Royal Netherlands
Meteorological Institute
*Ministry of Infrastructure
and Water Management*

Regime-based analysis of ocean-atmosphere coupling using Self Organizing Maps: Feature importance and spatio-temporal variability

Stavroula Biri and Ad Stoffelen

March 2026

Introduction

- Overall Goal: Evaluate missing variance in NWP model wind fields; attribute these to geophysical processes using physics-informed diagnostics
- Focus: Ocean coupling with the atmosphere through the oceanic **thermal feedback** (TFB) and the **current feedback** (CFB).
- TFB mechanics: SST structures modify turbulence and wind in the marine atmospheric boundary layer (MABL), which in turn affects the surface wind stress patterns ("top-down").
- CFB mechanics: it slows down the mean ocean currents by reducing the energy input from the atmosphere to the ocean ("bottom-up").
- The two feedback mechanisms are closely related

Introduction SOM

- Unsupervised neural network for pattern recognition and clustering
- Map high-dimensional data to 2D grid while preserving topology
- Each neuron (node) represents a cluster of similar patterns
- Competitive learning through Best Matching Unit (BMU) selection
- No physics imposed; training comes from the input variables

Application

Associate ocean-atmosphere feedback mechanisms by partitioning the ocean-atmosphere system into distinct SOM patterns, where coupling relationships vary systematically

- 5x5 SOM grid (25 total Nodes)
- Trained on daily normalized variables for one year (2020)
- Sequential training (start with core variables and add complexity)
- Each day build on the previous day’s training
- Evaluated on daily data for 2021; geographic binning applied on monthly/seasonal scale on 0.5°×0.5° resolution

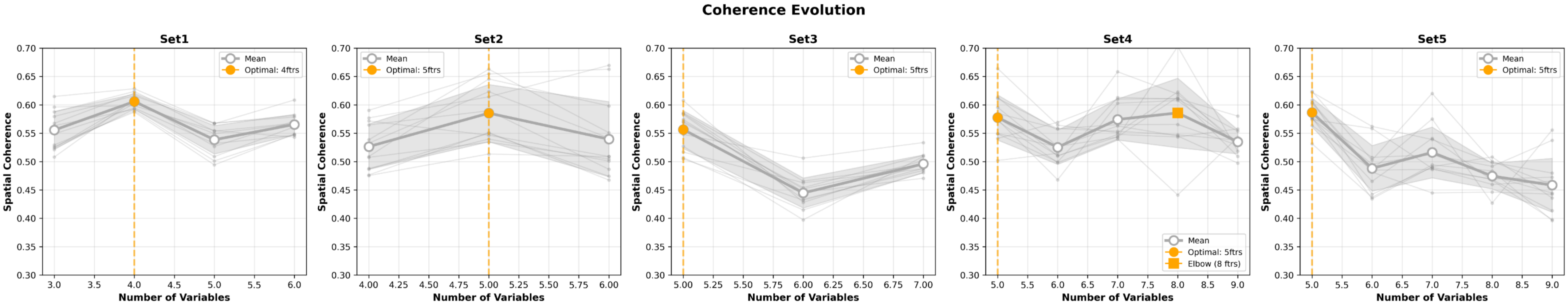
Set Definitions

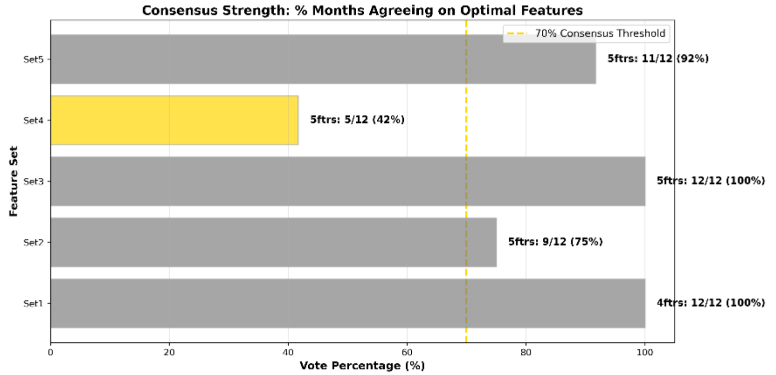
All Sets have u, v ASCAT-B level2

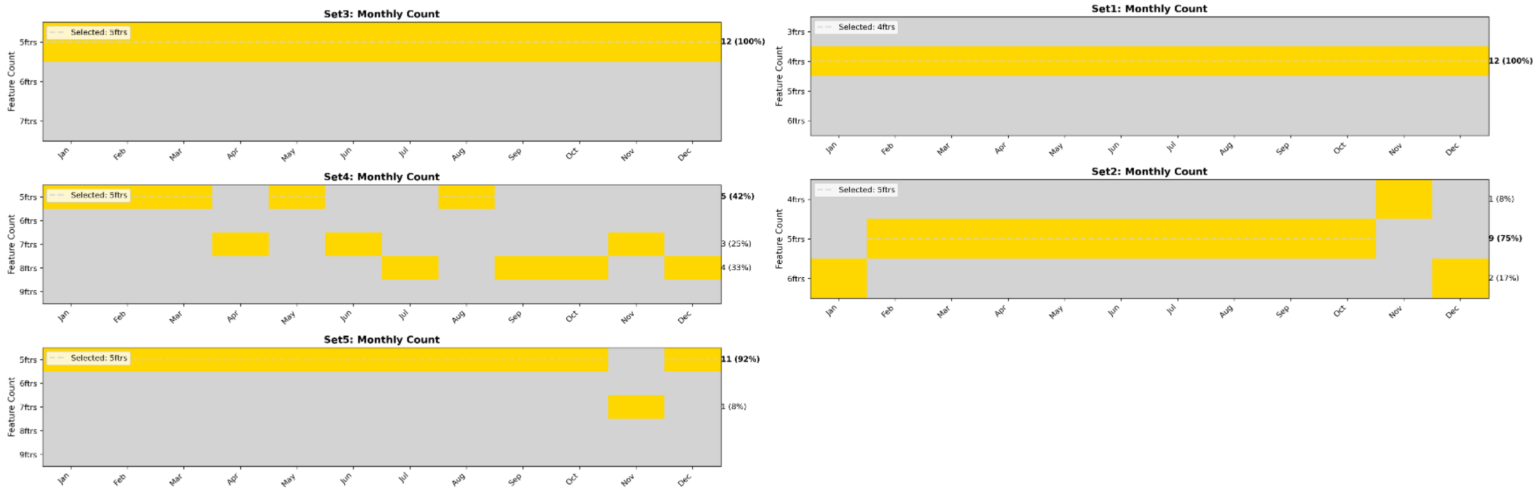
Set 1	Set2	Set 3	Set 4	Set 5
ζ	$\nabla\times\tau$	ζ	ζ	u_g
$\nabla\times\tau$	$\nabla SST\times\tau$	$\nabla\times\tau$	$\nabla_c SST$	z_o
∇U	$\nabla_c SST$	$\nabla_c SST$	$\nabla SST\times\tau$	Ri_b
$\nabla\times U$		τ	$\nabla\times\tau$	ζ
		$\nabla\times U$	z_o	$\nabla_c SST$
			u_g	$\nabla SST\times\tau$
			W_{ekman}	$\nabla\times\tau$

Once a variable is introduced in the training the sequence in which it was introduced does not matter

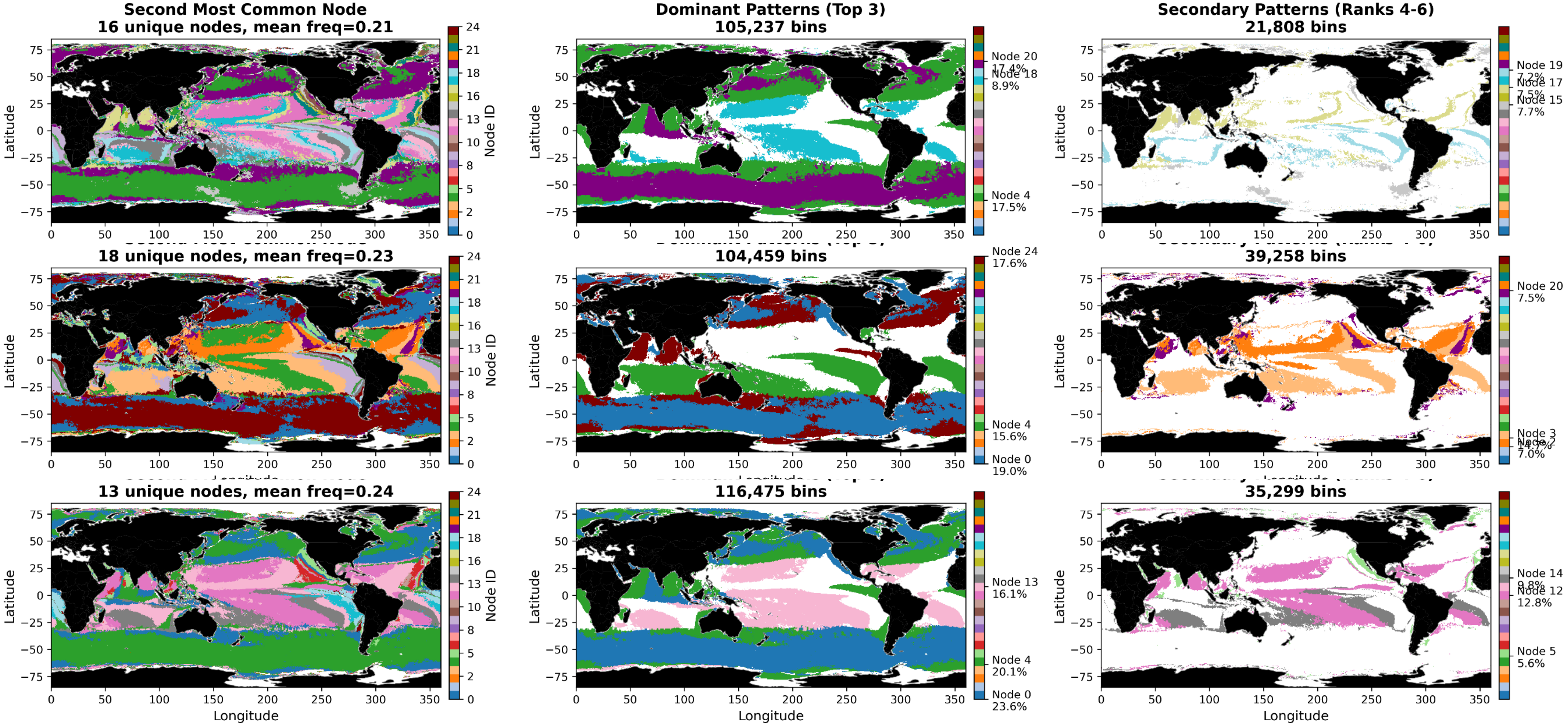
Identify Optimal #of variables



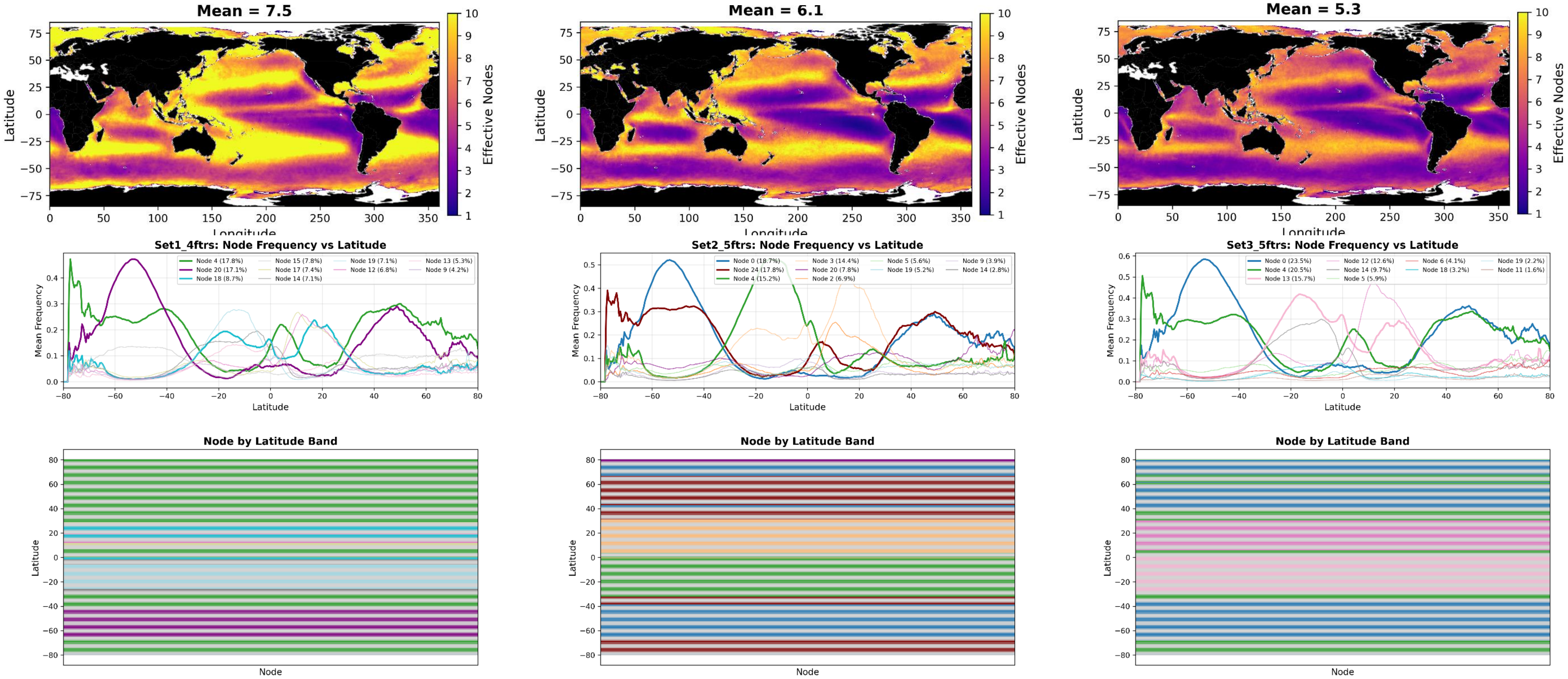




Spatial Patterns Sets: S1-4v, S2-5v, S3-5v

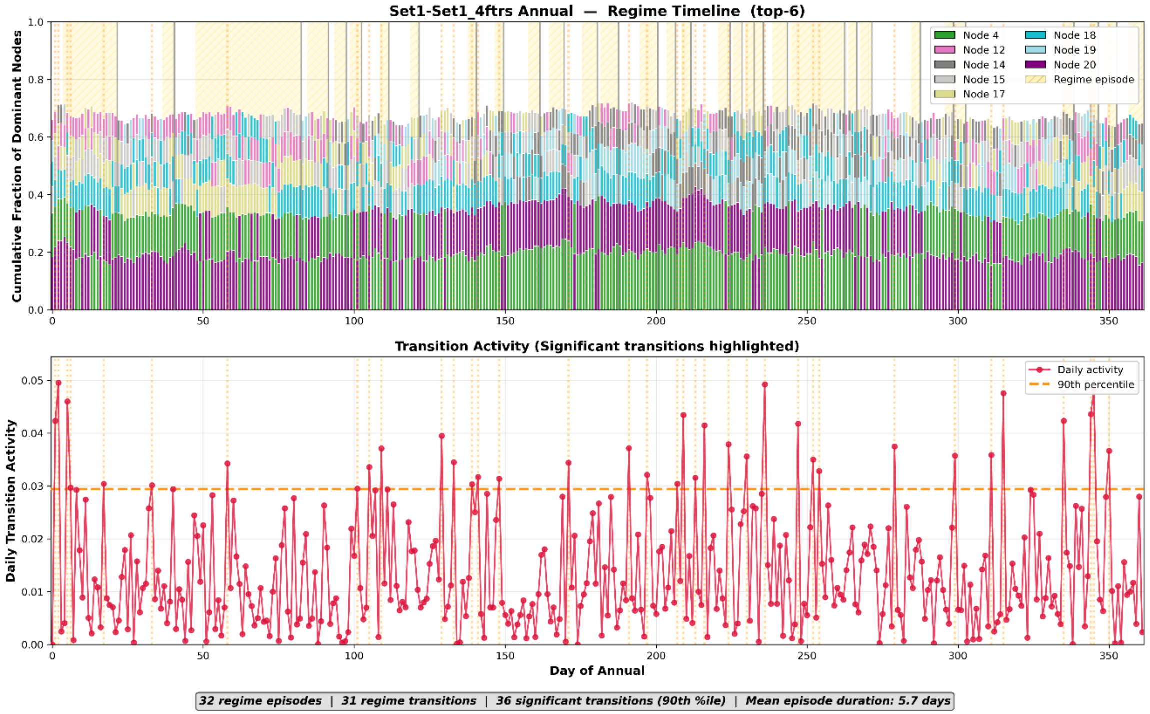


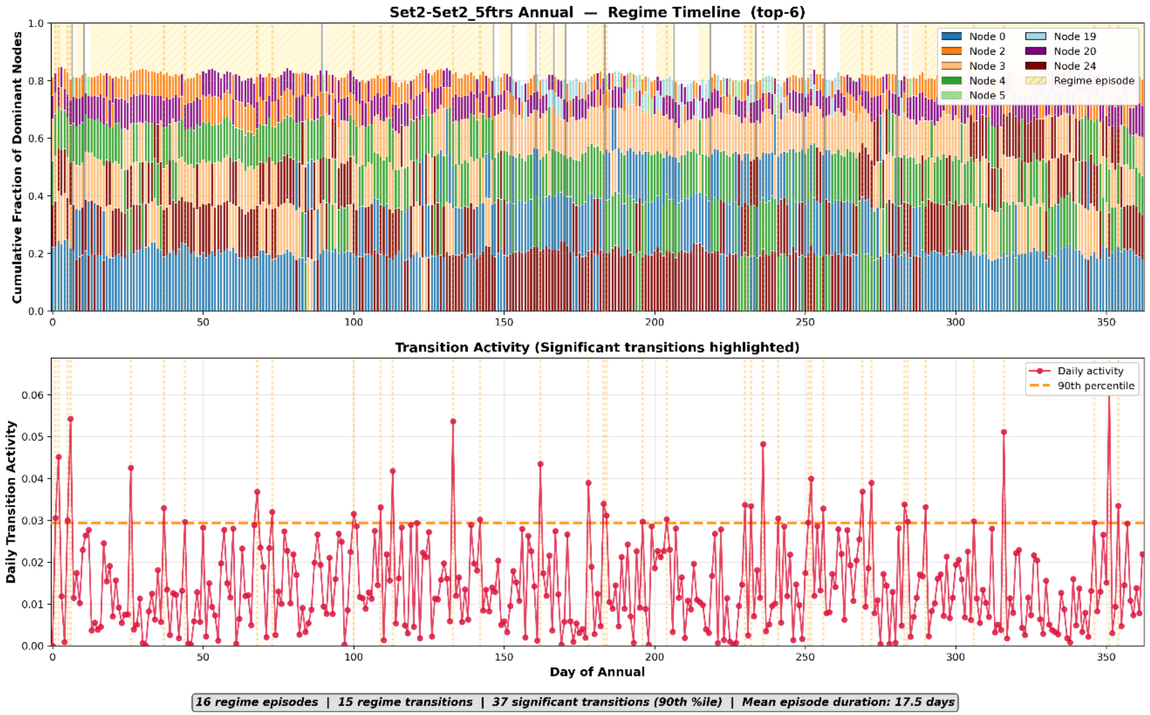
Spatial Patterns Sets: S1-4v, S2-5v, S3-5v

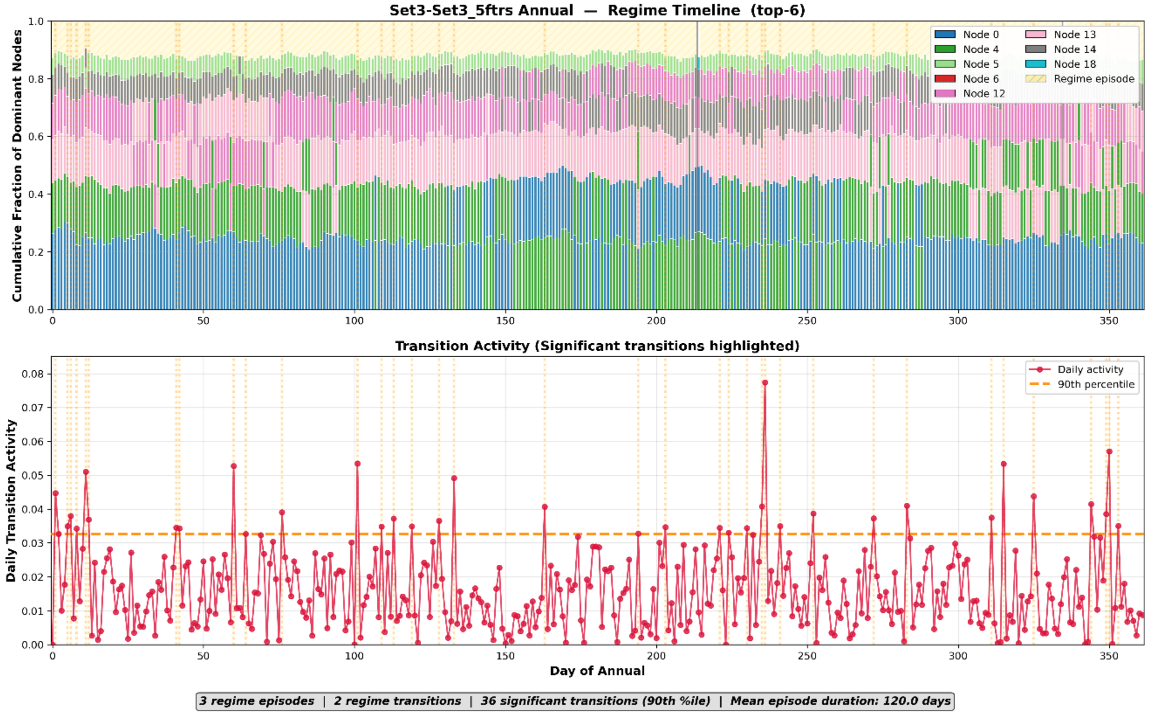


Nodes occupy preferred Lat. band - SOM learns meridional structure, not just random clusters

Temporal Patterns Sets



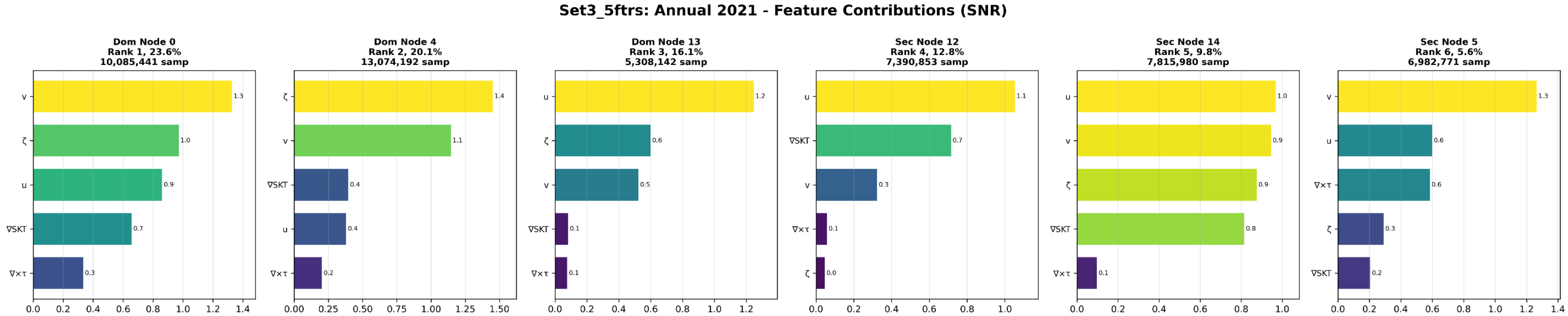




Transition: $T_{ij} = \left(\frac{1}{N_{days}}\right) \cdot \sum_t \delta(k^*_{ij}(t) \neq k^*_{ij}(t-1))$

Persistence: consecutive days with the same dom. node

Variable contributions: S1-4v, S2-5v, S3-5v



Set3 [illegible]

N4 [illegible] retains the broadest feature balance

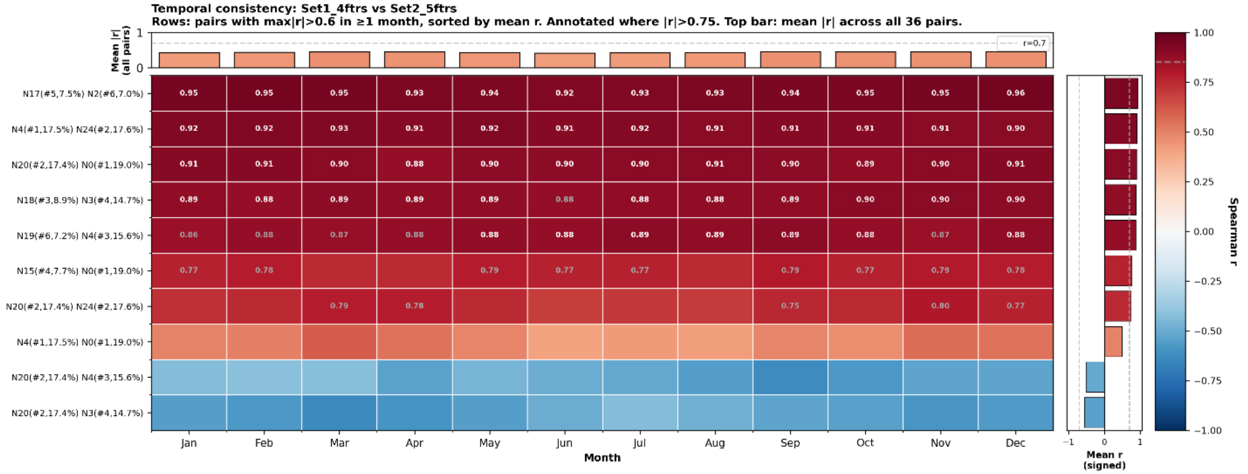
N0: mixed [illegible] combined CFB+TFB state

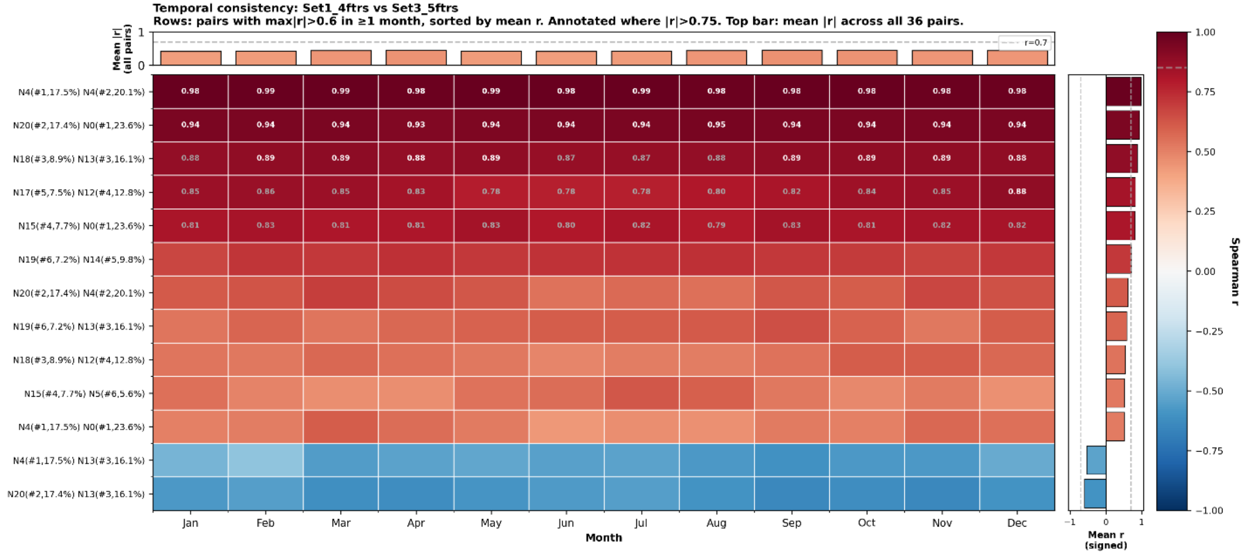
N19 [illegible] despite being in the subtropical gyre.

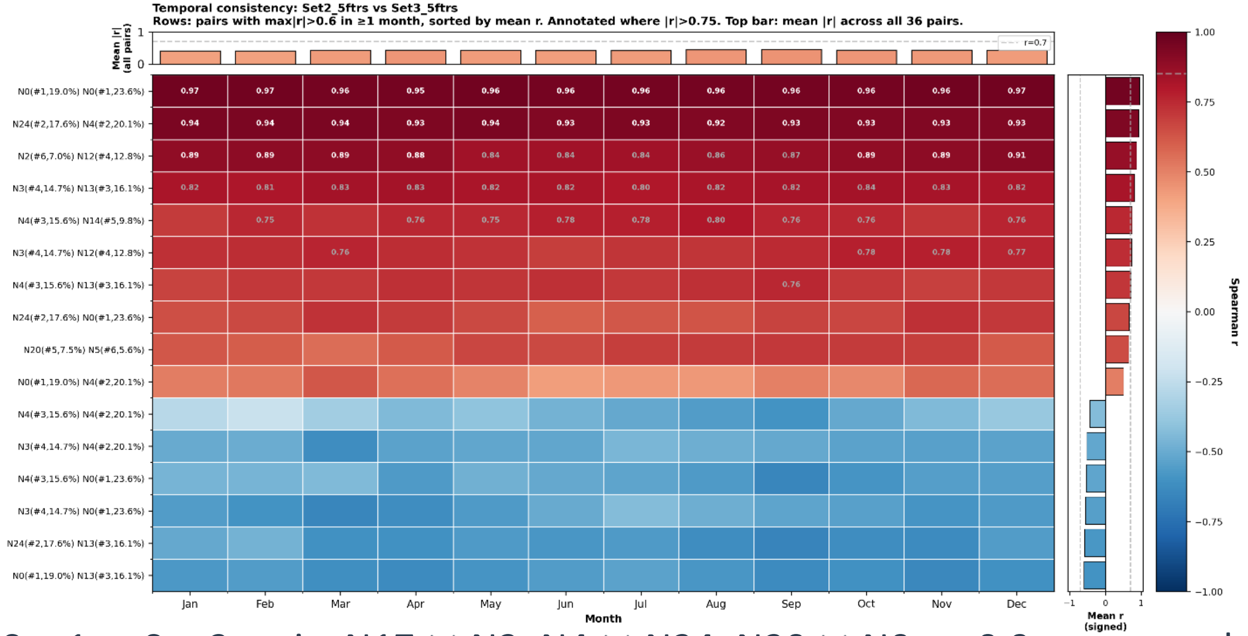
N13: kinematically driven with negligible TFB contribution.

N17: is the only secondary node with elevated ∇×τ, pointing to Ekman pumping relevance in the subtropics.

Temporal cross-set correlations







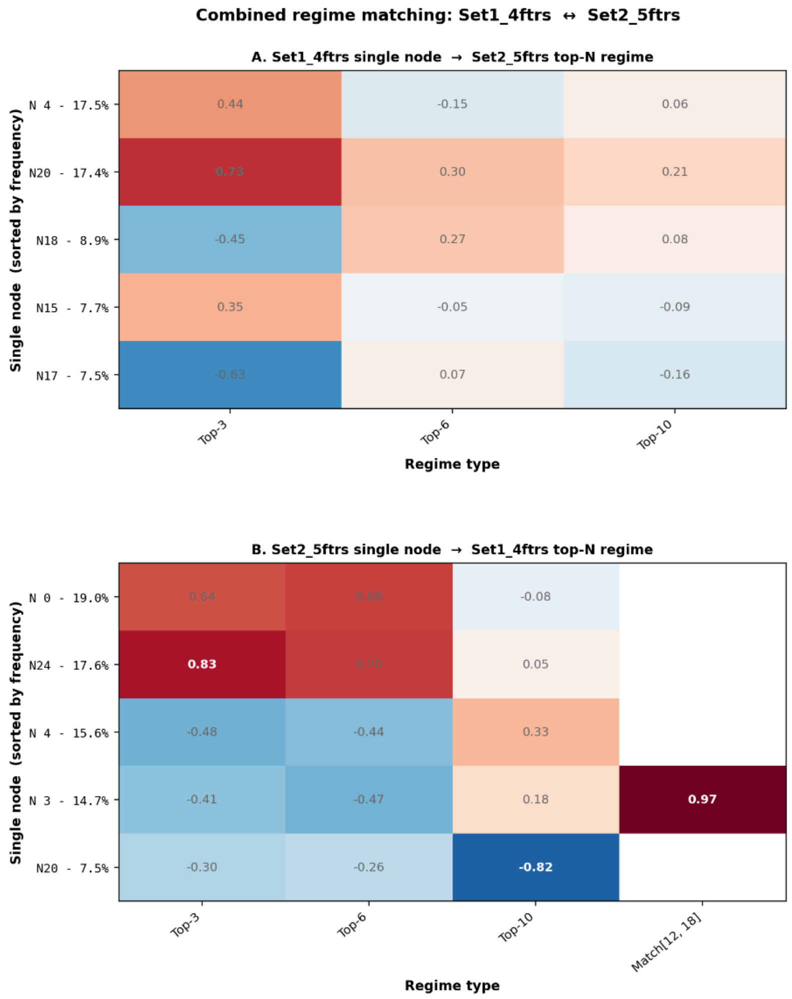
Set 1 vs Set 2: pairs N17↔N2, N4↔N24, N20↔N0 r > 0.9 year-round. The CFB-dominant and TFB-dominant sets agree on where the major coupling regimes are located.

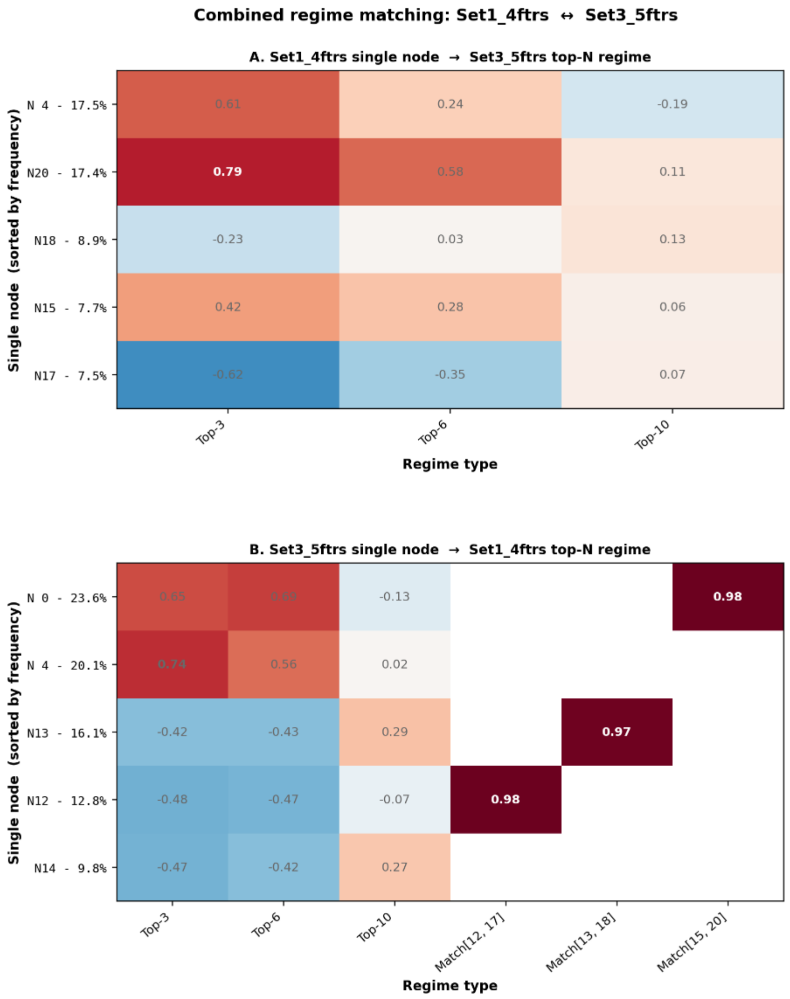
Set 2 vs Set 3: N0↔N0, N24↔N4, N2↔N12 show similarly high correlations.

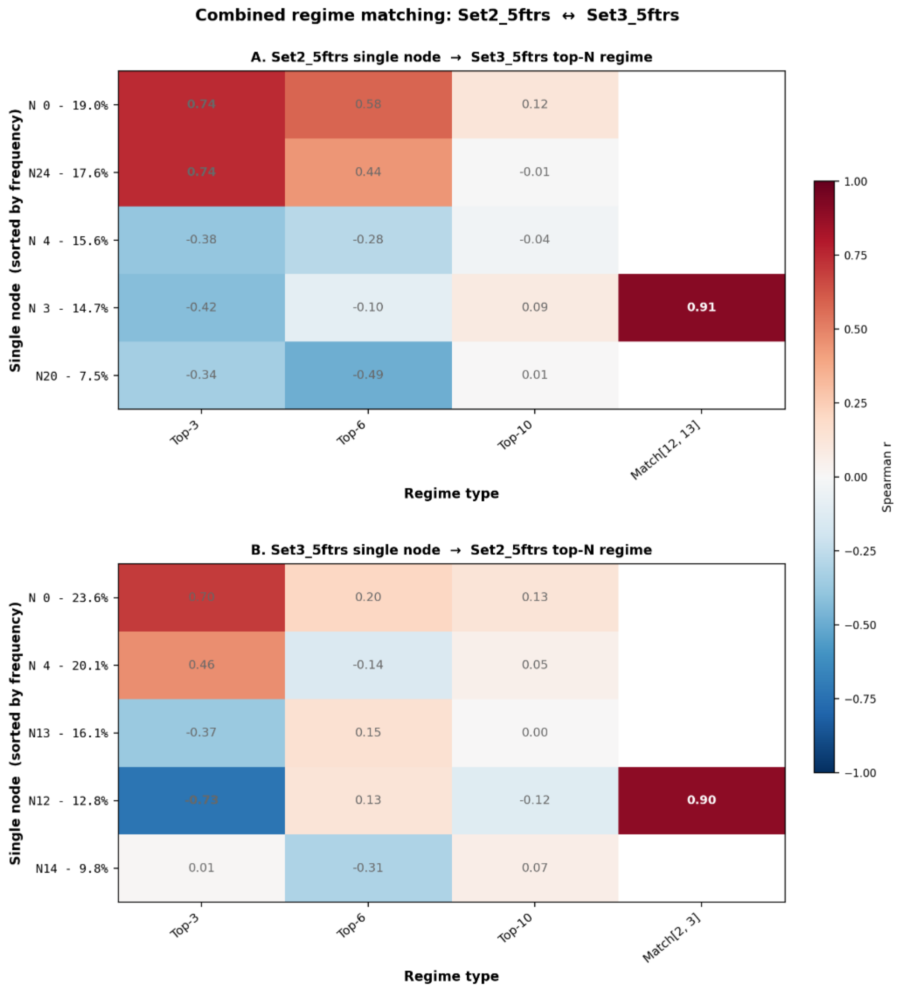
Two sets with overlapping but distinct variable combinations produce the same dominant spatial patterns.

Anti-correlated pairs mark regions where one mechanism is active and the other is suppressed

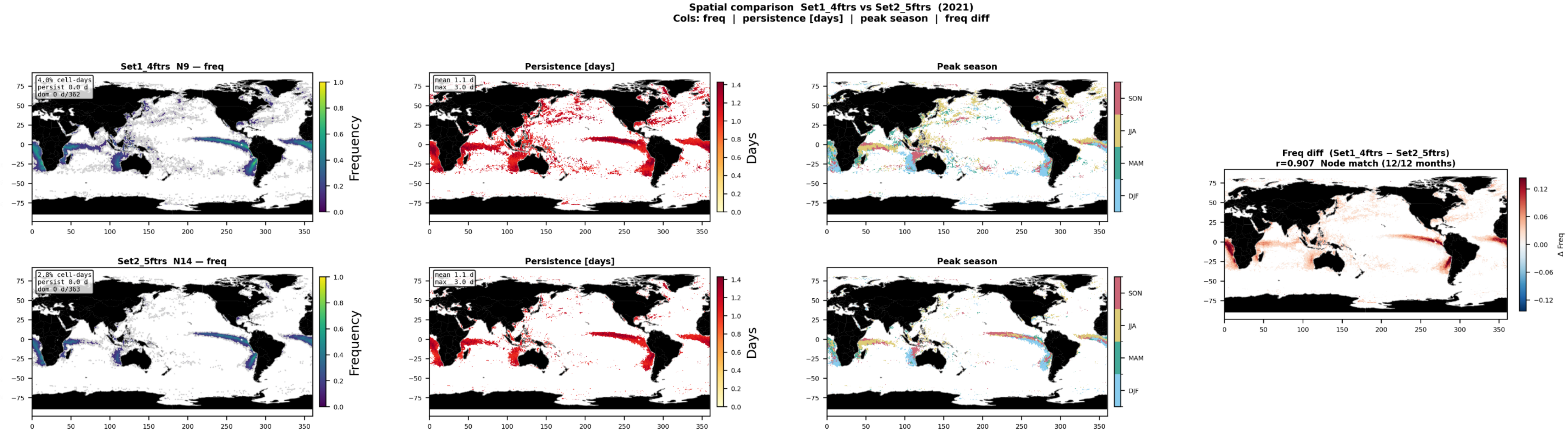
Combined patterns



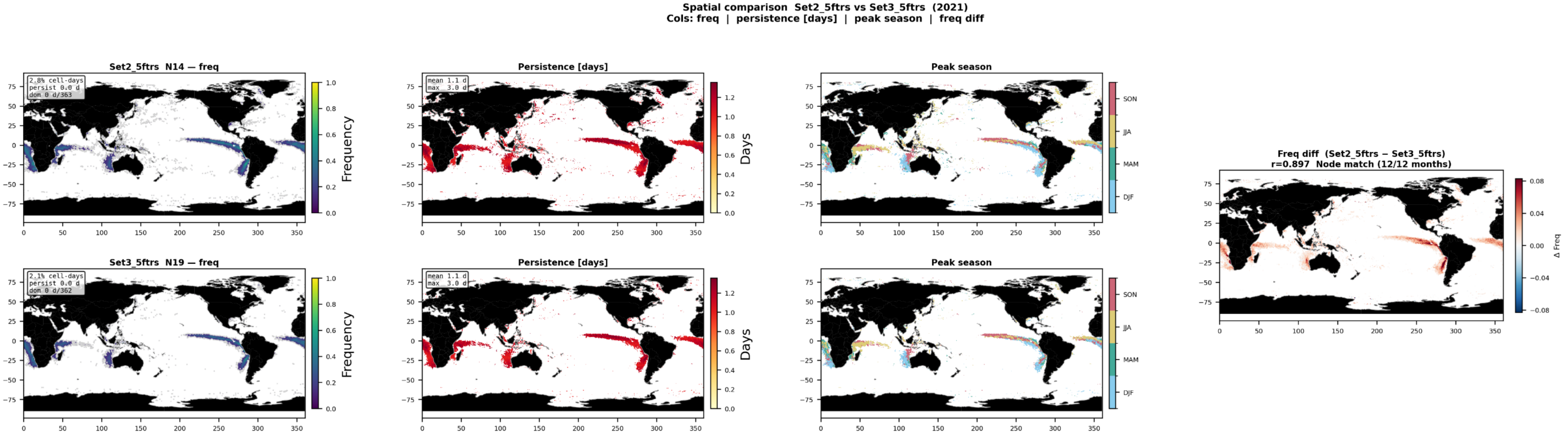




Spatial Patterns Pairs: S1 vs S2



Spatial Patterns Pairs : S2 vs S3



Summary

- A 5×5 SOM trained sequentially on physics-informed variables successfully partitions into 25 interpretable nodes representing distinct ocean-atmosphere coupling states
- **Optimal feature counts** identified objectively via spatial coherence: 4 variables for Set 1 (CFB), 5 for Sets 2–5 (TFB and mixed), confirmed by 75–100% monthly consensus
- **Spatial structure** is dominated by 3 primary nodes covering >100K grid bins, with 10–18 additional nodes capturing regional and secondary patterns in frontal zones and upwelling systems
- **Temporal behavior** differs by mechanism: CFB nodes (Set 1) transition every ~7 days; TFB nodes (Set 2) persist ~17 days; combined Sets persist up to 120 days, reflecting the different timescales of thermal vs dynamic forcing
- **Variable contributions** confirm physical interpretability: ζ and $\nabla\times\tau$ drive CFB nodes; ∇_cSST and ∇SST$\times\tau$ drive TFB nodes

Outlook

- **Mechanism attribution**: use the anti-correlated node pairs identified in the cross-Set analysis as spatial masks to isolate grid cells where CFB and TFB operate independently, enabling direct attribution of wind variance to each mechanism

- **Seasonal and regional focus**: deeper analysis of the DJF mid-latitude nodes where TFB dominates and the year-round western boundary current nodes where CFB is most active

- **Coupling to diagnostics**: link SOM node transitions to observed events to validate the physical interpretation of each regime

- **Operational relevance**: explore whether SOM node state provides predictive skill for surface wind stress errors in operational NWP systems