Reviewer 1

I thank the reviewer for their constructive comments and especially for the excellent suggestion of using a series of scrambled AO time-series in order to test the statistical significance of the frequency of decades with significant SC-AO correlations.

GENERAL COMMENTS

This study first uses the National Oceanographic and Atmospheric Administration (NOAA) 20th Century Reanalysis version 3 (20CRv3), together with some other observational data sets, to investigate trends in the Eurasian autumn (September to November) snow cover in years 1836-2015. Then, the variations in the interannual correlation between Eurasian autumn snow cover and the following winter Arctic Oscillation (AO) index are studied. The study adds to previous knowledge by (i) extending its analysis further back in time than has been done thus far and (ii) by also looking in some detail at the regional characteristics of the trends and the snow cover – AO relationship. In addition to the monthly mean snow cover extent in September, October and November, the rate of advance of snow cover during October and November is also analysed.

The main findings of the study include (i) an overall decreasing trend in snow cover extent in October and November since 1836, but particularly since 1966; (ii) a tendency for anticorrelated variations in the rate of snow cover advance between October and November, which is also reflected in the trends for 1966-2015 (slower advance in October but faster in November, due to a delay in the onset of the snow season); (iii) highly variable decadal correlations between the studied snow cover indices and the winter AO; and (iv) the rates of October snow advance in western and eastern Eurasia have typically opposite correlations with the subsequent winter AO index.

One may question the meaningfulness of analysing snow cover variations back to the year 1836, when little data was available to constrain the 20CRv3 reanalysis. However, I believe that this analysis still makes sense, particularly for the interannual variations. Assimilating surface pressure observations, the reanalysis should at least qualitatively capture the snow cover variability, which is largely forced by atmospheric circulation on the interannual time scale. For the same reason, the reanalysis also likely qualitatively captures the correlation between the autumn snow cover and the winter AO even in the 19th century, even though the number of pressure observations was much smaller than today.

My main concerns about this study relate to Section 4.2. Specifically, the manuscript suffers from a too uncritical interpretation of the snow cover (SC) – AO correlations. The finding that the decadal correlation between snow cover and AO indices has varied widely with time, between sometimes negative and sometimes positive values for all the five snow cover indices studied (Figure 7), allows for two diverging scenarios:

1. There is a real connection between autumn Eurasian SC and winter AO, but this connection varies with the changing background state of the climate system (e.g., low-frequency variations in SST). The connection may be causal (autumn SC affecting winter AO), or it may reflect a common underlying factor (e.g., interannual variations in SST) that affects both the autumn SC and the winter AO.
2. The connection between autumn Eurasian SC and winter AO is a statistical artefact. When the sample size is small, as it is for 10-year correlations of detrended SC and AO indices, strong correlations sometimes occur by pure chance, occasionally reaching the threshold of statistical significance.

   The manuscript does not discuss these two possibilities explicitly, but the spirit of discussion appears to favour the first explanation. However, the results in Figure 7 and Table 3 call for caution. In Table 3, the proportion of decadal correlations that reach the 10 % significance level varies from 4.2 % (SAI_NOV) to 14.1 % (SC_10). A pure chance would give 10 % on the average, but with a substantial variation around this expected number.

   I explored the potential importance of chance in creating apparently significant correlations by a Monte Carlo method. I created 179-year time series of pseudo-SC and pseudo-AO indices from normally distributed random numbers with no serial correlation (thus, pure white noise). Then I calculated the 170-year time series of running 10-year correlation between the two indices, after detrending them in the same way as in the manuscript, and further counted the proportion of correlations that exceeded the threshold for 10 % statistical significance. I repeated this 100 000 times.

   Within this large Monte Carlo sample, the proportion of statistically significant correlations was 14.1% or larger in 19.9 % of cases. Thus, this fraction of significant correlations could easily be explained by chance. The largest fraction of significant negative correlations (11.8 % for SC_09) turned out to be slightly more unusual, occurring in 5.9 % of the generated time series. However, the probability of getting either at least 11.8 % of negative or at least 11.8 % of positive correlations was nearly twice as large (11.6 %). Finally, the probability of getting a large fraction of significant correlations for at least one of the five SC indices is much larger than that for any specific index alone.

   Thus, based on the results presented in this manuscript, one cannot reject the null hypothesis that Eurasian autumn SC has no impact on winter AO at all. This is an important result as such.

   In the regional analysis, larger fractions of significant correlations are found, occasionally more than 20 % (Figure 8). The Monte Carlo tests described above indicate a chance of only 1 / 300 of getting either at least 20 % of significantly positive or at least 20 % of significantly negative correlations, provided that full 179 years of data are available (which, however, is not the case for all regions). Again, however, the probability of getting such a large fraction of significant correlations for some of the regions and some of the five SC indices is much larger than that for a single region and a single

SC index. Therefore, similar statistical concerns may apply to them as for the whole-Eurasia SC indices in Figure 7 and Table 3.

Obviously, absence of evidence is not evidence of absence. Some connection between Eurasian autumn SC and winter AO may exist, even if it is not convincingly seen in the correlation analysis. However, the lack of clear statistical evidence suggest that this connection is weak, explaining at best a modest fraction of the AO variability.

My recommendations for improving the statistical analysis in Section 4.2 are as follows:

1. Explicitly acknowledge the fact that much of the correlations and their apparent temporal variation may be due to pure chance.

Based on the statistics from the new 'scrambled data' methodology and new Eurasia region, the revised Table 3 is shown below. This demonstrates that the proportion of decades with a significant SC-AO relationship for Eurasia as a whole is statistically high or low for SCI_09 and SCI_10 for negative and positive relationships with the AO, respectively. In addition, the proportion of decades with a positive SAI_11-AO relationship is significantly small. Therefore, it appears likely that Eurasian snow cover does have some influence on the AO and the relationship is not purely stochastic. This is also the conclusion of Wegmann et al. (2020), who examined SC-AO relationships over the 20th Century between November SC and the NAO rather than AO and stated that, "We find evidence for a negative NAO-like signal after November … which is valid throughout the last 150 years."

Wegmann, M., Rohrer, M., Sanolaria-Otín, M., and Lohmann, G.: Eurasian autumn snow link to winter North Atlantic Oscillation is strongest for Arctic warming periods, Earth Syst. Dynam., 11, 509–524, https://doi.org/10.5194/esd-11-509-2020, 2020.

However, I will add the following sentence at an appropriate point, "We note that, despite the data providing statistical evidence for a connection between some autumn SC indices and the following winter AO, the possibility remains that the temporal variations in the strength of the SC-AO relationships are due to pure chance."

Table 3. Proportion of decades that the 20CRv3 SC indices have positive and negative correlations with the 20CRv3 AO, and the proportion that are statistically significant based on a one-tailed test ($p < 0.10$). The significance of the proportion of decades with a significant SC-AO correlation is determined based on a probability distribution function derived from 10000 scrambled AO time-series using a one-tailed test for each sign of correlation: *** is $p < 0.01$, ** is $p < 0.05$, * is $p < 0.10$. The period covered is 1831-40 to 2006-2015.

Snow Cover Index	Positive (%)	Negative (%)	Pos. (significant) (%)	Neg. (significant) (%)
SCI_09	32.35	67.65	2.35**	27.65***
SCI_10	48.82	51.18	4.12*	22.35**
SCI_11	52.94	47.06	11.18	10.00
SAI_OCT	38.82	61.18	11.18	13.53
SAI_NOV	47.06	52.94	2.94**	9.41

2. Shorten the discussion of the details in this section.

This will be achieved by reducing the number of subregions analysed in Fig. 8, as suggested by the reviewer, and primarily focusing the discussion on those regions with statistically significant high and low frequencies of significant decadal relationships for opposite signs of the SC-AO relationship in question (see #3 below).

3. Make the significance testing more rigorous. For the results in Figure 7 and Table 3, also report the statistical significance of the proportion of statistically significant correlations. Going beyond the simple Monte Carlo method described above, another choice could be tests in which the connection between SC and AO is broken by randomly scrambling the years of the AO index. For the regional analysis in Fig. 8, only show in colour those regions where the proportion of significant correlations is statistically significant and report the proportion of such regions in the figure itself or in a table.

   Many thanks for the excellent suggestion of scrambling the AO index in order to determine the statistical significance of the observed frequency of decadal SC-AO relationships. Following on from some sensitivity tests, I have undertaken this 10000 times for each SC-AO relationship. Importantly, this methodology also allows an examination of whether the frequency of decades with significant correlations is smaller as well as larger than one might expect through chance (one-tailed test). Thus, if the observed frequency is both significantly high for one sign of the SC-AO relationship and significantly low for the other sign then it is likely to be a more robust relationship than one in which the frequency is only significant for one sign. This new methodology has led to revised Figs. 8 (see below) and 9, Table 3 and additional new figures and tables. In combination with the previous analyses in the references suggested by Reviewer 3, using this new methodology has led to a different emphasis in the SC-AO results section and Discussion.

   I will add the following text in the methodology and include an example plot in the Supplementary Material:

   "To test whether the frequency of decades with a statistically significant SC-AO relationship was itself significant (either high or low), the AO time-series was scrambled 10000 times and probability distribution functions (PDFs) of the frequency of both significant positive and negative SC-AO relationships (based on a one-tailed test at $p < 0.10$) derived (see Fig. SX for an example). As might be expected, the PDFs for positive and negative relationships were very similar but not identical. The observed frequency value was then compared to the appropriate PDF to determine its level of statistical significance if any."

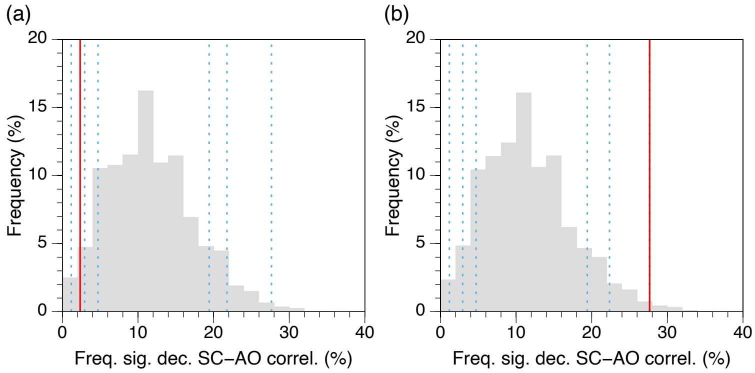


Figure SX. Probability distribution functions of the frequency of statistically significant ($p < 0.10$) decadal SCI_09-AO correlations based on 10000 scrambled AO time-series for (a) positive and (b) negative relationships. The dashed blue lines represent the values of statistical significance ($p < 0.10$, $p < 0.05$, $p < 0.01$) for high and low values. The red line is the value derived from the 20CRV3 data.

Based on the new Fig. 8 (see response to #14 below), the revised Section 4.2 will focus on regions of Eurasia that have both high and low statistically significant frequencies of significant decadal SC-AO correlations of different signs. Principally, these are (i) SCI_09 for north-east Eurasia (120-180°E, 60-70°N) and (ii) SCI_11 for south-west Eurasia (30-60°E, 40-55). The statistics are summarised in the new table and the figure below, the latter demonstrates some interesting decadal variability in the time-series of both SC-AO relationships that can be examined further in comparison to potential external forcing factors.

Table New. Proportion of decades that the 20CRv3 SC indices have positive and negative correlations with the 20CRv3 AO, and the proportion that are statistically significant based on a one-tailed test ($p < 0.10$): *** is $p < 0.01$, ** is $p < 0.05$, * is $p < 0.10$. The period covered is 1831-40 to 2006-2015.

Snow Cover Index	Positive (%)	Negative (%)	Pos. (significant) (%)	Neg. (significant) (%)
SCI_09_NE	26.47	73.53	1.76**	27.64***
SCI_11_SW	76.47	23.53	38.34***	4.12*

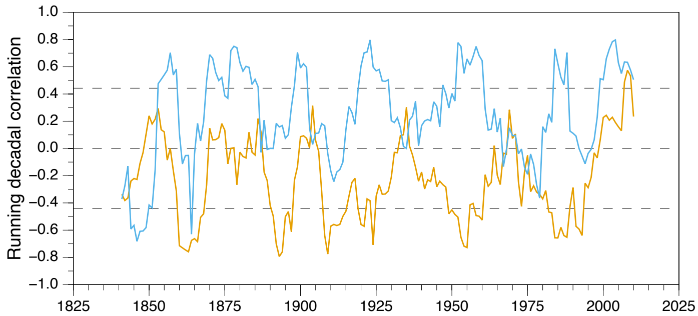


Figure New. Running decadal correlations between the SCI_09_NE and SCI_11_SW and winter AO. The dashed horizontal lines of the same colour indicate significance levels at $p < 0.10$ for a one-tailed test.

Apart from the statistical analysis, the manuscript is in reasonable shape. However, as described in the detailed comments below, there is room for improvement in some of the figures and some pieces of the text, and a few details in the methodology. There are also some typos and peculiarities in the wording.

DETAILED COMMENTS

1. Replace "positive (negative) relationships" with "statistically significant (positive) negative decadal correlations".

   Done

2. Summing of correlations makes no mathematical sense. For example, if both A and B are perfectly correlated with C (r = 1), the correlation of A + B with C is still 1 and not 2.

   It was not actually supposed to make 'mathematical sense', simply a qualitative device to demonstrate that when the SC-AO relationships of the two regions were of the same sign it gave a 'strong' SC-AO relationship for Eurasia as a whole. Note that following the new methodology this result has been removed anyway as the SAI_OCT dipole is no longer the principal SC-AO relationship.

3. L23-24. A new study suggests an underlying ratio of three when internal climate variability is filtered out from the Arctic temperature time series (Zhou et al.., 2024, Nature Geoscience, doi: 10.1038/s41561-024-01441-1).

   This reference will be added, and the sentence changed to, "It has long been known that the Arctic has been warming significantly faster than the global average, but recent studies have revealed that this 'Arctic amplification' is even greater than previously thought, being a factor of three to four over the past half-century (Rantanen et al., 2022; Zhou et al., 2024).

4. L37-39. A study published in 2002 gives no information of the recent decades, but rather the late 20th century.

   Indeed: the sentence will be amended to, "However, Kitaev et al. (2002) suggested that in the late 20th Century …"

5. at the end of November?

   Yes, this will be amended in the Fig. 1 caption.

6. L83-84. Please name explicitly the emission scenarios that lead to a -10 % vs. -38 % decrease in SC.

The sentence will be changed to, " … they varied from stabilization at ~2060 at –10 ± 5% (relative to 1995-2014) for Shared Socioeconomic Pathway (SSP)123 to continuous and ongoing SC losses that reach –38 ± 10% by 2100 in the least optimistic climate change scenario (SSP585)."

7. L155-156. It would seem better to explicitly exclude the fully oceanic regions from the analysis, as well as the regions with very little land (say less than 20 %).

   This has been done in the new analyses. Eurasia is now defined as 35-75°N rather than 35-80°N — all corresponding analyses have been redone and the text changed as appropriate — and an additional six subregions have been removed from the individual analysis, three of which were not included before due to being wholly ocean.

   The sentence will be changed to, "In order to examine spatial variability in the SC indices and their relationship to the AO, this region was subdivided into 40 areas, each of 5° latitude by 30° longitude. Note that some of these subregions actually contain little or no SC data as they are predominantly or wholly ocean (Fig. 2) and thus only 34 subregions are analysed individually."

8. Figure 2. The dashed lines are barely visible. Darker blue or thicker lines would be preferable.

   This has been done: dark blue for Eurasia and thicker cyan lines to denote the subregions.

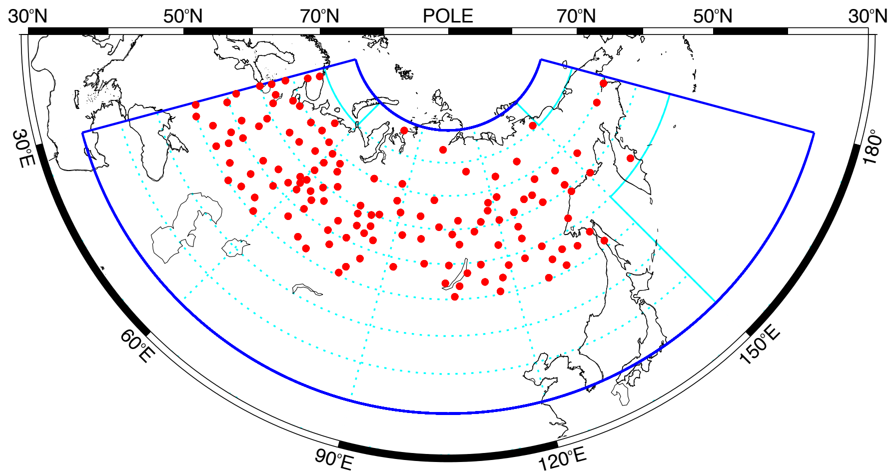


Figure 2. Red dots are the location of the 133 stations with daily snow depth observations used to validate ERA5. The thick blue line shows Eurasia, as defined in this study, while the dashed cyan lines denote the 34 subregions (5° latitude by 30° longitude) analysed individually.

9. I did not understand this. If SC is either 0 % or 100 % in all days of the month, there is no change with time and thus SAI = 0.

   I agree that this could have been phrased better: the sentence now reads, "Trends were only calculated when at least 50% of potential data were available: 'missing' reanalysis data result from a region either having 0% or 100% SC on all days in a month, giving an SAI of zero and thus making trend analysis problematical if there are a high proportion of such data."

10. Figure 5. The dashed line in (a) are very faint and could be made thicker. Also, there is a slight mismatch in the logic of colours between (a) and (b), since red is used for September in (a) but for October in (b)

All completed.

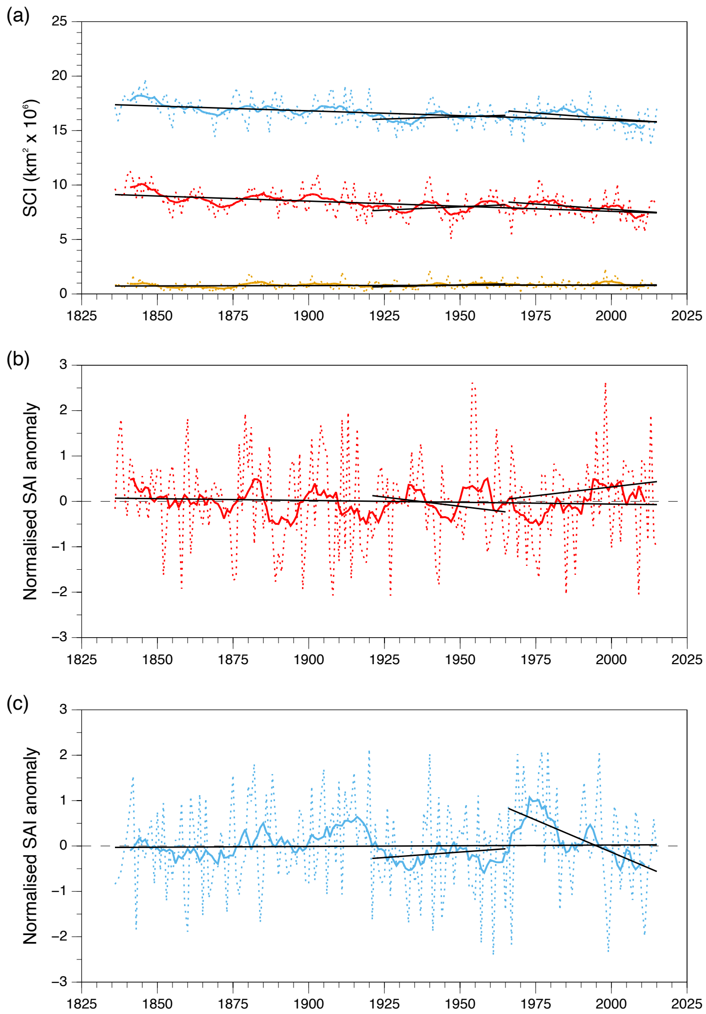


Figure 5. Annual Eurasian SC index data from 20CRv3. (a) SCI_09 (blue), SCI_10 (red) and SCI_11 (brown); (b) SCA_OCT and (c) SCA_NOV: annual data (dotted lines) and running decadal means (full lines) Black lines show trends for 1836-2015, 1921-1965 and 1966-2015.

11. There is very little land in this subregion and thus no reason to pay any attention to it. It might make better sense to totally omit such strongly ocean-dominated regions from the analysis (cf. comment 7 above).

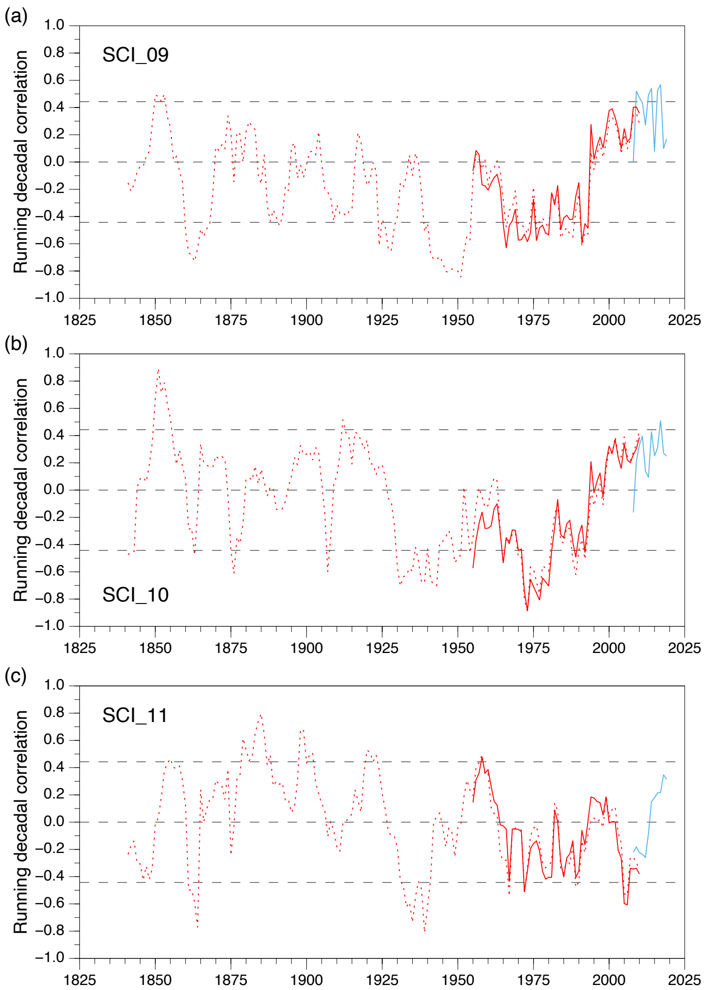
Please see response to comment 7.

12. Very little land in this region as well!

Please see response to comment 7.

13. Figure 7. Please include the index acronyms (SC_09, SC_10, SC_11, SAI_OCT, SAI_NOV) directly within the figure panels, e.g. in their top-left corners.

Completed.



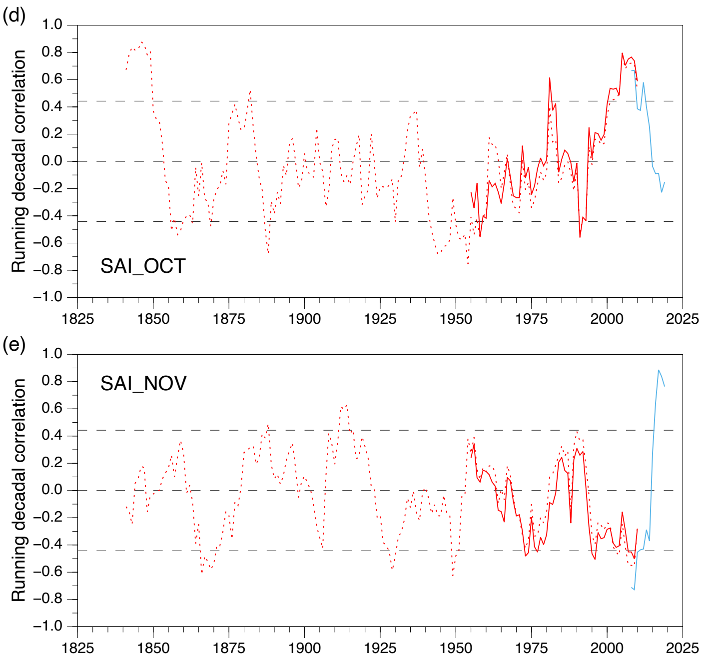


Figure 7. Running decadal correlations between snow cover indices and winter AO: (a) SCI_09; (b) SCI_10; (c) SCI_11; (d) SAI_OCT; (e) SAI_NOV. Snow cover indices from 20CRv3 are shown and ERA5 (2003-12 to 2012-21) as red and blue lines, respectively. Correlations with winter AO from CPC and derived from 20CRV3 are shown as full and dashed lines, respectively. The black dashed horizontal lines indicate significance levels at $p < 0.10$ and $p < 0.05$ for each decade considered independently.

14. Figure 8. As discussed in the General comments, it could be better to only use colours for those regions where the fraction of statistically significant decadal correlations is statistically significant at least at the 10 % level.

Completed

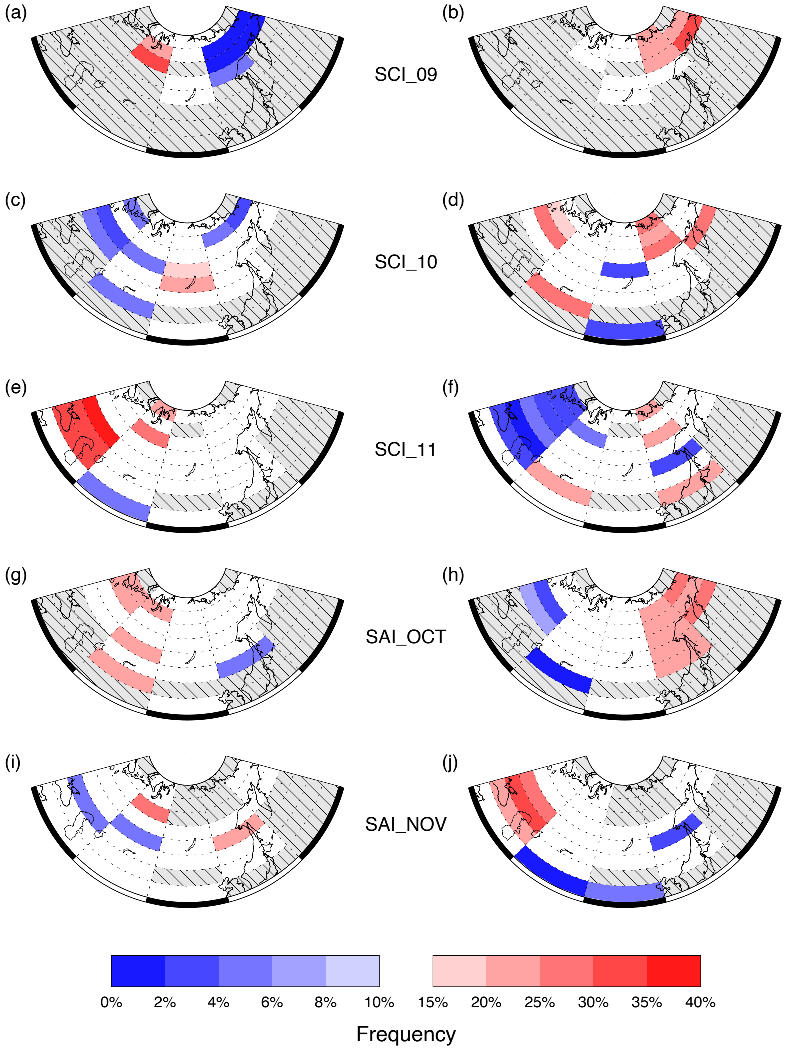


Figure 8. The frequency of decades with statistically significant correlations with the AO (p < 0.10) from the 20CRv3 data (1831-1840 to 2005-2014) for the five SC indices. Frequencies for positive correlations are shown on the left in red and negative correlations on the right in blue. Hatched subregions are where no value was calculated because data availability was less than 50%.

15. Beware of the caveats of multiple testing. The probability that a high fraction of significant correlations is found in some of many areas just by pure chance is far larger than the probability that this happens in any single region.

I will add something along the lines of, "We note that the statistical power of computing the correlations for the subregions separately is potentially compromised by the discrete nature of each test and ignores the fact that the fraction of true null hypotheses is unknown and that some false null hypotheses may be rejected (e.g., Wilks, 2006). Therefore, we focus on those subregions with statistically significant frequencies of SC-AO correlations for both signs of the relationship as likely having the most robust relationships

Wilks, D. S.: On "field significance" and the false discovery rate, J. Appl. Met. Climatol., 45, 1181–1189, https://doi.org/10.1175/JAM2404.1, 2006.

16. What is the correlation of SAI_OCT(west) and SAI_OCT(east)?

Following the revised analysis this comment is no longer relevant.

17. L555-559 and Figure 10. Summing of correlations makes no mathematical sense (cf. Comment 2). In contrast to correlations, covariances are additive, but only if the SAI indices are expressed in absolute (area) units rather than in normalized form. Thus, one option would be to make a plot that shows the absolute covariances of SAI_Oct(East), SAI_Oct(West) and SAI_OCT(East + West) with AO. Another option is to omit Figure 10.

    Please see response to comment 2.

    L591-592. The interpretation would be this if Figures 11c (11d) showed the correlation between SAI_OCT in west (east) and the 850 hPa geopotential height. However, what is shown is the correlation between the SAI_OCT – AO_winter *correlation* and 850 hPa geopotential height. Modify accordingly.

    Following the revised analysis this comment is no longer relevant.

18. L657-658. a trend from negative to positive correlations with winter AO?

    Following the revised analysis this comment is no longer relevant.

19. L679-682. What is the correlation between SAI_NOV(west) and AO for the whole period 1836-2015? What is the corresponding correlation between SAI_NOV(east) and AO?

    Following the revised analysis this comment is no longer relevant.

TYPOS ETC.

1. momentum or moisture? Changed to "… upwelling of longwave radiation and turbulent heat fluxes into the atmosphere …"
2. there have been Corrected, twice!
3. either SAI index? Yes, corrected.
4. Snow cover indices from 20CRv3 and ERA5 ( ) are shown as red and blue lines Yes, corrected.
5. Caption of Table 3. Check the years. The numbers in the table suggest a sample size of exactly 170. Yes, caption corrected.
6. October and November? Yes, corrected.