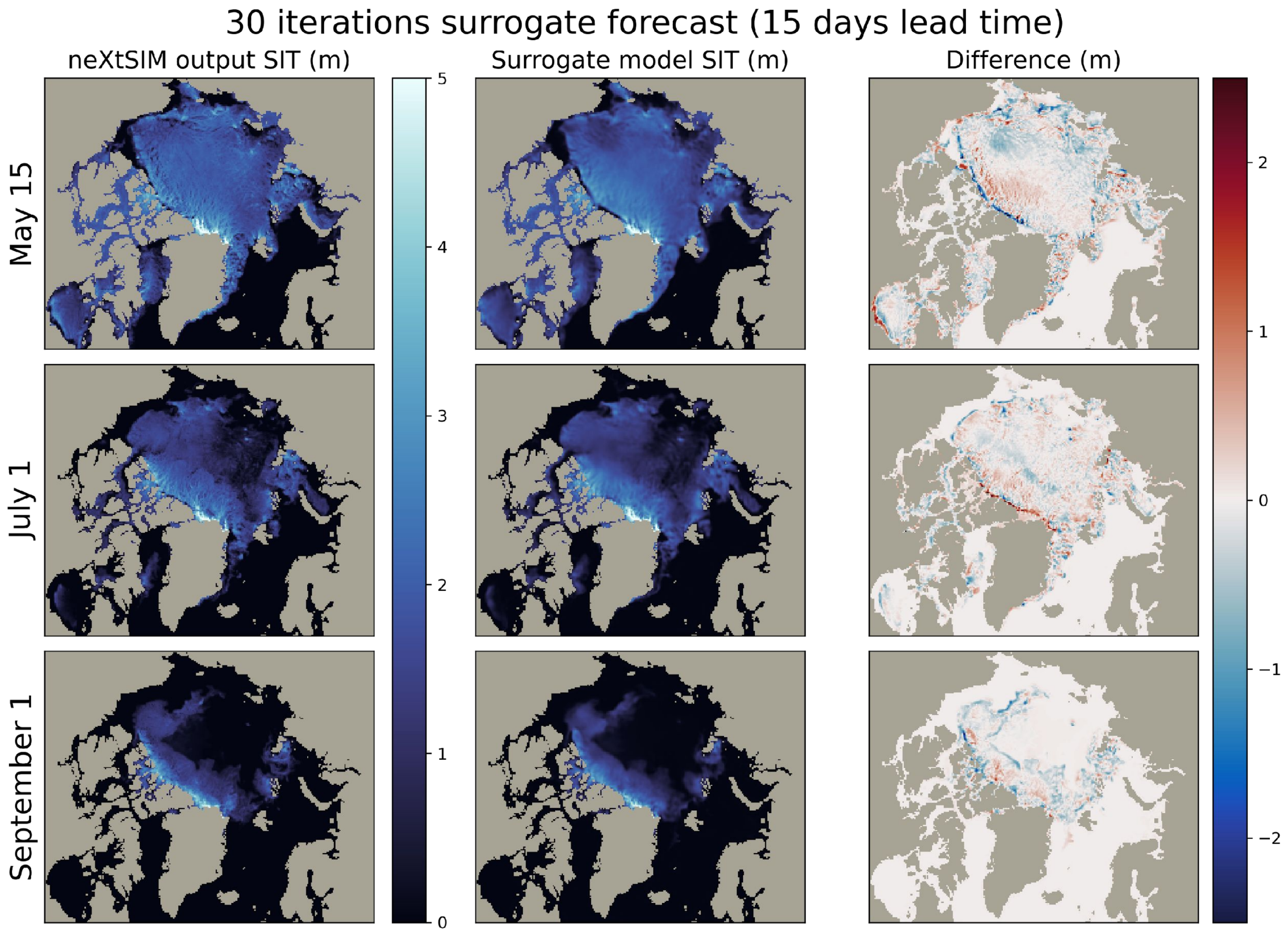
Averaged bias error :

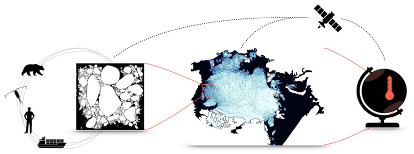
$$\mathrm{bias}(k) = \frac{1}{N_i} \sum_{n=0}^{N-1} \frac{1}{N_x \times N_y} \sum_{i,j}^{N_x,N_y} (\mathcal{M}^{\mathrm{p}}_{i,j}(t_{n+k}) - \mathcal{M}^{\mathrm{s}}_{i,j}(t_{n+k}))$$

Results – Surrogate modeling

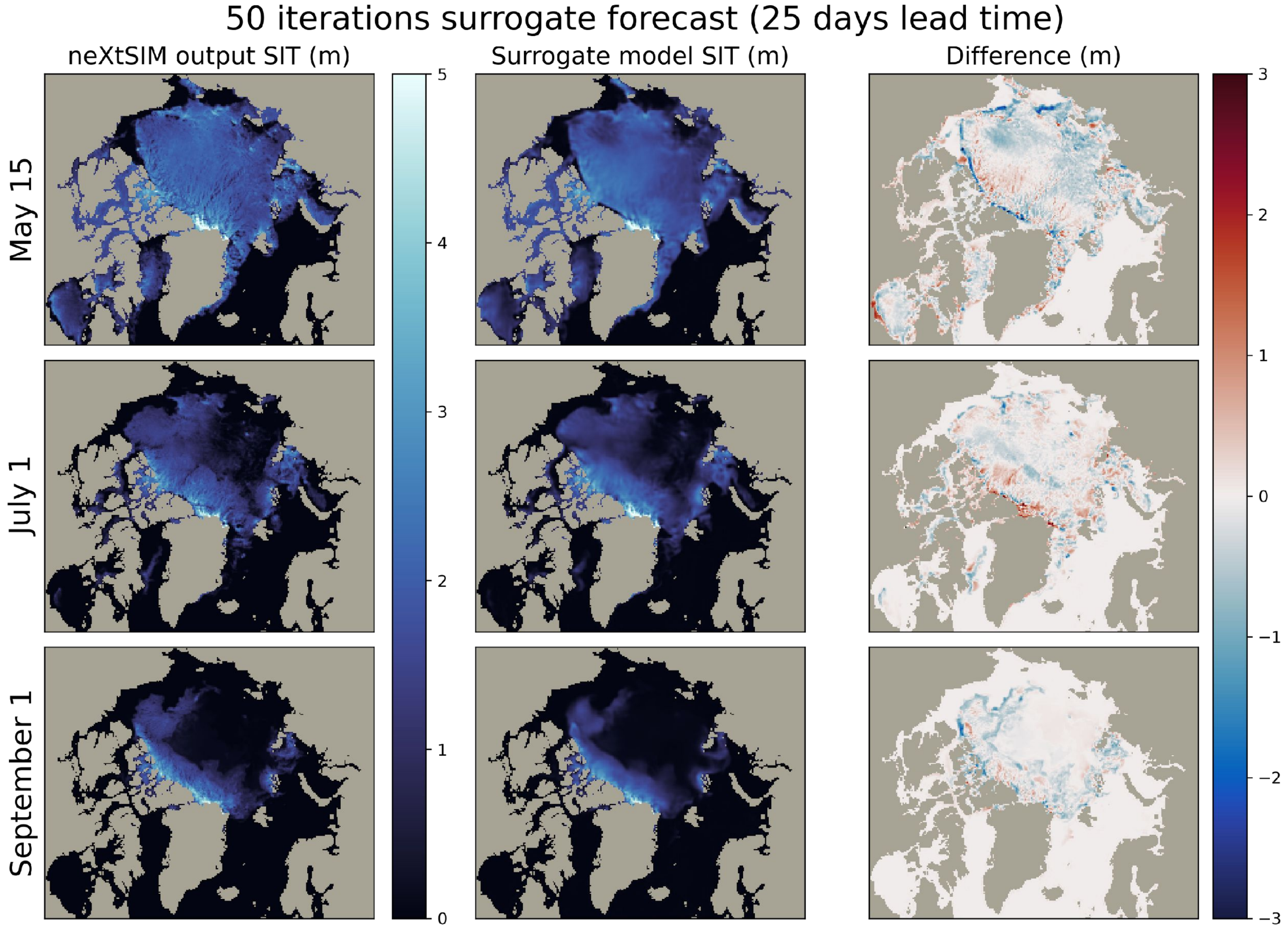


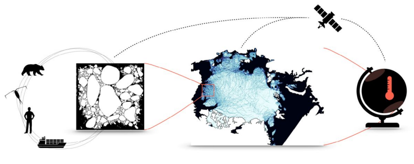
Results – Surrogate modeling





Results – Surrogate modeling





Results – Diffusion Analysis

Power spectral Density (PSD)

$$P(k_x, k_y) = |fft(X)|^2$$

In case of scaling properties, we observe a power-law, with the spectral exponent: β

$$P(k) \sim k^{-\beta}$$

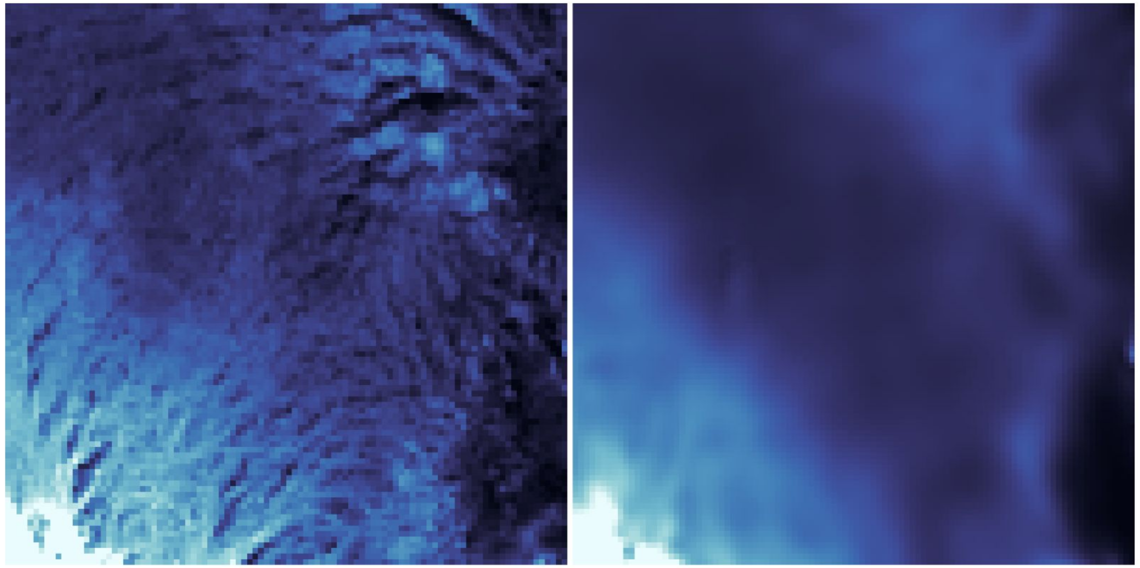
The spectral exponent is computed with the least square regression in log-log scale :

$$\log(P) \propto \log(k)$$

Can we quantify the smoothness of the surrogate modeling ?

neXtSIM

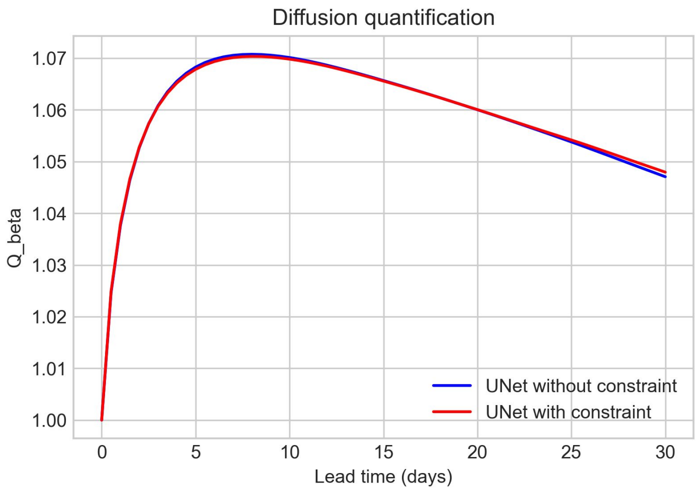
surrogate model after 60 iterations

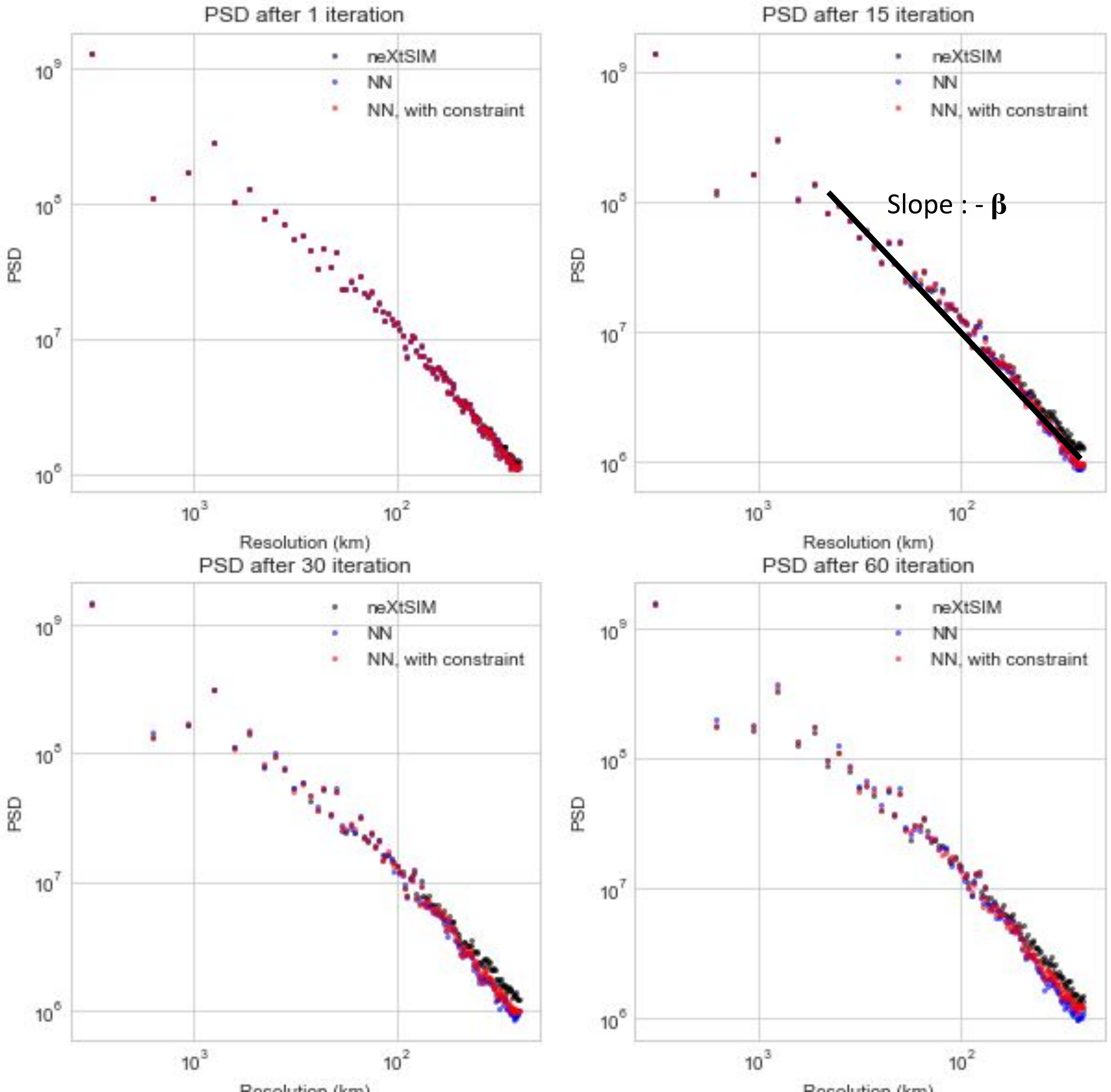


Results – Diffusion Analysis

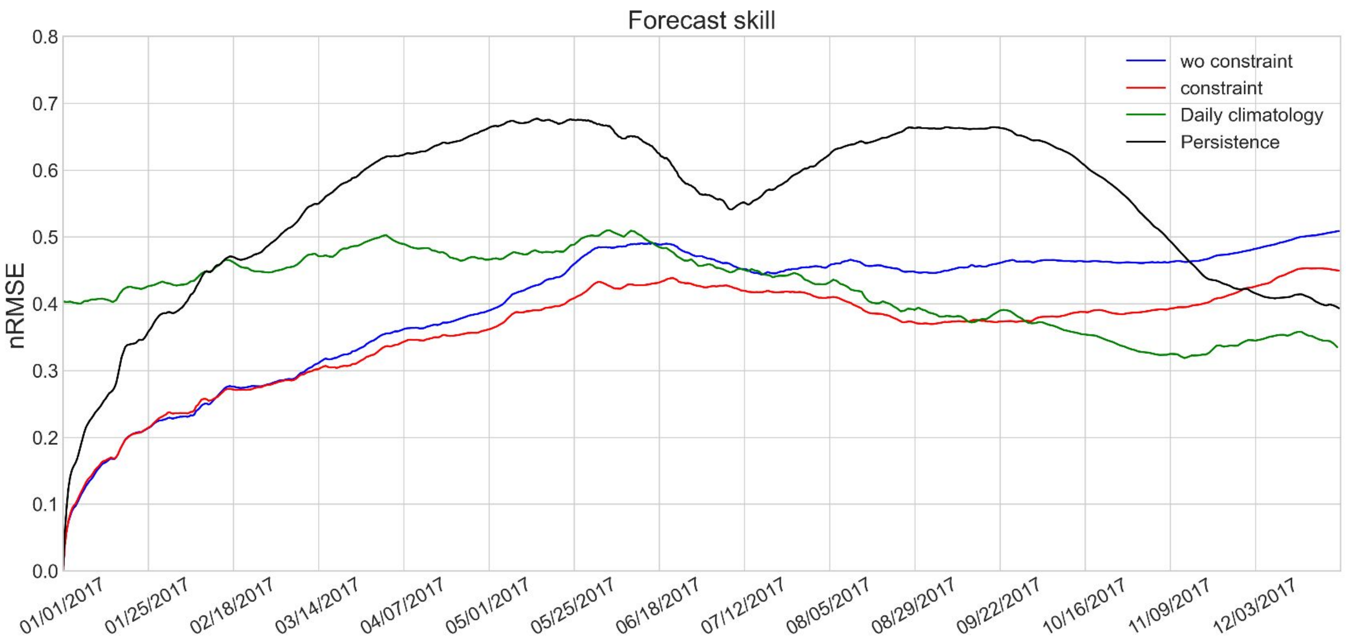
Analysis
Diffusion in the image = increase of the spectral exponent, until a plateau

$$Q_\beta(t) = \frac{1}{N_s}\sum_{n=0}^{N_s}\frac{\beta^n_{surr}(t)}{\beta^n_{neXtSIM}(t)}$$

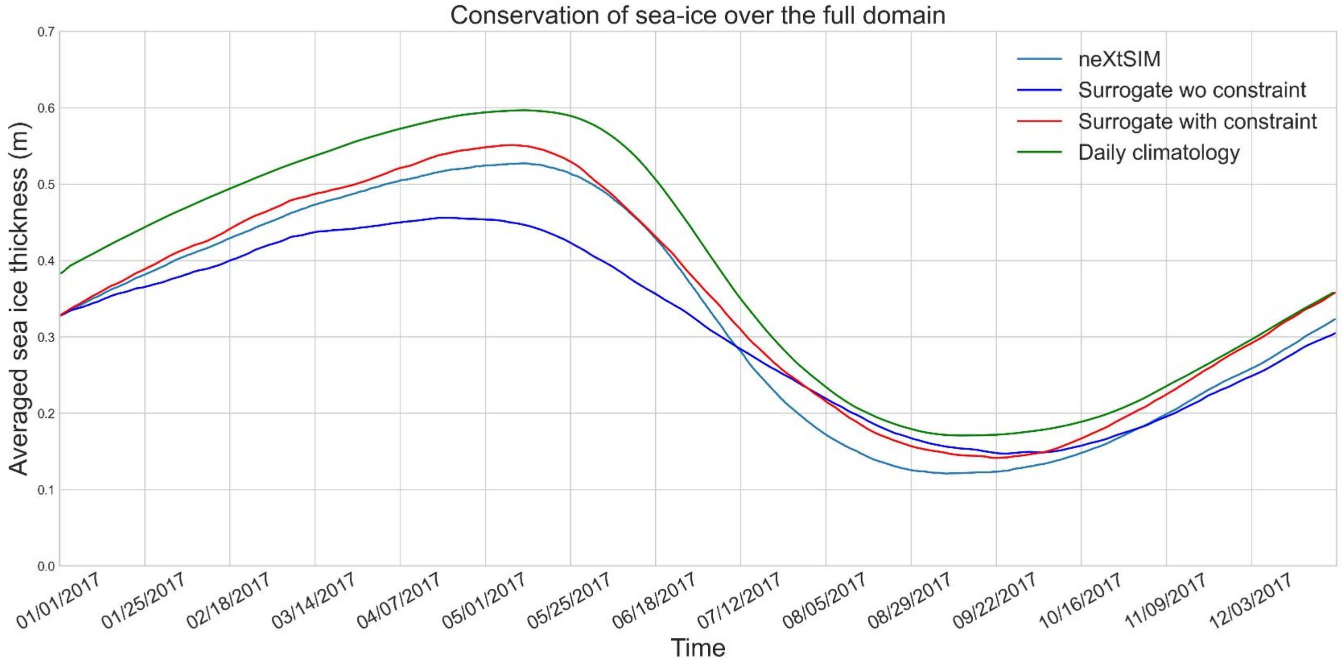


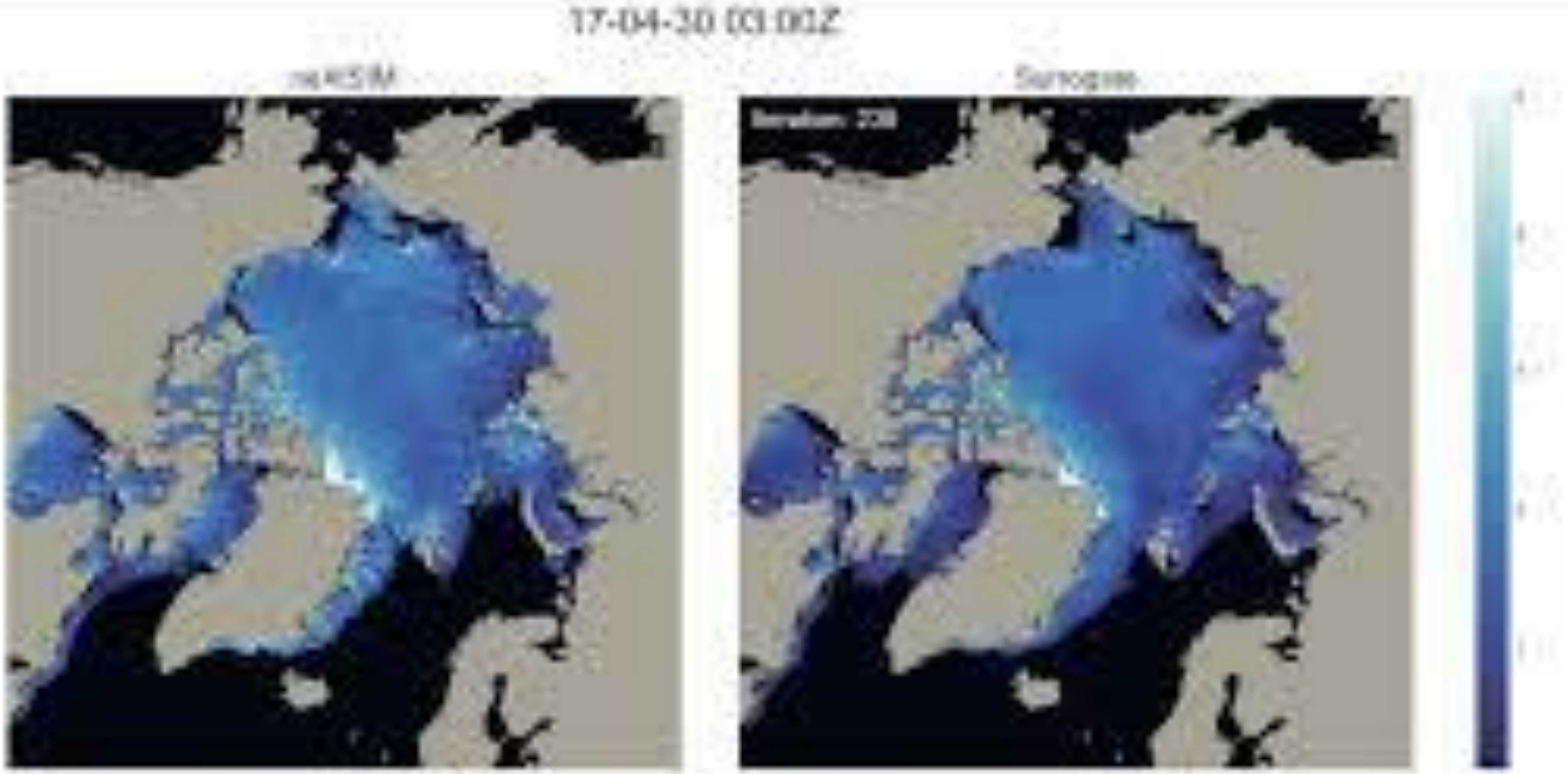


Seasonal forecast



Run starting on 1st of January, and we run the surrogate for 720 iterations, as described before





17-04-20 00:00Z
neXtSIM
Surrogate

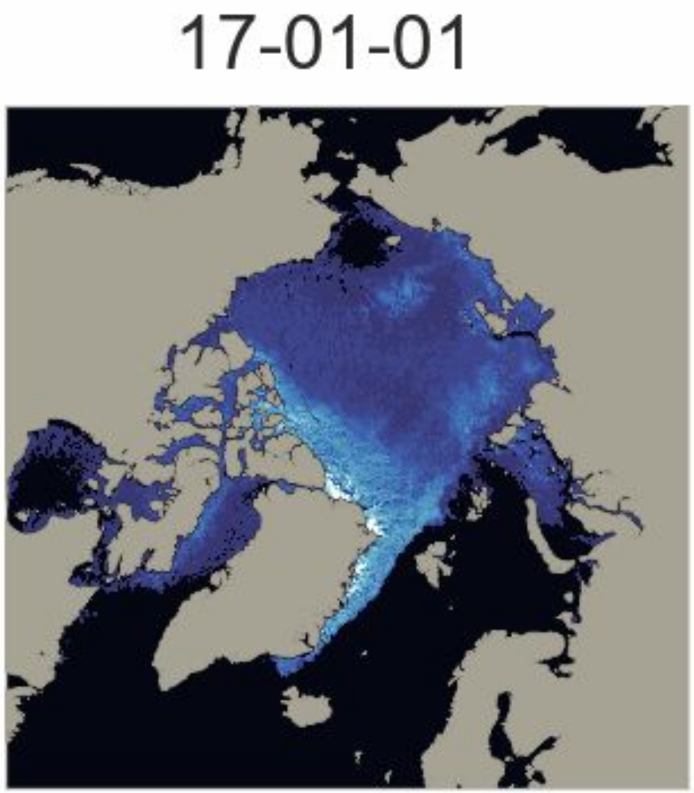
Take-home messages

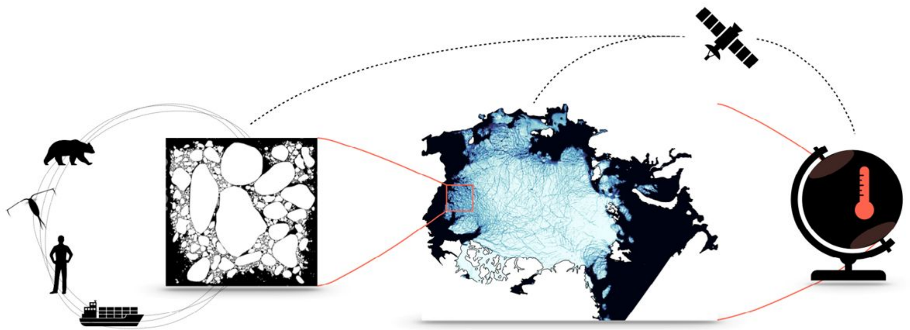
- A 2.5M parameters UNet neural network can surrogate sea-ice physical model and can predict sea-ice thickness, even on seasonal timescale

- Adding forcings, up to 12h, ensure a good forecast prediction and correct advection

- Smoothness of the surrogate can be quantified with PSD

- Stable surrogate model of neXtSIM could be a good basis for DA with access to a adjoint





Thank you for your attention!

Charlotte Durand
charlotte.durand@enpc.fr

This project is supported by Schmidt Futures – a philanthropic initiative that seeks to improve societal outcomes through the development of emerging science and technologies.

SCHMIDT FUTURES