Southern Ocean Dynamics Under Climate Change: New Knowledge Through Physics-Guided Machine Learning

William Yik
Dept. of Computer Science/Mathematics
Harvey Mudd College
Claremont, USA
wyik@hmc.edu

Maike Sonnewald
Dept. of Computer Science
University of California
Davis, USA
sonnewald@ucdavis.edu

Mariana C. A. Clare
ECMWF
Bonn, Germany
mariana.clare@ecmwf.int

Redouane Lguensat
IPSL, IRD
Paris, France
rlguensat@ipsl.fr

Abstract

Complex ocean systems such as the Antarctic Circumpolar Current play key roles in the climate, and current models predict shifts in their strength and area under climate change. However, the physical processes underlying these changes are not well understood, in part due to the difficulty of characterizing and tracking changes in ocean physics in complex models. To understand changes in the Antarctic Circumpolar Current, we extend the method Tracking global Heating with Ocean Regimes (THOR) to a mesoscale eddy permitting climate model and identify regions of the ocean characterized by similar physics, called dynamical regimes, using readily accessible fields from climate models. To this end, we cluster grid cells into dynamical regimes and train an ensemble of neural networks to predict these regimes and track them under climate change. Finally, we leverage this new knowledge to elucidate the dynamics of regime shifts. Here we illustrate the value of this high-resolution version of THOR, which allows for mesoscale turbulence, with a case study of the Antarctic Circumpolar Current and its interactions with the Pacific-Antarctic Ridge. In this region, THOR specifically reveals a shift in dynamical regime under climate change driven by changes in wind stress and interactions with bathymetry. Using this knowledge to guide further exploration, we find that as the Antarctic Circumpolar Current shifts north under intensifying wind stress, the dominant dynamical role of bathymetry weakens and the flow strengthens.

1 Introduction

Complex ocean systems such as the Antarctic Circumpolar Current (ACC) play a central role in the climate through mechanisms such as heat transport via powerful currents, large-scale overturning, and carbon exchange [1–3]. Such systems are known to exhibit a wide array of responses to anthropogenic forcing, but such changes and their underlying physics are poorly constrained [3–5]. This is, in part, a direct consequence of difficulties in both characterizing ocean physics and tracking its shifts under climate change. Model intercomparison projects, such as the Coupled Model Intercomparison Project (CMIP6) [6–8], have set up a vital framework for understanding differences between climate models

Preprint. Under review.

in both of these areas. However, modern models have only exacerbated challenges in characterizing and tracking ocean physics due to their complexity and size. As a result, key systems such as the ACC as well as their variability among models are often described using bulk metrics [1, 9]. For the first time, we apply the method Tracking global Heating with Ocean Regimes (THOR) proposed by [10] to the ocean of a mesoscale turbulence-permitting model, the Coupled Model 4 [11, 12], to both understand and track ocean dynamics under climate change. THOR is a three-step process beginning with unsupervised classification of ocean grid cells into dynamical regimes based on their physics. For this step, we use Native Emergent Manifold Interrogation (NEMI) [13]. The second component of THOR trains a deep ensemble of neural networks to predict these dynamical regimes from the previous step using readily accessible fields as inputs (e.g., sea surface height, depth, wind stress, and mass transport). Finally, the third step of THOR applies this trained deep ensemble to a new climate change scenario or entirely new model of interest to understand changes in dynamical regimes over time. This step also incorporates uncertainty quantification and eXplainable Artificial Intelligence (XAI) methods to guide further exploration of newly discovered ocean phenomena. Here we use THOR to investigate changes in the the region where the ACC meets the Pacific-Antarctic Ridge (PAR). THOR specifically reveals a shift in dynamical regime in this region under climate change, and XAI methods additionally show that these are related to changes in wind stress and interactions with the bathymetry. Using this new information provided by THOR, we further explore the ACC under climate change and find that the shifts in regional dynamical regime is explained by northward movement of the ACC driven by changes in wind stress. This ACC movement brings it into a new, less variable bathymetric region where its interactions with the sea floor are less strong, thus leading to a flow distributed over a larger area and concentrated at the surface. The remainder of this manuscript is structured as follows. Section 2 provides a more detailed overview of the THOR method. Section 3 details our exploration of the ACC guided by new knowledge revealed by THOR. Lastly, Section 4 concludes the paper with potential directions for future work.

2 Tracking global Heating with Ocean Regimes (THOR)

In this manuscript we apply THOR to the Modular Ocean Model version 6 (MOM6) [14, 12, 15], a component of the Coupled Model version 4 (CM4) [11]. CM4 operates at an ocean resolution of 0.25° which allows for mesoscale eddies, differentiating itself from the 1° ECCO model [16] studied by [10]. The first step of THOR addresses the key issue of unlabeled data in ML applications for climate sciences by leveraging unsupervised learning to create a labeled dataset from high dimensional ocean model data. Specifically, ocean grid cells are clustered into groups, called dynamical regimes, based on their mean balance of the barotropic vorticity (BV) equation, a characterization of local circulation of fluid flow, throughout a preindustrial control (piControl) run, illustrating the state of the ocean before the start of anthropogenically induced climate change. For more details, see [10, 17].

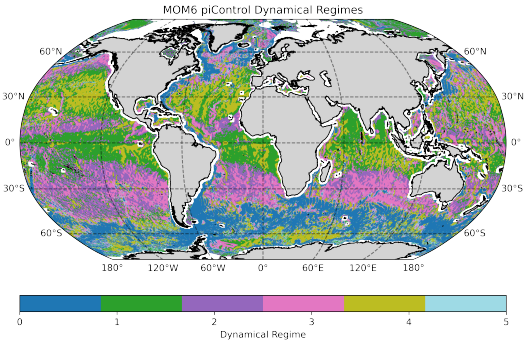

Figure 1: Six Dynamical regimes discovered by NEMI on the piControl run of MOM6.

Previous applications of THOR to the ECCO model [10, 18] used k-means clustering to reveal dynamical regimes. However, [13] demonstrates that k-means fails to converge for several information criteria metrics on the higher resolution MOM6 data. As such, for our first step of THOR we substitute k-means for Native Emergent Manifold Interrogation (NEMI), which has been shown by [13] to successfully cluster MOM6 BV data into a user-defined number of interpretable dynamical regimes. Here, we choose six labels, or dynamical regimes, following [10]. The global regimes found by NEMI for the piControl run of CM4 are shown in Figure 1. Notice the oceanographic features which are reflected in the dynamical regimes such as large wind stress-driven gyres and the the mid-Atlantic ridge.

While the first step of THOR is useful for revealing areas of the ocean characterized by similar physics, it requires that the BV budget be closed for each new ocean model and experiment in order

for NEMI clustering to be done. This is infeasible for a wide array of models and impossible with the data routinely made available through CMIP. As such, the second step of THOR trains a neural network (NN) to predict the clustered dynamical regimes from above from more accessible fields from an ocean model. Specifically, the inputs for each grid cell are the sea surface height above the geoid (ZOS), its x- and y-gradients, the depth relative to sea level (bathymetry), its x- and y-gradients, the curl of surface wind stress torque ($\nabla \times \tau_s$), the Coriolis parameter (f), and the depth-summed zonal and meridional mass transport (umo_2d and vmo_2d). The notable additions to this dataset (differentiating it from that of [10]) are the mass transport fields. These were added after preliminary testing showed that neural networks were unable to accurately classify pairs of regimes whose dominant physics were correlated with mass transport. Each of these inputs was chosen because they directly impact the values of BV equation terms, meaning that there is some physical relation between the chosen inputs and the dynamical regimes found in the first step of THOR. Thus, eXplainable artificial intelligence (XAI) methods may be applied to attribute NN predictions to specific physical inputs, allowing us to gain insight into the drivers of regime shift as demonstrated in the next section. During initial testing XAI also revealed spurious correlations between the input fields and output regimes which guided our choice to add the mass transport variables to the input.

We use the same MLP architecture described in [10] to predict the dynamical regimes, and an ensemble of 50 such MLPs are trained in order to regularize the latent loss space and gauge uncertainty in regime predictions. Specifically, we use entropy across the ensemble's predictions for each grid cell to quantify uncertainty following [19]. More detailed results from training can be found in Section A.1. Importantly, we find that though the NN does not achieve exceptionally high accuracy, the average entropy of incorrectly classified grid cells is higher than that of correctly classified grid cells. This indicates that the NN is making well-calibrated predictions, correctly represents its own uncertainty, and is not erroneously overconfident in its predictions.

3 Application to the Southern Ocean

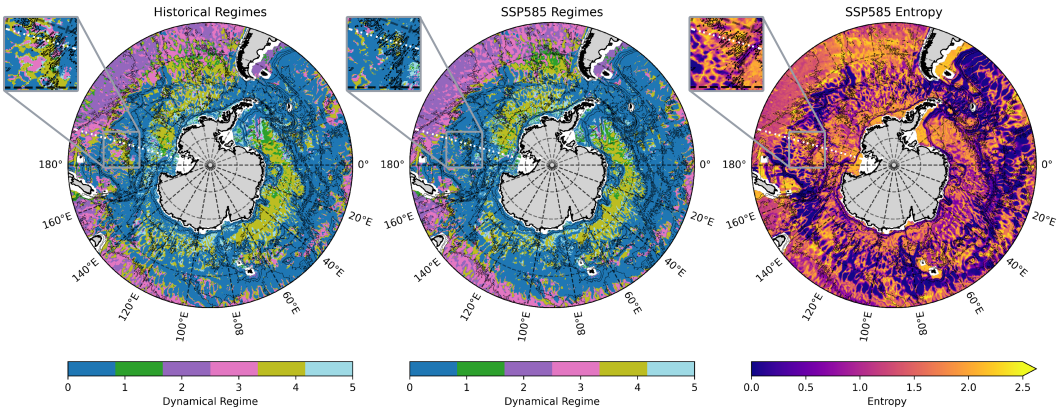

Figure 2: NN dynamical regime predictions for the Historical and SSP585 scenarios along with entropy (uncertainty) for the SS585 predictions. The contours show bathymetry. The inset shows the area of interest where the ACC meets the PAR, and the dashed white line shows the meridian where the transects of Figure 3 are taken.

The third and final step of THOR applies the trained NN ensemble from the previous section to predict dynamical regimes for entirely new experiments or ocean models. Here, we apply the ensemble to the Historical (post industrial revolution to 2015) and SSP585 (aggressive anthropogenic warming post 2015) runs of CM4 in order to track regime shifts in the Southern Ocean under climate change. Figure 2 shows the deep ensemble's mean Southern Ocean regime predictions for the Historical and SSP585 scenarios as well as the entropy for the SSP585 predictions as a proxy for the ensemble's uncertainty in its predictions under climate change. Acting as a tool to guide new discovery, THOR reveals a regime change where the ACC meets the Pacific-Antarctic Ridge (PAR), a divergent tectonic plate boundary characterized by rough bathymetry at around 60°S, 166°W. This region of interest is shown in the insets of Figure 2. The notable shift revealed by THOR from Regime 4 (light green) to Regime 0 (dark blue), combined with the high entropy motivates further investigation.

We use two XAI methods, the NN specific layer-wise relevance propogation (LRP) [20] and model agnostic Shapley Additive exPlanations (SHAP) [21], to determine which inputs help THOR's deep ensemble make its predictions where the ACC meets the PAR. The two values generally agreed on the relevance of the surface wind stress curl and the bathymetry, as well the sea surface height which is closely related to the wind (see Section A.2). Given that the ACC is largely driven by westerly winds surrounding Antarctica, the XAI methods may be indicating a shift in the wind stress and the interactions between the ACC and the bathymetry. Motivated by this, we investigate lateral (east-west) mass transport at 166°W, which represents the portion of the ACC which cuts through the PAR, as shown in the bottom two panels Figure 3. Here, we see a clear shift in the core of the ACC northward away from the PAR. Specifically, in the Historical scenario the eastward jets are closely connected to the bathymetric features of the PAR and concentrated south of diffuse circulation without a jet-like structure (intense red) at 53°S. In the SSP585 scenario, however, the area of intense flow around 50°S strengthens markedly and is associated with a strengthening of the wind stress. As such, the core of the ACC moves into a region of less dynamic bathymetry, allowing it to flow more freely and increase in intensity. Aligning with the new knowledge of wind stress relevance, the local wind stress curl strengthens just north of the PAR as shown in the top panel of Figure 3. In response, the ACC shifts in the same direction, pulling it away from the bathymetric influence of the PAR and fundamentally changing its governing physics, as reflected in the regime shift initially revealed by THOR in Figure 2.

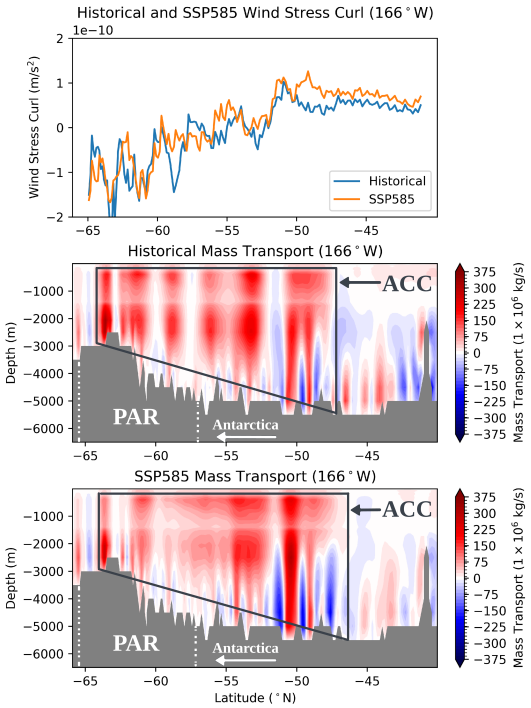

Figure 3: Wind stress curl and lateral (east-west) mass transport for both the Historical and SSP585 scenarios. The ACC and PAR are outlined in black and white, respectively. Red and blue indicte eastward and westward flow, respectively. Note that there is an artifact around -1500m due to regridding.

4 Conclusion

We extend the machine learning method THOR to a mesoscale eddy-permitting climate model in order to gain insight into the drivers underlying changes in ocean physics under climate change. Importantly, we use THOR not as an oracle, but rather a tool to discover new knowledge about ocean dynamical regimes and guide further investigation of shifts in the driving forces behind ocean circulation. Under this framework, THOR reveals that a fundamental shift in Southern Ocean physics occurs under climate change where the ACC meets the PAR. Using this new knowledge combined with XAI methods as a guide, we find that the wind stress curl increases in strength north of the PAR, pulling the ACC with it and effectively moving the ACC from a bathymetrically locked state over the PAR to one dominated by the wind stress where its flow increases in strength.

There are several possible directions for future work. First, different CMIP models have different representations of ocean physics, and the high-resolution version of THOR presented in this work could provide insight into how these differences affect predictions for various ocean systems under climate change. Such insight is at the heart of lowering model spread and uncertainty of future projections. Additionally, the NN ensemble applied in this work classifies the dynamical regimes of grid cells based on local information to that grid cell. Spatially-aware NN architectures may be applied, but the large continents and difficulties applying XAI methods to more complex architectures pose exciting future challenges. Regardless of the supervised technique used to classify dynamical regimes, we maintain that its predictive power should not trade off with its interpretability and transparency.

Acknowledgments and Disclosure of Funding

The authors would like to thank Stephen Griffies, Tarun Verma, and Isaac Held for their helpful discussions and feedback. WY acknowledges funding from the NOAA Ernest F. Hollings Undergraduate Scholarship. MS funding: Cooperative Institute for Modeling the Earth System, Princeton University, under Award NA18OAR4320123 from the National Oceanic and Atmospheric Administration, U.S. Department of Commerce.

Data Availability Statement

The data and code are currently unavailable to preserve anonymity but are available upon request.

References

[1] Christopher C Chapman, Mary-Anne Lea, Amelie Meyer, Jean-Baptiste Sallée, and Mark Hindell. Defining southern ocean fronts and their influence on biological and physical processes in a changing climate. *Nature Climate Change*, 10(3):209–219, 2020.

[2] Adele K Morrison and Andrew McC Hogg. On the relationship between southern ocean overturning and acc transport. *Journal of Physical Oceanography*, 43(1):140–148, 2013.

[3] Claus W Böning, Astrid Dispert, Martin Visbeck, SR Rintoul, and Franziska U Schwarzkopf. The response of the antarctic circumpolar current to recent climate change. *Nature Geoscience*, 1(12):864–869, 2008.

[4] Clothilde E Langlais, Stephen R Rintoul, and Jan D Zika. Sensitivity of antarctic circumpolar current transport and eddy activity to wind patterns in the southern ocean. *Journal of Physical Oceanography*, 45 (4):1051–1067, 2015.

[5] Sarah M Larson, Martha W Buckley, and Amy C Clement. Extracting the buoyancy-driven atlantic meridional overturning circulation. *Journal of Climate*, 33(11):4697–4714, 2020.

[6] Gerald A Meehl, George J Boer, Curt Covey, Mojib Latif, and Ronald J Stouffer. The coupled model intercomparison project (cmip). *Bulletin of the American Meteorological Society*, 81(2):313–318, 2000.

[7] Veronika Eyring, Sandrine Bony, Gerald A Meehl, Catherine A Senior, Bjorn Stevens, Ronald J Stouffer, and Karl E Taylor. Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, 9(5):1937–1958, 2016.

[8] Brian C O'Neill, Claudia Tebaldi, Detlef P Van Vuuren, Veronika Eyring, Pierre Friedlingstein, George Hurtt, Reto Knutti, Elmar Kriegler, Jean-Francois Lamarque, Jason Lowe, et al. The scenario model intercomparison project (scenariomip) for cmip6. *Geoscientific Model Development*, 9(9):3461–3482, 2016.

[9] KA Donohue, KL Tracey, DR Watts, María Paz Chidichimo, and TK Chereskin. Mean antarctic circumpolar current transport measured in drake passage. *Geophysical Research Letters*, 43(22):11–760, 2016.

[10] 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.

[11] IM Held, H Guo, A Adcroft, JP Dunne, LW Horowitz, J Krasting, E Shevliakova, M Winton, M Zhao, M Bushuk, et al. Structure and performance of gfdl's cm4. 0 climate model. *Journal of Advances in Modeling Earth Systems*, 11(11):3691–3727, 2019.

[12] Alistair Adcroft, Whit Anderson, V Balaji, Chris Blanton, Mitchell Bushuk, Carolina O Dufour, John P Dunne, Stephen M Griffies, Robert Hallberg, Matthew J Harrison, et al. The gfdl global ocean and sea ice model om4. 0: Model description and simulation features. *Journal of Advances in Modeling Earth Systems*, 11(10):3167–3211, 2019.

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

[14] Stephen M Griffies. Elements of the modular ocean model (mom). *GFDL Ocean Group Tech. Rep*, 7(620): 47, 2012.

[15] Stephen M Griffies, Alistair Adcroft, and Robert W Hallberg. A primer on the vertical lagrangian-remap method in ocean models based on finite volume generalized vertical coordinates. *Journal of Advances in Modeling Earth Systems*, 12(10):e2019MS001954, 2020.

[16] GAEL Forget, J-M Campin, P Heimbach, CN Hill, RM Ponte, and C Wunsch. Ecco version 4: An integrated framework for non-linear inverse modeling and global ocean state estimation. *Geoscientific Model Development*, 8(10):3071–3104, 2015.

[17] Hemant Khatri, Stephen M Griffies, Benjamin A Storer, Michele Buzzicotti, Hussein Aluie, Maike Sonnewald, Raphael Dussin, and Andrew Shao. A scale-dependent analysis of the barotropic vorticity budget in a global ocean simulation. *ESS Open Archive*, 2023.

[18] Maike Sonnewald, Redouane Lguensat, Aparna Radhakrishnan, Zoubero Sayibou, Venkatramani Balaji, and Andrew Wittenberg. Revealing the impact of global warming on climate modes using transparent machine learning and a suite of climate models. In *ICML 2021 Workshop on Tackling Climate Change with Machine Learning*, 2021. URL `https://www.climatechange.ai/papers/icml2021/13`.

[19] 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.

[20] Grégoire Montavon, Alexander Binder, Sebastian Lapuschkin, Wojciech Samek, and Klaus-Robert Müller. Layer-wise relevance propagation: an overview. *Explainable AI: interpreting, explaining and visualizing deep learning*, pages 193–209, 2019.

[21] María Vega García and José L Aznarte. Shapley additive explanations for no2 forecasting. *Ecological Informatics*, 56:101039, 2020.

A Supplemental Material

A.1 Neural Network Training Details

For the second step of THOR, we train an ensemble of 50 MLP NNs each with the same architecture as [10] but different weight and bias initializations. These NNs are trained to predict dynamical regimes for individual ocean grid cells from accessible ocean model fields. Specifically, the 10 inputs are the sea surface height above the geoid (ZOS), its x- and y-gradients, the depth relative to sea level (bathymetry), its x- and y-gradients, the curl of surface wind stress torque ($\nabla \times \tau_s$), the Coriolis parameter (f), and the depth-summed zonal and meridional mass transport (umo_2d and vmo_2d). The input size of the NNs was resized from 8 in [10] to 10 to account for our addition of the mass transport fields. The outputs are the dynamical regime classifications given by NEMI in step 1 of THOR.

We train THOR's deep ensemble to predict the dynamical regimes of single grid cells in a preindustial control (piControl run), which represents the state of the ocean before anthropogenic climate change. Thus, the inputs for each grid cell are the piControl mean values of the 10 inputs described above. Similarly, the outputs are one of six dynamical regimes found by NEMI's clustering, which was based on the mean balance of the BV equation during the piControl run. We split the training, validation, and test data by regions of the ocean to avoid autocorrelation in the data. Specifically, we designate the Atlantic ocean as test data, the eastern basin of the Pacific ocean as validation, and the remainder of the ocean as training. This is illustrated in Figure 4.

Each NN ensemble member was trained for 100 epochs to minimize categorical cross-entropy loss using the Adam optimizer with a learning rate of 1×10^{-4} and batch size of 32. To prevent overfitting, we enforce early stopping with a patience of 5 epochs. The final validation accuracy of ensemble members ranged from 65-70%, and the majority early stopped before 25 epochs. Figure 5 shows a visual summary of the training results. Notice that the entropy is higher in regions where THOR's deep ensemble makes more incorrect predictions, particularly in the Weddell Sea and south of Greenland. Separating grid cell entropy by correct and incorrect predictions reveals that, indeed, the average entropy for incorrect predictions is greater than that of correct predictions. This indicates that the ensemble is making well-calibrated measurements of uncertainty and that it is not overstating its own confidence in its predictions.

A.2 XAI for the Southern Ocean

We apply two XAI methods to understand the relevance of each input field in the ensemble's predictions. The first is layer-wise relevance propagation (LRP), which is specific to neural network models and uses the magnitudes of internal weights as a proxy for relevance. These relevance values are backpropogated through the network to determine the importance of each input. The relevance for each input is a value between -1 and 1, inclusive. Positive values indicate that the input was actively helpful for the NN in making its prediction, negative values indicate that the input was actively unhelpful, and 0 represents neutral relevance. See [20] for more details. We apply LRP to each of the 50 NNs of THOR's ensemble. The LRP values whose sign is consistent between the 25th and 75th percentiles of the NN ensemble in the Southern Ocean for the SSP585 (aggressive anthropogenic warming) scenario are shown in Figures 7 and 8. See [19] for more information on applying LRP to ensembles of NNs.

The second XAI method we apply is Shapley Additative exPlanations (SHAP), which is a model agnostic method for determining the relevance of each input. SHAP is a type of occlusion analysis which determines the output effect of removing/adding features from the input. For a given input and predicted output, a positive SHAP value indicates that the input increases the probability of the model predicting the output. Similarly, a negative SHAP value indicates that the input decreases the probability, and 0 indicates no change. As with LRP, we apply SHAP to each of the 50 NNs in the ensemble. The SHAP values whose sign is consistent between the 25th and 75th percentiles of the NN ensemble in the Southern Ocean for the SSP585 scenario are shown in Figures 9 and 10. See [19] for more information on applying SHAP to ensembles of NNs.

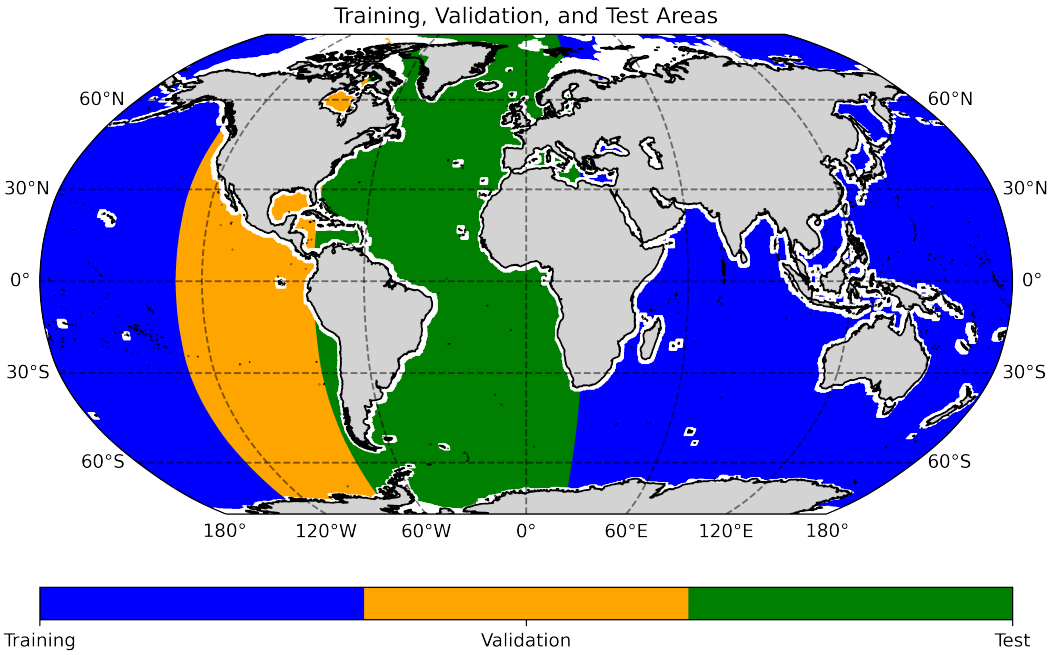

Figure 4: The training, validation, and test areas used for THOR's deep ensemble.

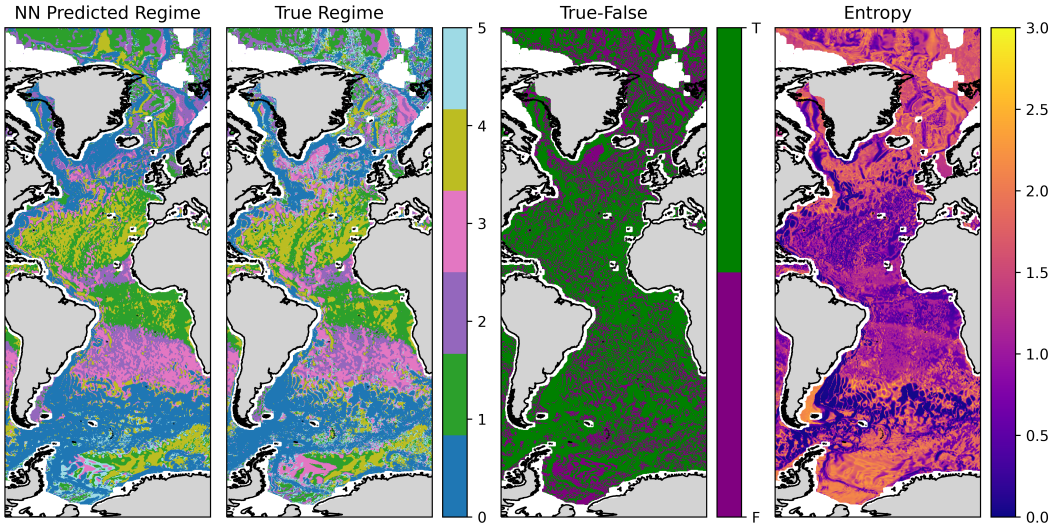

Figure 5: The NN ensemble's mean dynamical regime predictions for the test region and the true regimes. The true-false mask shows where the ensemble's mean prediction was the correct regime and where it was not. Finally, the entropy quantifies the regime's uncertainty in its predictions over the test region.

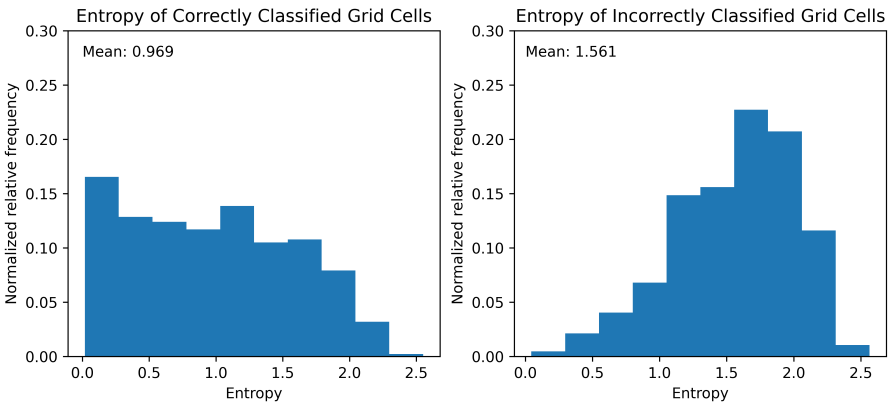

Figure 6: Entropy distributions for the correctly classified grid cells (left) and incorrectly classified grid cells (right).

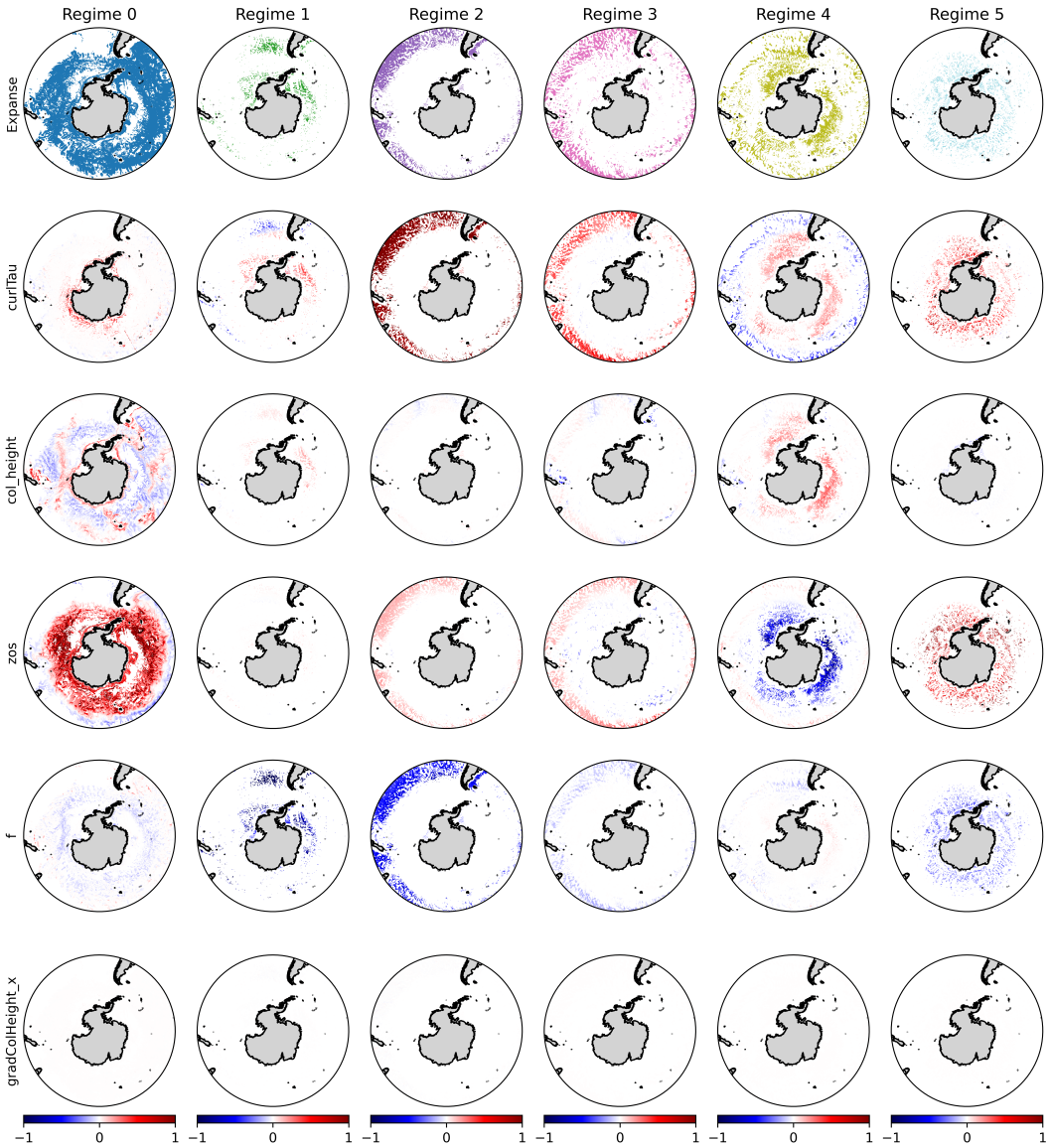

Figure 7: SSP585 scenario LRP values in the Southern Ocean for five of the NN inputs: wind stress curl, bathymetry, sea surface height, Coriolis parameter, and x-gradient of bathymetry. The first row shows the areas where THOR's NN ensemble predicted each regime, and each subsequent row shows the relevance of an input for predicting each regime.

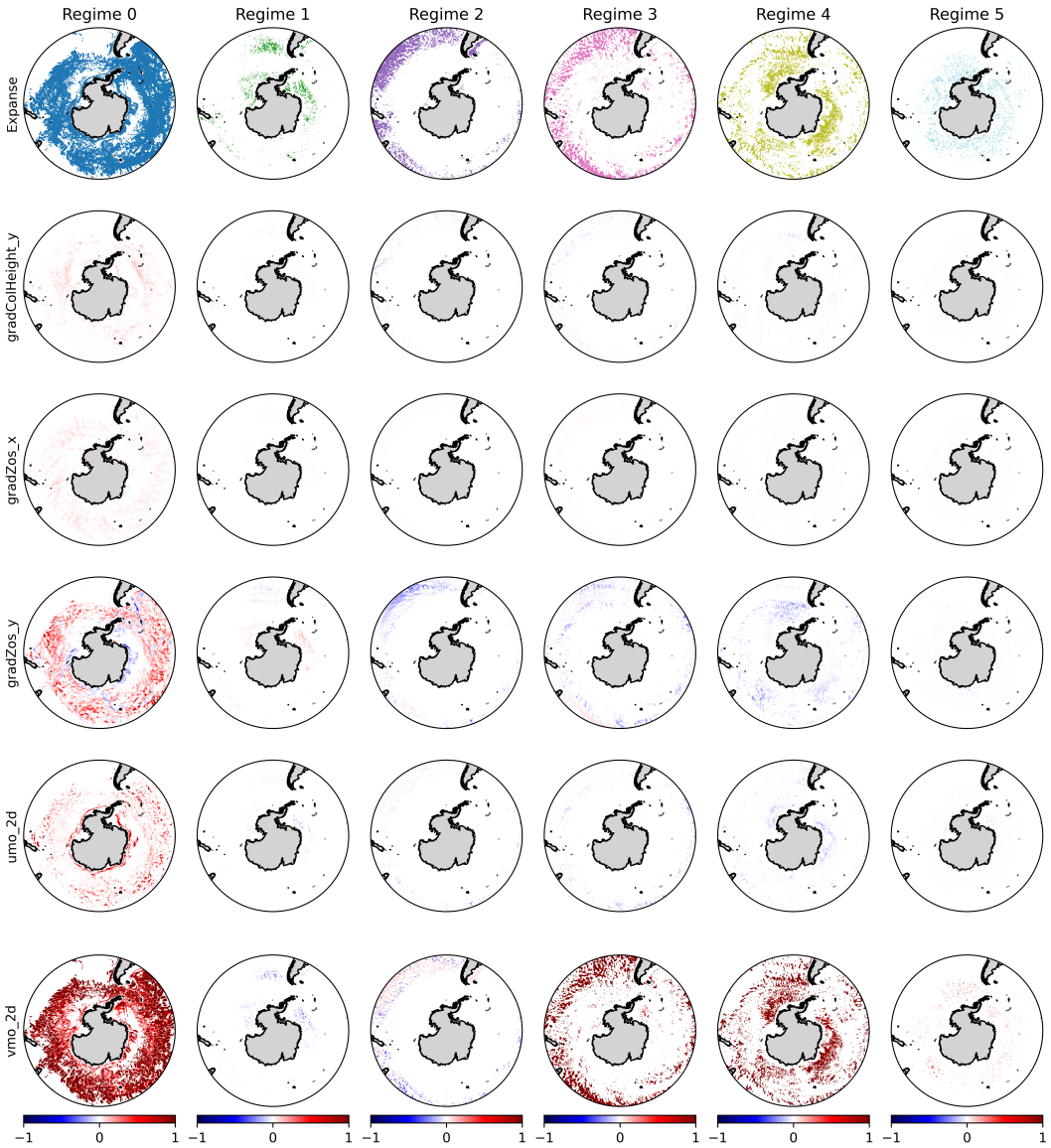

Figure 8: SSP585 scenario LRP values in the Southern Ocean for five of the NN inputs: y-gradient of bathymetry, x- and y-gradients of sea surface height, and depth-summed lateral and meridional mass transport. The first row shows the areas where THOR's NN ensemble predicted each regime, and each subsequent row shows the relevance of an input for predicting each regime.

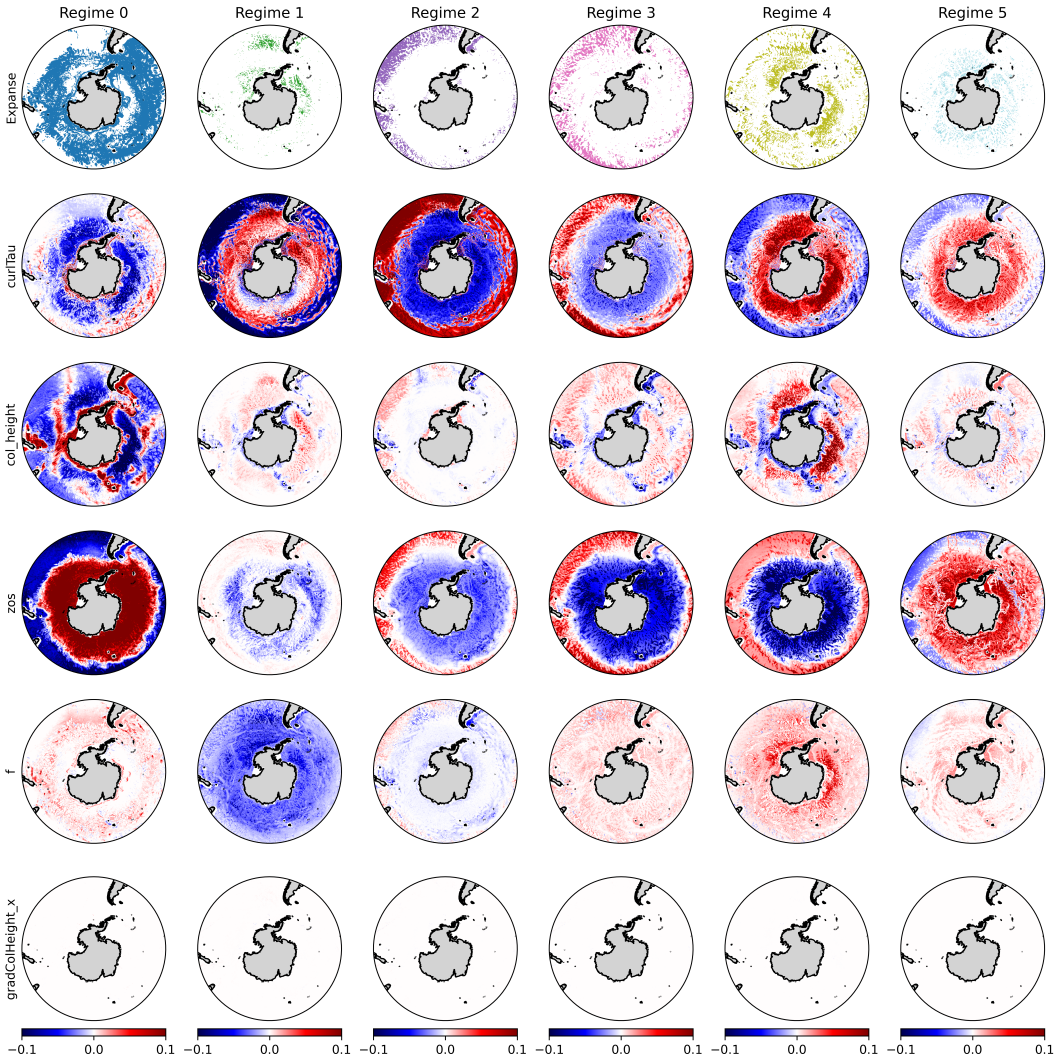

Figure 9: SSP585 scenario SHAP values in the Southern Ocean for five of the NN inputs: wind stress curl, bathymetry, sea surface height, Coriolis parameter, and x-gradient of bathymetry. The first row shows the areas where THOR's NN ensemble predicted each regime, and each subsequent row shows the relevance of an input for predicting each regime.

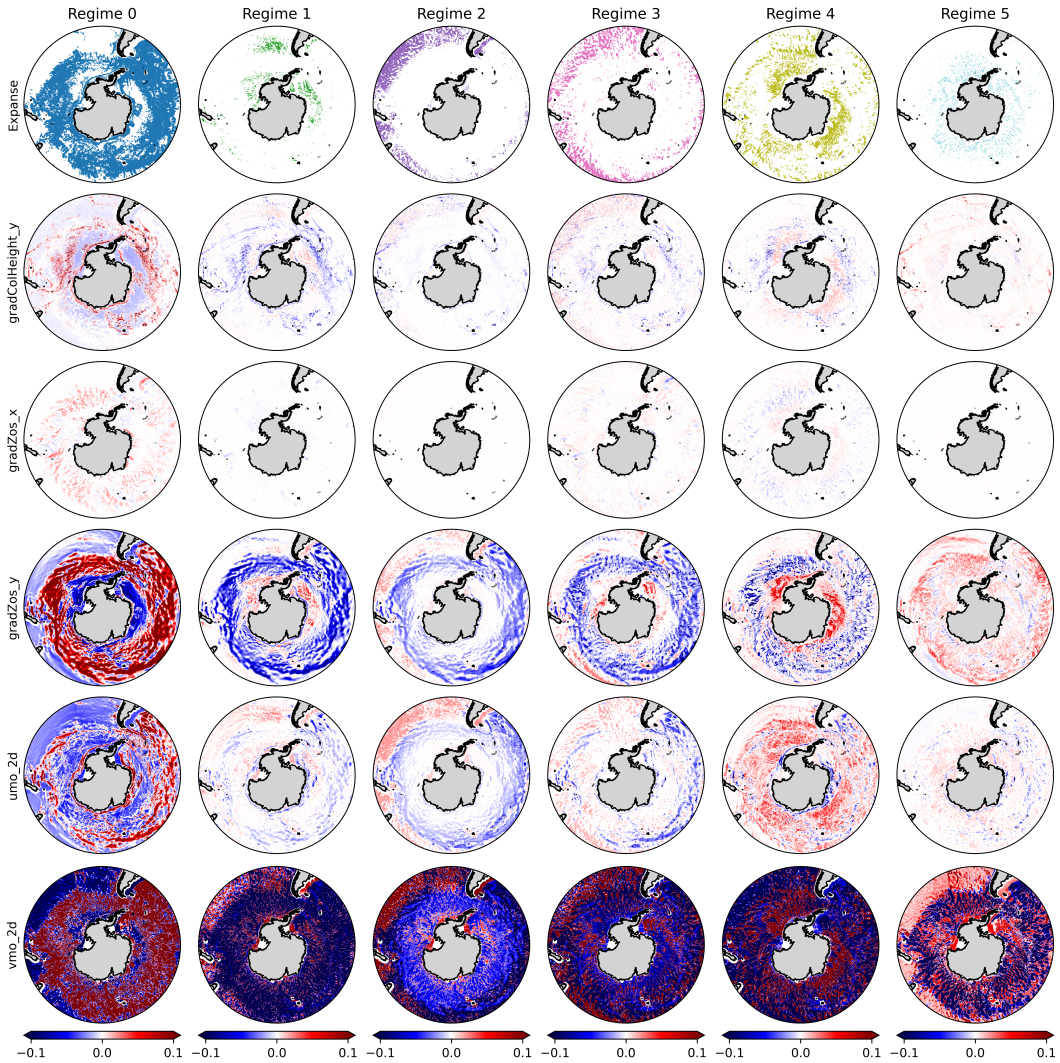

Figure 10: SSP585 scenario SHAP values in the Southern Ocean for five of the NN inputs: y-gradient of bathymetry, x- and y-gradients of sea surface height, and depth-summed lateral and meridional mass transport. The first row shows the areas where THOR's NN ensemble predicted each regime, and each subsequent row shows the relevance of an input for predicting each regime.