Response to EC1: MS No.: egusphere-2026-1250

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

May 25, 2026

Jie Feng: I have now received the comments from all three reviewers. All of them expressed concern that the manuscript lacks critical information, such as details on the methodology. In particular, Reviewers 1 and 2 both emphasized that the results are presented too qualitatively and that key physical interpretations are clearly insufficient. Given that addressing these weaknesses would require considerable time, I recommend rejecting the manuscript at this stage. However, I would encourage resubmission after the authors have substantially improved the overall quality and readability of the paper.

Authors: Based on the submission distributed to the reviewers, we understand the decision to reject encouraging resubmission of our manuscript, however, from specific references to line numbers by the reviewers, it is certain that the version of the manuscript distributed was a preliminary draft that appeared on the ESSOR archive and not the version intended for review which appears on the archive as v2. The confusion of paper versions on the archive appears to have come about at the submission stage. The error was picked up almost immediately and the updated version intended for review was uploaded to the archive within a day. This was communicated to the editorial support team but, was clearly not communicated to the reviewers.

We have however, prepared detailed responses to the reviews and made additional revisions to address the points raised, noting that many, if not most, of the concerns were addressed in the manuscript intended for review.

Rather than resubmit with a new set of reviewers, we would request that we be allowed to respond and to continue the review process as we have a substantially revised and augmented paper already prepared. Most importantly, the revised submission contains many quantitative diagnostics, extended methods sections, and physical analysis not present in the variant incorrectly supplied to the reviewers.

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.

Response to RC2: MS No.: egusphere-2026-1250

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

May 25, 2026

RC2: This manuscript investigates the relationship between persistent synoptic-scale atmospheric features and recent Antarctic sea-ice anomalies, with particular emphasis on the low sea-ice years 2016, 2021, and 2023. The authors develop a reduced-order coupled atmosphere-sea-ice model using reanalysis data and apply the FEM-BV-VAR framework together with a non-stationary transition-matrix approach to identify persistent background states of varying duration. The manuscript argues that persistent synoptic patterns strongly influence the rate and spatial structure of sea-ice variability. During periods of relatively weak synoptic variability, a large-scale Antarctic warming pattern becomes the dominant influence on sea-ice retreat. The topic is interesting and potentially important. However, in its current form, I do not think the manuscript is yet ready for publication. My main concern is that the physical interpretation remains insufficiently developed, and several key methodological and diagnostic issues need to be addressed before the conclusions can be considered robust.

Authors: We thank the reviewer for the effort taken to examine the manuscript.

Comments:

1. The most important weakness of the manuscript is that the results are primarily presented as statistical or pattern-based associations, but the underlying physical processes are not sufficiently explained. At present, the analysis supports physically plausible relationships, but it does not yet provide a sufficiently clear process-based interpretation of how the identified synoptic states lead to sea-ice growth or retreat. The paper would be much stronger if the authors could provide more detailed discussion of the relevant physical mechanisms, for example through thermodynamic, dynamic, or regional process diagnostics.

Authors: The paper has been revised to better emphasise and quantify a process-based interpretation of the mechanisms by which the identified synoptic states lead to sea-ice growth or retreat. This includes additional references to recent literature and earlier process based modelling studies by some of the authors and additional quantitative measures of the contribution of relevant teleconnections to the observed persistent states. To facilitate this there has also been some reorganisation of the methods to appendices with additional technical details as requested by another reviewer. A key point in applying regression learning to data is that physical processes are inferred based on the methodological approach. Our approach is to infer plausible physical relationships based on the manifestation of persistent synoptic states. Where an emergent structure has a significant pattern correlation to a canonical teleconnection pattern we argue it is reasonable to assume that teleconnection is dominant over the period identified given the averages are over reanalysed daily anomalies and the reference patterns are the commonly accepted descriptors of the particular teleconnections. Thus, our approach allows us to study multiple teleconnections and their influence at once rather than studying the individual influence of each teleconnection.

2. Related to the previous point, the manuscript often goes beyond what the analysis can firmly support. The current framework is useful for identifying covariation and persistent background states, but it does not by itself establish strict causal attribution. Some conclusions therefore appear overstated. The authors should be more cautious in distinguishing between statistical association, dynamical interpretation, and formal attribution.

Authors: Here we are applying optimisation methods to fit a stochastic model to reanalysis data then applying Markov transition matrices to affiliations of each successive data point to states determined from an optimal linear vector autoregressive model. We make no claims regarding causality. We note that a future study could use the generated Markov model to measure predictive information allowing for estimates of Granger causality; we only conjecture around the physical implications of identified persistent atmospheric events preceding a sea-ice loss event. We have carefully revised our paper to ensure there is no confusion about this point.

3. The manuscript refers to a "climatological anthropogenic warming pattern" as a dominant influence during periods of weak synoptic variability. However, the present study does not perform a formal attribution analysis separating externally forced and internally generated variability. Therefore, this terminology is too strong in the current context. The authors should either soften this statement or provide stronger justification for using such wording.

Authors: We agree on this point. We have revised the manuscript to soften the statement simply noting there are periods of sea ice loss coincident with periods of relatively quiescent winds which are also coincident with anomalously warm surface temperatures.

4. Data: The study uses daily SIC from NCEP, but the comparison with NSIDC appears to be limited to monthly-scale analysis. Monthly comparison alone is not sufficient, especially given that the manuscript focuses on synoptic events. The authors should provide a more detailed validation of the daily SIC fields, including spatial patterns and event-scale behavior, rather than relying only on monthly anomaly comparison. For atmospheric fields, it is not clear how robust the results are to the choice of reanalysis. The manuscript should discuss whether the same persistent patterns and event structures would be obtained using ERA5 instead of NCEP, or at least assess the differences between the two products for the key variables used in the analysis. At present, the sensitivity of the conclusions to the reanalysis dataset remains unclear.

Authors: A detailed comparison of differences between NSIDC and NCEP is beyond the scope of this study, however, we note that the NSIDC website states "NSIDC currently uses the National Centers for Environmental Prediction-National Center for Environmental Research (NCEP-NCAR) Reanalysis 1 for assessments of atmospheric circulation and temperature". On that basis we assume that NNR1 is a sufficiently reliable and consistent dataset to base our study, and thus it is the most appropriate to consider in the first instance. In terms of differences between the NNR1 atmosphere versus ERA5 or other products, there are many varied reanalysis intercomparison studies, many relevant to this study are focussed on particular variables e.g., the intercomparison study of surface air temperatures over Antarctica [5], or for the Southern Hemisphere troposphere over the pre-satellite record including as in Appendix A of [9]. It is not unreasonable to expect differences in the metastable states between different products. For example [1] report major artifacts in ERA5 2-m air temperature trends over Antarctica prior to and during the modern satellite era which would undoubtedly lead to differences to those reported here. However, attribution of differences in reanalysis requires understanding the assimilation systems and underlying model biases [11, 2, 8, 4], and relationships between teleconnections [3] such that a detailed intercomparison that includes attribution of the cause of any differences between products is not within the scope of this study.

5. The manuscript spends considerable space describing the methodology, while the physical interpretation of the results is comparatively brief. I suggest streamlining the methodological description in the main text and moving some technical details to an appendix or supplement, in order to leave more room for discussion of the physical implications of the identified patterns.

Authors: Agreed. The FEM-BV-VAR methodology has been moved to an appendix to address another reviewer's request for more details on the methodology and to streamline the main body of the manuscript. Additional analysis and diagnostics have been added to improve physical interpretation and correspondence to recent literature.

6. The identified events have very different durations. This raises an important question: are the events with different lengths dynamically comparable? Do they represent similar mechanisms, or are they fundamentally different types of processes being grouped within the same framework? The authors should discuss whether the same interpretation applies across events of very different duration, and whether duration itself carries physical significance.

Authors: The duration of given events is only an indication of their relative persistence. For example, two distinct events of similar amplitude and persistence associated with the same statistically stationary state may exhibit quite distinct local dynamics. This is a key point and can easily be understood from the perspective of PCA or k-means. As an example, an EOF is an invariant pattern associated with a statistically stationary state that explains a specified fraction of the total variance. FEM-BV-VAR performs a global optimisation to assign sequential data instances for given embeddings to a finite number of locally stationary states. This means the patterns themselves are not invariant but time evolving with a probability that any particular data instance resides in one or more metastable states. Hence in post processing patterns are determined directly from the reanalysis anomalies. Therefore the correct interpretation is that varying local synoptic processes are being grouped within a common global clustering framework. That is locally diverse dynamics can have common global affiliations to given teleconnections.

7. The manuscript focuses strongly on atmospheric driver, but the role of the ocean and sea-ice dynamics is not sufficiently considered. Given the current understanding of Antarctic sea-ice variability, it is important to clarify to what extent the identified atmospheric patterns act alone, and to what extent oceanic forcing or sea-ice dynamical processes may also be essential. At minimum, the limitations of excluding these factors from the main analysis should be discussed more explicitly.

Authors: As the revised title, "Covariations between persistent synoptic features and Antarctic sea ice via unsupervised regression learning" suggests, this work is solely focusing on the role of the atmosphere on case studies of recent record low sea-ice events. Given the paucity of subsurface ocean observations of sufficient density to adequately constrain reanalysis of subsurface ocean dynamics, coupled model simulations are required and beyond the scope of the current study. We now include additional discussion of such modelling studies for example the earlier general modelling studies of [10] & [6] and the most recent study of record low sea ice using an ocean model constrained by observed surface forcing [7]. A natural follow-up study could include applying this methodology to ocean-SIC coupled data to find specific oceanic drivers on longer timescales. This, however, would require more careful consideration of the ocean data used as psuedo-observations as mentioned above.

8. The wind vectors in the figures are very difficult to read. In their current form, I cannot clearly determine their direction or magnitude. The figure design should be improved. Moreover, some figures appear to lack units. All plotted variables should include clear and consistent units in the color bars, captions or axis labels.

Authors: We have modified the figures to be more legible with regards to the arrows and checked all units are accounted for.

References

[1] D .H. Bromwich, A Ensign, S-H Wang, and X Zou. Major artifacts in era5 2-m air temperature trends over antarctica prior to and dduring the modern satellite era. *Geophysical Research Letters*, 51(21):e2024GL111907, 2024. e2024GL111907 2024GL111907.

[2] G. Compo and Coauthors. The twentieth century re-analysis project. *Quart. J. Roy. Meteor. Soc.*, 137(654):1–28, 2011.

[3] D. Harries and T. J. O'Kane. Dynamic Bayesian Networks for Evaluation of Granger Causal Relationships in Climate Reanalyses. *Journal of Advances in Modeling Earth Systems*, 13:1–37, 2021.

[4] H Hersbach, B Bell, P Berrisford, S Hirahara, A Hor'anyi, J Mu noz-Sabater, J Nicolas, C Peubey, R Radu, D Schepers, A Simmons, C Soci, S Abdalla, X Abellan, G Balsamo, P Bechtold, G Biavati, J Bidlot, M Bonavita, G De Chiara, P Dahlgren, D Dee, M Diamantakis, R Dragani, J Flemming, R Forbes, M Fuentes, A Geer, L Haimberger, S Healy, R .J Hogan, E H'olm, M Janiskov'a, S Keeley, P Laloyaux, P Lopez, C Lupu, G Radnoti, P de Rosnay, I Rozum, Fr Vamborg, S Villaume, and J Th'epaut. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020.

[5] B Huai, Y Wang, M Ding, J Zhang, and X Dong. An assessment of recent global atmospheric reanalyses for antarctic near surface air temperature. *Atmospheric Research*, 226:181–191, 2019.

[6] R. J. Matear, T. J. O'Kane, J. S. Risbey, and M. A. Chamberlain. Sources of heterogeneous variability and trends in antarctic sea-ice. *nature communications*, 6(8656):1–9, 2015.

[7] A Narayanan, H Ayres, M .H England, F .A Haumann, M .R Mazloff, A Silvano, T Spira, S Zhou, and A .C Naveira Garabato. Compound drivers of Antarctic sea ice loss and Southern Ocean destratification. *Science Advances*, 12(eaeb0166):1–14, 2026.

[8] T .J O'Kane and Coauthors. CAFE60v1: A 60-year large ensemble climate reanalysis. Part I: System design, model configuration, and data assimilation. *J. Climate*, 34(1):5153–5169, 2021.

[9] T .J O'Kane, J .S Frederiksen, C .S Frederiksen, and I Horenko. Beyond the first tipping points of Southern Hemisphere climate, 2024.

[10] T. J. O'Kane, R. J. Matear, M. A. Chamberlain, J. S. Risbey, B. M. Sloyan, and I. Horenko. Decadal variability in an OGCM Southern Ocean: Intrinsic modes, forced modes and metastable states. *Ocean Modelling*, 69:1–21, 2013.

[11] S Saha, S Moorthi, H-L Pan, X Wu, J Wang, S Nadiga, P Tripp, R Kistler, J Woollen, D Behringer, H Liu, D Stokes, R Grumbine, G Gayno, J Wang, Y-T Hou, H-Y Chuang, H-M .H Juang, J Sela, M Iredell, R Treadon, D Kleist, P Van Delst, D Keyser, J Derber, M Ek, J Meng, H Wei, R Yang, S Lord, H van den Dool, A Kumar, W Wang, Long C, M Chelliah, Y Xue, B Huang, J-K Schemm, W Ebisuzaki, R Lin, P Xie, M Chen, S Zhou, W Higgins, C-Z Zou, Q Liu, Y Chen, Y Han, L Cucurull, R .W Reynolds, G Rutledge, and M Goldberg. The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, 91:1015–1058, 2010.

Response to RC3: MS No.: egusphere-2026-1250

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

May 25, 2026

Authors: We thank the reviewer for the effort taken to examine the manuscript and constructive comments.

RC3: The manuscript considers three years (2016, 2021, 2023) with anomalously low Antarctic sea-ice extends and aims to identify atmospheric patterns that played a role in these events. The approach of the authors is based on the FEM-BV-VAR learning algorithm. The analysis seems interesting , however, the methodology is explained insufficiently to the point that it is hard to pinpoint exactly what the results of the manuscript are. The main aspects that remain unclear are as follows:

1) The FEM-BV-VAR learning algorithm is mentioned, and Axelsen et. al. (2025) is cited, but there is no description of what the algorithm does. It is not necessary to describe the algorithm in detail; for details a reference is sufficient; but the following point should be clear: What is the input to the algorithm? What is the output of the algorithm? What are the parameters K, M, p, and what role to they play in the algorithm? Similarly, the principal component analysis should be elaborated on: What objects are the principal components? How are they determined? Overall, the introduction of the methodology would greatly benefit from some more rigorous mathematical notation.

Authors: In the updated manuscript, we provide an expanded description of the FEM-BV-VAR method in its own appendix, as well as additional details on the principal component analysis and dimension reduction.

2) The paragraph 167-174 seems to describe the construction of some matrices which would become much more comprehensible with some more mathematical notation.

Authors: In the updated manuscript we provide additional details on the PCA, including construction of the multivariate singular value decomposition and baseline climatology for calculation of anomalies and dimension reduction. To paraphrase: 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. 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 PCs retained 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 treated as noise. This approach to dimension reduction has been shown to be appropriate for increasing signal to noise in the Southern Hemisphere tropospheric flow [3, 4].

3) Equation (1) defines the root mean square error. It seems that the model output o_j and the training data s_j have not been mentioned before. It should be clarified exactly what the model output is and what is being used as the training data.

Authors: FEM-BV-VAR is used to fit a model to NNR1 data in order to assign data instances to locally stationary states. The model outputs a set of parameters Θ and probabilities $\Gamma(t)$ details of which we provide in an Appendix. We also provide a detailed description of the cross validation in an appendix, including specific use of RMSE weighted by the affiliation probabilities. Here we use the same process as mentioned in Appendix B of [5]. To quote 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 [2]. We clarify this process and include the reference in the revised manuscript.

4) In 186-189 the affiliation sequence is mentioned. Apart from $\Gamma(t)$ being a probability vector, it is unclear what this vector describes and how it is obtained.

Authors: The vector $\Gamma(t) = [\gamma_1(t), \gamma_2(t), \ldots, \gamma_K(t)]$ is an output of the FEM-BV-VAR method that represents the affiliation (i.e., likelihood) for the data on a given day to reside in a given cluster state based on the appropriate optimal model parameters. The method used to determine these affiliations will be described in detail in an appendix of the revised manuscript.

5) In the paragraph 203-234 a transition probability matrix in a time window $[t_0, t_1]$ is considered. They are defined by Equation (3) but it is unclear where the probabilities $Pr(k_{t+1} = j | k_t = i)$ come from. Additionally, the right-hand side of the definition depends on $t \in [t_0, t_1]$ while the left-hand side does not depend on t. Additionally, it is unclear how the time windows $[t_0, t_1]$ are determined.

Authors: Equation (3) is simply the number of times we transition from state i to state j over the window $[t_0, t_1]$ divided by the number of times we were in state i. The window $[t_0, t_1]$ is a sliding 30-day window that is applied across the entire dataset (so one window is January 1^{st} through January 31^{st}, then January 2^{nd} through February 1^{st}, etc). These will be clarified in the updated manuscript.

6) It should be elaborated what type of LOWESS method is used and with which parameters.

Authors: We use Python's `statsmodels.nonparametric.lowess` method (an algorithm that uses local linear estimates to generate the smoothed curve) on the raw stationary distribution curves with the fraction parameter set to the window length (30 days) divided by the number of days in the dataset (24,046 days) such that they follow the stationary curves as closely as possible. We will clarify this in the updated manuscript.

Some other remark regarding the manuscript:

7) I recommend to make clearer what the results of the paper are. It seems that the qualitative discussion of the driving factors for anomalously low sea-ice extend is based largely on a direct observation of the data presented in Figures 4-6. Are the GPH and temperature values which are shown in Figures 4-6 output of the model or simply the NNR1 data? If this is the NNR1 data, then it seems that the main contribution of the algorithmic approach is the determination of certain events within the considered timespan. Either way, I recommend including a discussion about which results of the manuscript can be tested in some way (e.g. against some null-hypothesis), and which results are 'validated' based on phenomenological observations.

Authors: The updated manuscript will include quantitative interpretations of the results, including pattern correlation analyses relative to canonical modes of variability (SAM, PSA1&2, etc) and amplitude analyses of the Amundsen Sea Low where appropriate in order to establish confidence that the patterns we obtain correspond to known modes. The output of Figures 4 through 6 are composites of the raw NNR1 data over periods determined by the FEM-BV-VAR method in conjunction with the transition matrix method. We note that the manuscript is a case study of the record low years and the findings of this study cannot be generalised to other years. That said, a preliminary analysis of several years of average sea ice extent indicates the zonal SAM pattern tends to be more dominant throughout the year relative to the record low years examined here.

Authors: We now include a quantitative interpretation of the atmospheric drivers via pattern correlations for each event in 2016 with the relevant teleconnections. As an example we show in Table 2 values for 2016.

8) In Figure 2 it seems that increasing K improves the RMSE. Why have no values of K larger than 4 been considered?

Authors: We build on our recent study of tropospheric dynamics [1] in which 4 metastable states were optimal. The challenge is to find the most parsimonious number of states that have plausible physical interpretability. For a given K, m, p the model class is determined by p and are independent for specified p values. Hence choice of optimal model is not simply dependent on RMSE alone. The optimal number of states can also vary according to metric chosen. For example application of Aikaike of Bayesian inference criteria will more strongly penalise K, favouring fewer states with $K = 2$ most common. Cross validation is less restrictive resulting in $K = 3, 4$.

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.

References

[1] A .R Axelsen, T .J O'Kane, C .R Quinn, and A .P Bassom. Hyperbolicity and Southern Hemisphere persistent synoptic events. *Journal of Advances in Modeling Earth Systems*, 17(4):e2024MS004834, 1–25, 2025.

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

[3] 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.

[4] 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.

[5] 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.