Response to RC1: MS No.: egusphere-2026-1250

Andrew Axelsen, Terence O'Kane, Courtney Quinn, Andrew Bassom

May 25, 2026

Authors response to RC1:

We thank the reviewer for their careful examination of the manuscript.

RC1: The manuscript applies a FEM-BV-VAR reduced-order framework combined with a sliding-window transition-matrix approach to investigate persistent atmosphere–sea-ice states during three recent Antarctic low sea-ice years (2016, 2021, and 2023). The topic is timely, and the effort to move beyond fixed temporal averaging is valuable, with a potentially useful methodological contribution in identifying variable-length persistent events. However, the presentation and organization of the manuscript are difficult to follow, and the authors are encouraged to substantially revise the text to improve clarity.

General Comments:

1) The methodological description needs to be improved. At present, it is hard to understand what FEM-BV actually is, how it is implemented, and how the resulting states should be physically interpreted.

Authors: In the revised manuscript much of the technical information pertaining to the FEM-BV-VAR methodology, justification for cross validation approach and determination of hyperparameters, has been moved to Appendix A with additional explanation to improve clarity. We still present the motivation for using the FEM-BV-VAR approach in the main body of text.

2) The manuscript lacks clarity regarding the training and validation strategy of the machine learning framework. In general, ML-based approaches require a clear separation between training and validation (or testing) datasets to ensure robustness and avoid overfitting. While the authors mention the use of cross-validation in terms of RMSE minimization, it remains unclear how the data are actually split (e.g., temporally, randomly, or using block cross-validation), and whether the temporal dependence in the data is properly accounted for. This is particularly important for climate time series, where autocorrelation can bias standard validation approaches. A more detailed description of the data partitioning strategy and its implications for model robustness is needed.

Authors: The process we use is the same as mentioned in Appendix B of [4]. To paraphrase what was mentioned there: *We select a single set of values for the hyperparameters K, m, and p, we use the following cross-validation method. The observed sample is divided into $N_{fold}+1$ approximately equal length segments $\mathcal{T}_1, \ldots, \mathcal{T}_{N_{fold}+1}$, and each model is refit N_{fold} times, where on the ith iteration the first i segments are used as the training sample. Holding the obtained state parameters $\hat{\Theta}$ fixed, the optimal affiliations are calculated by minimizing the cost function evaluated* *over the $(i+1)$th segment, adjusting the upper bound C_T as appropriate for the length of the segment with fixed p. The weighted root-mean-square error*

$$\mathrm{RMSE}_i = \sqrt{\frac{1}{d(T_i - m_{max})} \sum_{t \in \mathcal{T}_{i+1}} \sum_{j=1}^{K} [\gamma_t]_j ||\mathbf{x}_t - \hat{\mathbf{x}}_t^{(j)}||^2} \tag{1}$$

is then evaluated for each test segment, where $\hat{\boldsymbol{x}}_t^{(j)}$ denotes the expected value under state j. The mean reconstruction RMSE over the set of test sets provides a measure of the model's ability to generalize to future data, which we use in lieu of estimates of out-of-sample prediction error, with good performance on this measure involving a compromise between model flexibility and overfitting the training data. We note that the more standard cross-validation approach, that is estimation of the out-of-sample forecast error, would require an additional model for the dynamics of the hidden switching process, which we here leave to future work. Alternatively, in-sample measures based on information criteria could be used when combined with an appropriate likelihood model. However, this similarly requires an appropriate probabilistic model to be specified for the switching and noise processes, and, moreover, the very large number of estimated degrees of freedom in comparison to the available sample size may lead to concerns as to their suitability [1]. We clarify this process and include the reference in the revised manuscript.

3) The presentation of the results is currently somewhat limited and relies heavily on qualitative interpretation of composite maps. While the figures provide useful visual information, the analysis would benefit from more quantitative diagnostics to support the main conclusions.

Authors: We have incorporated a much more detailed analysis and discussion regarding atmospheric drivers by performing pattern correlations with respect to the canonical definitions of the respective modes (SAM, PSA, etc) and amplitudes of the Amundsen Sea Low. The pattern correlation tables in Appendix B of the revised manuscript will provide quantitative measures to aid in interpretation of the slow dynamics. Teleconnection indices such as the SAM Index are typically defined by a three-month running mean thereby removing highly variable daily fluctuations. Here, we contend that a meaningful quantitative measure of the relative contribution of a teleconnection to a given metastable state is the correlation between the observed pattern calculated directly from daily reanalysis fields and that of a given canonical teleconnection index.

As an example, we show in Table 1 pattern correlations for each event in 2016

4) The analysis is limited to three specific years (2016, 2021, and 2023), which represent extreme low sea-ice conditions. While these case studies are relevant, the sample size is relatively small, raising concerns about the generality and robustness of the conclusions. It would strengthen the manuscript to extend the analysis to a broader range of years, including both extreme and more typical conditions. In particular, a quantitative comparison between extreme low sea-ice years and climatologically normal years would help to better assess whether the identified patterns are robust features or case-specific behavior.

Authors: We did not intend to make general statements from our case study-specific findings, so to that end, we have changed the title of the updated manuscript to "Covariations between persistent synoptic features and record low Antarctic sea ice events via unsupervised regression learning". We will note that a more general study attempting to identify common drivers for extreme low, high or neutral years is an interesting topic, but our main focus of this study is

r	SAM	PSA1	PSA2	wave-4a	wave-4b	ASL	ZW3A	ZW3B
event 1	0.464	0.339	0.439	0.082	-0.082	0.102	-0.383	0.044
event 2	0.615	0.178	-0.148	0.156	0.178	0.885	-0.053	-0.194
event 3	0.779	-0.006	0.189	0.069	0.037	0.946	0.063	0.078
event 4	0.440	0.455	0.041	-0.194	-0.068	-0.269	-0.159	-0.202
event 5	0.694	-0.115	0.230	0.072	-0.098	0.915	0.122	0.139
event 6	0.774	0.042	-0.114	-0.093	-0.159	0.739	0.216	-0.069
event 7	0.386	-0.191	-0.119	0.075	-0.483	0.590	0.239	0.020
event 8	0.010	0.396	0.370	0.041	0.145	-0.384	-0.034	0.304
event 9	0.719	-0.345	-0.044	-0.013	-0.039	0.921	0.312	0.124
event 10	-0.548	-0.197	0.099	0.127	-0.214	-0.072	-0.112	0.047

Table 1: Pattern correlations (PSA1 & 2, SAM, wave-4a & b, ASL, ZW3A & B for each event occurring in 2016 compared to each of the atmospheric modes mentioned in appendix B of the revised manuscript. We color red all correlation coefficients with a magnitude greater than 0.4 to highlight those of sufficient significance that we can be confident that the pattern is embedded in the event. In the cases where we have a significant positive correlation with ASL, we refer to the amplitude of the surface pressure anomaly to confirm confidence in that particular atmospheric mode.

introducing a methodology that can identify the relevant drivers on a case-specific basis. We also note that the record lows have only occurred very recently in the observational record imposing a severe restriction on the number of available samples. A preliminary analysis of several years of overall neutral sea ice extent anomalies indicate the zonal SAM pattern tends to be more dominant throughout those years relative to the record low years examined here. The methodology is general, though in the sense that it is able to pull out multiple atmospheric drivers across a range of temporal and spatial scales with a single method.

Specific Comments:

1) Lines 131-152: Please clarify the provenance and reliability of the NNR1 sea-ice concentration data before 1979. Was the FEM-BV model trained on the full 1959-2024 period?

Authors: We have performed a sensitivity analysis comparing the model trained on the both the full 1959-2024 reanalysis period, as used in the manuscript, and when restricted to the satellite era, finding no significant difference in the final root mean square errors of those two models or the optimal model class in general.

2) Lines 155-173: The preprocessing needs more detail. Please specify how anomalies are computed, what climatological base period is used, how the daily/unit matrix normalization affects amplitude information, and how many PCs are retained and why?

Authors: Anomalies are calculated relative to a base period of 1959-2024. For atmospheric variables and sea-ice concentrations, we take the leading 20 PCs and perform a multivariate singular value decomposition (SVD). We normalise the anomalies prior to performing the SVD by making the datasets have a unit matrix norm in time and level as the SVD includes geopotential height and sea-ice concentration with disparate units without which the atmosphere dominates to the exclusion of any sea ice variability. The 20 leading PCs correspond to 56% of the total explained variance across all variables. This is sufficient to capture the major low-frequency modes on timescales synoptic and longer, (e.g. blocking events), while faster-smaller scale features less relevant to persistent states are ignored. This approach to dimension reduction has been shown to be appropriate for increasing signal to noise in the Southern Hemisphere tropospheric flow [2, 3]. For the multivariate case, it is expected that there is more noise to filter due to fast subscale processes not associated to the subsystem coupling.

3) Lines 211-219: The choice of the final day as the representative day is reasonable; however, it would be helpful to assess how sensitive the results are to this assumption. For example, would using the first day instead lead to different identified patterns?

Authors: If we set t_F to be the first day of the window, the temporal sequence of events will be shifted by 30 days such that an event that takes place over the 1^{st} of June to the 15^{th} of June, say, will now span between the 2^{nd} of May and the 16^{th} of May (a full 30 days earlier). We find that our results from the manuscript are overall insensitive to the choice of t_F with the pattern correlations generally identifying the same dominant drivers in events. For two given equally-sized sets of "pattern" data B_i and C_i, with $i \in \{1, ..., V\}$, we define the pattern correlation coefficient $r(B, C)$ between the two patterns by

$$r(B,C) := \frac{\sum_{i=1}^{V} B_i C_i}{\sqrt{\sum_{i=1}^{V}(B_i)^2}\sqrt{\sum_{i=1}^{V}(C_i)^2}}, \tag{2}$$

and we consider $|r| > 0.4$ to be significant. When changing t_F to the first day of the window in 2016, we find that the southern annular mode (SAM) and Amundsen sea low (ASL) still play a role in the record low at the end of the year (e.g. Event 6 in the Winter correlates with SAM [$r = -0.718$] and ASL [$r = 0.915$, $\min(Z'_{1000}) = -3,338$]). The year of 2021 still has a persistent ASL throughout the latter half of the year that drives the record low (For events 8 through 11, we have $r \in [0.718, 0.934]$ with $\min(Z'_{1000}) \in [1,519, 1,653]$). Finally 2023 still has a transition between various coherent structures during the latter half of the year, transitioning between PSA patterns (Event 6 $r = -0.430$ for PSA2, Event 7 $r = 0.649$ for PSA1) before ending with a SAM event ($r = 0.831$ for the final event).

Ultimately, setting t_F as the final day in the window is more consistent with our use of the measure in that the most dominant state is calculated from previous observations rather than being a predictor for future behaviour. The choice of final day also results in smooth state affiliation probabilities that more closely match those found from application of LOWESS filtering as in [2]. We have revised the manuscript to discuss the implications of this choice which apply to any application of Markov transition matrices to time series.

References

[1] K .P Burnham and D .R Anderson. *Model selection and multimodel inference: a practical information-theoretic approach.* Springer, 2002.

[2] C. L. E Franzke, T. J O'Kane, D. P Monselesan, J. S Risbey, and I Horenko. Systematic attribution of observed Southern Hemispheric circulation trends to external forcing and internal variability. *Nonlin. Processes Geophys.*, 2:675–707, 2015.

[3] T. J. O'Kane, J. S. Risbey, C. L. E. Franzke, I. Horenko, and D. P. Monselesan. Changes in the metastability of the midlatitude Southern Hemisphere circulation and the utility of non-stationary cluster analysis and split-flow blocking indices as diagnostic tools. *Journal of the atmospheric sciences*, 70(3):824–842, 2013.

[4] C Quinn, D Harries, and T .J O'Kane. Dynamical analysis of a reduced model for the North Atlantic Oscillation. *Journal of the Atmospheric Sciences*, 78(5):1647–1671, 2021.