Towards generating stationary realizations of simulated Antarctic ice shelf melt rates from limited model output

Jacquelyn A. Shelton[1], Alexander Robel[2], Matthew Hoffman[3], and Stephen Price[3]

[1] Hong Kong Polytechnic University [2] Georgia Institute of Technology [3] Los Alamos National Laboratory

- Antarctic Ice Sheet ice loss accelerated by surrounding ocean's extreme warming over last 30 years → dominant contributor to global sea level rise
- Unknown: how much ice loss due to anthropogenic changes and to internal variability
- End Goal: construct a machine learning (ML) approach for data generation using limited model output (single realization) from expensive state-of-the-art Earth System model (E3SM)
- Current Aim: build dynamic hierarchical clustering approach that tackles core challenges BEFORE data generation possible: identify stationary subspaces of the input data that are realistic (physically consistent) and representative of its complex spatiotemporal dynamics
- Results: our approach identifies stationary subspaces of the input data that are realistic (physically consistent) and representative of its complex spatiotemporal dynamics
- This work marks promising first steps towards generating additional realizations of variable Antarctic sub-ice shelf melt in a computationally affordable fashion, and consequently, towards addressing the general and pervasive problem of data scarcity in the climate sciences

Approach: Dynamic Hierarchical Agglomerative Clustering

Prerequisites prior to data generation: identify individual subspaces in data that are:

1. *representative* of its spatiotemporal dynamics,
2. *realistic* in terms of consistency with physically observed dynamics, and
3. *stationary* over the entire time-series, regardless of the behavior of the data within each subspace relative to any other's (may vary independently over arbitrary time-scales)

Idea: construct dynamic hierarchical clustering pipeline to adaptively learn stationary subspaces, the number of which can grow or shrink in a data-driven fashion according to the data dynamics while simultaneously incorporating relevant prior domain knowledge (e.g. physical observations, problem setting).

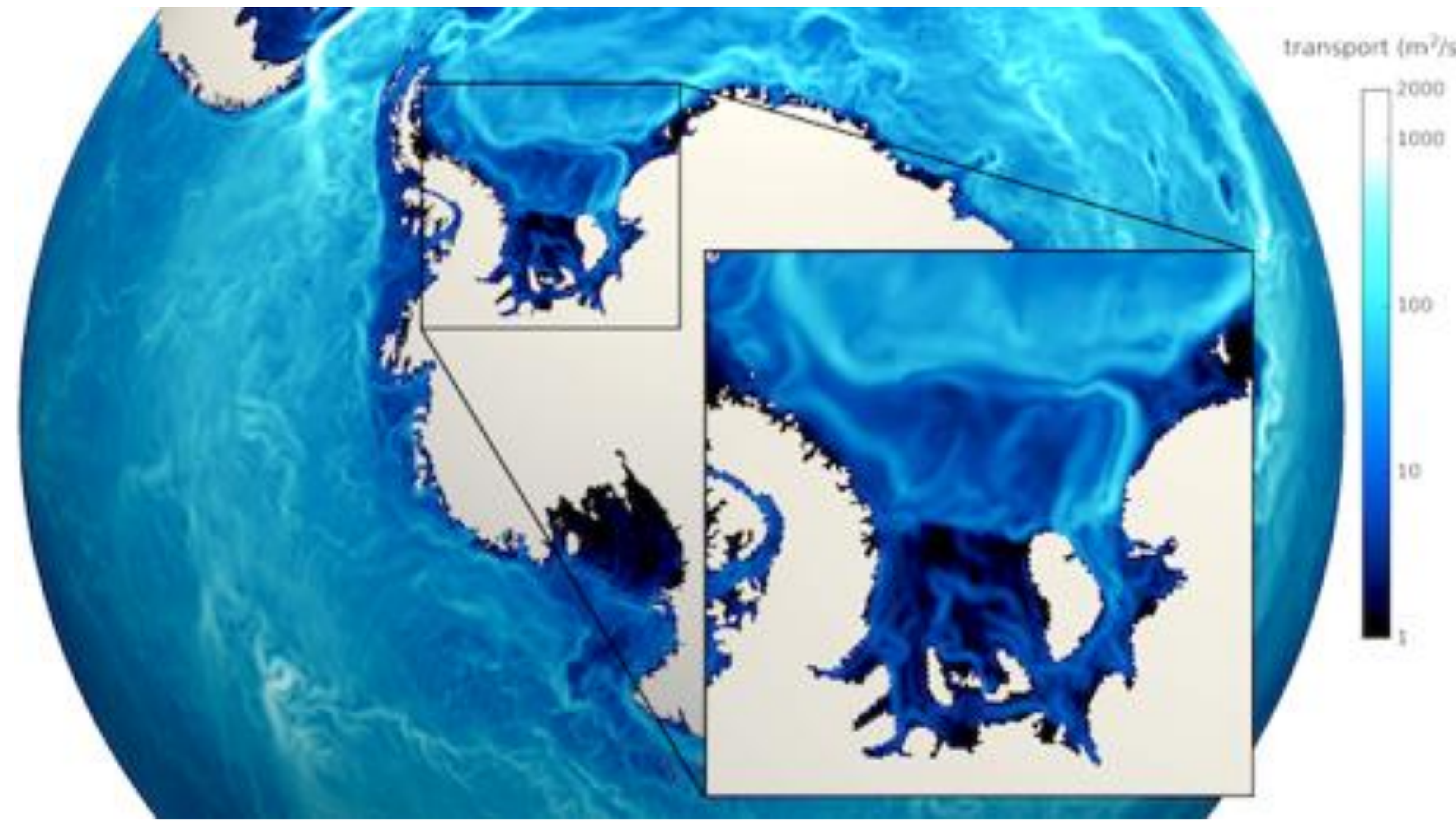
Notation: Let $p_1 = (x_1, y_1)$ and $p_2 = (x_2, y_2)$ be 2 pixel locations given by their 2D coordinates on the Antarctic Ice Sheet. Note that each pixel spans *10* km. Ice melt flux time-series at these locations: $F_1 = (f_1^1, \ldots, f_1^M)$ and $F_2 = (f_2^1, \ldots, f_2^M)$, where M denotes the number of time-steps in the simulation, for which for this data is a monthly resolution over *150* years, for *M = 1800* time-steps. Normalized versions denoted: $\widehat{F_k} = \frac{F_k}{max\ F_k}$ for $k \in \{1,2\}$, and the spatial-distance threshold denoted: d_{thr}^s.

Aggregate spatiotemporal distance criterion

$$d_{s,t}(p_1, p_2) = \begin{cases} \sum_{k=1}^{M} |\hat{F}_1(k) - \hat{F}_2(k)| & \text{if} \quad ||p_1 - p_2|| \leq d_{thr}^s \\ +\infty & \text{else} \end{cases}$$

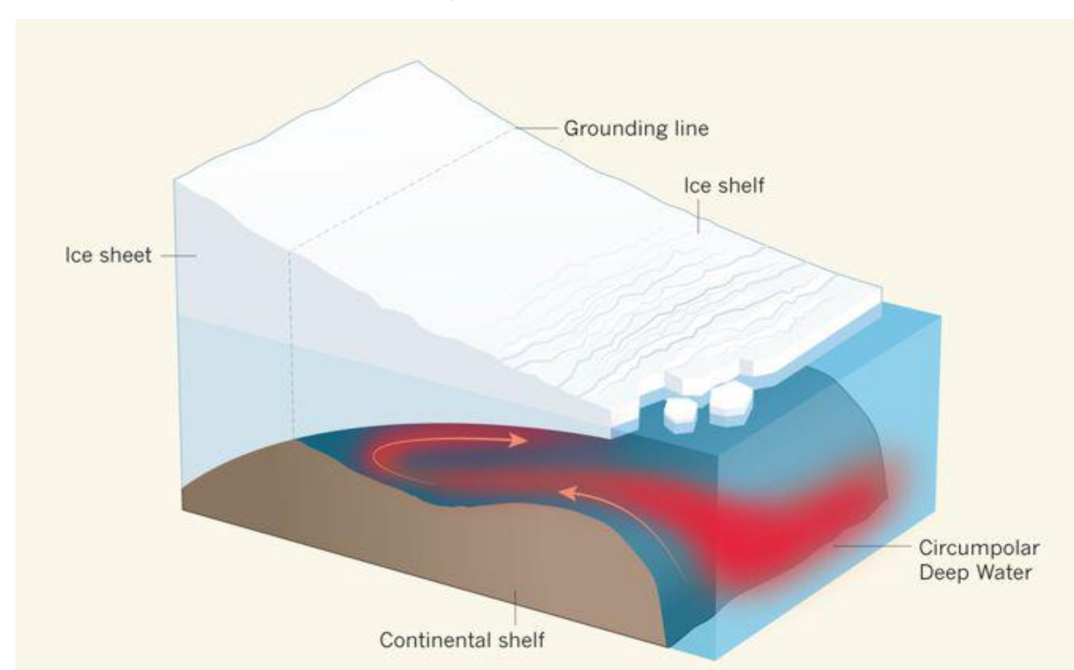
Stationarity: Evaluate the *stationarity of each identified cluster* using the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) hypothesis test with the null hypothesis that the time-series is stationary around the mean. A cluster is considered to be stationary if the majority of its pixels pass the KPSS test.

Data: Energy Exascale Earth System Model (E3SM)



Ocean currents around Antarctica simulated by E3SM

This model has the capability of simulating global modes of climate variability while also including regions of fine spatial resolution of particular interest. The simulation used in this study is particularly focused on resolving ocean-induced melting beneath ice shelves, the floating appendages of the Antarctic Ice Sheet which are in contact with the ocean. For initial simulations, we use a re-gridded version of the model output at 10 km resolution from a native resolution of 30 km (considered "fine" by Earth System model standards, though still relatively coarse for resolving km-scale ice sheet features). We only consider a single field, the freshwater flux at the upper boundary of the ocean surface where it is in contact with ice sheets. The simulation represents a constant climate pre-industrial run beginning in 1850 and spanning 150 years with output generated monthly, for a total of 1800 time-steps in the data set. Because we only have one realization from this model, we will construct a machine learning approach to identify and characterize its stationary spatiotemporal dynamics, thereby making it possible to generate additional data that is both realistic and representative.



Anatomy of a melting ice shelf: Ice shelves are suspended on the ocean from the grounding line, where they connect to the ice sheets laying on Antarctic continental land. Warm water flows under the ice shelf, leading to increased rates of ice melt

References

[1] Robel, A., Seroussi, H., Roe, G. Marine ice sheet instability amplifies and skews uncertainty in projections of future sea-level rise. In Proceedings of the National Academy of Sciences, 116(30). 2019.

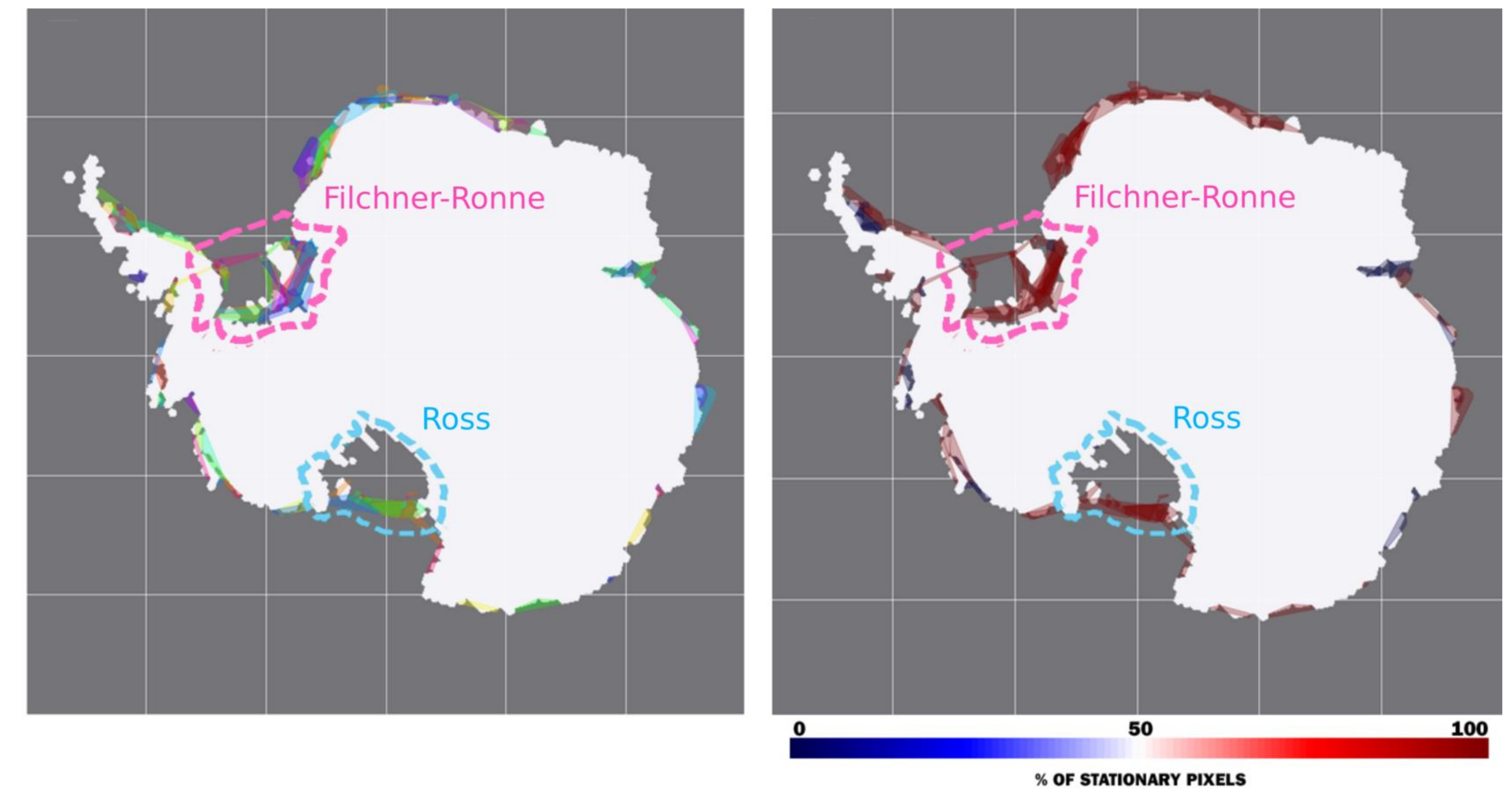
[2] Zepeda-Mendoza, M., Resendis-Antonio, O. Hierarchical Agglomerative Clustering. In Proceedings of Encyclopedia of Systems Biology, 2013.

[3] Chu, M., Xie, Y., Mayer, J., Taixe, L., Thuerey, N. Learning temporal coherence via self-supervision for GAN-based video generation. In Proceedings of ACM Transactions on Graphics, Volume 39, Issue 4, 75(1-13). 2020.

[4] Binkowski, M., Sutherland, D., Arbel, M., Gretton, A. Demystifying MMD GANs. In Proceedings of the ICLR 2018 Conference Blind Submission, 2018.

E3SM Energy Exascale Earth System Model

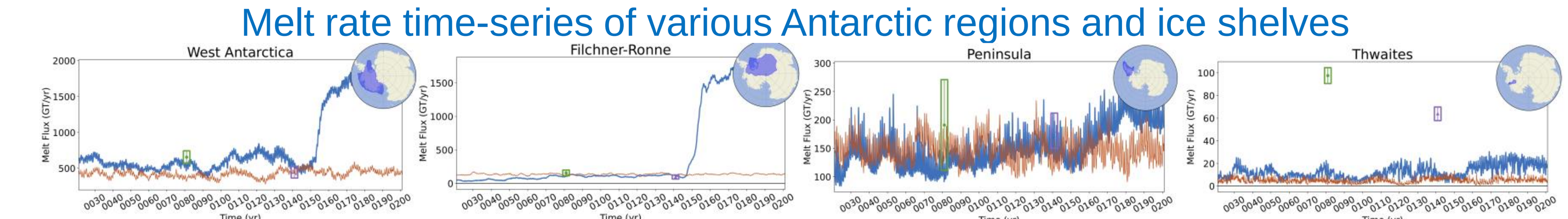
Results: clusters identified are representative of spatiotemporal dynamics and realistic



Results of the proposed dynamic hierarchical clustering pipeline applied to the Antarctic Ice Sheet: The two largest ice shelves are outlined: Filchner-Ronne (pink) and Ross (cyan). Left: clusters identified (46) with hierarchical agglomerative clustering, with the convex hulls representing regions connecting pixels with maximum distance of 50 km (as set by d_{thr}^s = 5 pixels) between pixels. Clusters identified are physically consistent. Right: Stationarity analysis of the clusters in (left) via KPSS hypothesis testing per pixel, shown as the percentage of pixels with very high stationarity in red. Approximately 70% of the clusters were deemed stationary.

Clusters physically consistent: - Several clusters encompass entire single small (<100 km) ice shelves, which indicates that the most important variability on these ice shelves occurs over significant fractions of the entire ice shelf --> larger polygons along the northern coast - Whereas at the two largest Antarctic ice shelves (Ross and Filchner-Ronne), there are smaller clusters within the ice shelf that generally occur near the grounding line where the ice shelf is deepest and melt rates highest, or near the ice shelf calving front where the seasonal cycle in sea ice is a strong driver of sub-shelf melt rates --> numerous small polygons in Ross and Filchner-Ronne

Melt rate time-series of various Antarctic regions and ice shelves



Conclusions / impact

- This work builds a dynamic hierarchical clustering approach that tackles core challenges to the generation of data using limited model output (a single realization) from a state-of-the-art Earth System model, E3SM – the approach identifies stationary subspaces of the input data that are realistic (physically consistent) and representative of its complex spatiotemporal dynamics.
- Promising first steps towards generating additional realizations of variable Antarctic sub-ice shelf melt in a computationally affordable fashion, and consequently, towards addressing the general and pervasive problem of data scarcity in the climate sciences.