UNIVERSITY OF WISCONSIN
MILWAUKEE

A Higher-Order Linear Inverse Model Diagnostic Across Matched-Resolution Simulations

Raphaël Hébert[1], Sergey Kravtsov[1]

[1]University of Wisconsin–Milwaukee

Broadening Earth System Research Across Scales with High-Resolution Modeling
Sept. 15th, 2026

hbert@uwm.edu, raphael.hebert@awi.de, rphlhbrt@gmail.com

1. Motivation, Method & Data

Background: the classical LIM assumes AR(1), Markovian dynamics from the current state alone. Higher-order effectively augments the state space, letting the model draw on correlation structure from unobserved variables embedded in the observations' own history. A dense AR(p) for r EOFs, though, needs $p \times r^2$ parameters: regularization needed. We separate diagonal (each mode) and off-diagonal (mode coupling) entries: truncating the off-diagonal saves the most parameters for the least real skill lost. The result is a parsimonious diagnostic of each dataset's own inherent (linear) stochastic predictability.

Motivating question: does horizontal resolution genuinely change the *stochastic predictability* of daily near-surface temperature, or only its spatial detail?

Hypothesis: Why would resolution change predictability at all? Several distinct, partly-opposing mechanisms could produce a resolution-dependent signal here, independent of whether the *true* predictability actually differs:

- ↓ **More resolved noise:** finer grids resolve genuine small-scale/mesoscale chaotic variability that may be less predictable.
- ↑ **Newly-resolved slow signal:** some slow, large-scale predictable processes may only emerge once the small-scale process generating them is resolved.
- ↑ **Spurious oscillations:** a resolution-specific numerical artifact (grid-imprinting, aliased short waves) can masquerade as a coherent, predictable oscillation with no physical counterpart.

Method: Mixed-order AR(p) LIM in EOF space. We fit a Linear Inverse Model on EOF-truncated daily anomalies:

$$\mathbf{z}(t) = \sum_{l=1}^{p} \mathbf{A}_l \mathbf{z}(t-l) + \boldsymbol{\xi}(t),$$

where $\mathbf{z}(t) \in \mathbb{R}^r$ are EOF coefficients and $\boldsymbol{\xi}$ is white noise. Increasing p captures longer memory, but adds r^2 free parameters per lag level, regularized here by a **mixed-order structure**, motivated by the physical separation of timescales between self- and cross-mode coupling:

Entry	Physical role	Memory depth
Diagonal $[\mathbf{A}_l]_{ii}$	persistence, oscillations	p_{diag} lags *(long)*
Off-diagonal $[\mathbf{A}_l]_{ij}$	mode interactions, teleconnections	p_{off} lags *(short)*

Design: the identical LIM setup is applied to MERRA-2 first for validation, then to each HighResMIP resolution pair unchanged. Since resolution is the only variable that differs within a pair, any difference in diagnosed skill, optimal EOF truncation, or spectral structure can be attributed to resolution itself.

Data. MERRA-2 (1980–2021) validates the LIM before applying it, unchanged, to three matched resolution pairs: CNRM-CM6-1/-HR, MPI-ESM1-2-HR/XR, CESM-LR/-HR..

2. Model & Reanalysis Configurations

Dataset	Atmosphere	Ocean	Ref.
MERRA-2	GEOS-5, 0.5°×0.625° (∼50 km), 72 levels	Atmosphere-only reanalysis	G17
CNRM-CM6-1	ARPEGE 6.3, T127 (∼150 km), 91 levels	NEMO 3.6, eORCA1 (∼100 km), 75 levels	V19
CNRM-CM6-1-HR	ARPEGE 6.3, T359 (∼50 km), 91 levels	NEMO 3.6, eORCA025 (∼25 km), 75 levels	V19
MPI-ESM1-2-HR	ECHAM6.3, T127 (∼100 km), 95 levels	MPIOM1.6, TP04 (∼40 km), 40 levels	M18
MPI-ESM1-2-XR	ECHAM6.3, T255 (∼50 km), 95 levels	MPIOM1.6, TP04 (∼40 km), 40 levels	G19
CESM-LR	CAM5-SE, ne30 (∼100 km), 30 levels	POP2, gx1v6 (∼100 km)	C20
CESM-HR	CAM5-SE, ne120 (∼25 km), 30 levels	POP2, (∼0.1°, ∼10 km)	C20

G17: Gelaro et al. (2017), *J. Climate*, 30(14), 5419–5454, doi:10.1175/JCLI-D-16-0758.1
V19: Voldoire et al. (2019), *JAMES*, 11(7), 2177–2213, doi:10.1029/2019MS001683
M18: Müller et al. (2018), *JAMES*, 10(7), 1383–1413, doi:10.1029/2017MS001217
G19: Gutjahr et al. (2019), *GMD*, 12, 3241–3281, doi:10.5194/gmd-12-3241-2019
C20: Chang et al. (2020), *JAMES*, 12(12), e2020MS002298, doi:10.1029/2020MS002298

Note: MPI-ESM1-2-HR/XR use the same ocean grid, only atmosphere resolution differs.

3. EOF Truncation Sensitivity

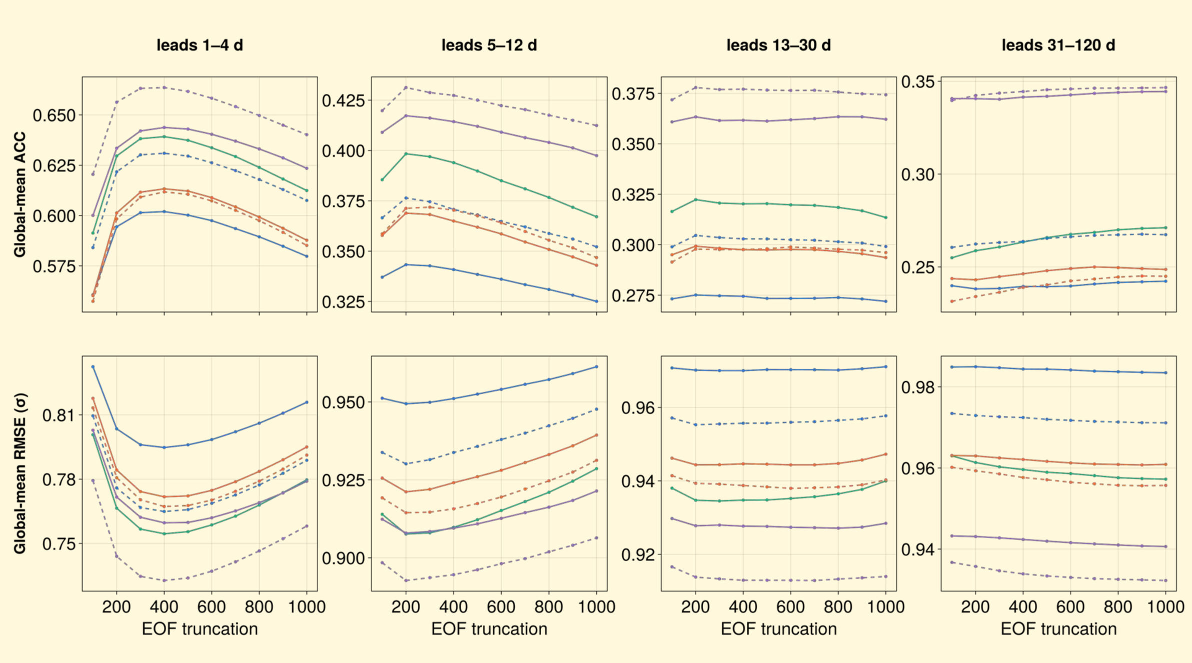

Figure: Global-mean ACC (top) and normalized RMSE (bottom) vs. EOF truncation ($r = 100$–1000), $p24/2$ configuration, across four lead windows. Short-lead skill peaks near $r \approx 300$–400 then declines; the longest-lead window (31–120 d) instead keeps rising out to $r = 1000$. Our $r = 500$ choice sits past the short-lead peak but before that long-lead tail.

4. Mixed-Order Configuration Sweep (500 EOFs)

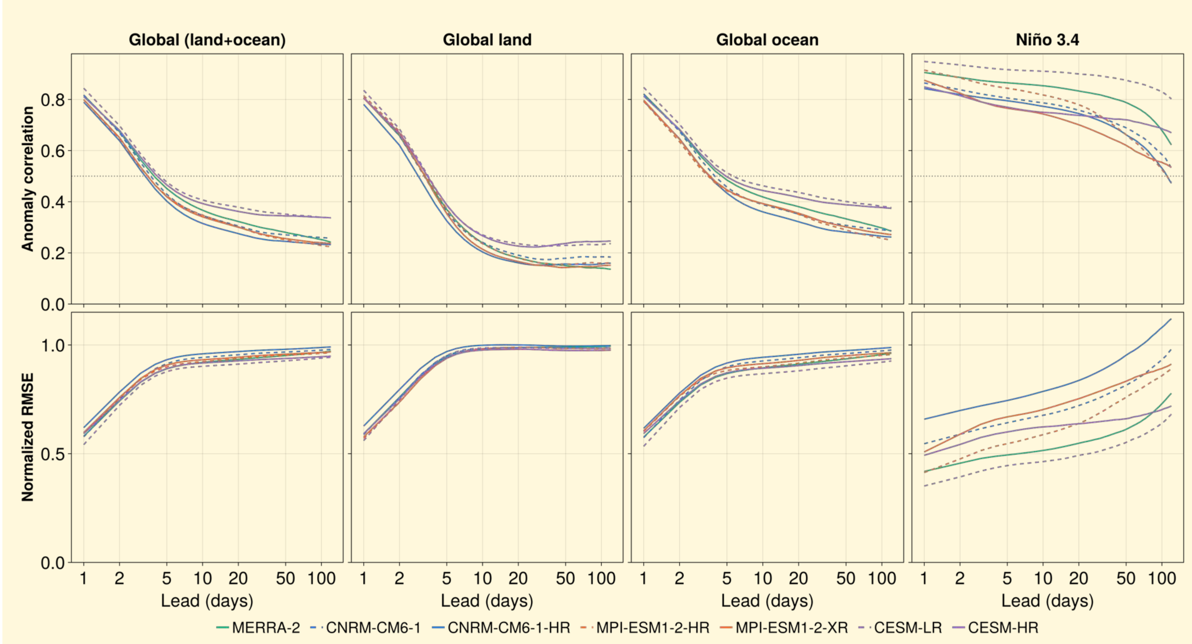

Figure: LIM forecast skill (top: anomaly correlation; bottom: normalized RMSE) vs. lead time for the selected $p24/2$ configuration (500 EOFs), across four regions.

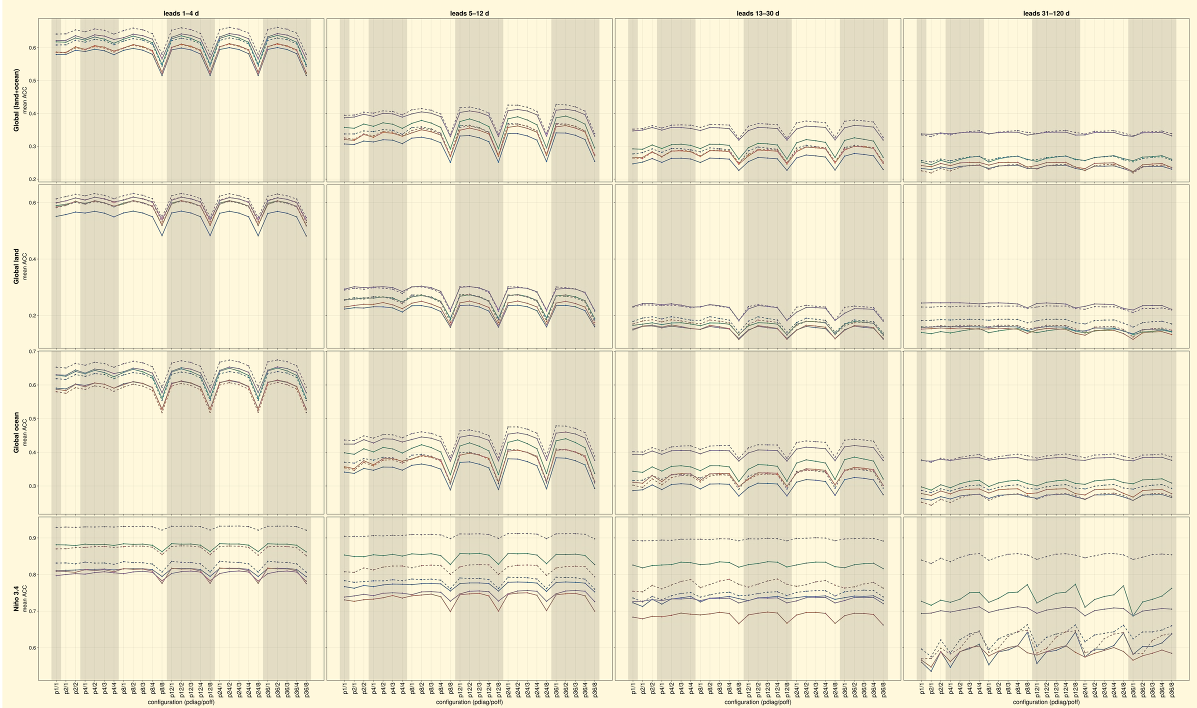

Figure: Mean ACC vs. mixed-order configuration, across regions and lead windows (500 EOFs).

5. Selecting a Configuration: $p24/2$

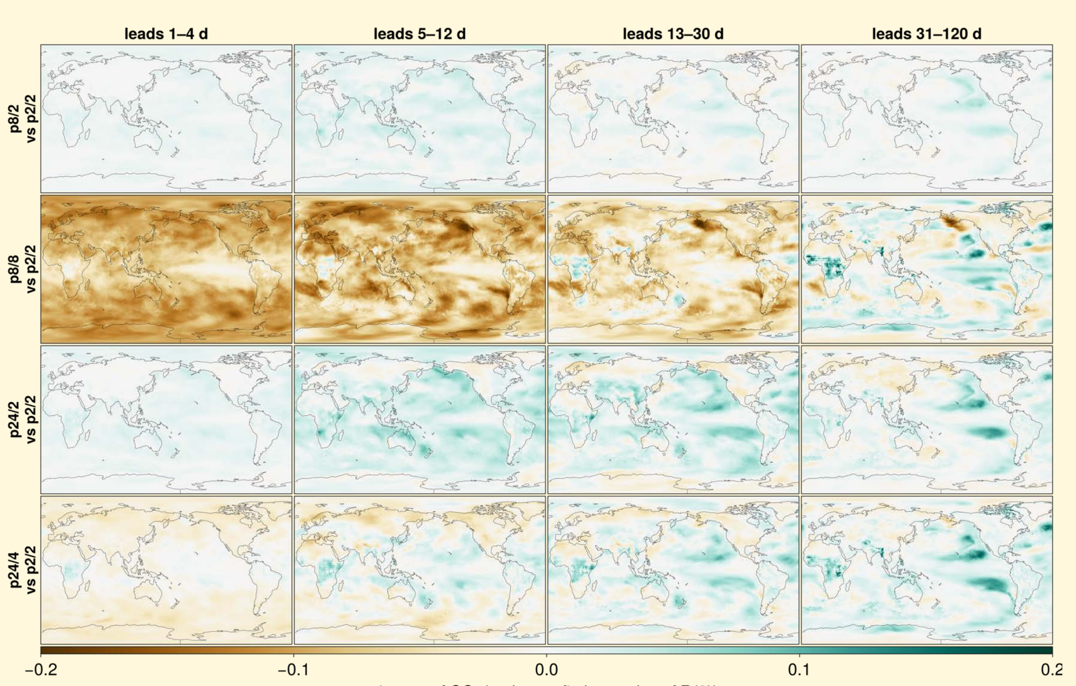

Figure: MERRA-2: ACC change relative to the dense $p2/2$ baseline for four alternative configurations.

9. HR − LR Teleconnection Strength: Europe & Niño 3.4

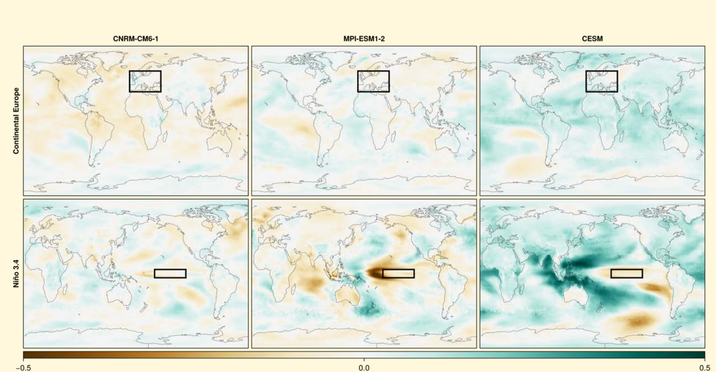

Figure: Change in correlation with the reference regions (HR − LR), Europe (top) and Niño 3.4 (bottom). Teal = High-Res ↑ correlated; Brown = High-Res ↓ correlated.

10. Conclusion

- **Overall:** models are less predictable than MERRA-2 *except* the CESM pair.
- **ENSO (Niño 3.4):** only CESM-LR exceeds MERRA-2's predictability; CESM-HR's ENSO predictability is *lower* than CESM-LR's, spectral bump has lower frequency and a reduced spatial signature.
- **Labrador Sea:** CNRM-CM6-1 (Low-Res) shows high low-frequency scaling and large predictability, higher than MERRA-2; CESM-HR becomes more predictable in line with MERRA2; MPI-ESM1-2's High-Res is more variable at low-frequency but not more predictable.
- **Norwegian Sea:** All models show a large predictability gain with higher resolution, but Low-Res are closer to MERRA2. CNRM's HR has less variance below 5-year timescales; MPI's HR gains variance at all timescales, CESM's HR lose variance at all timescales, very similar to MERRA2 until 10-year.
- **North Pacific:** Higher-resolution systematically decrease predictability in all models, and low-frequency variance, becoming more similar to MERRA2.
- **CESM stands apart:** it is the only pair whose High-Res reaches mesoscale-eddy-resolving ocean resolution (∼10 km), and predictability increases almost everywhere over the ocean and over Continental Europe (through correlation with the oceans) at HR.

6. HighResMIP Resolution-Pair Comparison

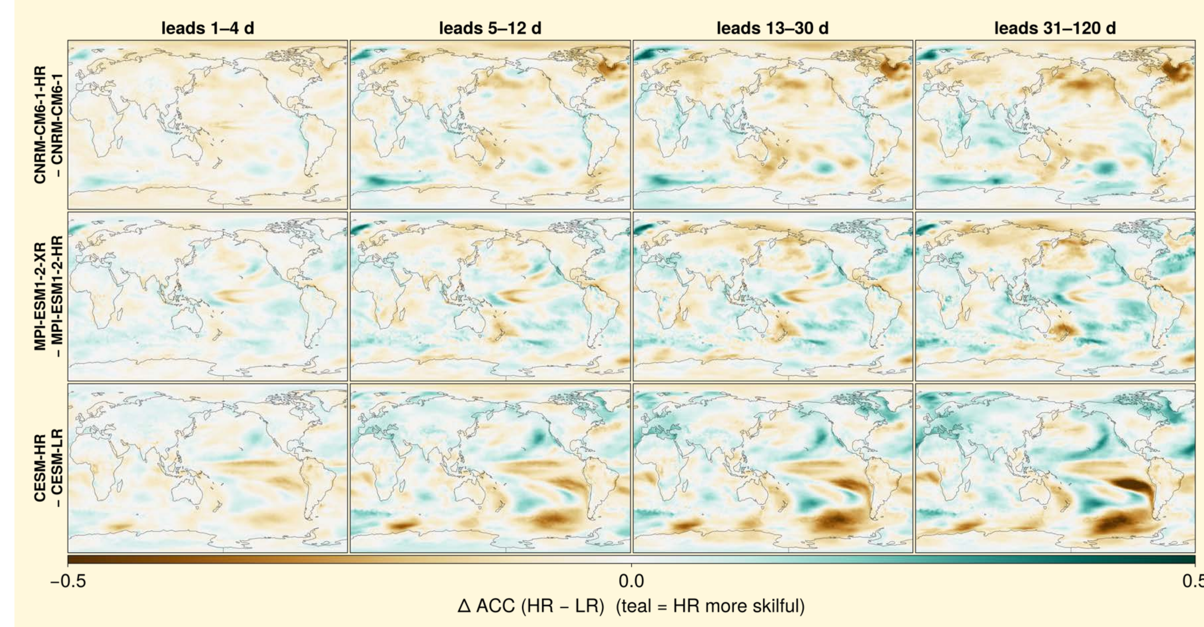

Figure: Resolution effect on skill: ACC(high-res) − ACC(low-res) for model pairs ($p24/2$, 500 EOFs). Teal = High-Res ↑ predictable; Brown = High-Res ↓ predictable.

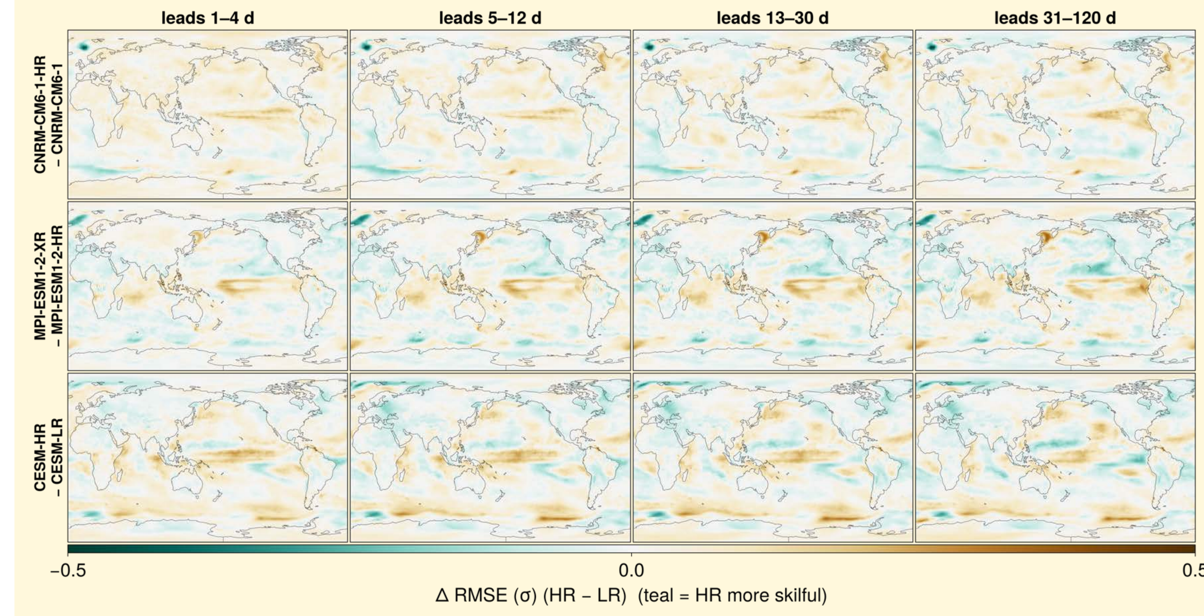

Figure: Same comparison for normalized RMSE (σ).

7. Spectral Structure: High-Res/Low-Res PSD Ratio

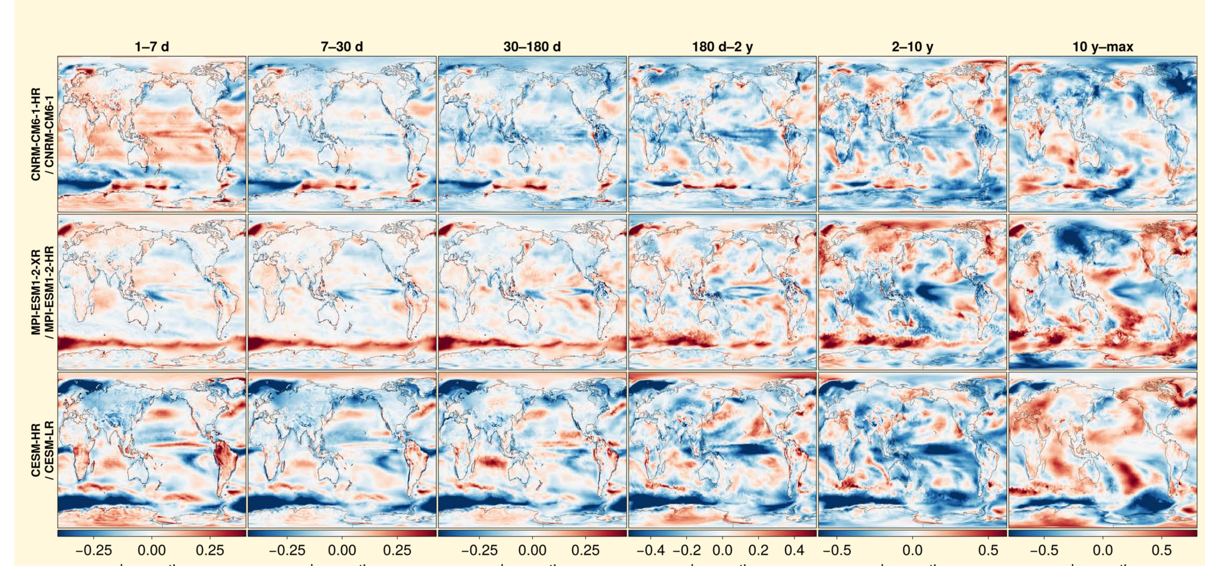

Figure: HR/LR ratio of band-averaged power spectral density, all three pairs, across six timescale bands (each band has its own color scale, shared across pairs). Red = High-Res ↑ variance; blue = High-Res ↓ variance.

8. Regional Spectra & LIM Forecast Skill

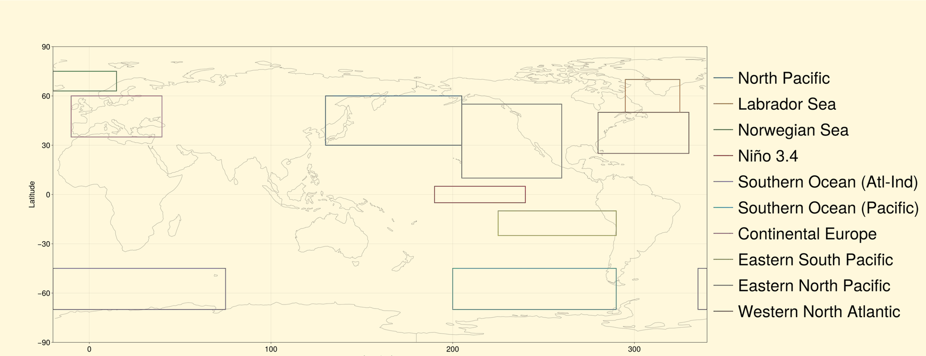

Figure: Regions used in the spectra/skill figure below.

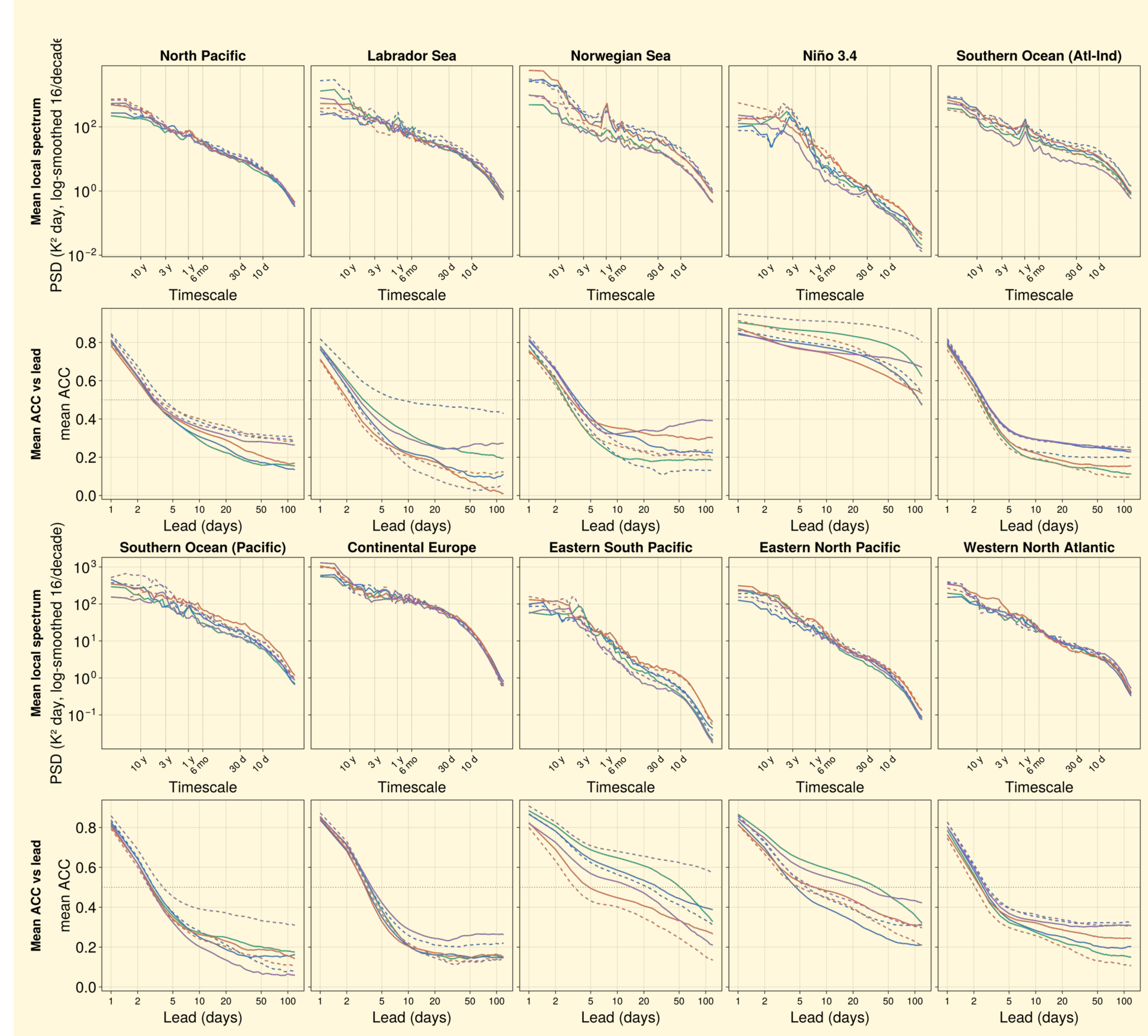

Figure: Mean annual power spectrum (odd rows) and LIM ACC vs. lead (even rows), $p24/2$, 500 EOFs for regions appearing as most resolution-sensitive. Ocean-point means throughout except Continental Europe (land-point mean).

hbert@uwm.edu, raphael.hebert@awi.de, rphlhbrt@gmail.com