Supplemental Material

Supplemental Material in support of 'Observational evidence for a regime shift in summer Antarctic sea ice'

Supplement section 1 – Further information on regression model

In this section we provide more detail on the regression model presented in *Section 3.1* of the main manuscript; this model was developed to predict summer (DJF) Antarctic sea ice anomalies using monthly indices of large-scale climate variability (SAM, ASL, ZW3, SOI, IOD, and SIA as outlined in manuscript Table 1). The main steps are as follows:

1. Prepare climate indices and convert to Principal Components (PCs)
2. Select PCs to include in model using cross-validation
3. Fit model using only the selected PCs

The model preparation comprised converting each monthly index into separate indices for each calendrer month leading up to the summer (DJF) sea ice period, i.e, the SAM index was separated into a March SAM index, April SAM index, and so on. Note that for the sea ice area (SIA) record, only the nine months preceding summer (i.e., March to November inclusive) are included. This gives a total of 81 timeseries (12 for each of the 6 atmospheric indices, and 9 for the SIA). These 81 timeseries were then transformed into Principal Components (PCs), to account for the fact that some of the indices are highly correlated (e.g., SAM and ASL have strong correlations for each month). Since the indices have different units, the PCs were derived from the correlation matrix rather than the covariance matrix (effectively standardizing each unit).

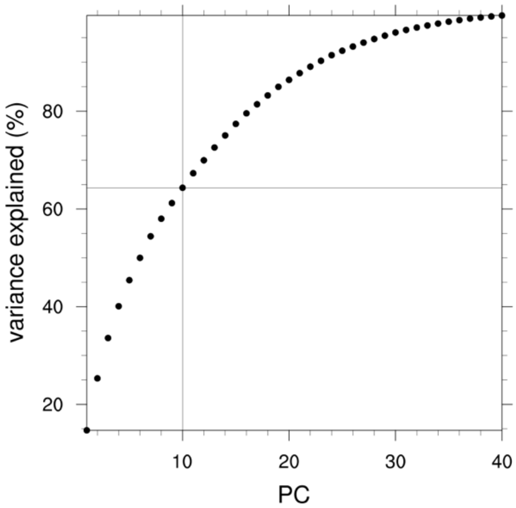

Figure S1 – Cumulative variance-explained amongst the input predictor variables, for all 40 Principal Components. The leading 10 PCs that are used in the final model are highlighted.

The selection of how many PCs to retain was made using a cross-validation approach(Picard and Cook, 1984). Each PC has 44 records (1979-2022). The model was fit omitting one of these records, and then used to make an out-of-sample prediction of the omitted record's summer SIA (e.g., the model was fit using the 1980-2022 sample, and then used to predict 1979). The model was fit using Ordinary Least Squares regression, and before fitting the SIA data were standardized (i.e. divided by the standard deviation after subtracting the mean). Predictions were made for all levels of truncation (i.e., retaining different numbers of PCs). Model fit was made using multiple Least Squares Regression. The process was repeated to make out-of-sample predictions for all 44 years and for different numbers of retained PCs, providing records that can be compared with the observed DJF SIA record (Figure S1). This analysis demonstrates that the model fit is optimized when either 10 or 20 PCs are retained, where optimization is indicated by the maximum fraction of variance explained (R^2) and minimum root-mean-square error (RMS) compared to the observed summer SIA record. Since using more predictors generally increases the possibility of overfitting (e.g. Akaike, 1974), and retaining 20 PCs does not greatly improve the skill above 10 PCs, all model fits presented in the main manuscript were made using the 10 leading PCs. The loading patterns of those PCs are shown in Figure S2).

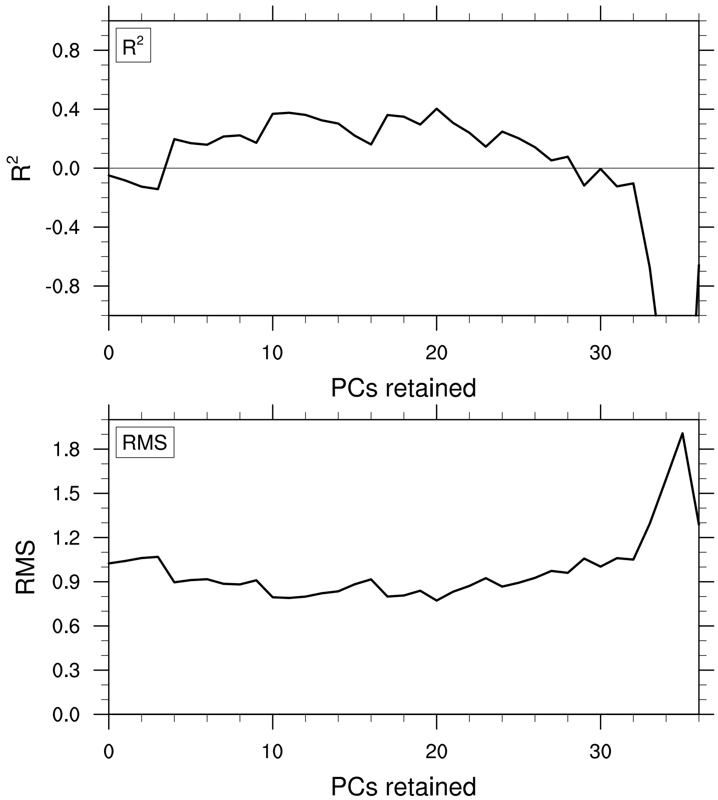

Figure S2 – a) Fraction of variance explained (R^2) and b) root-mean-square error (RMS) between observed and out-of-sample prediction of 1979-2022 Summer (DJF) total Antarctic sea ice area (SIA), for different numbers of retained PCs (x-axes). PC=0 means that only the bias term (unity) was used. Model skill is indicated by maximizing R^2 and minimizing RMS error.

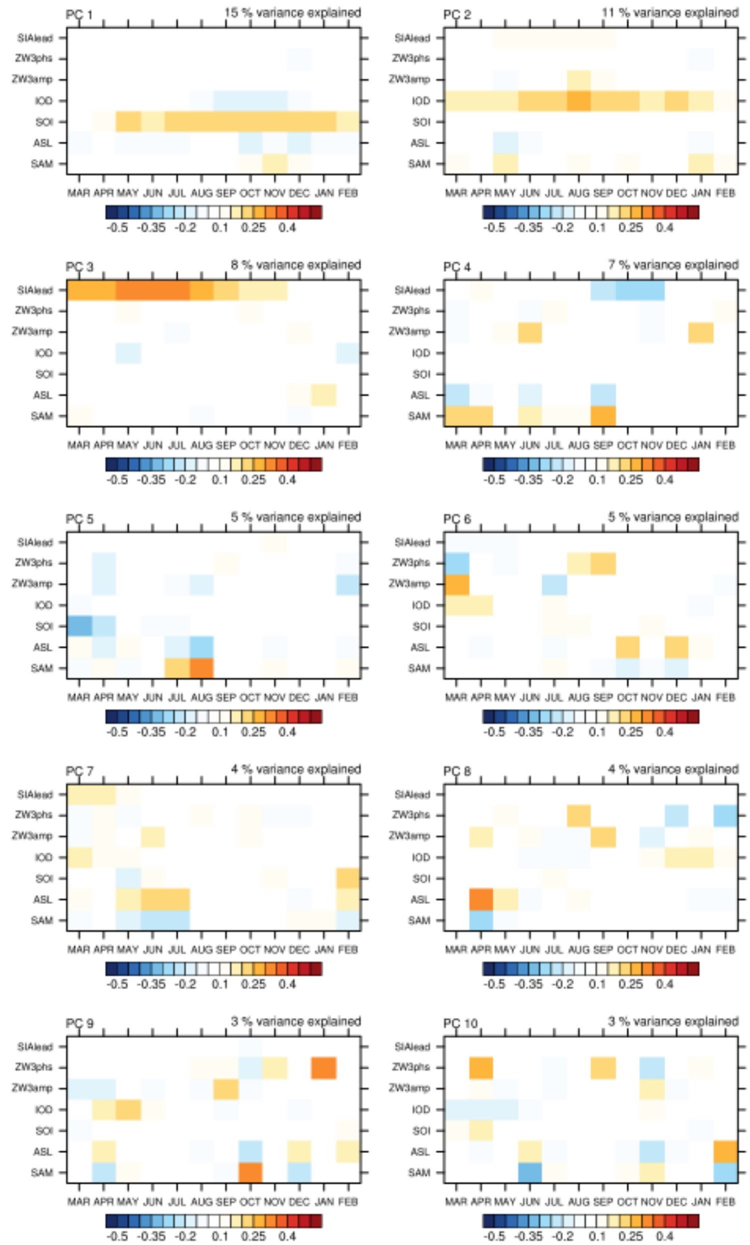

Figure S3 – Loading patterns for the 10 leading Principal Components (PCs) that were included in the regression model in Section 3.1. These patterns indicate the indices and months that are represented by each PC. (For example, panel a) shows that PC1 is dominated by the winter Southern Oscillation Index (SOI), and PC2 (panel b) by the Indian Ocean Dipole (IOD). Variance-explained by each PC is shown in the top right of each panel. The PCs were calculated from the correlation matrix, hence the patterns represent correlations (1 ≥ r ≥ -1); the sign convention is arbitrary.

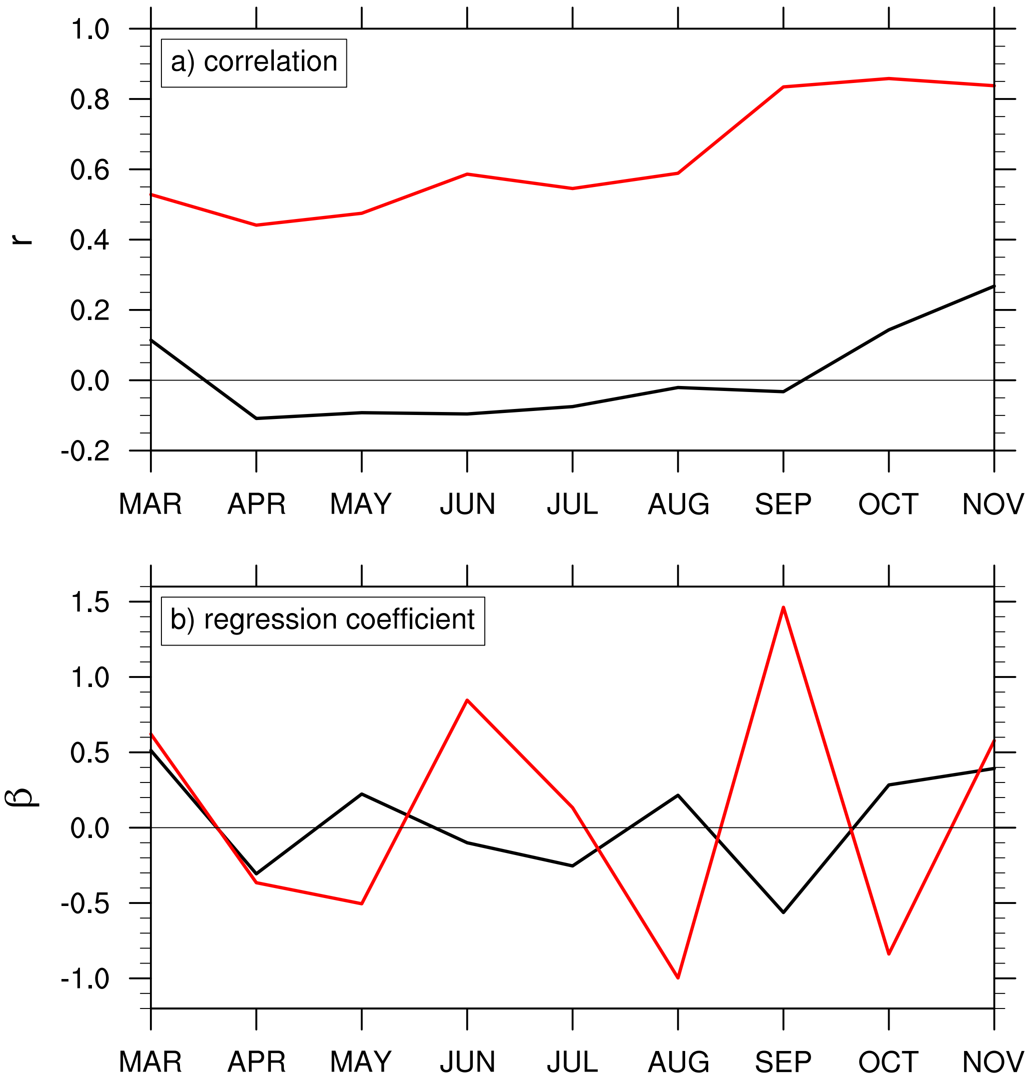

Figure S4 – a) Correlation coefficient and b) regression coefficient between each month's SIA record and the subsequent summer (DJF) SIA, for 1979-2007 (black line) and 2006-2022 (red line).

References

Akaike, H., 1974: A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19,** 716-723, 10.1109/TAC.1974.1100705.

Picard, R. R., and R. D. Cook, 1984: Cross-Validation of Regression Models. *Journal of the American Statistical Association*, **79,** 575-583, 10.1080/01621459.1984.10478083.