Explainable Machine Learning for Inferring Subsurface Ocean Dynamics

William Yik,[a,c] Maike Sonnewald[b,c]

[a]Department of Computer Science/Department of Mathematics, Harvey Mudd College, Claremont, CA [b]Department of Computer Science, University of California, Davis, CA

[c]Ocean and Cryosphere Division, NOAA Geophysical Fluid Dynamics Laboratory, Princeton, NJ

Introduction

- Complex ocean systems such as the **Antarctic Circumpolar Current (ACC)**, which play key roles in the Earth's climate, are known to change in strength and location under climate change
- These shifts are not well constrained and their physical drivers are not well understood
- We use the machine learning-driven method **Tracking Heating with global Ocean Regimes (THOR)** to both identify and track regions of the ocean characterized by similar physics, revealing drivers of ocean dynamical shifts under climate change

Tracking Heating with global Ocean Regimes (THOR)

We extend THOR, originally developed by Sonnewald and Lguensat [1], to a 0.25°, mesoscale eddy permitting ocean model, the **Modular Ocean Model version 6 (MOM6)**, a component of the Coupled Model version 4 (CM4). THOR consists of two components.

Step 1: Unsupervised clustering of ocean grid cells

- **Dynamical regimes** are regions of the ocean characterized by similar physics as defined by the barotropic vorticity (BV) equation

$$\overbrace{\beta V}^{\text{Advection}} = \underbrace{\nabla\times(p_b\nabla H)}_{\text{Bottom pressure torque}} + \overbrace{\nabla\times\tau}^{\text{Wind and bottom stress}} + \underbrace{\nabla\times\mathbf{A}}_{\text{Non-linear torque}} + \overbrace{\nabla\times\mathbf{B}}^{\text{Lateral viscosity}}$$

- **Native Emergent Manifold Interrogation (NEMI)** [2] is used to cluster ocean grid cells based on their average balance of the BV equation during a pre-industrial control (piControl) run

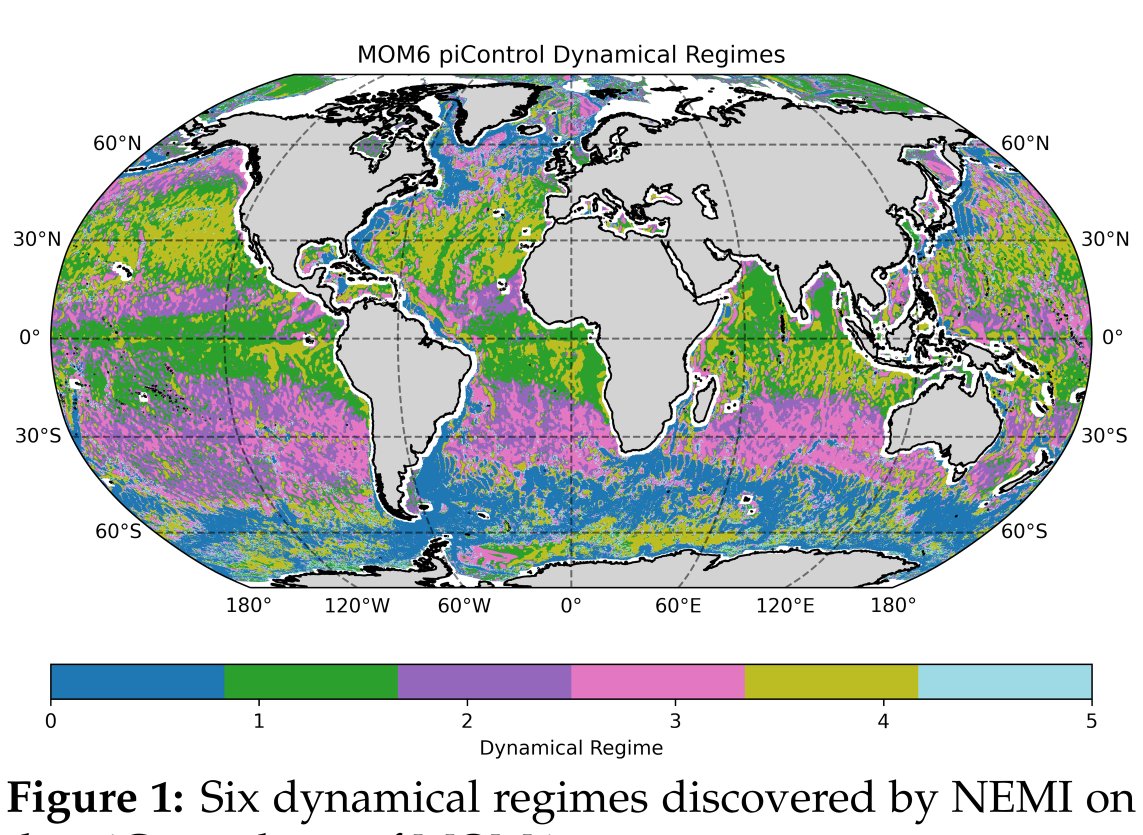

Figure 1: Six dynamical regimes discovered by NEMI on the piControl run of MOM6.

Step 2: Supervised learning of dynamical regimes

- An ensemble of **neural networks (NNs)** is trained to predict dynamical regimes from more accessible input fields for seamless application to other scenario experiments or entirely different ocean models
- **Inputs**: sea surface height above the geoiod (ZOS) + lat/lon gradients, depth relative to sea level (bathymetry) + lat/lon gradients, curl of surface wind stress torque ($\nabla\times\tau_s$), Coriolis parameter (f), depth-summed zonal and meridional mass transport (umo 2d and vmo 2d)
- **Entropy** is used to quantify the NN ensemble's uncertainty in its predictions [3]

$$H_i = -\sum_{j=1}^{N_l} p_{ij}\log(p_{ij})$$

Application to the Southern Ocean

We apply THOR's NN to the **Historical** and **SSP585** runs of MOM6 to track dynamical regimes in the Southern Ocean under climate change.

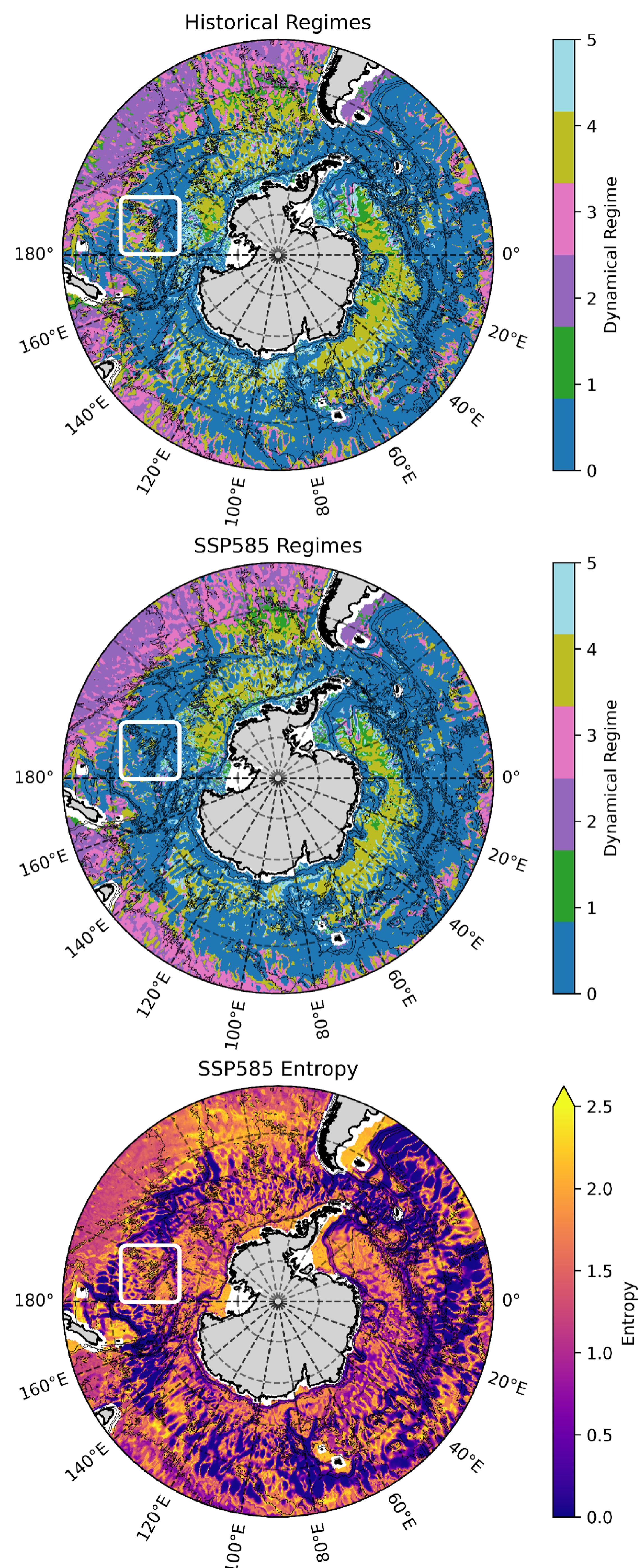

Figure 2: NN dynamical regime predictions for the Historical and SSP585 runs along with entropy (uncertainty) for the SS585 predictions. The contours show bathymetry, and the region of interest where the ACC meets the Pacific-Antarctic Ridge is outlined in white.

- We focus on the region where the ACC meets the **Pacific-Antarctic Ridge (PAR)**, a divergent tectonic plate boundary characterized by rough bathymetry at around 60°S, 166°W
- Specifically, between the Historical and SSP585 runs we see a **shift in dynamical regime** from Regime 4 (light green), which is characterized by a large wind stress, to Regime 0 (blue), which is characterized by flow free of bathymetric influence
- Two **eXplainable Artificial Intelligence (XAI) methods**, layer-wise relevance propagation (LRP) and SHapley Additive exPlanations (SHAP), reveal that the curl of surface wind stress torque ($\nabla\times\tau_s$) and the bathymetry actively help the NN make its regime predictions where the ACC meets the PAR

Guided by the new knowledge revealed by THOR, we find that the **wind stress maximum shifts northward**, which changes the ACC's interactions with the bathymetry of the PAR.

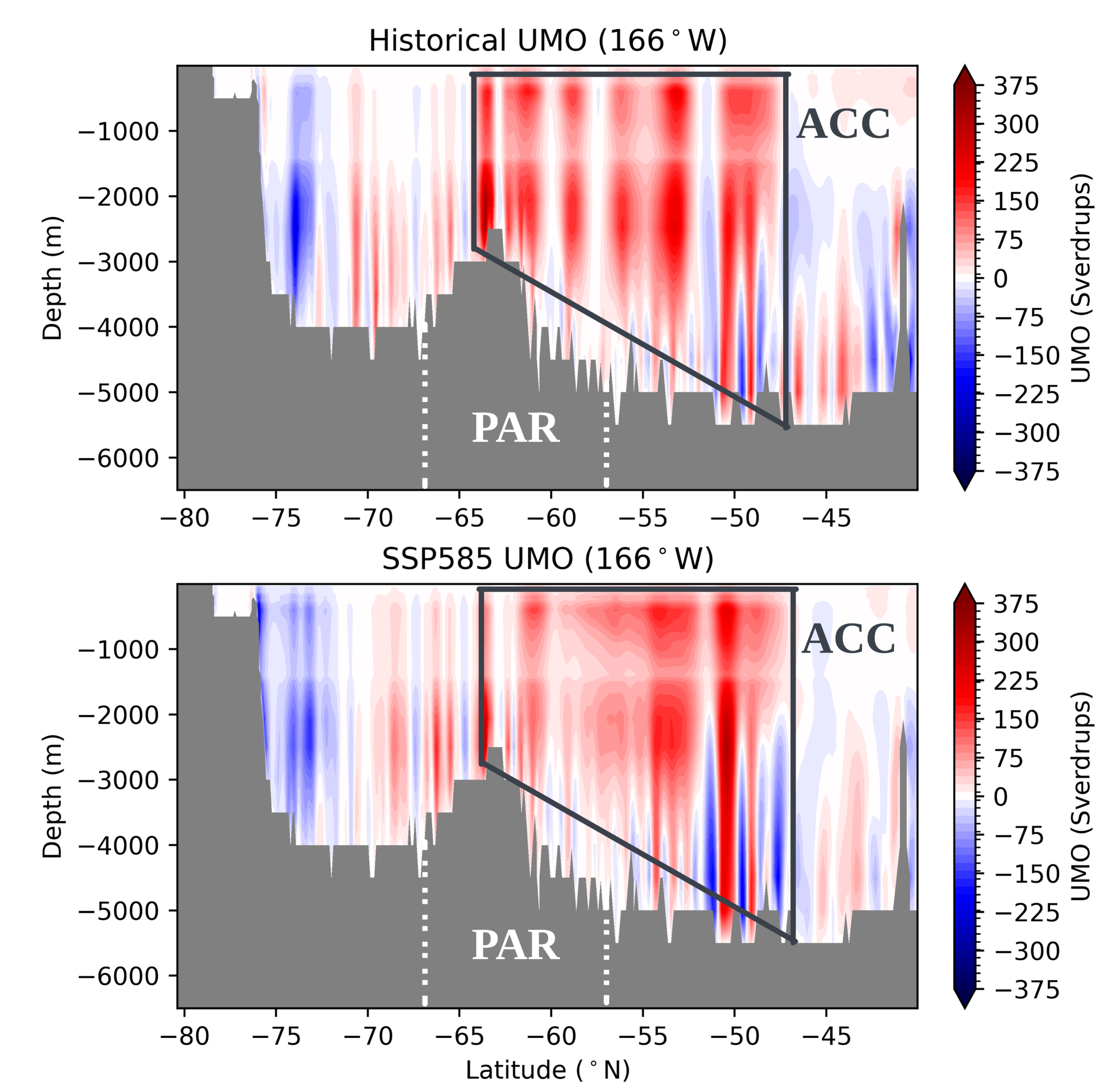

Figure 3: Transects at 166°W of zonal mass transport (umo 2d) for both the Historical and SSP585 runs. The ACC and PAR are outlined in black and white, respectively.

- ACC-PAR regime shifts are caused by a **northward shift in the ACC** driven by changes in wind stress
- This ACC movement brings it away from the PAR into a new, less variable bathymetric region where its interactions with the sea floor are less strong, thus leading to **stronger baroclinic flow**

Conclusion

- We extend THOR to a mesoscale eddy permitting climate model, which allows us to precisely identify and track ocean dynamical regimes under climate change
- Future work will include applying THOR to other climate models to understand differences in their ocean physics parameterizations

Acknowledgements

The authors would like to thank Mariana Clare, Redouane Lguensat, Stephen Griffies, Tarun Verma, and Isaac Held for their helpful discussions and feedback. WY acknowledges funding from the NOAA Ernest F. Hollings Undergraduate Scholarship.

References

[1] Maike Sonnewald and Redouane Lguensat. Revealing the impact of global heating on north atlantic circulation using transparent machine learning. *Journal of Advances in Modeling Earth Systems*, 13 (8):e2021MS002496, 2021.

[2] Maike Sonnewald. A hierarchical ensemble manifold methodology for new knowledge on spatial data: an application to ocean physics. *Authorea Preprints*, 2023.

[3] Mariana CA Clare, Maike Sonnewald, Redouane Lguensat, Julie Deshayes, and Venkatramani Balaji. Explainable artificial intelligence for bayesian neural networks: toward trustworthy predictions of ocean dynamics. *Journal of Advances in Modeling Earth Systems*, 14(11):e2022MS003162, 2022.

Correspondence to wyik@hmc.edu