ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Schmidt Sciences

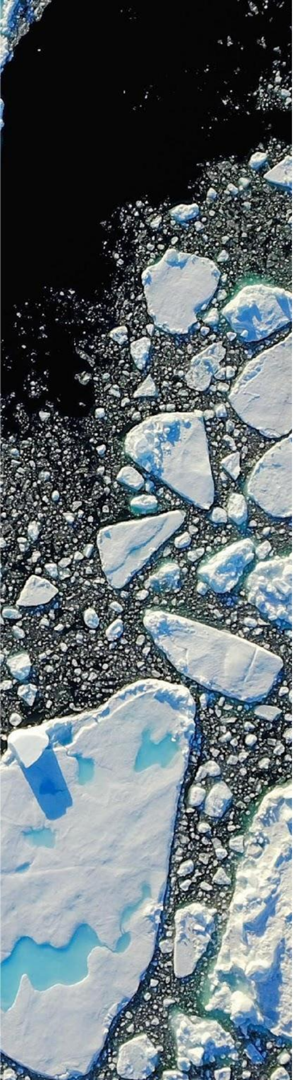
The Scale-Aware Sea Ice Project
SASIP

MELTED

Giovanni De Cillis

Online bias-correction in sea-ice models using machine learning and data assimilation

A. Carrassi, J. Brajard, L. Bertino, M. Broccoli, T. Finn, M. Bocquet, D. Iovino



Motivation and Objectives

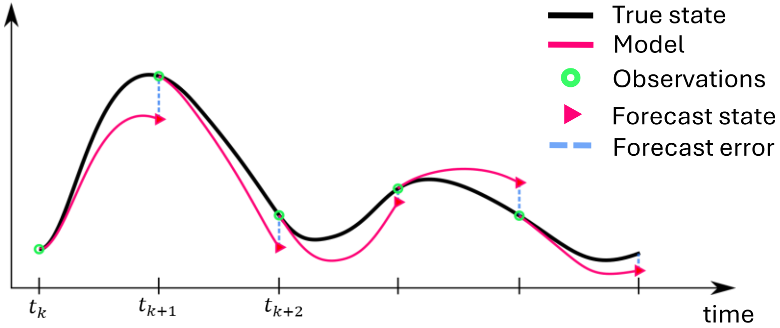
- Physics-based climate predictions are inherently affected by errors of different nature.
- Sea ice modeling presents additional complexities due to its multiscale, multiphase, discontinuous nature.

- Development of **hybrid physics-data-driven models** for sea ice predictions
  - State-dependent parametrizations of model errors, <u>learning from the analysis increments</u>

$$x_{k+1} = M(x_k) + \hat{\varepsilon}$$

Physics-based model

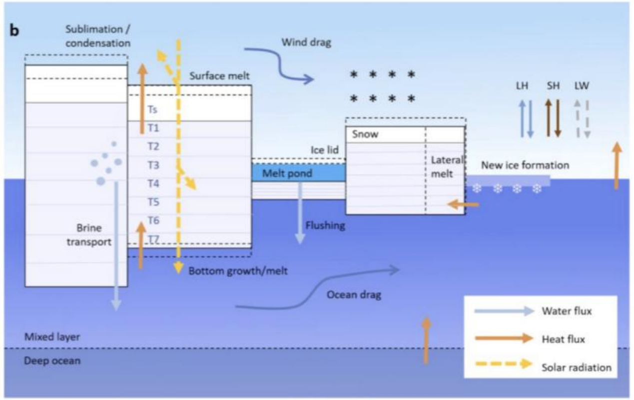
ML model



Two modeling scenarios of increasing complexity

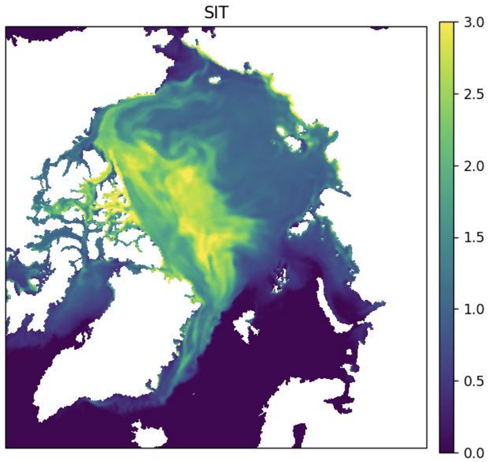
Icepack – 1D ice column model

- State-of-the-art column physics module (CICE)
- Thermodynamics and Ice Thickness Distribution transport
- No advection



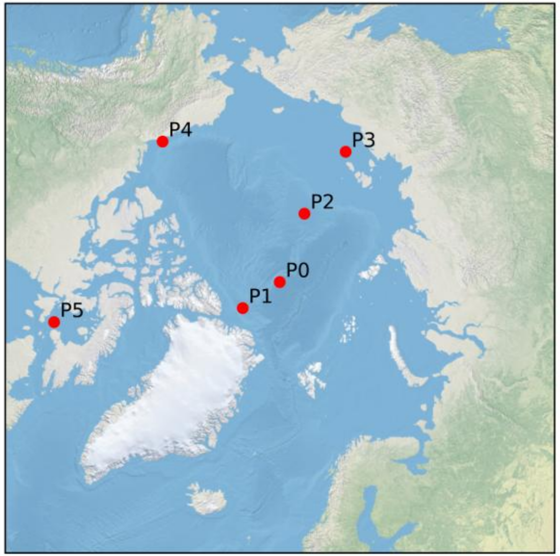
C-GLORS – CMCC Global ocean-sea ice reanalysis system

- Ocean component: NEMO v4.2
- Sea ice component: CICE6
- 3D-Var



Experiments with Icepack

6 locations simulated



- 7 ice layers + 1 snow layer
- Coupled slab ocean model
- 5 ice thickness categories
- Atmospheric forcing from ERA5 (1993-2007)

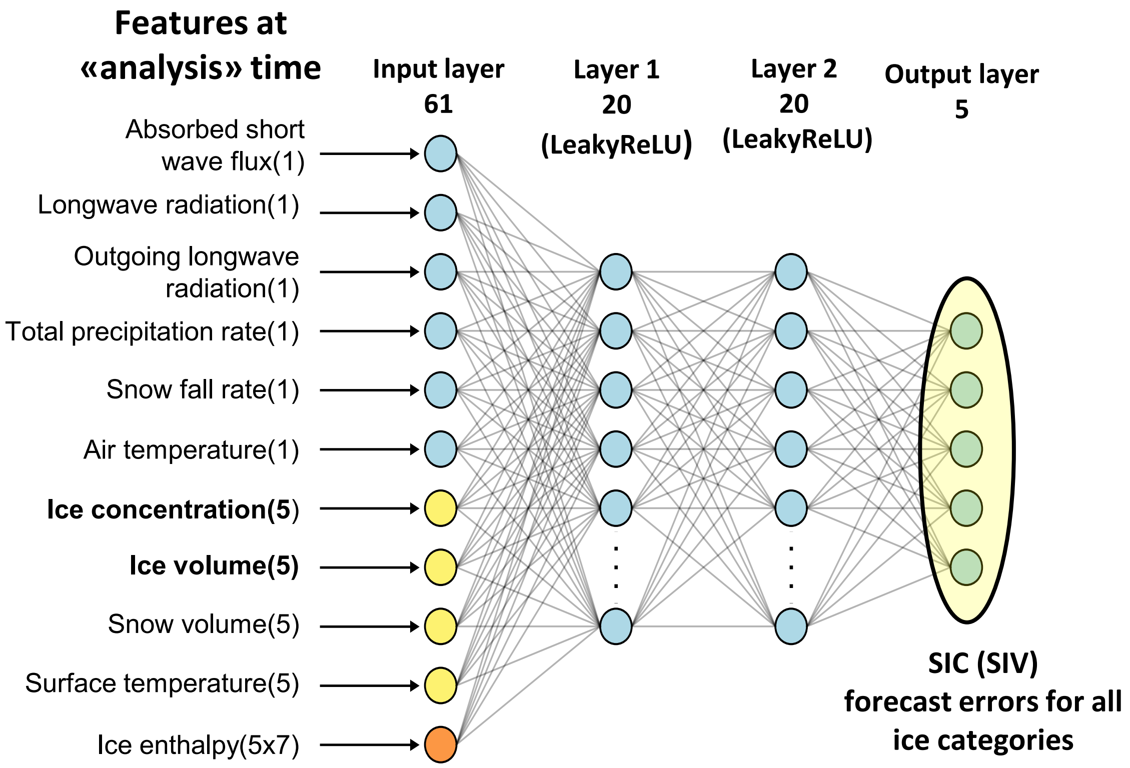
Training against parametric errors

Parameter description	Default (true) value	Distribution	Min value	Max value	Mode
Thermal conductivity of snow (W m^{-1}K^{-1})	0.3	Uniform	0.03	0.65	
Max. melting snow grain size (μm)	1500	Triangular	250	3000	1500

Urrego-Blanco et al. 2016, *Uncertainty quantification and global sensitivity analysis of the Los Alamos sea ice model*

- 30 different configurations were generated sampling the parameters independently from the distributions in the table.
- Default parameters for the true run
- Forecast errors at 60-days lead time

Training data and neural network architecture



- **2 MLP for each member** (for SIC and SIV with the same architecture)
- Dataset size: 3132 instances (**10 years, weekly frequency, 6 locations**)

1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
Training	Training	Training	Training	Training	Training	Training	Training	Validation	Test

Training   Validation   Test

- MSE loss
- Early stopping based on validation loss

The hybrid Icepack-NN model

$$x_{k+1} = M(x_k^h)$$
$$x_{k+1}^{h,*} = x_{k+1} - g(\theta^*, x_{k+1})$$
$$x_{k+1}^h = \Pi^{con}(x_{k+1}^{h,*})$$

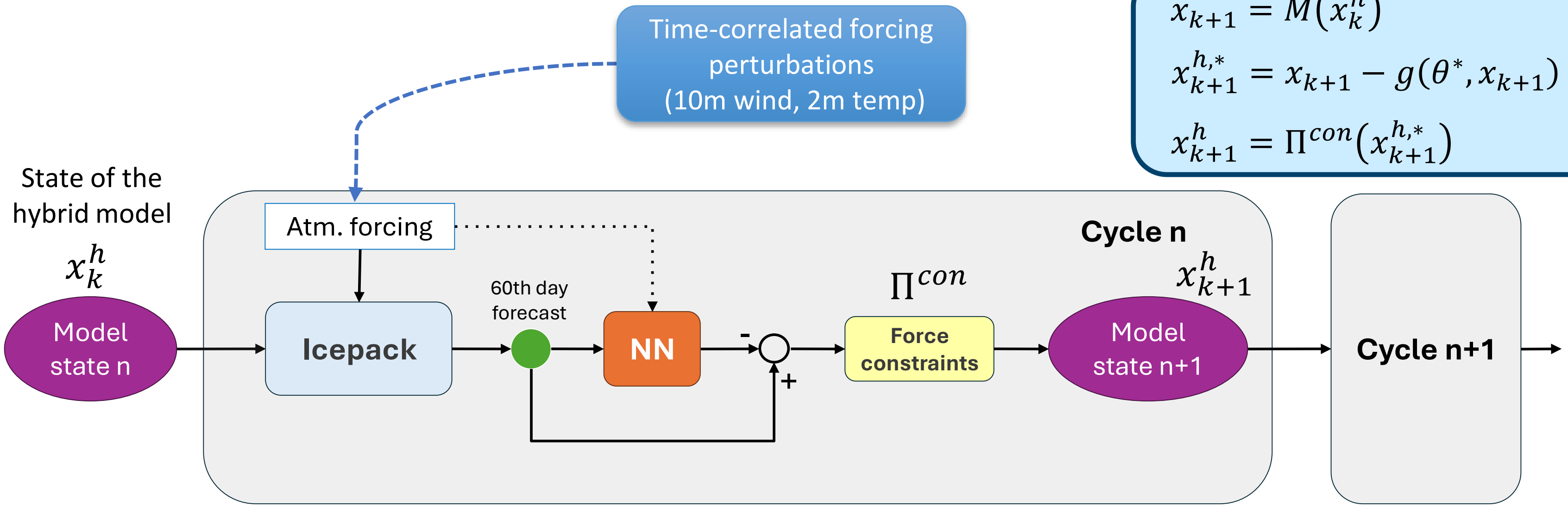
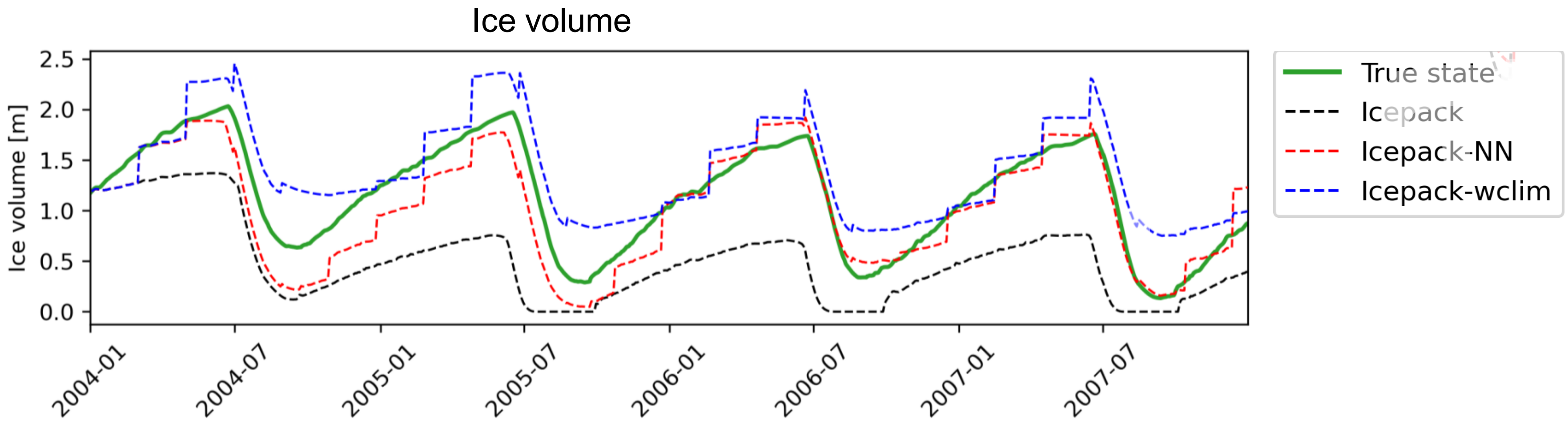


Figure: Schematic of NN correction integrated in the model's online execution, forming the hybrid model loop.

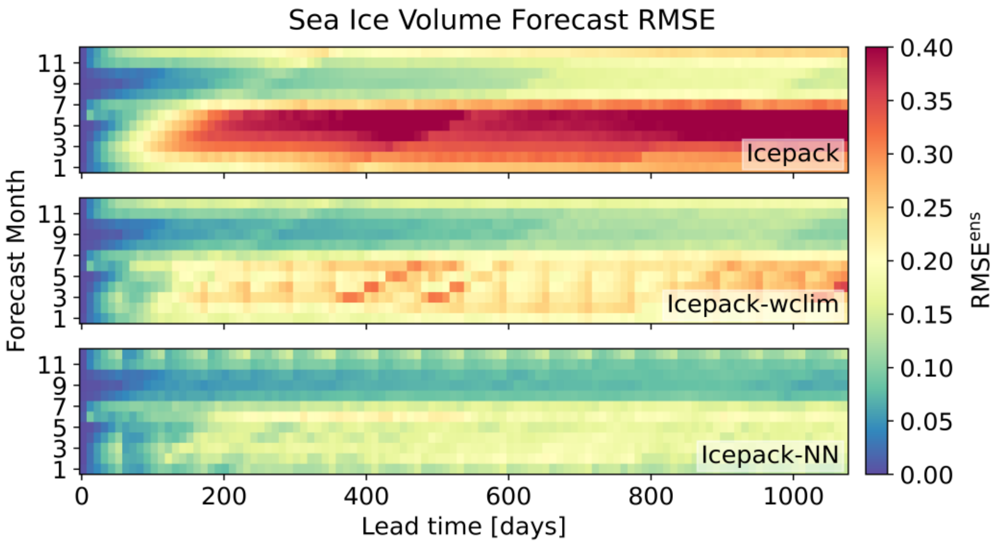
Forecast models	Correction type
Icepack-NN	NN-based
Icepack	N/A
Icepack-wclim	Error climatology

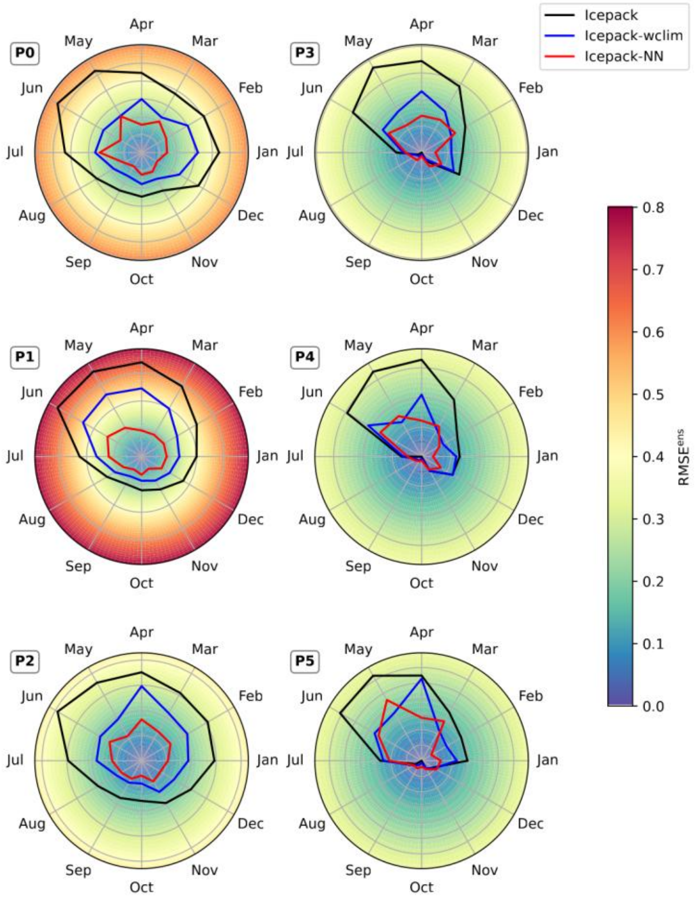
Hybrid model execution

- 3 systems simulated (**Icepack, Icepack-NN, Icepack-wclim**)
- 30 configurations tested.
- Time-correlated perturbations applied to the atmospheric forcing
- 6 locations (same as for training data)
- 52 start dates (1 per week in 2004)
- End date: 31-12-2007



Seasonal and geographical error patterns





- Errors strongly depend on the forecast month
- Larger errors during spring months
- Icepack-NN works better than Icepack-wclim (60% vs 40% RMSE reduction)
- Error vanishes in southernmost locations from August to November.

Transfer learning experiments

Collect data to correct a specific model

⇩

Train a NN to correct model error

⇩

Deploy the hybrid model

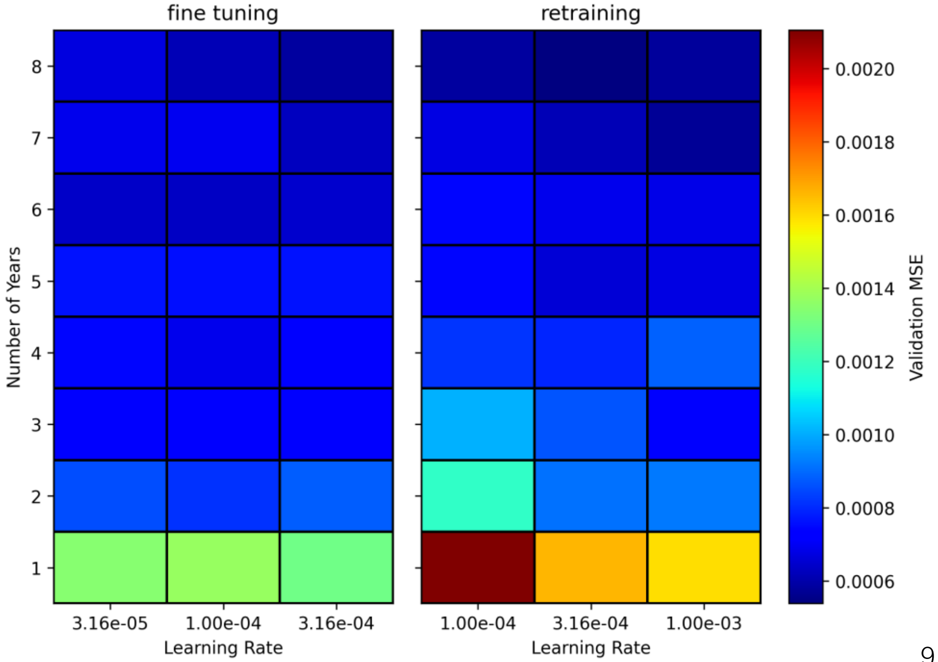
⇩

The model is updated

?

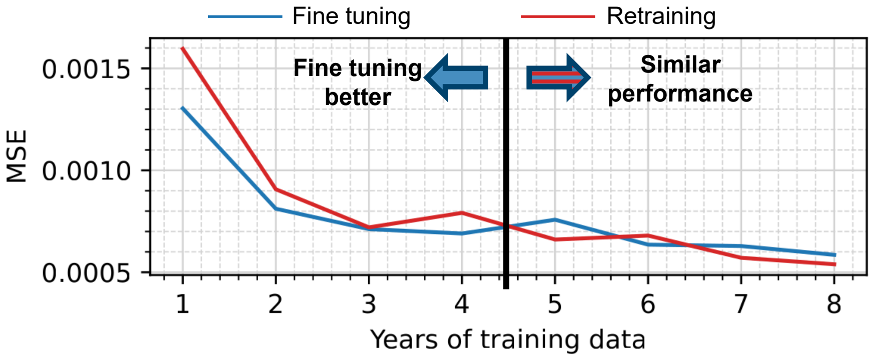
Can we take advantage of the pretrained NN? To what extent?

- Selected source-target model pairs
- Applied retraining and fine-tuning using increasing amounts of new training data
- Optimized the learning rate for each case

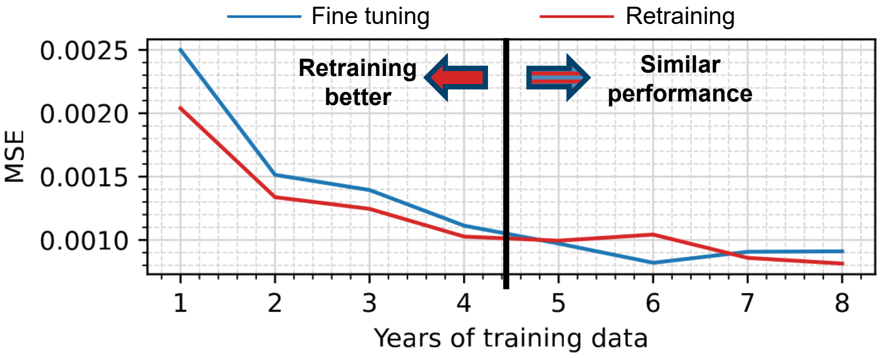


Transfer learning experiments

A successful case: Transfer learning works



A negative result

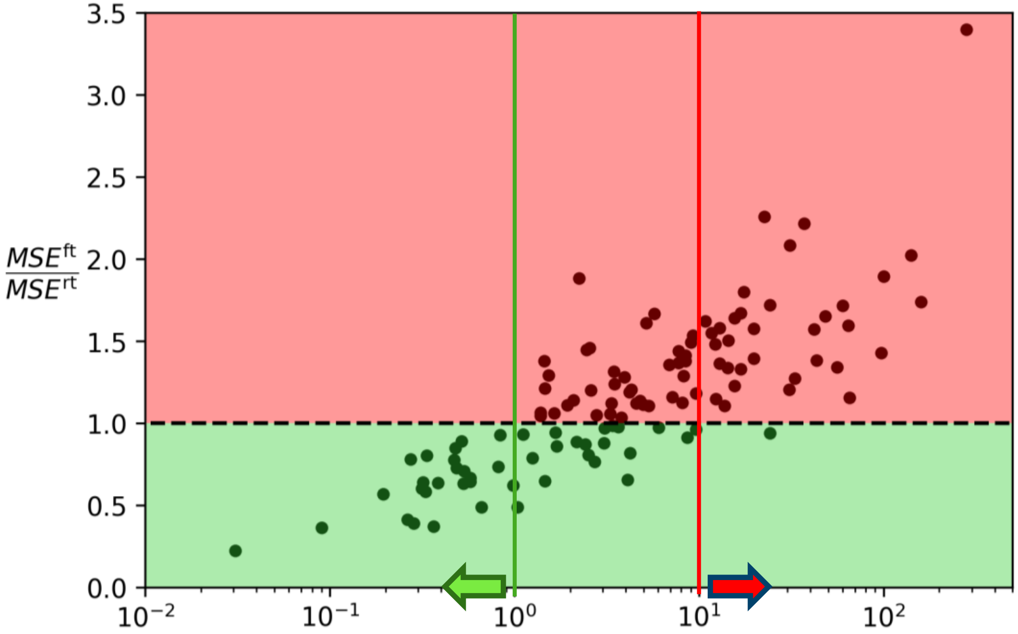


- **Fine-tuning can be beneficial** when a limited amount of new training data is available.
- As **more years of new training data** are introduced, the performance of the two approaches tends to **converge**.

- **Direct transfer results provide information on fine-tuning potential**

Assessing transfer learning potential

- Fine-tuning vs Retraining using 1 year of training data
- Several source-target model pairs tested



$$\frac{\text{Direct transfer MSE}}{\text{Uncorrected Forecast MSE}} \quad (\text{nMSE}^{\text{pre}})$$

- When nMSE^{pre} < 1, **fine-tuning** is preferred
- When nMSE^{pre} > 10, **retraining** is better
- When 1 < nMSE^{pre} < 10, both approaches should be tested

De Cillis et al. 2026
accepted in QJRMS

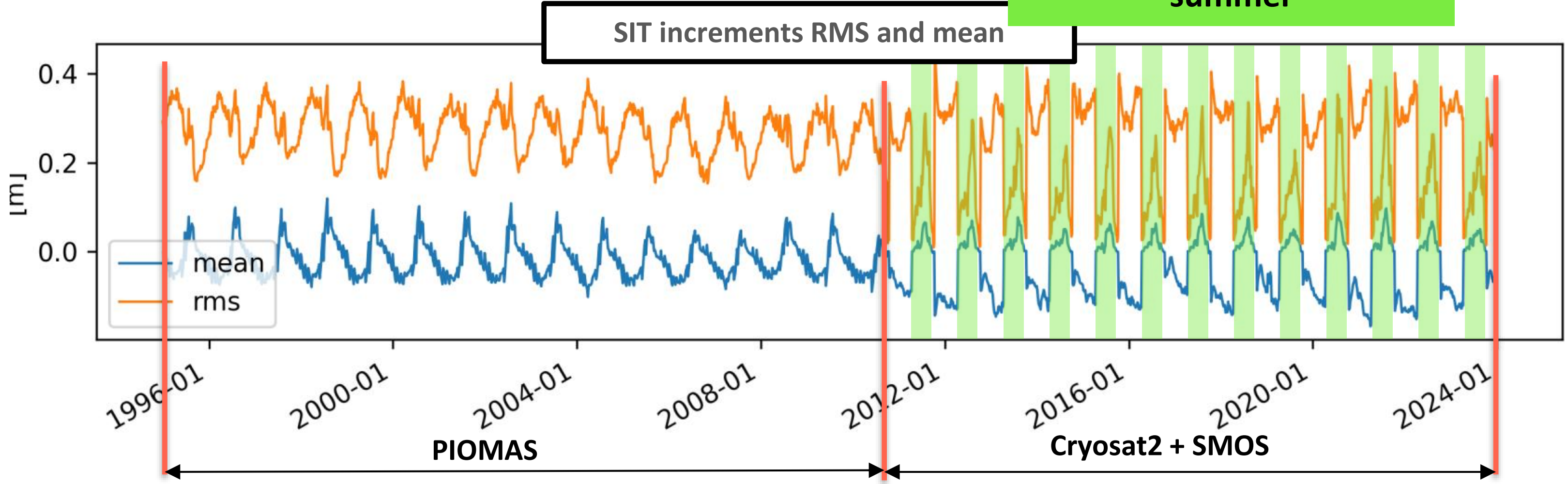
C-GLORS reanalysis data

CMCC Global Ocean Physical Reanalysis System

- 1995 – 2023, 5 days assimilation cycle
- Sea ice assimilated data:
  - Sea ice concentration: OSISAF
  - Sea ice thickness: PIOMAS, Cryosat2 (CS2), SMOS

SEA ICE THICKNESS DATA
- 1995 - 2010/09: assimilation of PIOMAS dataset
- 2010/10 - 2024:
  - mid-Oct - mid-Apr assimilation of CS2 + SMOS
  - **mid-Apr - mid-Oct no direct thickness observations**

Missing SIT observations in summer



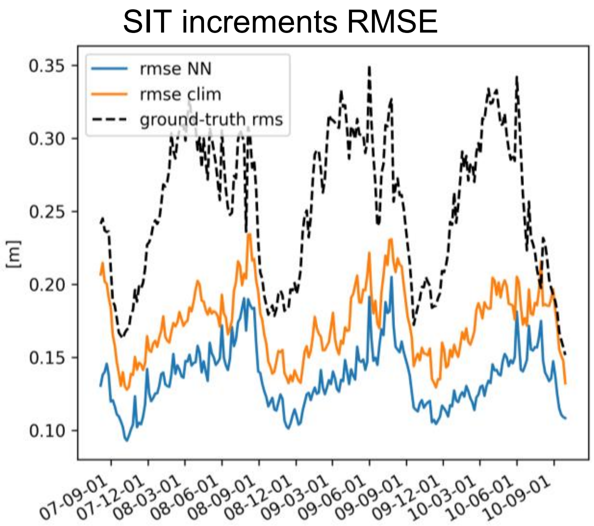
NN development and validation

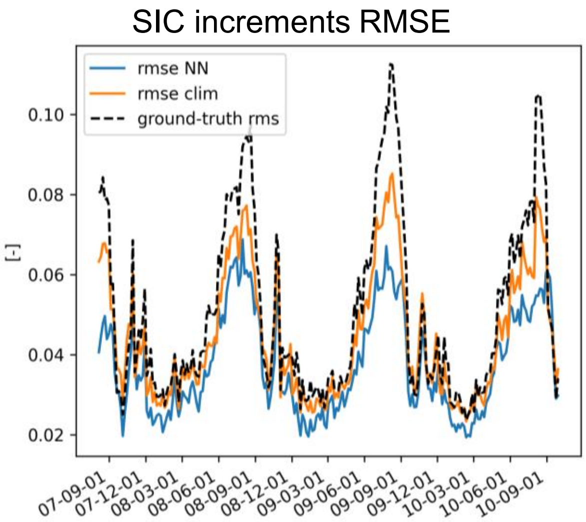
MODEL DEVELOPMENT STRATEGY:
- Split the dataset into the two periods
- Use the **PIOMAS period** for initial training and validation
- Benchmark against climatological predictions

MODEL
- **U-Net** predicting SIC and SIT increments
- Two output heads for SIC and SIT
- **Partial convolution** for the land mask

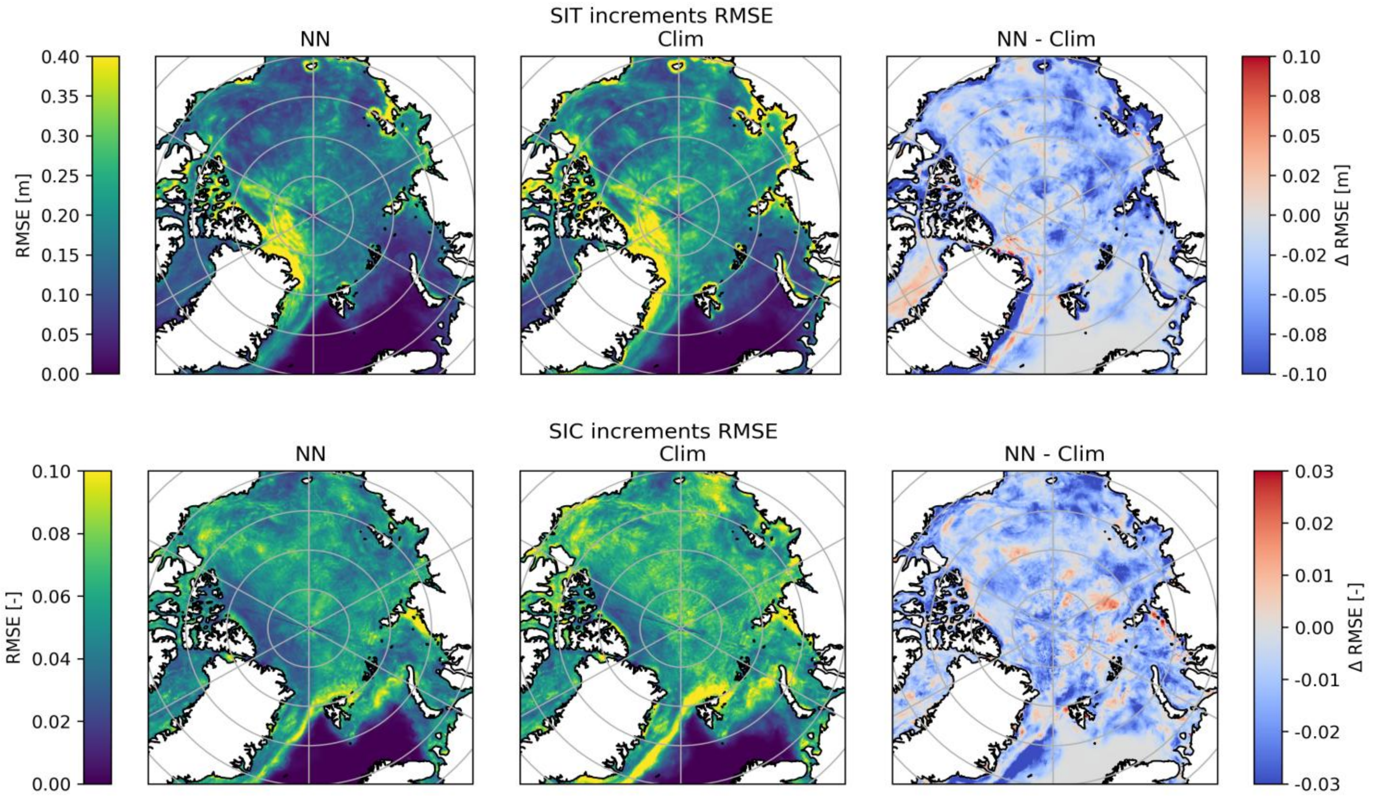
INPUT FEATURES:
- **Sea ice, ocean, and atmospheric** predictors

The U-Net outperforms climatology on the PIOMAS period





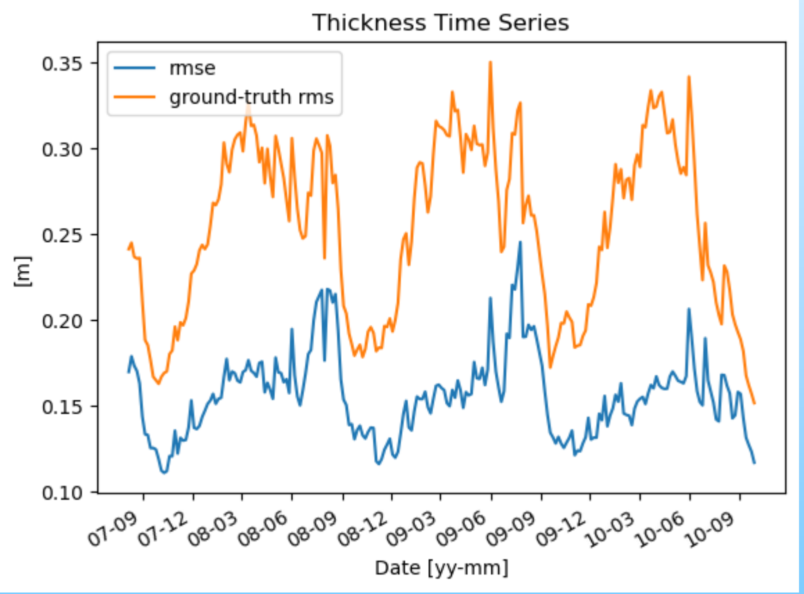
Spatial RMSE maps



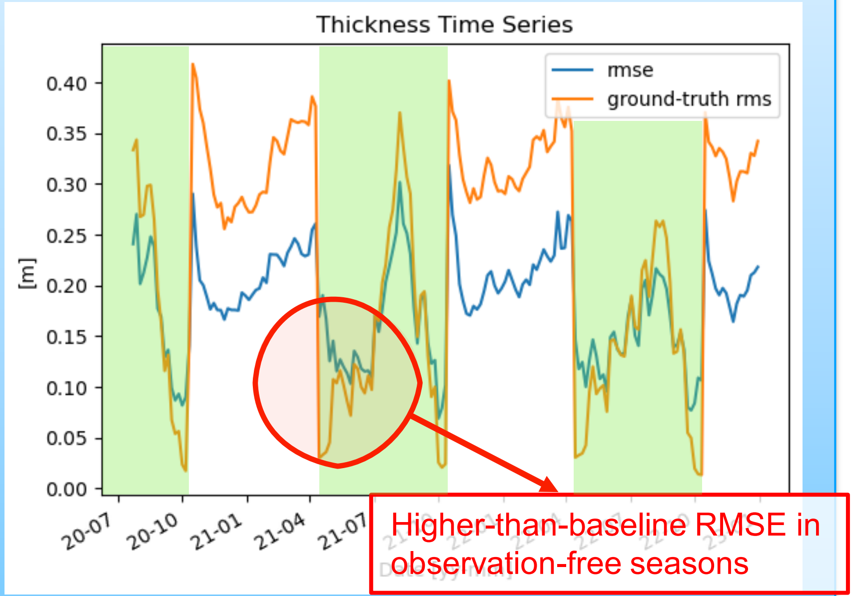
Gaps in SIT observations challenge NN training

Same U-Net architecture, trained separately on two periods

1995-2010/09 data (**PIOMAS** SIT assimilation)

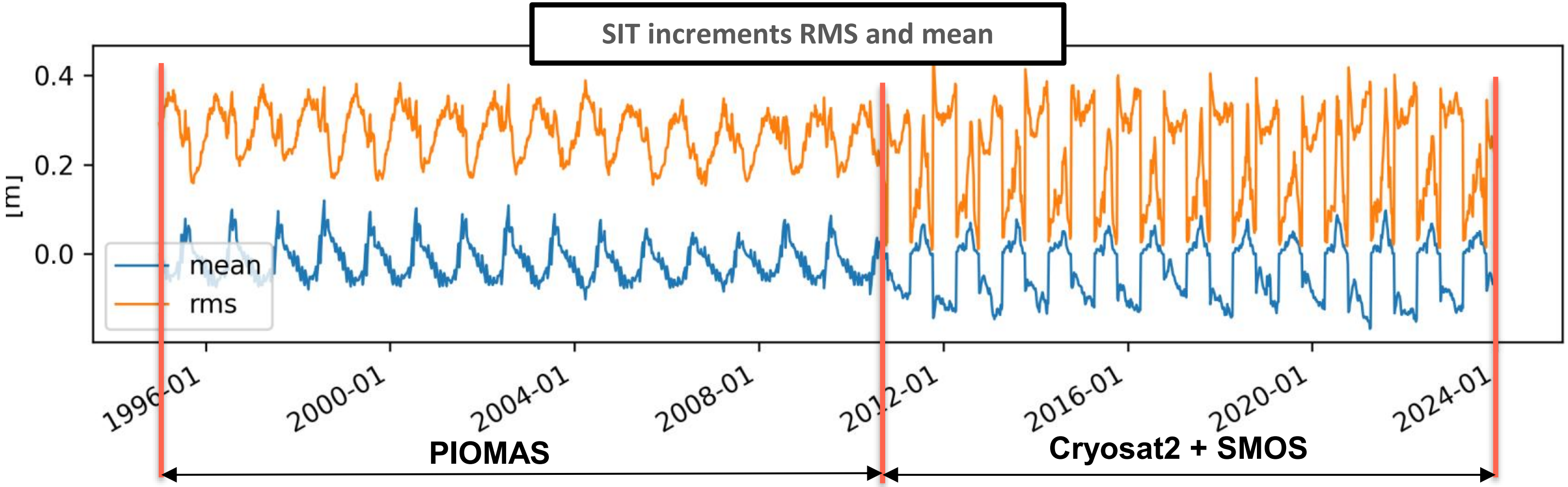


2010/10-2024 data (**CS2SMOS** SIT assimilation)



- **Key point**: In seasons without SIT assimilation, increments are a weaker proxy for model error, so lower scores may not indicate poorer bias prediction.

Exploiting transfer learning to learn from the full dataset



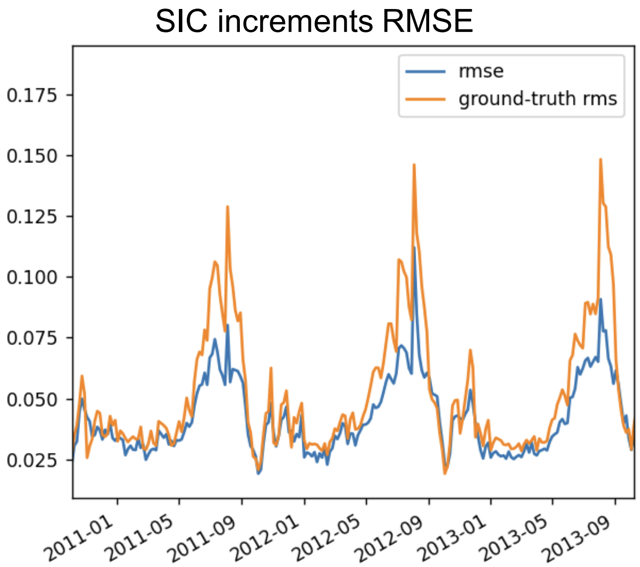
Learn increments from the PIOMAS period

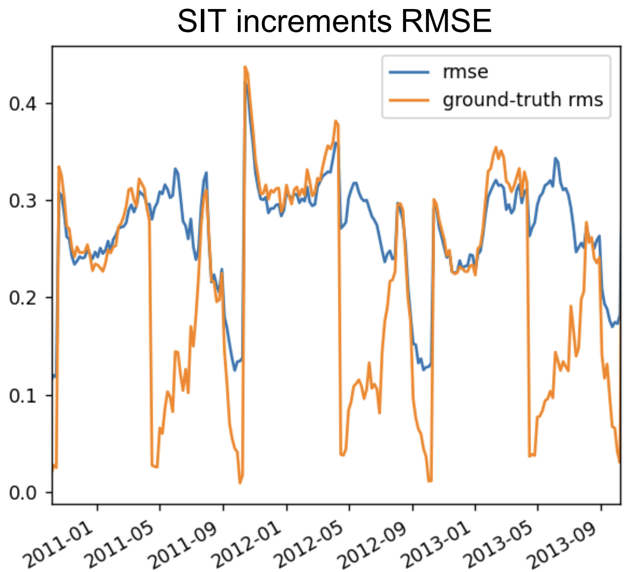
Learn increments from the CS2SMOS period

- More robust feature maps
- Extrapolate to periods without SIT obs

Direct transfer as an indicator of task similarity

U-Net trained on the PIOMAS period, applied without retraining to the CryoSat2/SMOS period

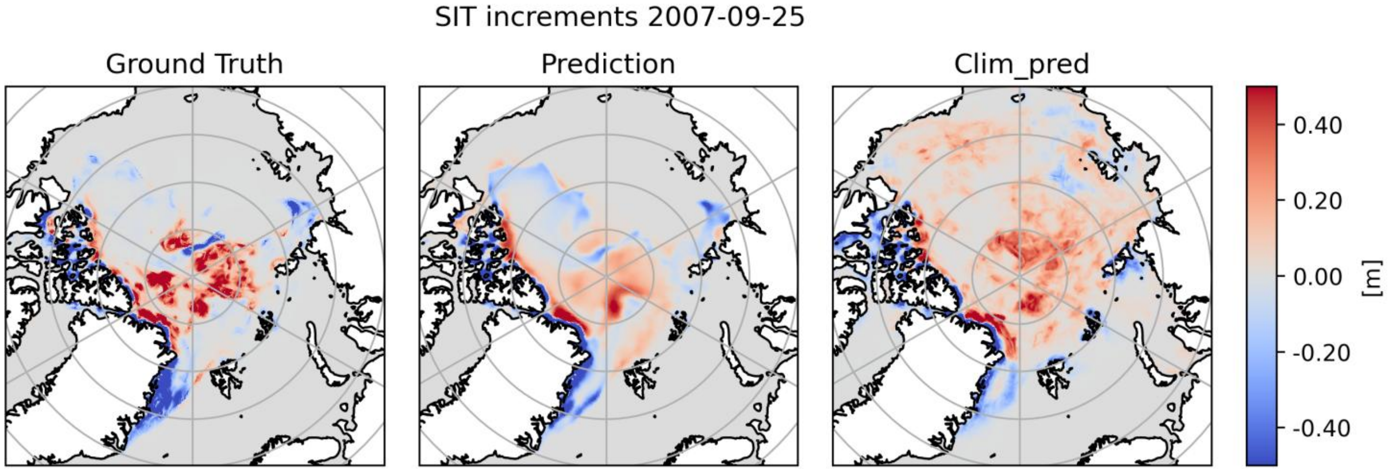


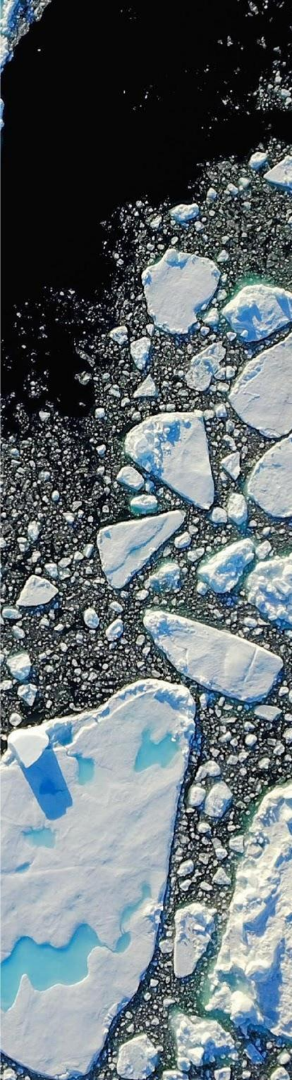


- SIC skill remains good under direct transfer
- SIT skill is limited, but does not fall below the zero baseline
- These results motivate fine-tuning to extend SIT bias prediction to observation-free seasons.

Next steps for the 2D C-GLORS system

- Test fine-tuning strategies to make predictions in the CryoSat2 period and extrapolate to observation-free seasons
- Integrate the U-Net into the DA system for online bias correction



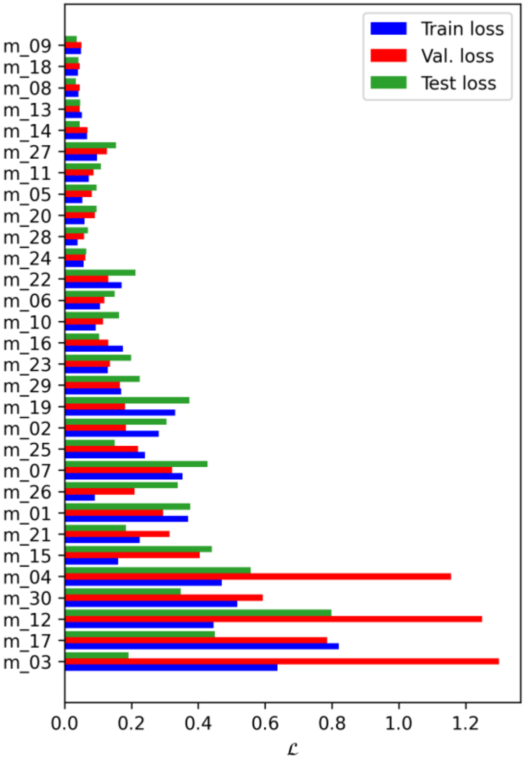


Thank you for your attention!

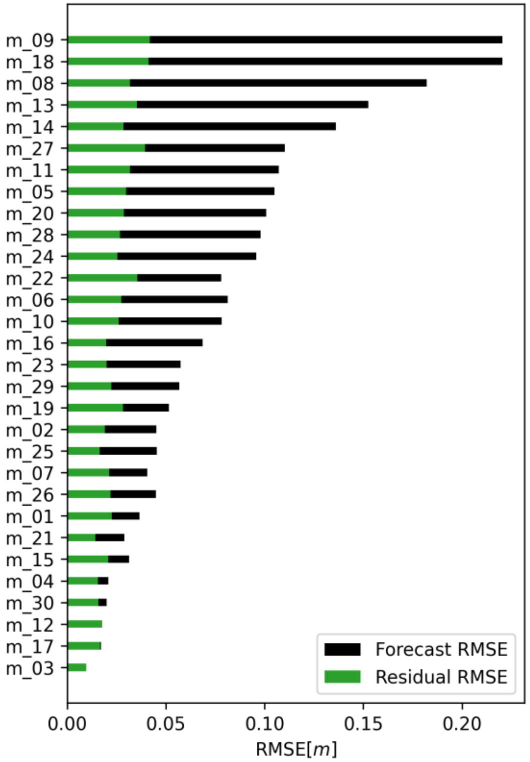
NN training and offline performance

Forecast RMSE on the training set

Loss function

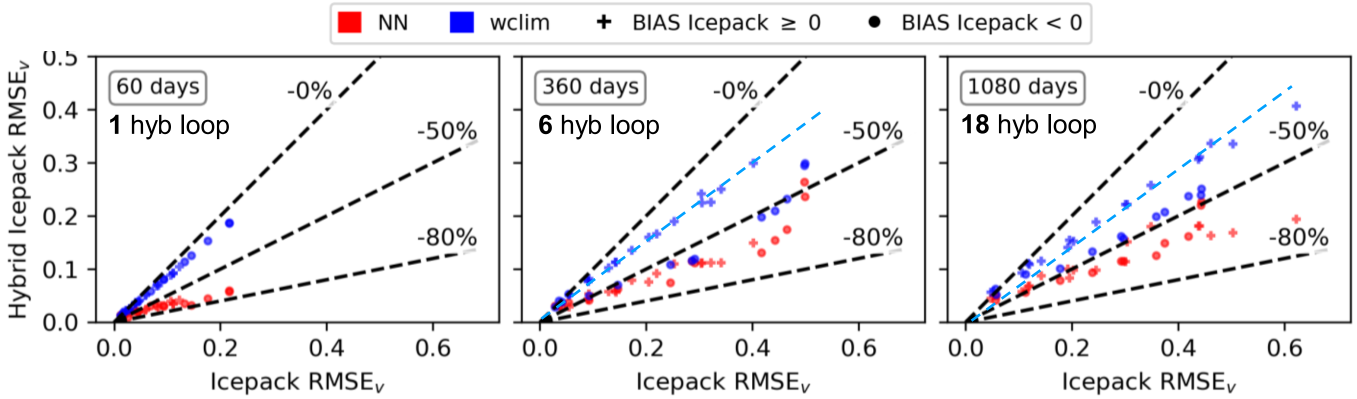


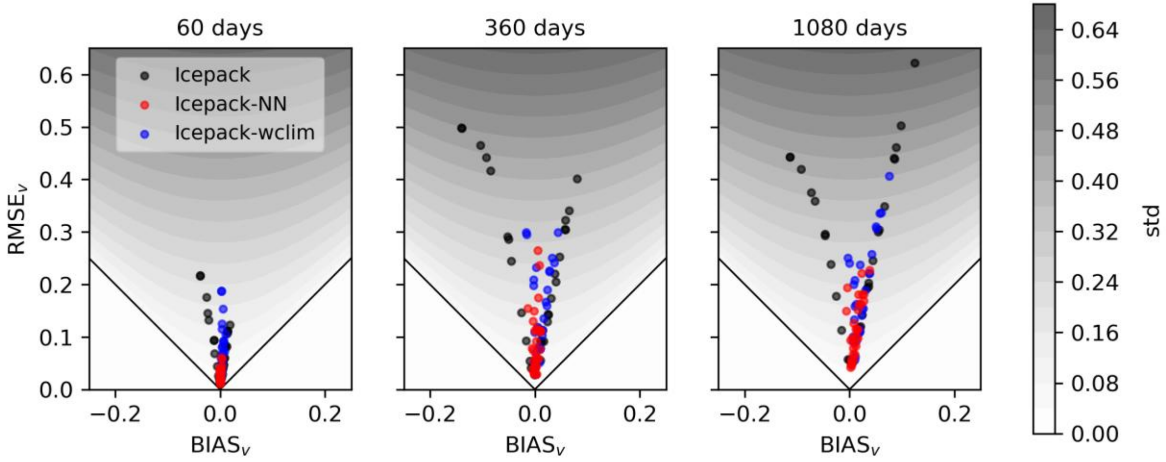
Ice volume RMSE on the test set



- The NN works better when there are large errors to correct
- Up to 80% ice volume RMSE reduction

Hybrid model forecasting performance

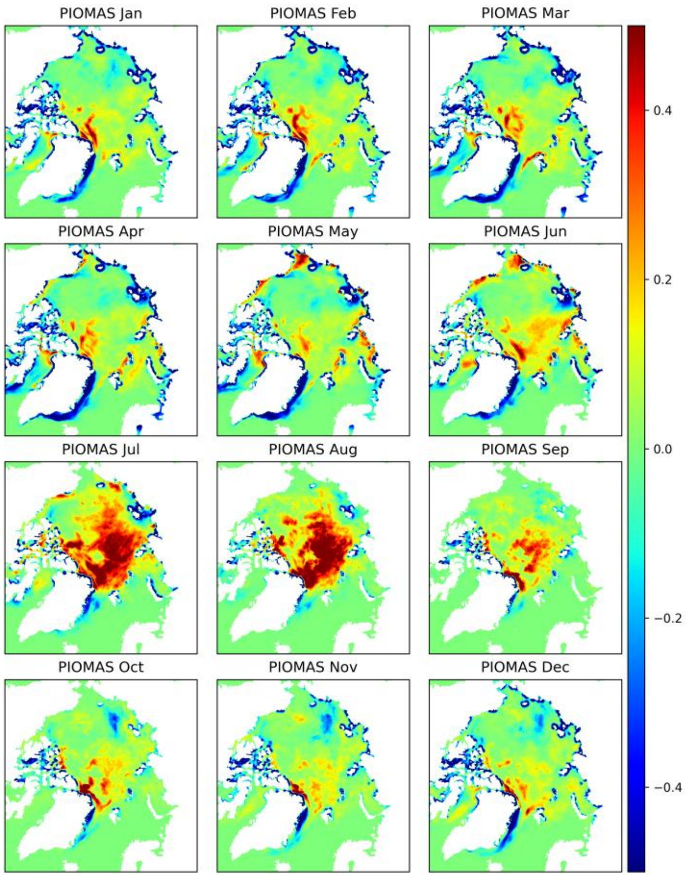




- Both Icepack-NN and Icepack-wclim reduce RMSE and BIAS
- Icepack-NN shows better performance at all lead times
- Icepack-wclim struggles to correct models with positive bias

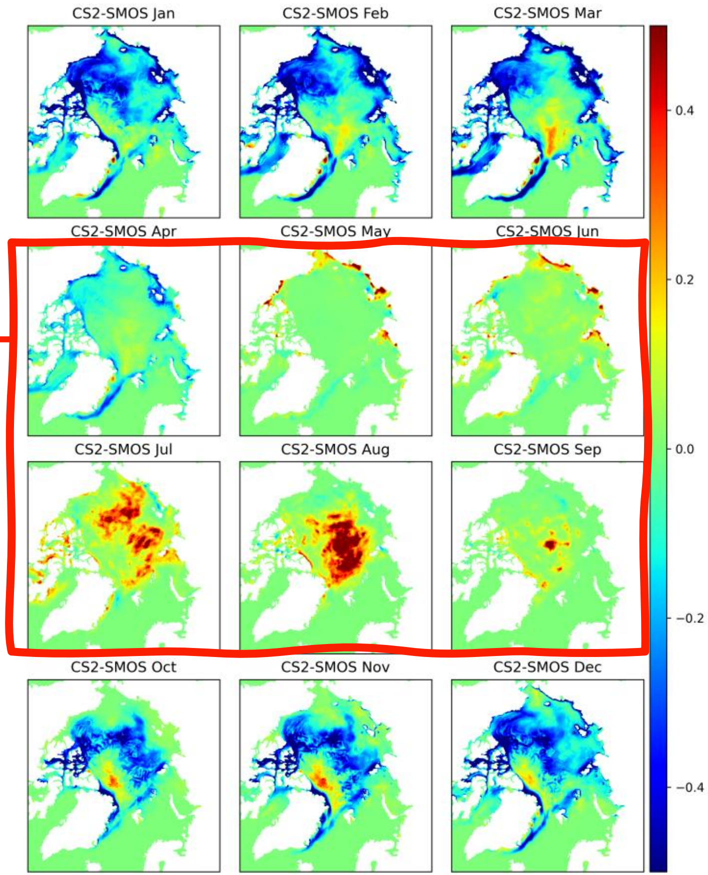
C-GLORS SIT analysis increments monthly climatologies

PIOMAS



CS2-SMOS

No SIT observations



Feature selection

Permutation Importance-based Recursive Feature Elimination

Rows represent successive iterations (top to bottom). Square colors denote the increase in validation loss caused by permuting each feature, relative to the reference obtained without permutations. Reference validation loss values are shown in the right panel. At each iteration, the feature whose permutation produces the smallest increase is removed, indicated by hatching in the subsequent row and labeled on the left.