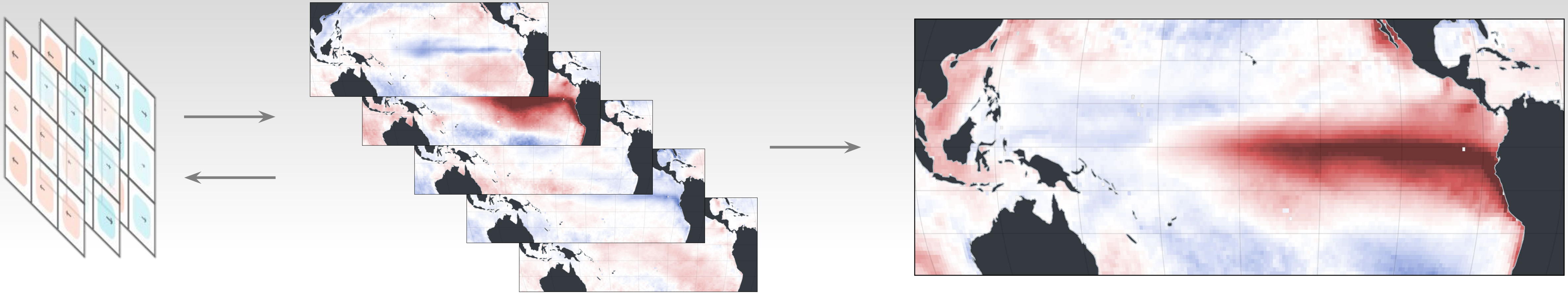
Hybrid deep learning and model-analog forecasting of ENSO



Kinya Toride

Matthew Newman, Andrew Hoell, Antonietta Capotondi, Jakob Schlör, Dillon Amaya

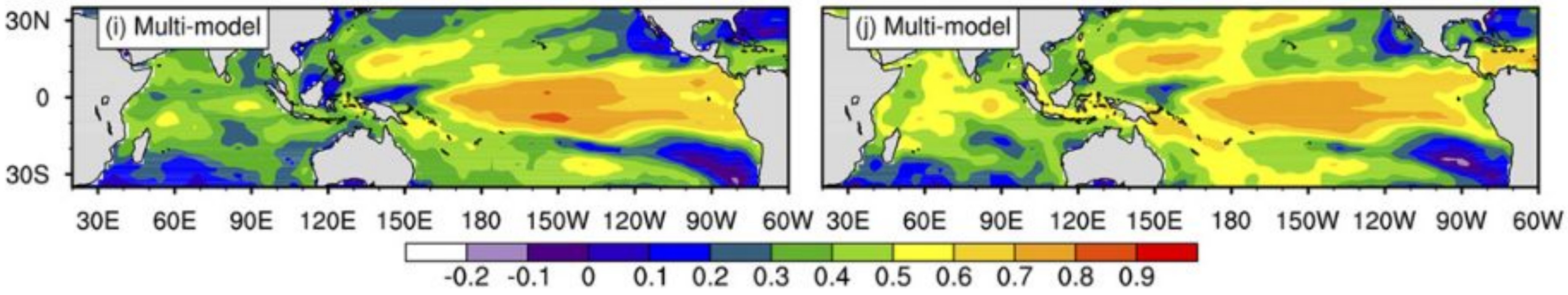
CIRES & NOAA PSL

S2S Workshop | June 2024

“Model-Analog” forecasts

- Analogs are drawn from a repository of climate simulations
- Model-analog provides a comparable hindcast skill to dynamical models

Model-analog NMME (dynamical models)

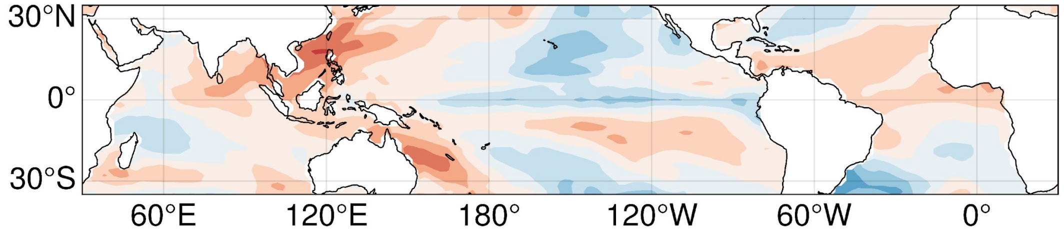


Anomaly correlation of SST forecast at 6 months lead

Ding et al. (2018)

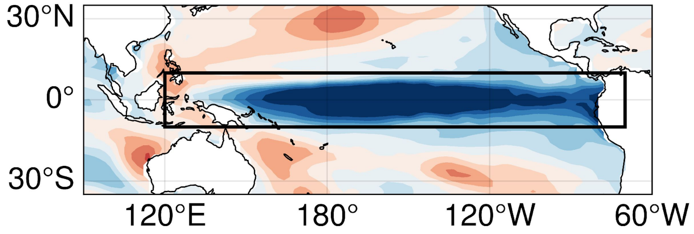
Issue: Initial analogs can evolve to very different states

Initial condition

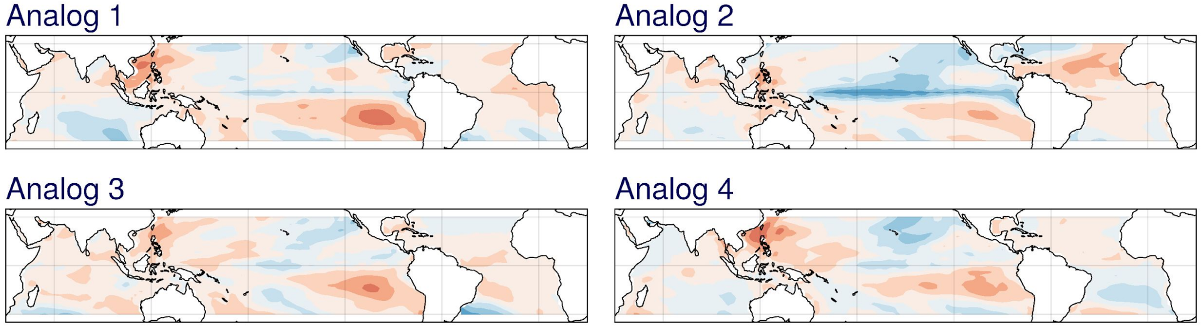


12 months

Target



Analogs (closest conditions)

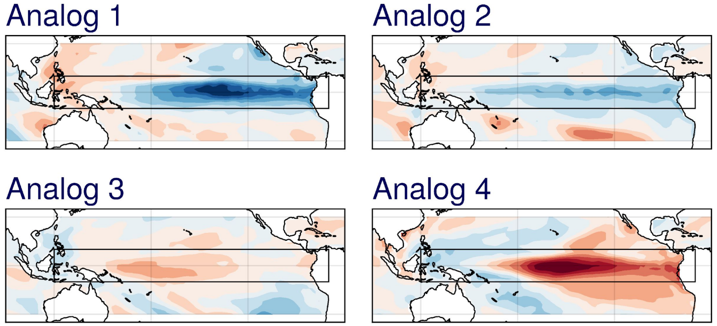


Analogs based on the entire tropical region.

??

12 months

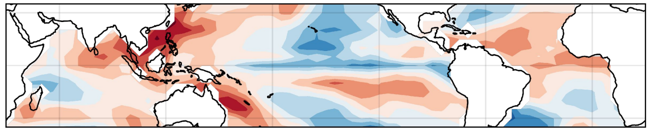
Analog forecasts



Small errors can grow significantly

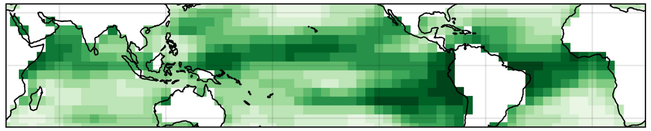
Aim: Use deep learning to constrain error growth

Input: initial condition



Deep learning

Output: weights used to select analogs

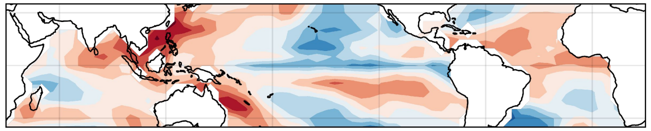


Weights show “sensitive region” where initial error growth is significant

- Large weight regions: analogs need to match closely
- Small weight regions: analogs can differ

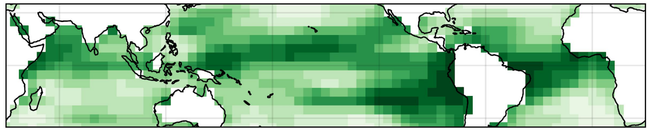
Hybrid deep learning and model-analog (DL + MA)

Input: initial condition



Deep learning

Output: weights used to select analogs



Analog forecasting

Interpretable-by-design

- Estimated weights show important (sensitive) regions.
  Objectively evaluated by forecasting skill

- Analog forecasting provides evolution of the entire system.
  Fully based on physical models

Data and model

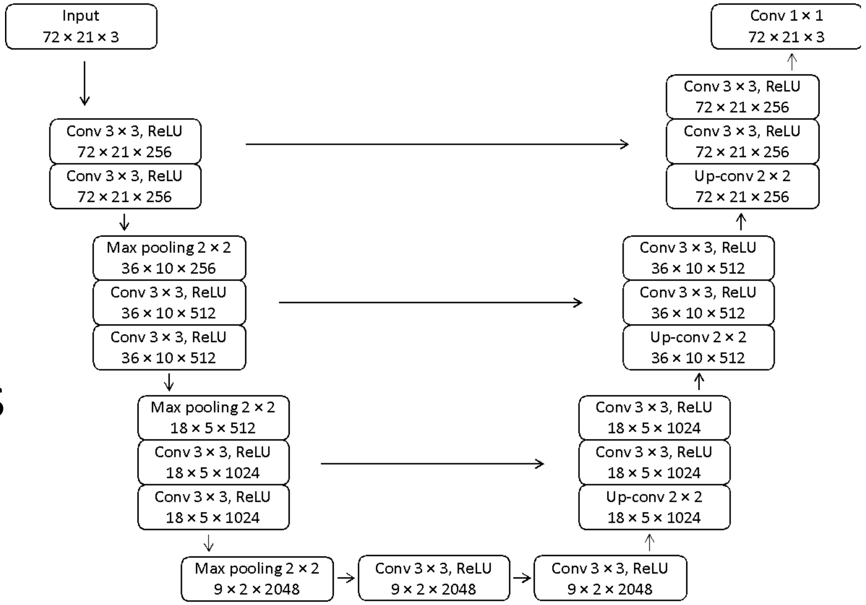
CESM2

- 1850–2014, 100 ensembles
- Monthly anomaly
- Sea surface temperature (SST)
- Sea surface height (SSH)
- Zonal wind stress (TAUX)
- 50°S–50°N

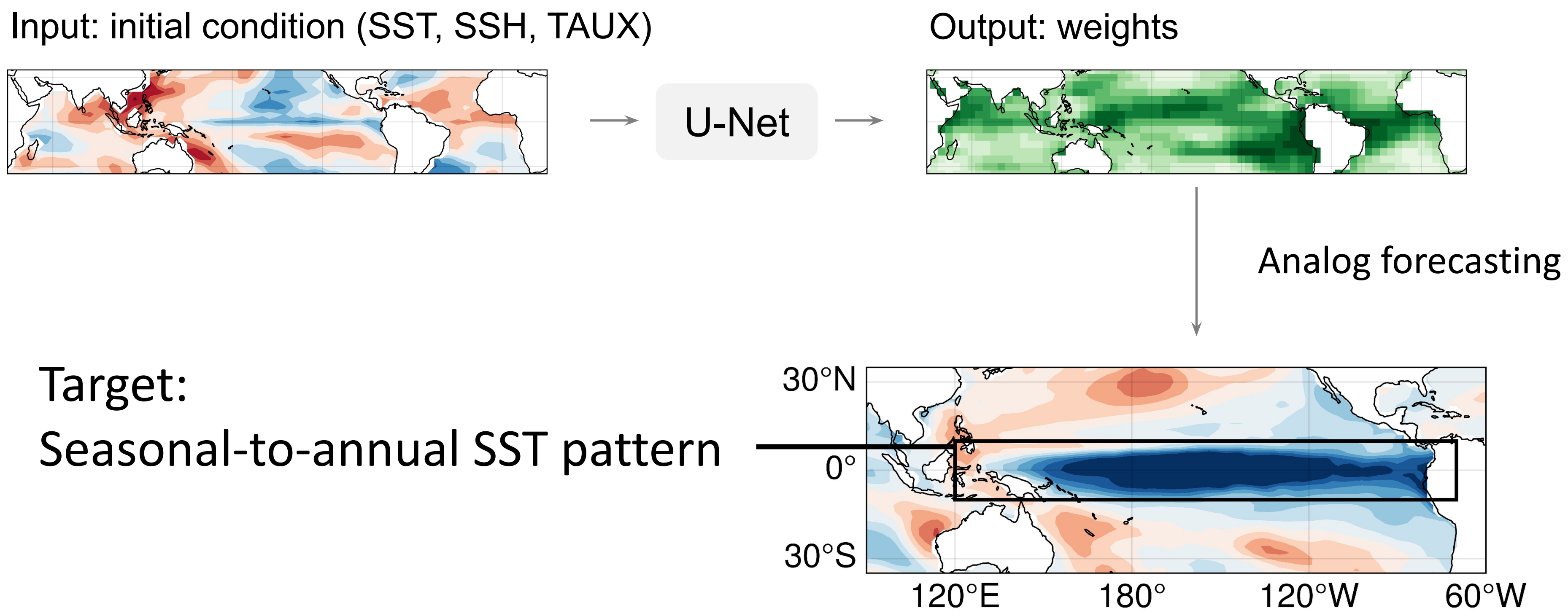
	Period	Sample size
Training (library)	1865–1958	94 y × 100 (70%)
Validation	1959–1985	27 y × 100 (20%)
Test	1986–1998	13 y × 100 (10%)

U-Net

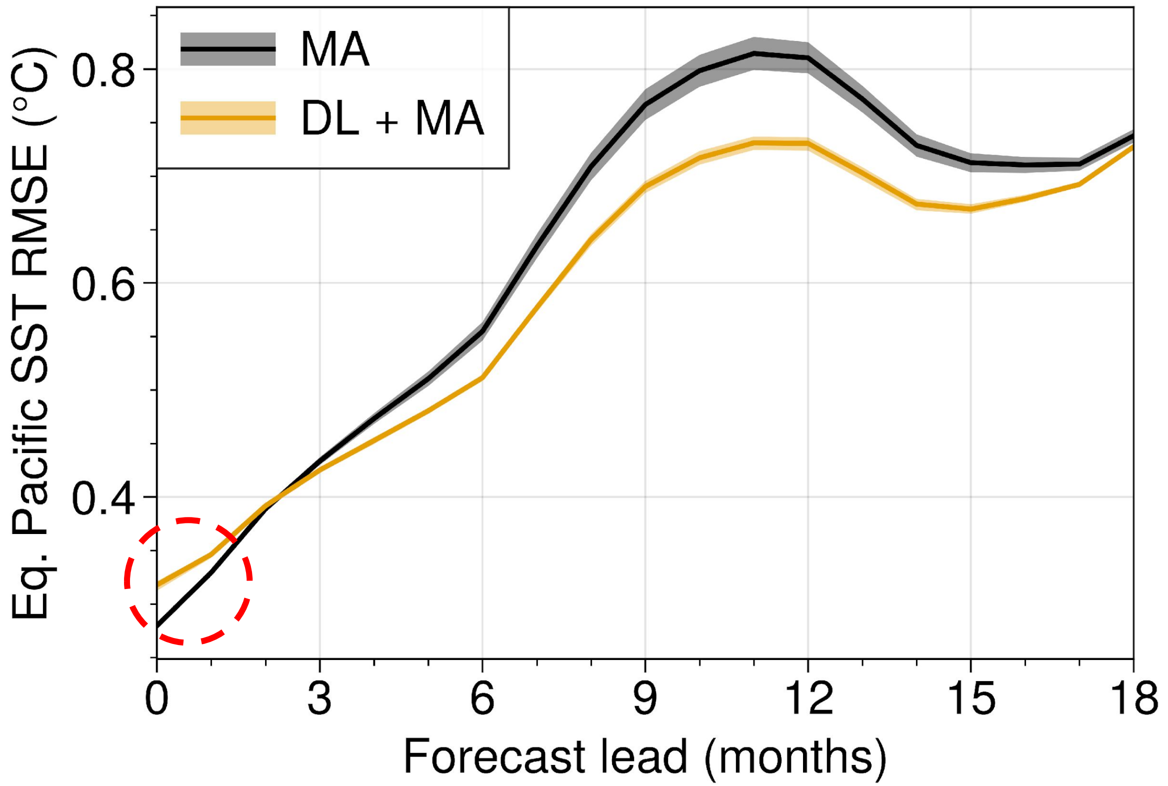
- Downsampling, upsampling, skip connections



Target: SST pattern over the equatorial Pacific

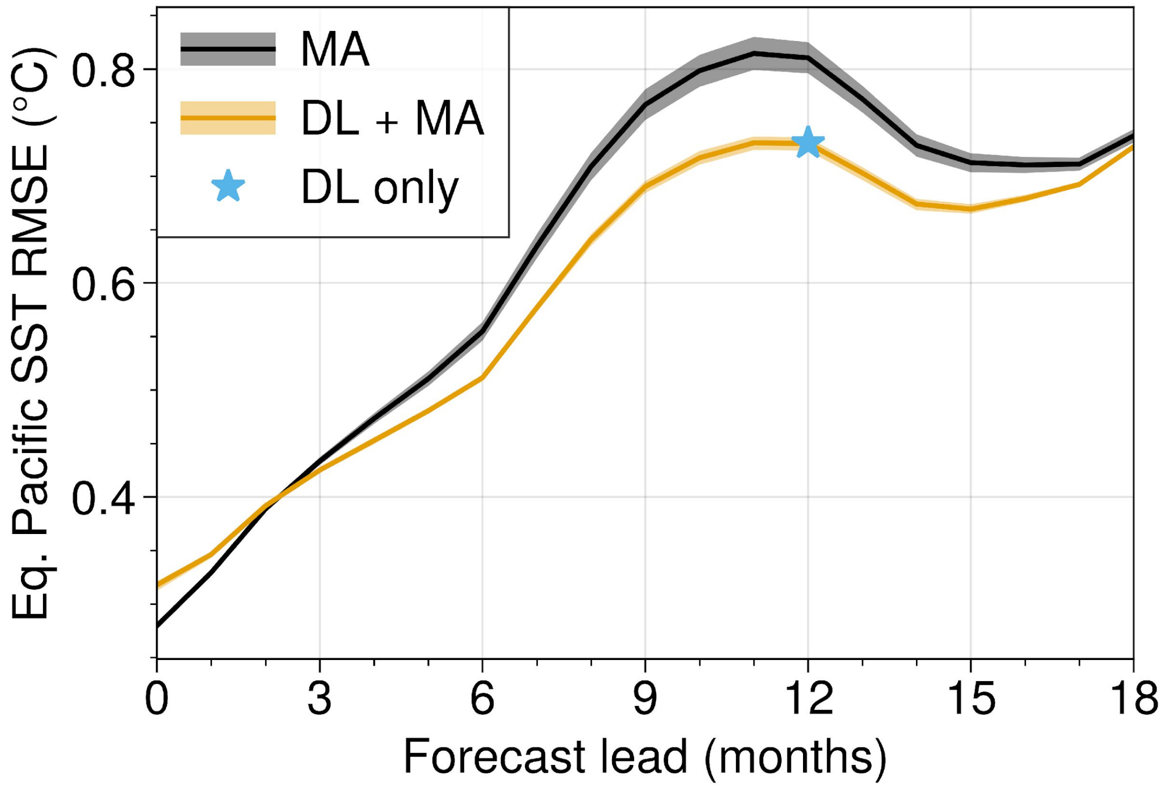


10% improvements at 9-12 months lead



Smaller weights are assigned to the target area Analogs can differ

Comparable skill to an equivalent DL-only method



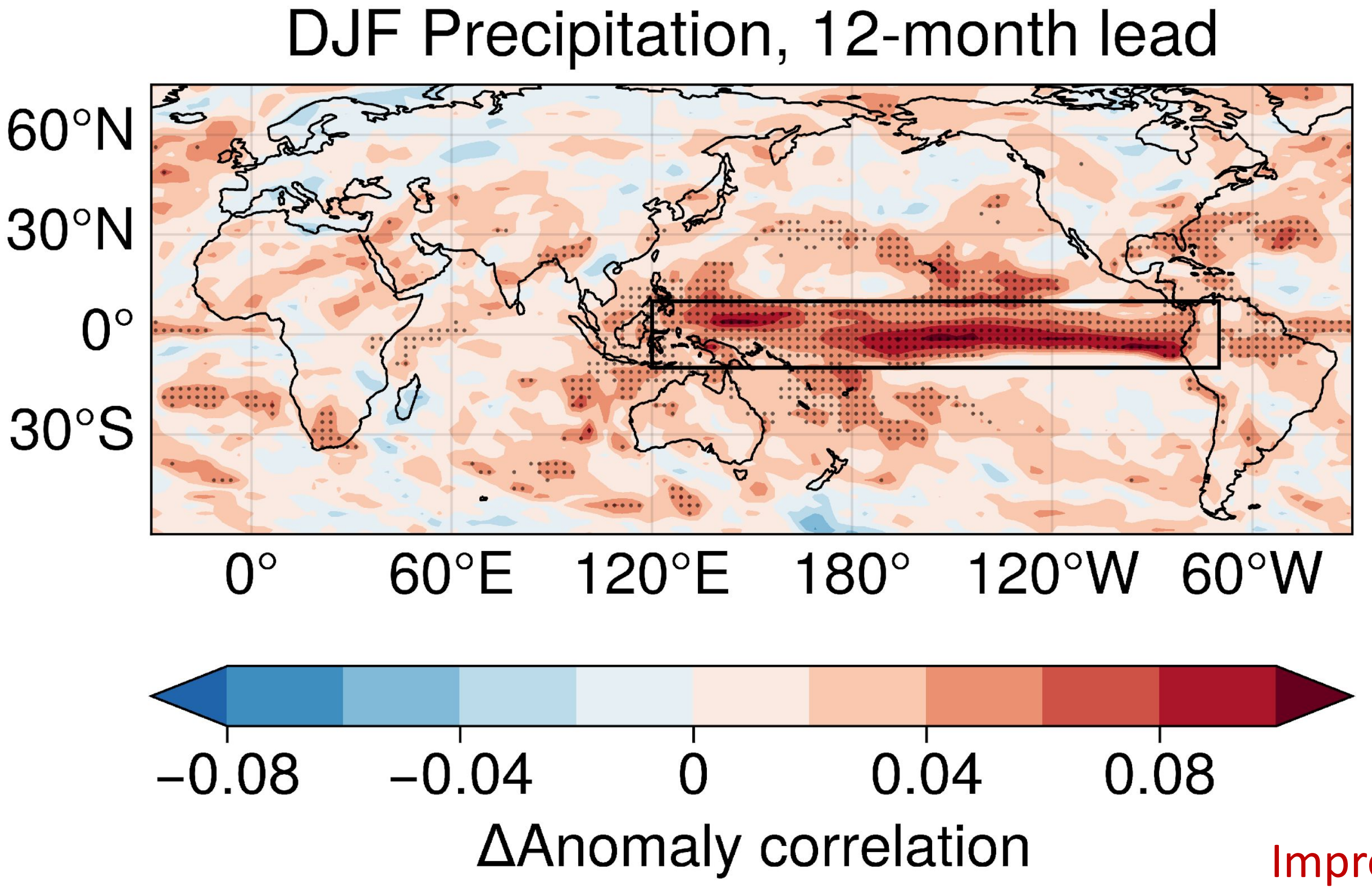
DL-only method

- Same network except for the last layer which predicts equatorial Pacific SST directly.
- Needs to be trained for each lead.

The hybrid method enhances interpretability and captures the time evolution of entire system without compromising DL skill.

Improvements in precipitation forecasts

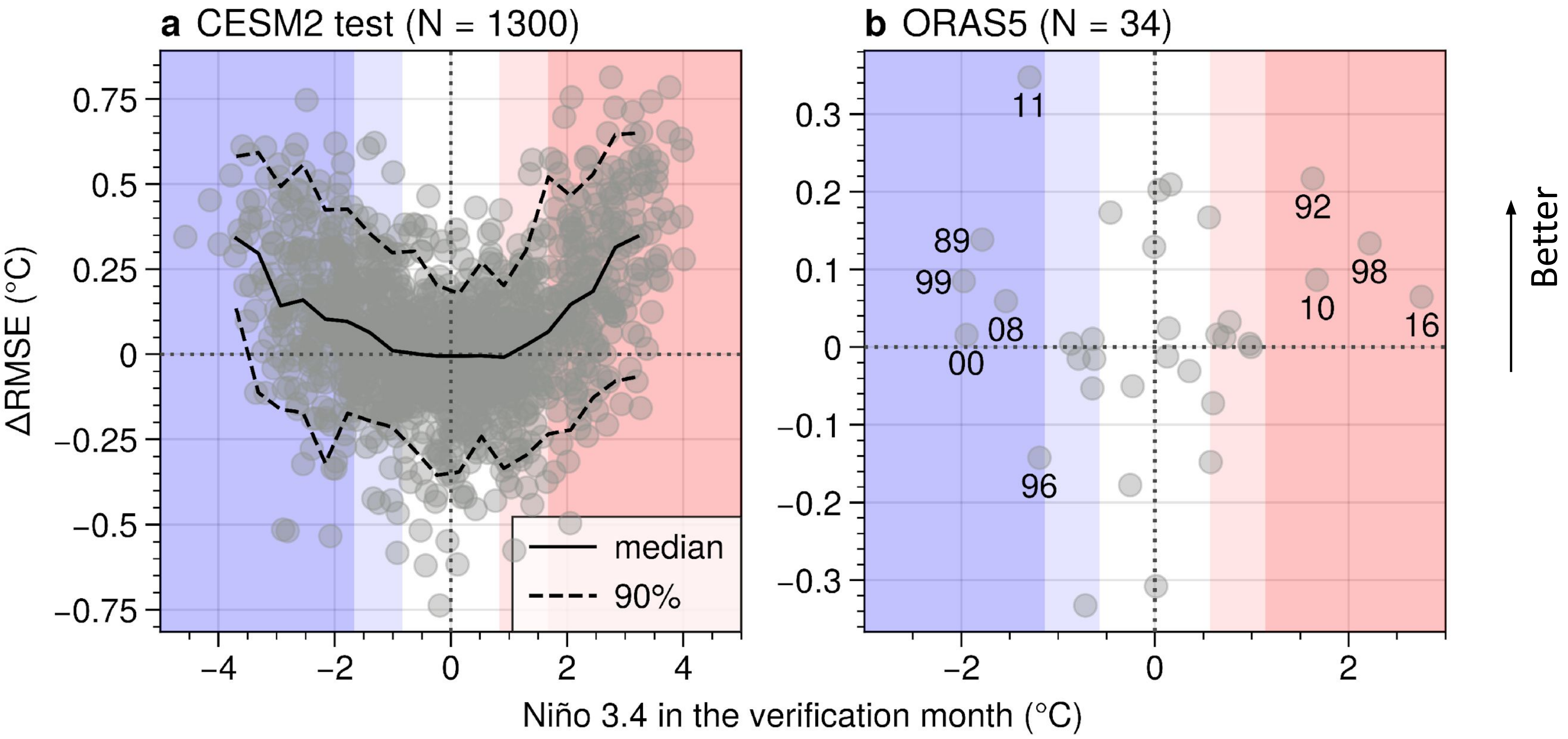
Once analogs are identified, forecasting can be extend to any field available.



Improvements

Better improvements for extreme events

Application to a reanalysis

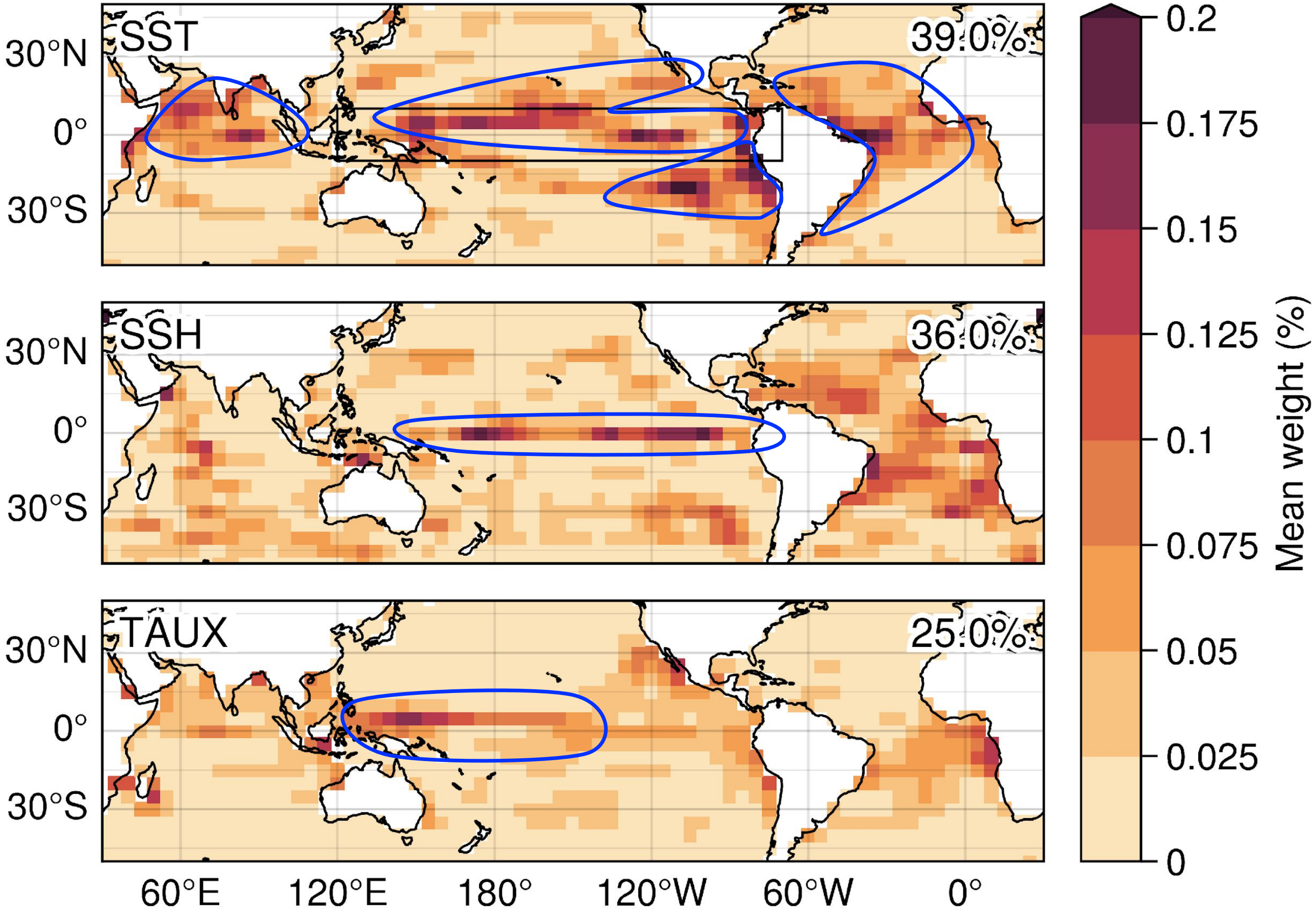


Light background: > 0.5 σ
Dark background: > 1 σ

Analyzing “sensitive regions”

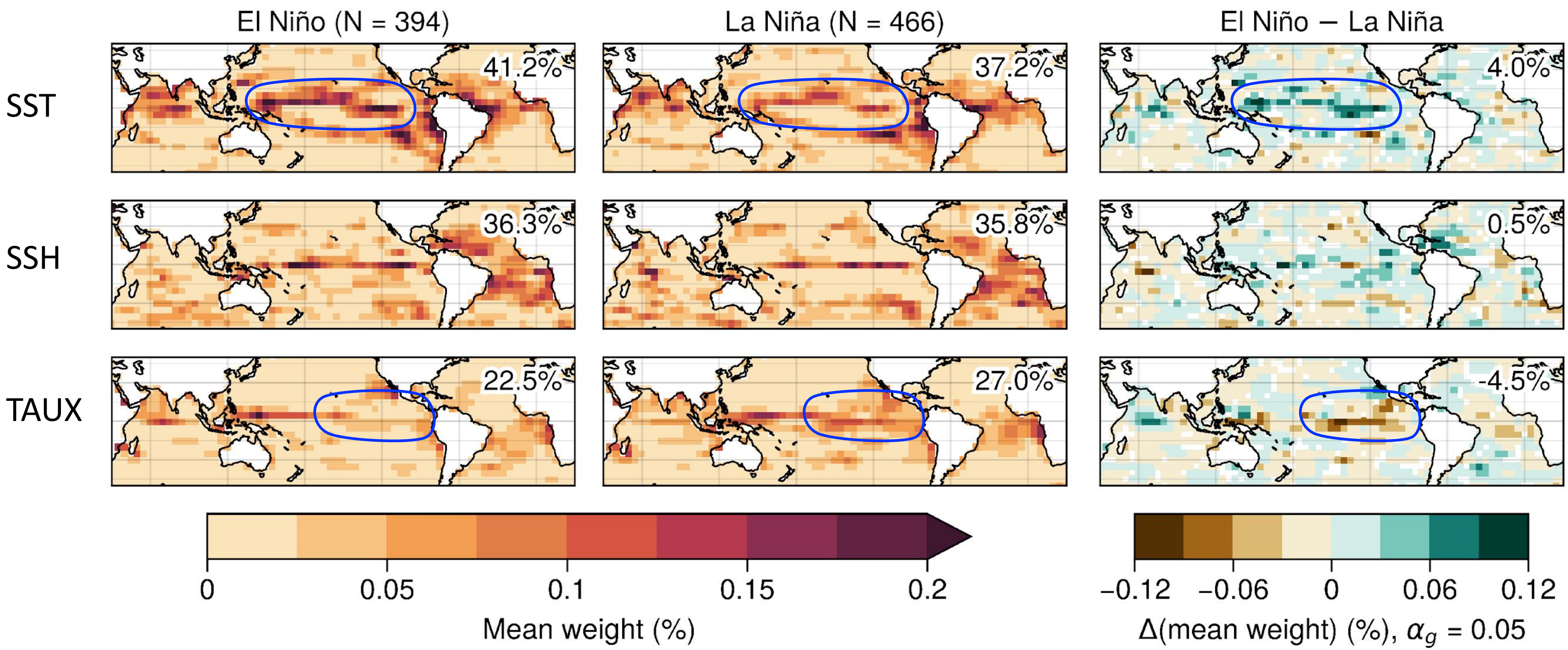
“Sensitive regions” are linked to various physical processes

Mean weights of all events (n = 1300), January



- SST ~ SSH > TAUX
- SST: Off-equatorial weights
  Pacific meridional modes
- SSH: Thermocline slope
  Recharge-discharge state
- TAUX: Westerly wind event

Asymmetry in El Niño and La Niña forecasts



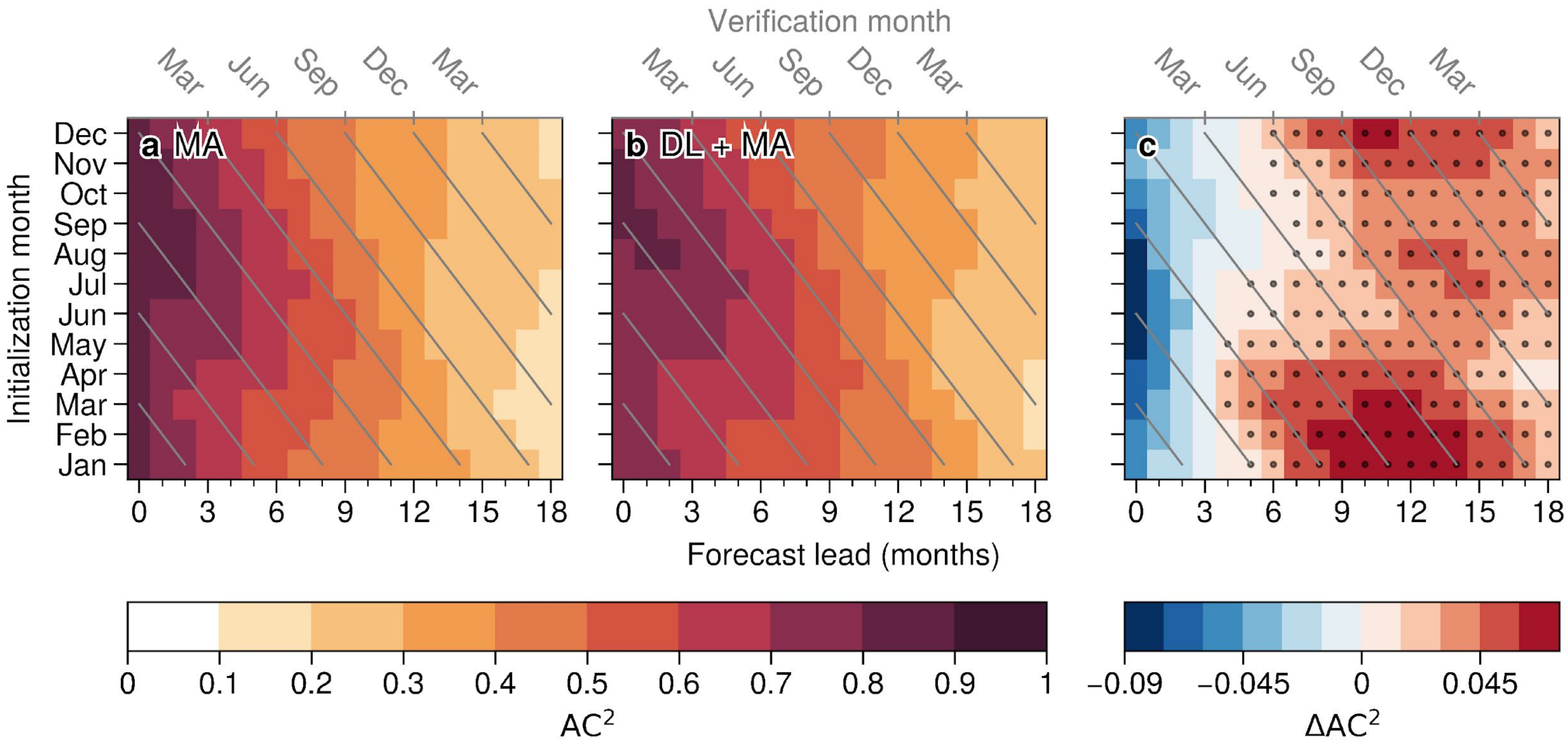
Conclusions

- Deep learning improves analog ENSO forecasting by 10%.
  - Better improvements for forecasting extreme events.
- This approach reveals state-dependent sensitivity linked to the Pacific Meridional Modes, equatorial recharge oscillator, and westerly wind bursts.
  - For El Niño forecast: Pacific SST is more sensitive.
  - For La Niña forecast: Pacific wind stress is more sensitive.
- Broad implications for forecasting diverse climate phenomena.

Toride, K., M. Newman, A. Hoell, A. Capotondi, J. Schlör, D. Amaya, Using Deep Learning to Identify Initial Error Sensitivity for Interpretable ENSO Forecasts, https://arxiv.org/abs/2404.15419

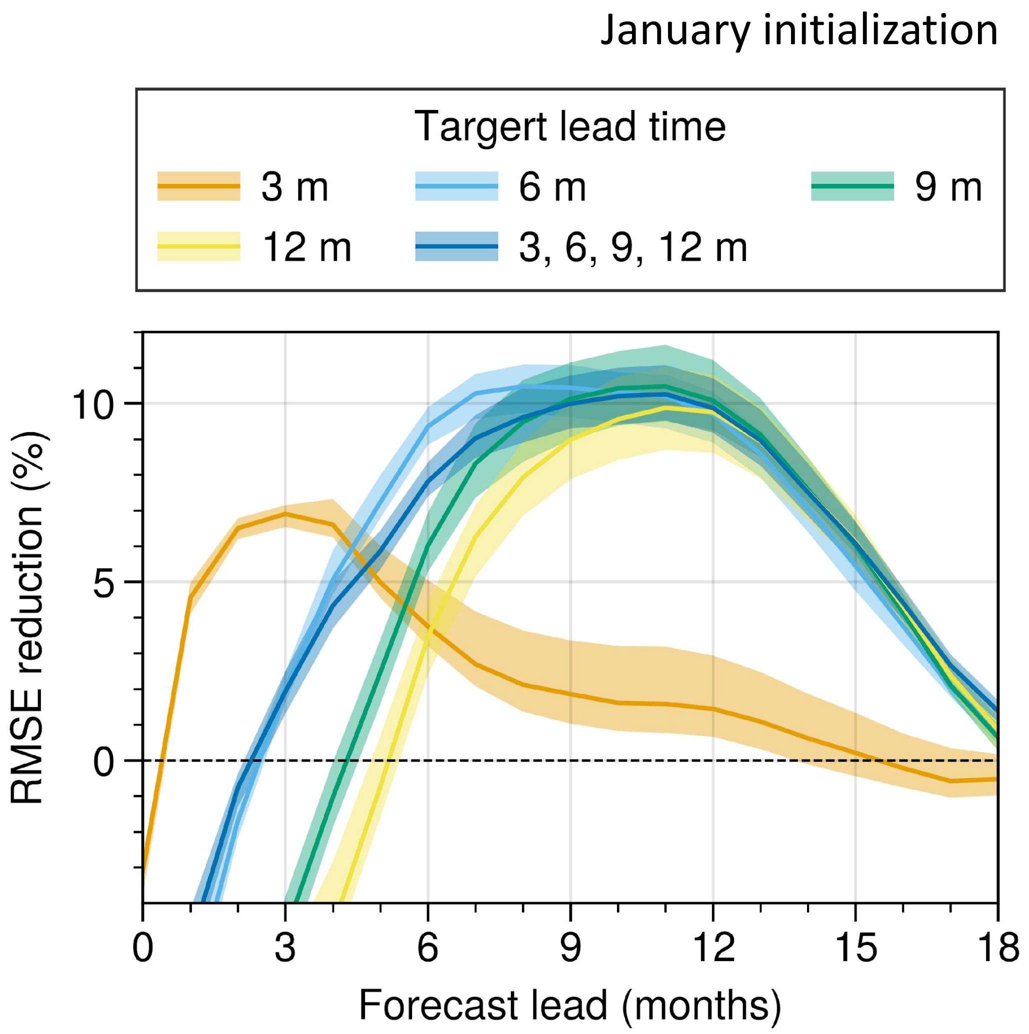
Contact: Kinya Toride (kinya.toride@noaa.gov)

Seasonal skills



Significantly improves 6-18 months forecasts

Do we need to train the model for each lead time?



The averaged distance of 3-, 6-, 9-, 12-months ≈ 6 months lead

→ Train for trajectories that closely shadow the reference, rather than snapshot-like analogs.

Continuous Ranked Probability Score

MSE of the predicted CDF and the true CDF

$$\mathrm{CRPS} = \int_{-\infty}^{\infty} [F(y) - F(y_o)]^2 dy$$

= Integral of the Brier Score over all possible threshold values

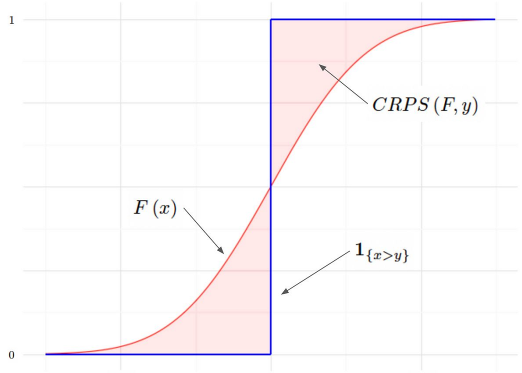
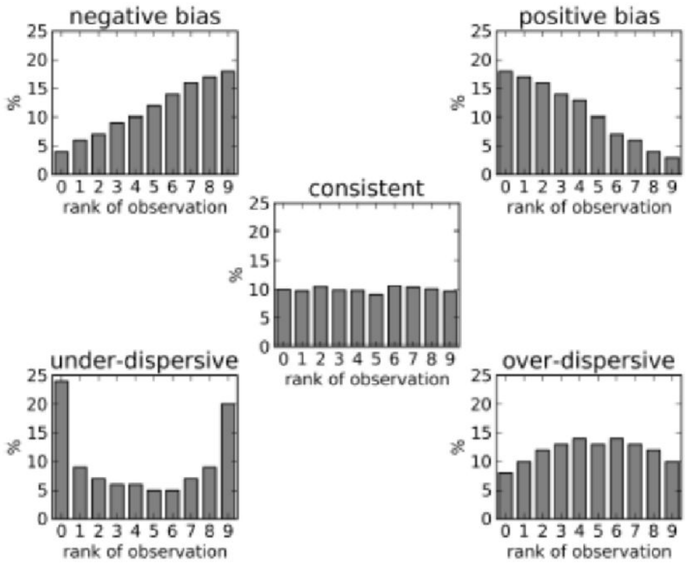


Image credit to Itamar Faran

<u>Decomposition of CRPS (Hershbach 2000)</u>

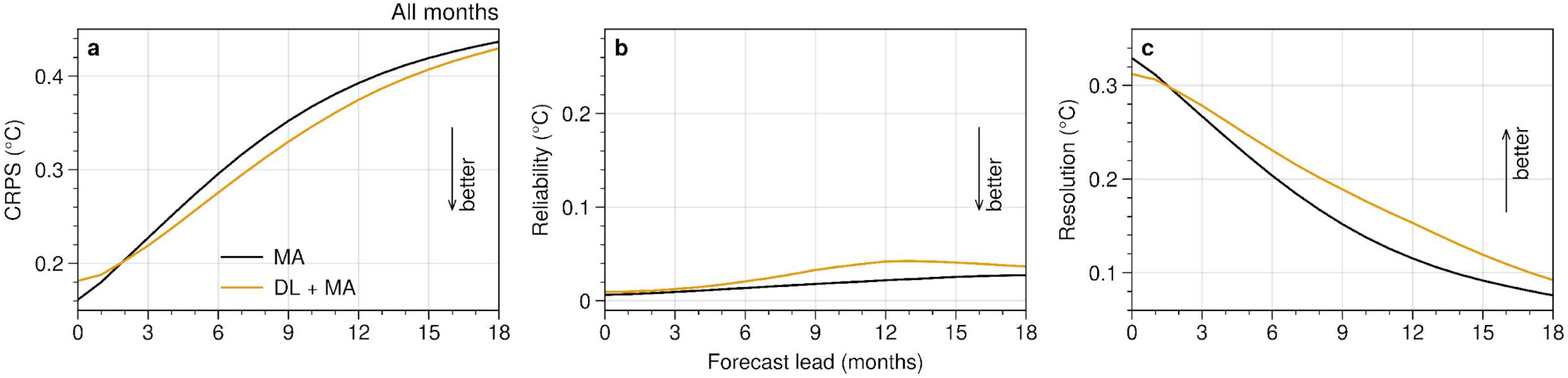
$$\mathrm{CRPS} = \textcolor{red}{\mathrm{Reliability}} - \textcolor{blue}{\mathrm{Resolution}} + \mathrm{Uncertainty}$$

Reliability = flatness of the rank histogram
Resolution = sharpness

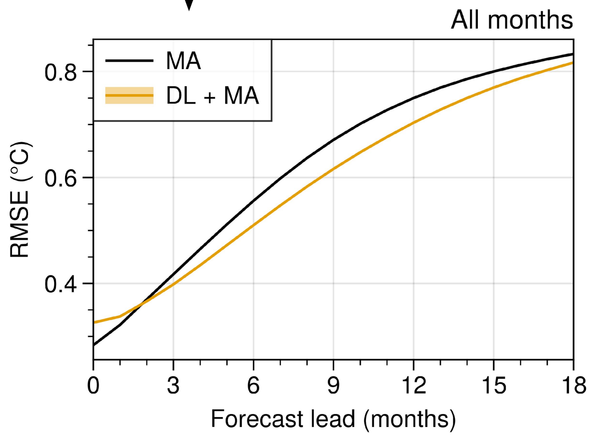


Schroedter-Homscheidt et al. (2016)

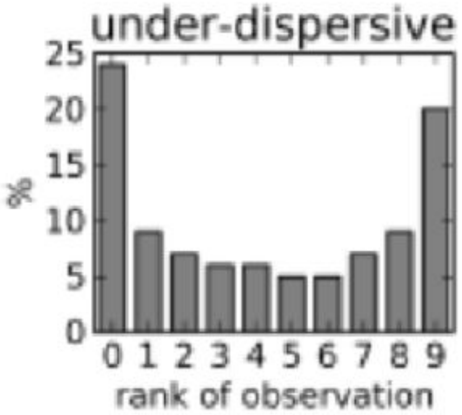
Overall probabilistic skill improves, but reliability worsens



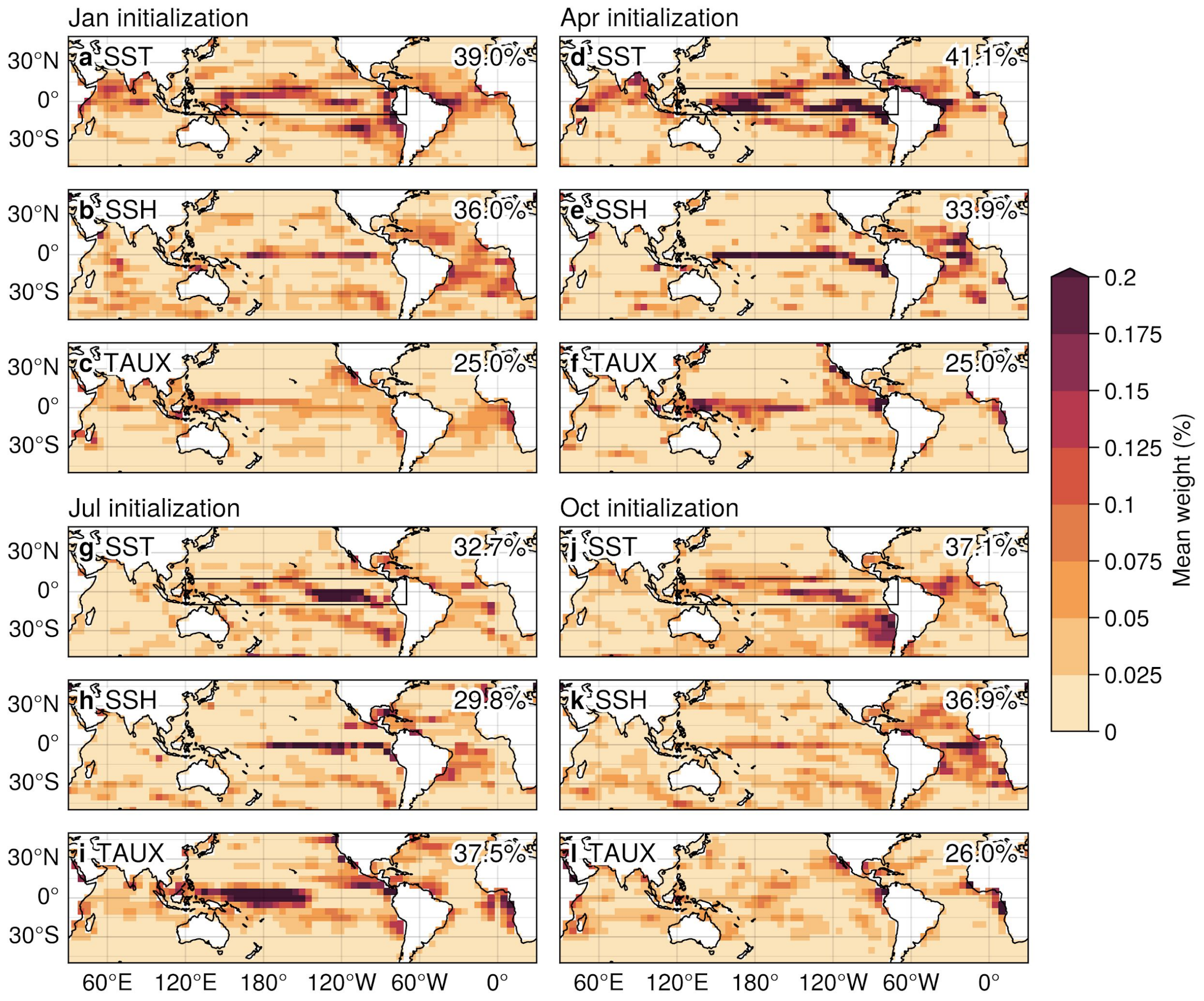
Similar to the ensemble mean skill

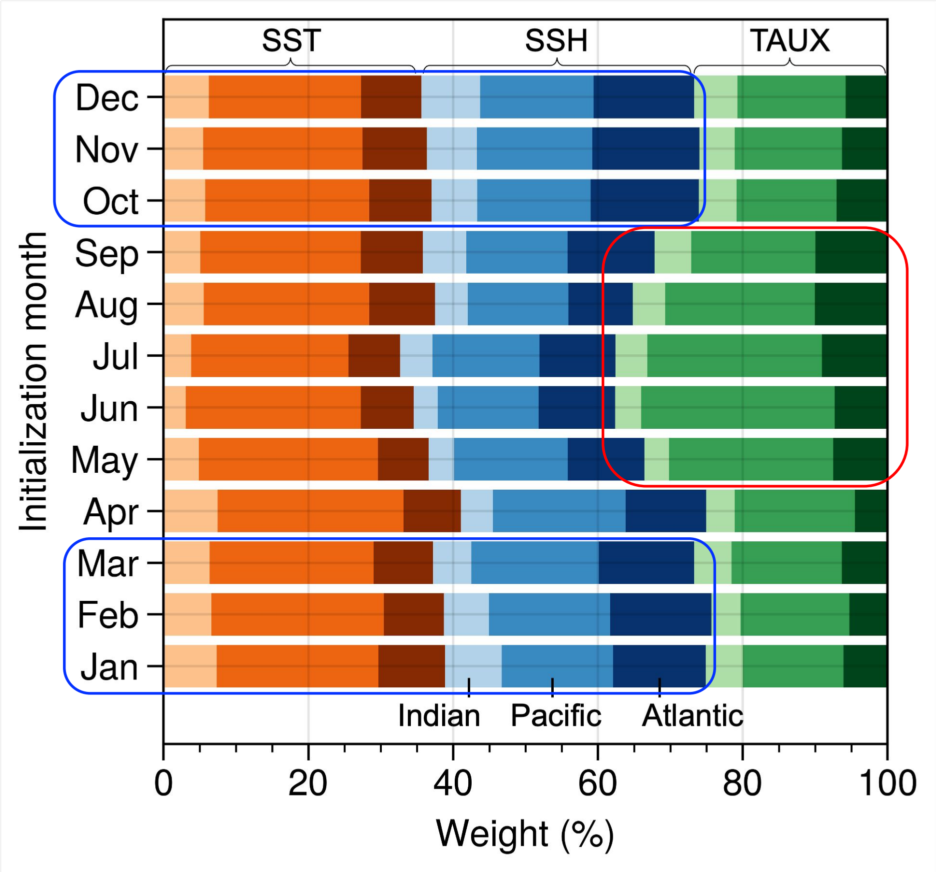


Ensembles get sharper and become under-dispersive.

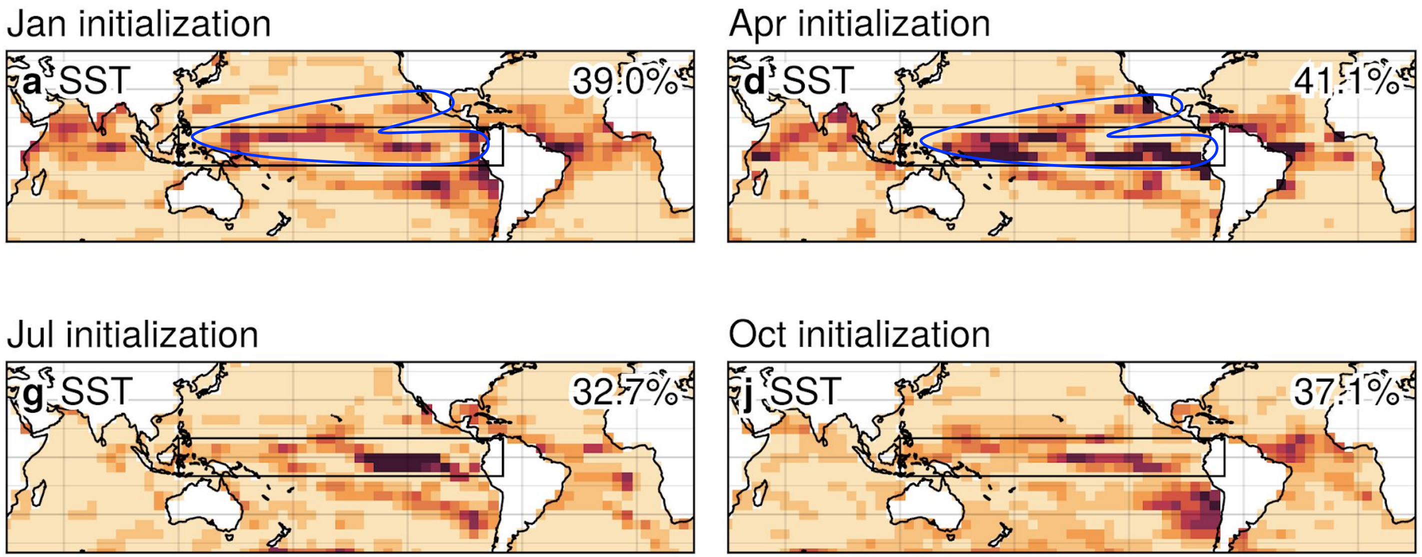


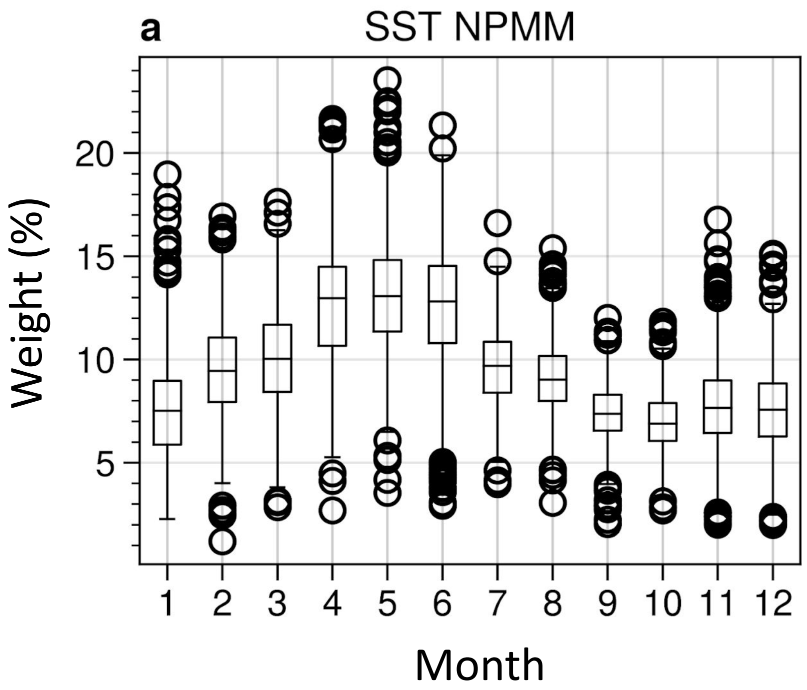
Seasonal variation of the weights



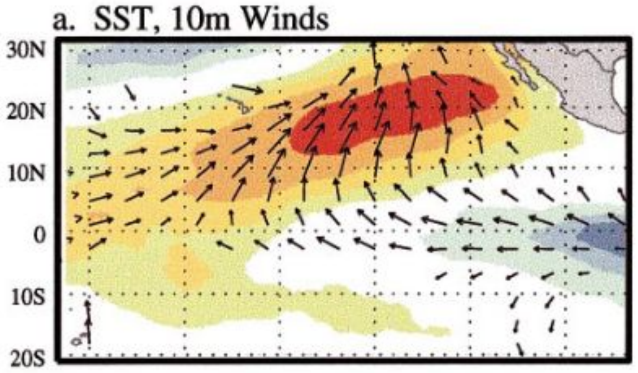


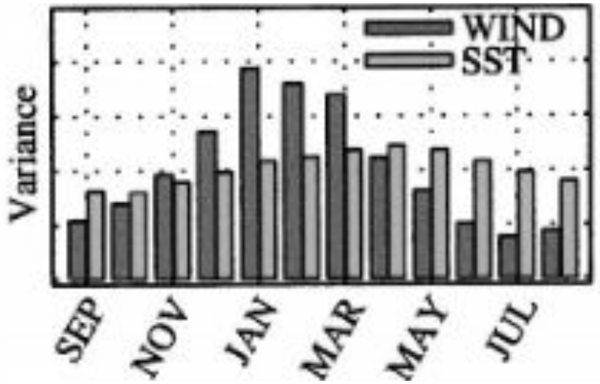
Seasonal variation of the weights



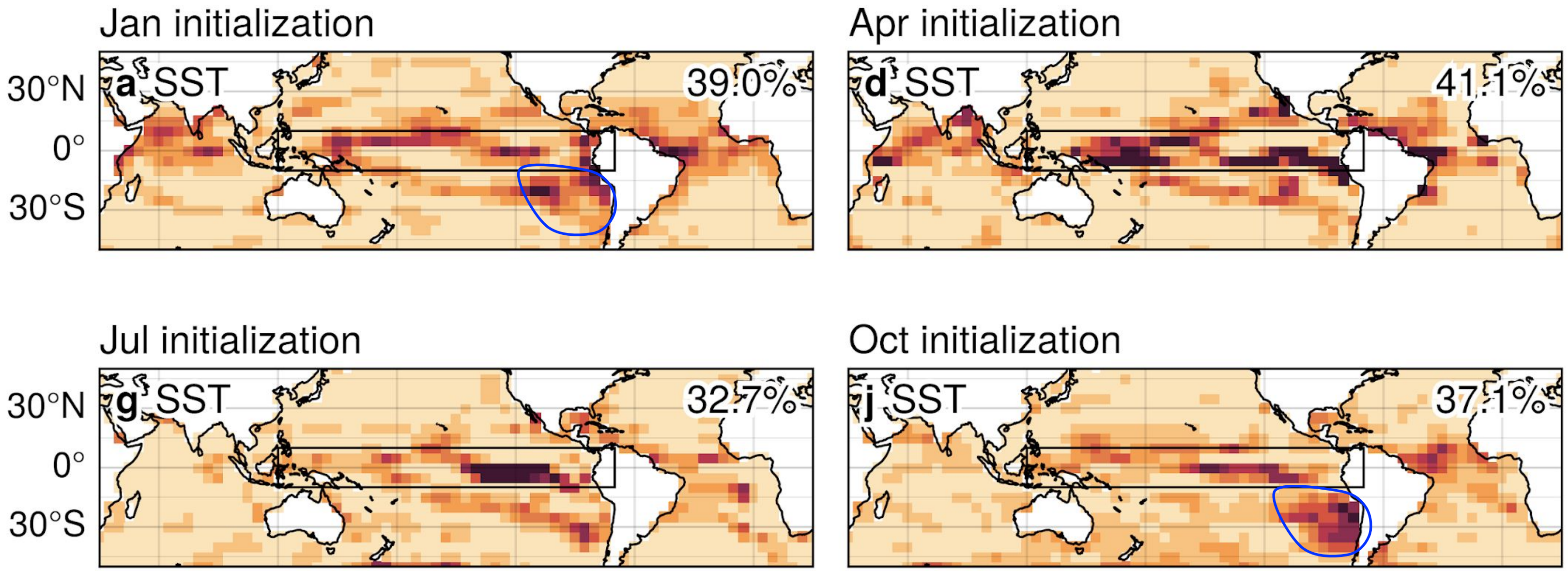


The North Pacific Meridional Mode (NPMM) typically peaks in boreal spring. (Chiang and Vimont 2004)

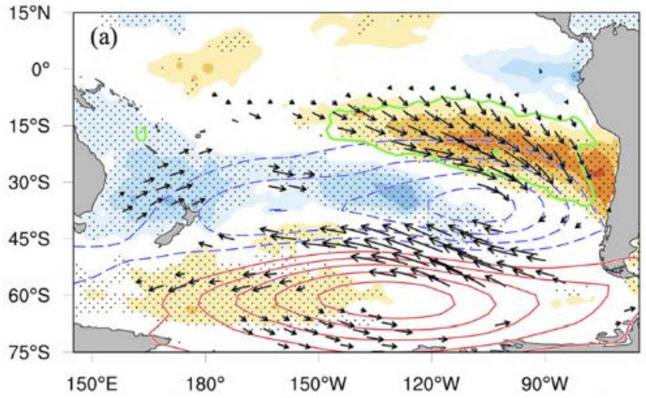


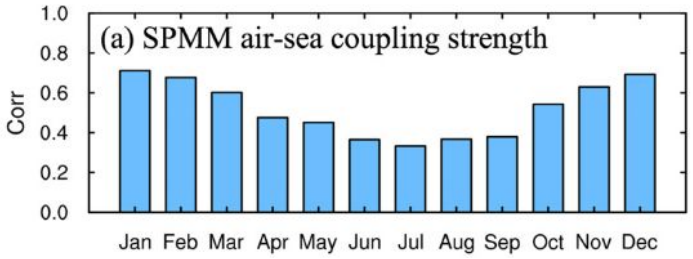


Seasonal variation of the weights

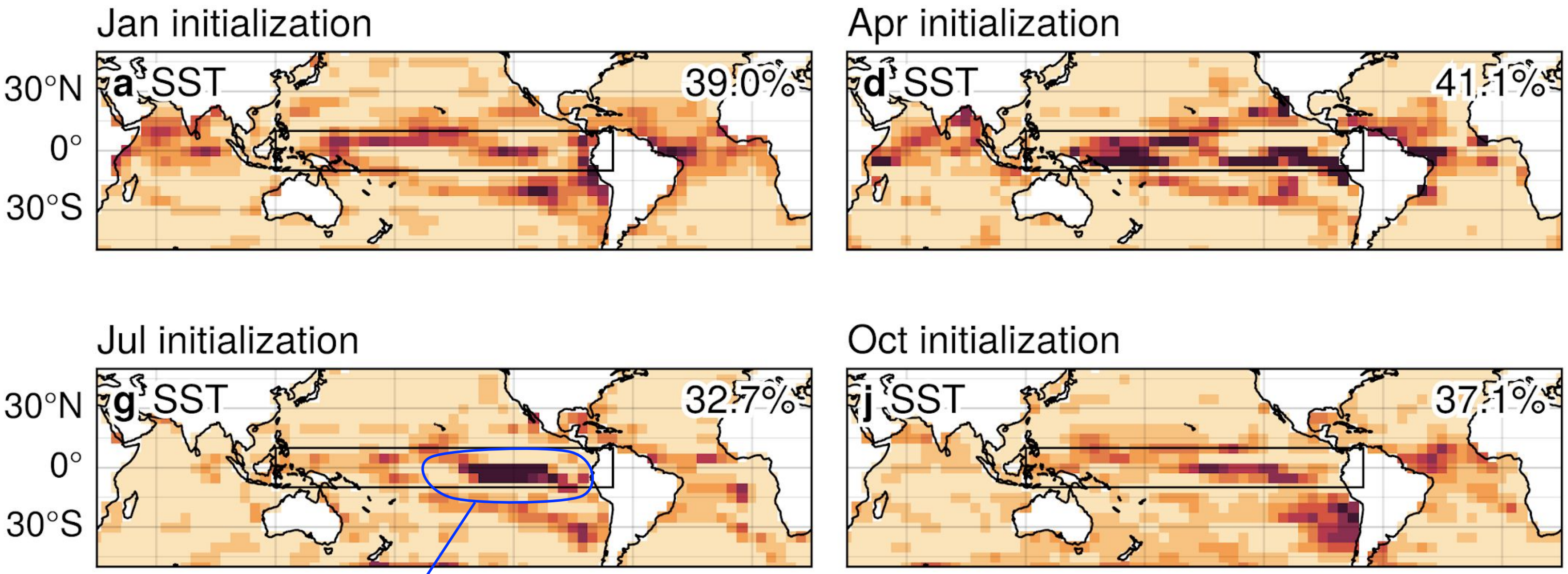


The South Pacific Meridional Mode (SPMM) typically peaks in boreal winter.
(You and Furtado 2018)





Seasonal variation of the weights

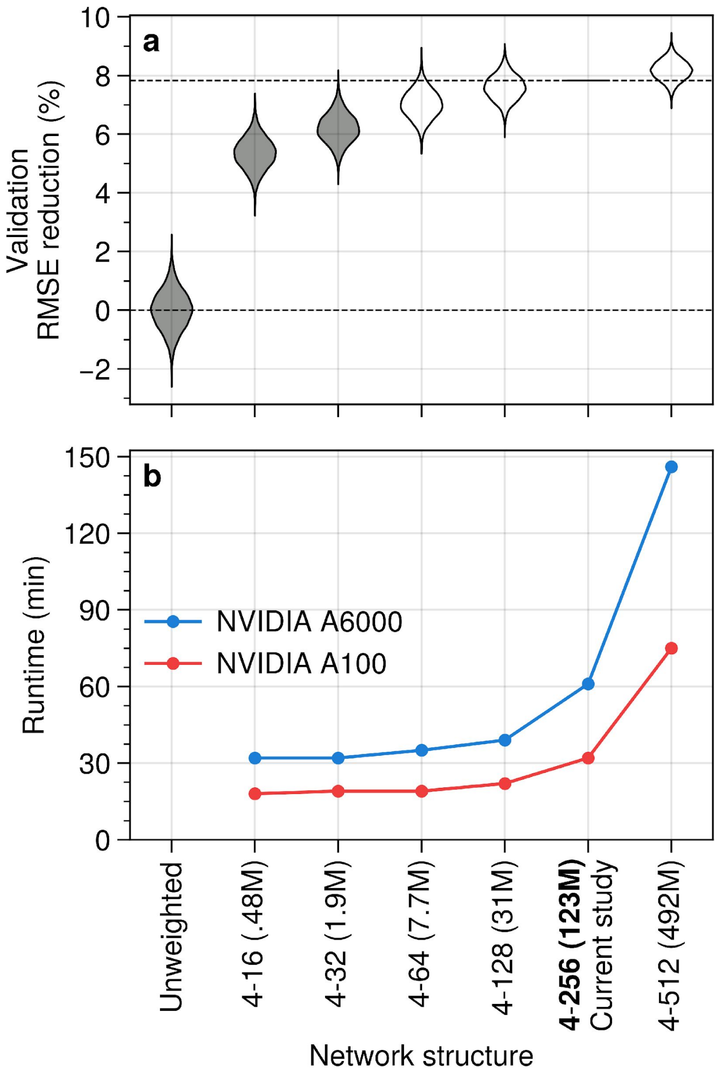


Concentrated over the eastern equatorial Pacific

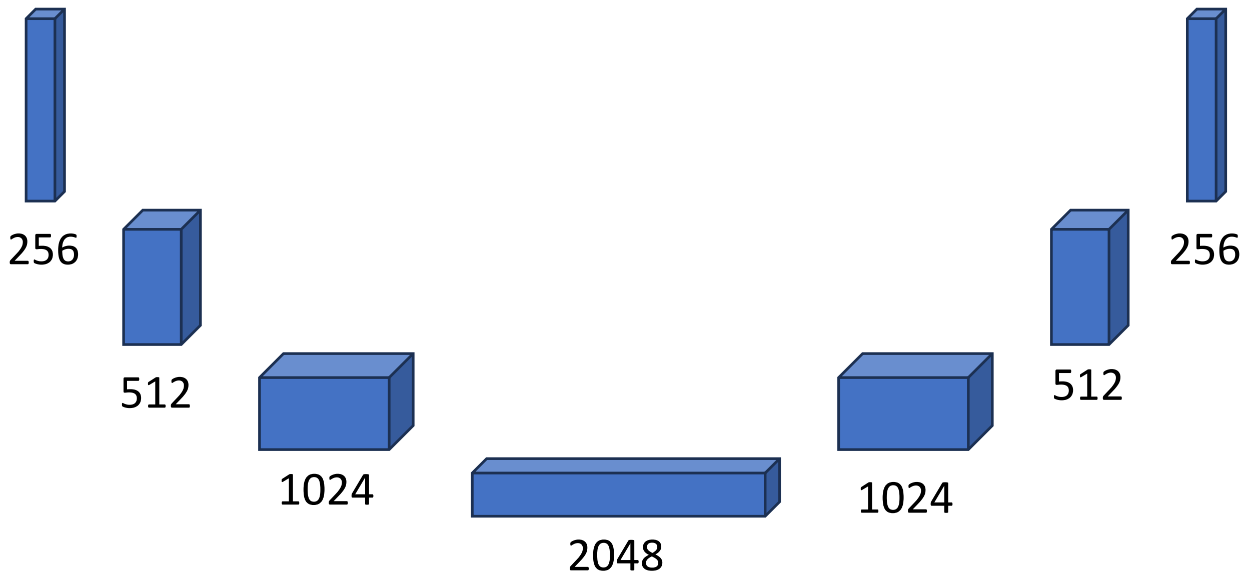
 ENSO events in boreal winter and their influences on subsequent seasons

Note: These weights improves 6-18 months forecasts

Network size & training time



UNet: 4-layer with initial channel size of 256



Channel = number of kernels used in convolution

e.g.) a color image has 3 channels (RGB)