

# MORE-PERSISTENT WEAK STRATOSPHERIC POLAR VORTEX STATES LINKED TO COLD EXTREMES

MARLENE KRETSCHMER, DIM COUMOU, LAURIE AGEL, MATHEW BARLOW, ELI TZIPERMAN, AND JUDAH COHEN

Over recent decades, the stratospheric polar vortex has shifted toward more frequent weak states, which can explain Eurasian cooling trends in boreal winter in the era of Arctic amplification.

Despite global warming, recent winters in the northeastern United States, Europe, and especially Asia were anomalously cold. Some mid-latitude regions like central Asia and eastern Siberia even show a downward temperature trend in winter over past decades (Cohen et al. 2014a; McCusker et al. 2016). In contrast, the Arctic has been warming rapidly, challenging scientists to explain the so-called warm Arctic–cold continents pattern in boreal winter

(Shepherd 2016). Though there is general agreement that sea ice loss contributed to the warming of the Arctic via ice–albedo feedbacks (Screen and Simmonds 2010), it remains controversial whether observed midlatitude cooling is related to internal atmospheric variability (Sun et al. 2016; McCusker et al. 2016), to tropical (Palmer 2014) or Arctic (Cohen et al. 2013; Cohen 2016) trends in teleconnection indices, or to a combination of those.

**AFFILIATIONS:** KRETSCHMER—Potsdam Institute for Climate Impact Research, Earth System Analysis, and Department of Physics, University of Potsdam, Potsdam, Germany; COUMOU—Potsdam Institute for Climate Impact Research, Earth System Analysis, Potsdam, Germany, and Institute for Environmental Studies, VU University Amsterdam, Amsterdam, Netherlands; AGEL—Department of Environmental, Earth, and Atmospheric Sciences, and Intercampus Marine Science Graduate Program, University of Massachusetts Lowell, Lowell, Massachusetts; BARLOW—Department of Environmental, Earth, and Atmospheric Sciences, University of Massachusetts Lowell, Lowell, Massachusetts; TZIPERMAN—Department of Earth and Planetary Sciences, and School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts;

COHEN—Atmospheric and Environmental Research, Lexington, Massachusetts

**CORRESPONDING AUTHORS:** Marlene Kretschmer, kretschmer@pik-potsdam.de; Judah Cohen, jcohen@aer.com

*The abstract for this article can be found in this issue, following the table of contents.*

DOI:10.1175/BAMS-D-16-0259.1

A supplement to this article is available online (10.1175/BAMS-D-16-0259.2)

In final form 25 May 2017

©2018 American Meteorological Society

For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#).

Previous research showed that a weak stratospheric polar vortex (hereafter also referred to as “polar vortex” or “vortex”) can affect surface weather via a downward influence of planetary waves (Baldwin and Dunkerton 2001; Hitchcock and Simpson 2014), which leads to cold-air outbreaks in the midlatitudes and a negative surface Arctic Oscillation signal (Cohen et al. 2013; Kolstad et al. 2010; Butler et al. 2014; Baldwin and Dunkerton 2001; Sigmond et al. 2013; Kretschmer et al. 2016). Moreover, it was shown that sudden stratospheric warmings (SSWs) can modulate the tropospheric flow for up to 2 months (Baldwin and Dunkerton 2001; Hitchcock and Simpson 2014), which can even offset the impact of El Niño–Southern Oscillation (ENSO) events (Polvani et al. 2017). Consequently, including stratosphere activity in climate models significantly improves seasonal forecast skill for winter weather (Scaife et al. 2016; Sigmond et al. 2013). Despite this key role of the polar vortex for winter circulation and surface temperature, a quantitative analysis of the potential stratospheric role for the recent cooling trends has yet been lacking.

There are several metrics to describe polar vortex variability, extreme states, and its coupling with the troposphere, but the different indices do not necessarily capture all of these aspects. Often, the stratospheric impact on surface temperatures is analyzed in the context of SSWs (Polvani et al. 2017; Butler et al. 2014). Detection of SSWs is, however, sensitive to their exact definition, which varies throughout the literature (Butler et al. 2015). Moreover, SSWs are individual rare events and thus do not describe the overall behavior of the vortex. The tropospheric response of SSWs depends, however, on their temporal evolution and persistence in the stratosphere (Kodera et al. 2016; Runde et al. 2016). To study the recovery phase of extreme stratospheric events, Hitchcock et al. (2013) identified polar night jet oscillation (PJO) events. These describe long-lasting anomalous warm temperatures in the stratospheric polar cap and are often preceded by SSWs, but approximately half of the SSWs recover rapidly from the abrupt warming (Hitchcock et al. 2013).

Recently, machine learning approaches such as clustering algorithms have successfully been applied to study impacts of and changes in circulation patterns (Feldstein and Lee 2014; Horton et al. 2015; Lee and Feldstein 2013; Cheng and Wallace 1993), providing a promising data-driven tool to classify atmospheric fields. Motivated by these results, we perform cluster analysis on the daily extratropical stratosphere to identify its dominant spatial patterns and temporal evolution. This way, we can study

different vortex states as well as persistence of specific events. We analyze how long-term changes in polar vortex variability might have affected surface warming patterns.

**DATA.** We use daily mean European Centre for Medium-Range Weather Forecasts (ECMWF) interim reanalysis (ERA-Interim; Dee et al. 2011) data from January 1979 to December 2015, leap days excluded. Data that were used to characterize the stratospheric polar vortex (geopotential height and zonal wind velocity at 10 hPa) were provided on a  $0.75^\circ \times 0.75^\circ$  latitude–longitude grid. To study precursors and lagged effects of different polar vortex cluster events, we use gridded ( $3^\circ \times 3^\circ$ ) data of sea level pressure, near-surface temperature, and poleward heat flux ( $v^*T^*$ ) at 100 hPa, where  $v$  is the meridional wind velocity,  $T$  is the temperature, and the asterisks denote the deviation from the zonal mean. We further use daily mean Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2; Molod et al. 2015), data from 1980 to 2015 to perform sensitivity analyses on the reanalysis product and clustering technique used.

**METHODS.** We employ hierarchical clustering (Cheng and Wallace 1993) on the daily mean zonal wind velocity field poleward of  $60^\circ\text{N}$  at 10 hPa. We chose this domain and level for consistency with most other SSW definitions and polar vortex studies (Butler et al. 2015). We limit the cluster analysis to the months January and February (JF) over the period 1979–2015, as these months show the strongest polar vortex variability. First, we calculate the climatological anomalies for each day by subtracting their multiyear mean. Additionally, to account for the denser grid toward the pole, we apply area weighting. There are  $n = 2,183$  daily observations ( $37 \text{ years} \times 59 \text{ days}$ ), each corresponding to a vector of length 19,680 (number of grid points in our domain) representing the state of the polar vortex on a particular winter day. The cluster algorithm groups days with similar extratropical stratospheric wind fields in one cluster, which can be represented by the composite of all days assigned to it (see appendix and supplemental information for more details: <https://doi.org/10.1175/BAMS-D-16-0259.2>).

We determine time series of the seasonal occurrence frequencies for each cluster, which ranges from zero (absent) to one (every day of the winter was assigned to that cluster). Linear trends in occurrence frequency are calculated using a least squares fit regression model, and the slope was tested for significance using a two-sided Student’s  $t$  test. We define

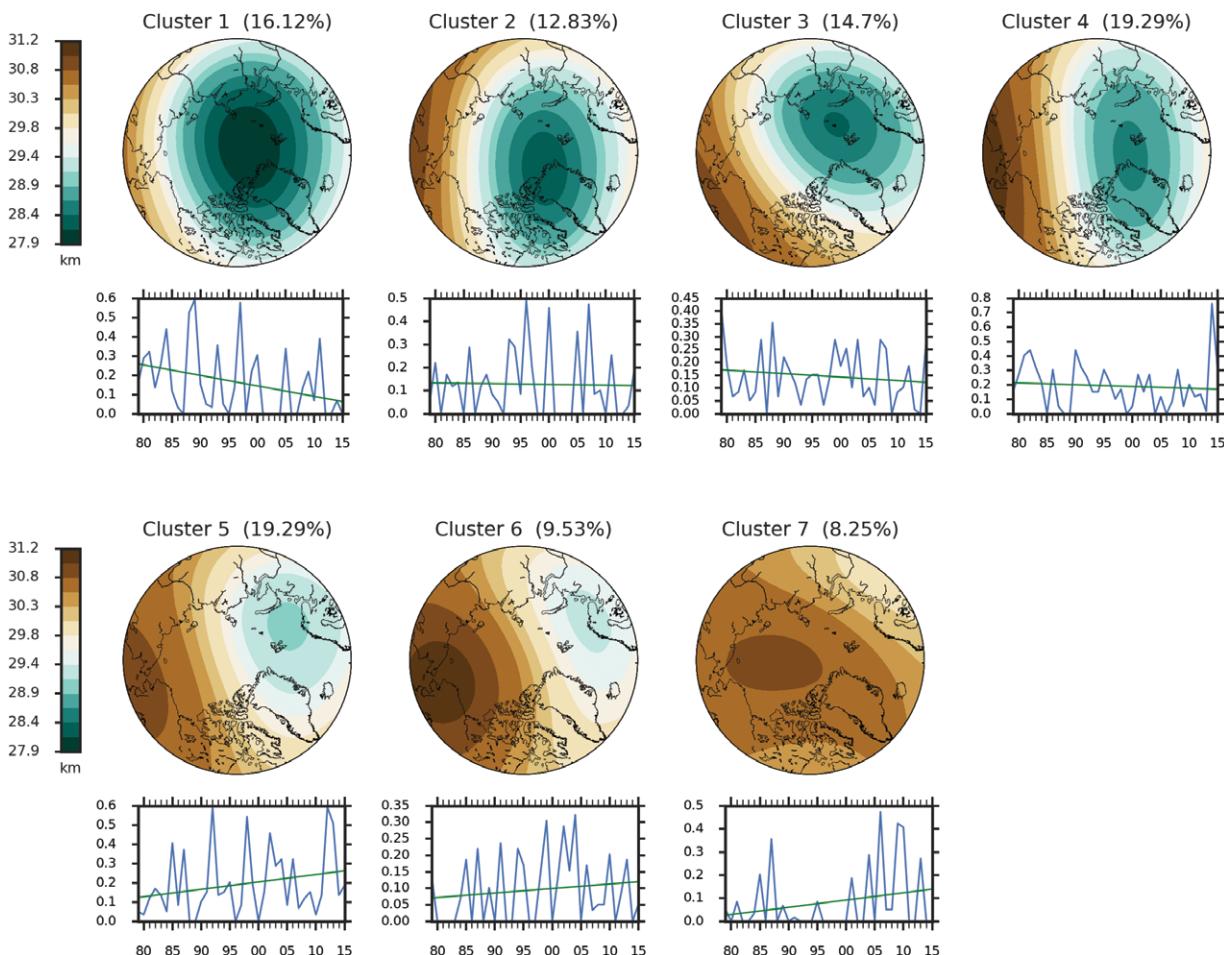
a cluster “event” as a period of consecutive days for which the same cluster is identified.

**MORE FREQUENT WEAK POLAR VORTEX STATES.**

Our analysis reveals that seven is an appropriate choice for the number of clusters, providing a sufficiently detailed overview of the spectrum of different polar vortex patterns, while still allowing each pattern to describe a significant part of the total polar vortex phase space (see appendix and supplemental information). This is also demonstrated by the relatively high mean pattern correlation of 0.59, which is used to estimate how well the clusters represent the original data: the area-weighted pattern correlation of each daily field to its cluster composite is calculated, and the average over all days represents a global measure of similarity.

Figure 1 shows the composite mean of the 10-hPa geopotential height field for all seven clusters,

ordered by polar cap height (i.e., the area-weighted mean at 10-hPa geopotential height north of 60°N), starting with the strongest polar vortex cluster (thus with the lowest polar cap height). Though clustering was performed on the zonal wind field, we present geopotential heights for easier visualization of the different polar vortex shapes. The associated zonal wind plots are given in Fig. ES3 in the online supplement. The patterns range from a strong circumpolar vortex (cluster 1) to a slightly less strong polar vortex (cluster 2), to progressively weaker polar vortices with displaced vortex centers toward Eurasia (clusters 3, 5, and 6) and North America (cluster 4), and finally a weak distorted vortex (cluster 7). Below the cluster composites, time series of their seasonal frequency with a linear least squares fit trend line are displayed for each cluster. The strong vortex cluster (cluster 1) has a significant ( $p = 0.047$ ) downward linear trend of  $-0.2 (37 \text{ yr})^{-1}$ , whereas the weak vortex clusters 5, 6,



**FIG. 1. Polar vortex clusters and their frequency trends. Composite mean of 10-hPa geopotential height values over all days that were assigned to the same cluster (clustering performed with zonal wind anomalies) and time series of normalized occurrence frequency in winter (JF) with least squares fit line. The number in parentheses denotes the total frequency occurrence (%) for the studied period.**

and 7 increased in frequency, the last with a trend of  $0.12 (37 \text{ yr})^{-1}$  ( $p = 0.146$ ).

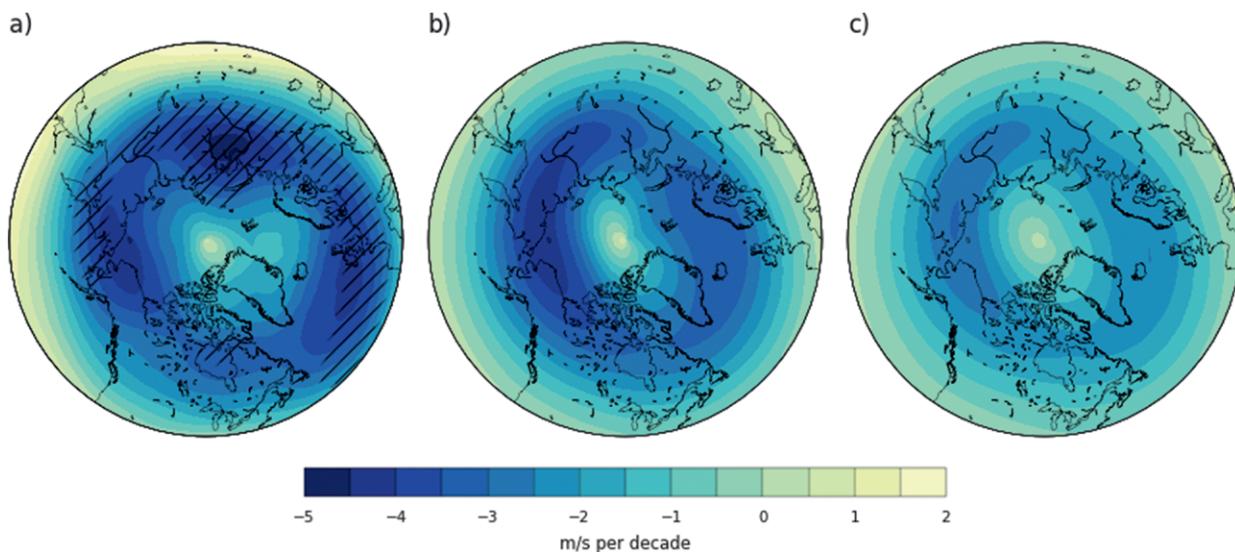
In principle, it is possible that trends in (seasonal) frequency are only the result of two or more similar clusters with opposing trends that would cancel each other out if those clusters were merged. To test this possibility, we calculate for each day the pattern correlation with the composite mean of each cluster (Fig. ES4, see supplemental information for details). This thus quantifies how the daily polar vortex patterns resemble the different clusters at each time step. We find that the strong vortex clusters (clusters 1 and 2) exhibit a downward trend in pattern correlation ( $p \approx 0.07$ ). In contrast, the weak vortex clusters (clusters 6 and 7) have upward trends ( $p \approx 0.07$ ). Thus, over the last 37 winters, the daily polar vortex state shifted toward the weaker cluster patterns. This is consistent with the overall weakening of the stratospheric zonal wind field, especially at the vortex edge over the continents (Fig. 2a, Fig. ES5 for the polar cap mean). South of  $60^\circ\text{N}$ , the trends in zonal wind velocity are even upward, indicating an equatorward shift and broadening in addition to the weakening of the vortex.

To test how well our cluster analysis reflects observed trends, we multiply the zonal wind composite mean of each cluster with the slope of its frequency trend (Lee and Feldstein 2013). Summed for all clusters (Fig. 2b), this shows how much of the seasonal-mean change is explained by the change in frequencies, and we find that it compares well with the actual trend field (Fig. 2a). In fact, approximately

72% of the observed weakening north of  $60^\circ\text{N}$  is already explained by the less frequent occurrence of the strong vortex cluster 1 and the more frequent occurrence of the weak polar vortex cluster 7 (Fig. 2c).

To further test how the frequency of cluster events changes over time, we count the mean seasonal occurrence in the first half (1979–96) and the second half (1998–2015) of the studied time period for each cluster (Fig. 3a). We find that the frequency of cluster 7 increased significantly (using a bootstrapping approach; see appendix) by 140% from on average  $\sim 3$  days per winter up to roughly 7 days ( $p < 0.01$ ). In contrast, the frequency of cluster 1 halved from approximately 12 days per season to just 6 days ( $p < 0.05$ ). The increased frequency of cluster 7 days results from an increase in the persistence of cluster 7 events (consecutive days assigned to cluster 7). Whereas in the first half of the studied time period the mean persistence of cluster 7 events was 5.3 days, it was significantly ( $p < 0.01$ ) longer in the second half, with events persisting on average 14.1 days (an increase by more than 160%). In contrast, the mean persistence of cluster 1 events was approximately 9 days in both periods, but their occurrence dropped notably from 27 events in the first half to just 11 events in the latter half. Thus, the increase in cluster 7 days is due to longer events and the decrease in cluster 1 days is due to less events.

**ROBUST CLASSIFICATION OF WEAK POLAR VORTEX STATES.** Our finding of



**FIG. 2.** Trend in strongest and weakest polar vortex clusters explain the overall trend of the polar vortex. (a) Seasonal-mean (JF) trend in zonal wind poleward of  $40^\circ\text{N}$ . Significant values ( $p < 0.1$ ) according to two-sided Student’s *t* test are shown in hatches. (b) Sum of all seven polar vortex cluster representatives multiplied by their trend in seasonal frequency. (c) As in (b), but only for clusters 1 and 7.

more (less) frequent weak (strong) polar vortex days over the past winters is robust and insensitive to the total number of clusters (from 2 to 20 clusters). Furthermore, the cluster representatives and frequency trends of the strongest and the weakest cluster are robustly identified and are mostly insensitive to the dataset (MERRA-2 instead of ERA-Interim), clustering technique (using *k*-means or self-organizing maps instead of hierarchical clustering), clustered variable (geopotential heights instead of zonal

wind velocity), and pressure level (100 hPa and the mean over 10–50 hPa). Generally, clustering over lower pressure levels results in higher seasonal frequencies of weak polar vortex states. This is consistent with previous studies showing that disturbances of the upper stratospheric flow persist for longer when they descend to lower levels (Hitchcock et al. 2013) and also with the fact that strong lower-stratospheric anomalies often coincide with tropospheric circulation anomalies (Baldwin and Dunkerton 2001) that are not necessarily observed at higher levels. More precise information on how the different tests compare can be found in the supplementary information (Figs. ES6–ES15).

Our clustering methodology is also consistent with other metrics to classify extremely weak states of the stratospheric polar vortex. All starting days of major SSWs in January and February, as detected by Charlton and Polvani (2007), are assigned to the weak vortex clusters 6 and 7 (Fig. ES17), which also coincide with polar night oscillation events (Fig. ES16; Hitchcock et al. 2013). In summary, the different sensitivity tests show that a cluster approach applied at 10 hPa provides a robust and appropriate methodology to study the occurrence and persistence of weak polar vortex events as well as their coupling with lower-stratospheric pressure levels.

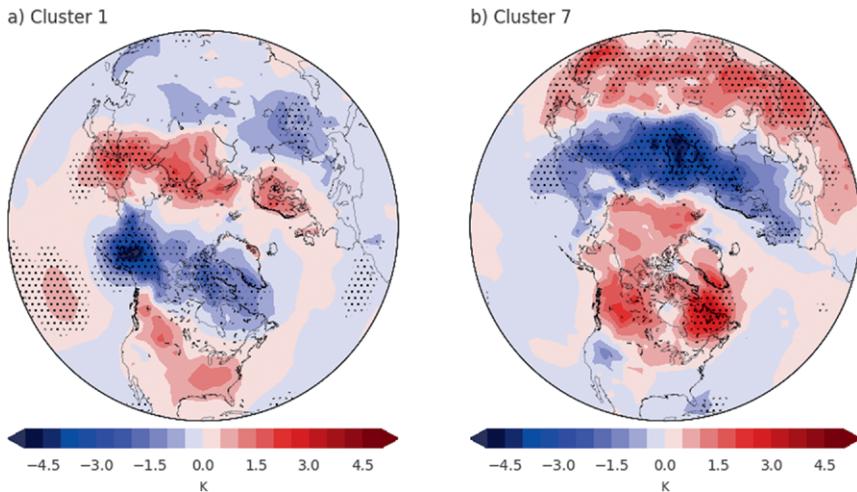
**LINKS TO SURFACE TEMPERATURE.** The tropospheric response to weak polar vortex states can influence surface weather for up to 2 months (Baldwin and Dunkerton 2001; Hitchcock and Simpson 2014; Sigmund et al. 2013). Further, the tropospheric



**FIG. 3.** Average occurrence (days) per winter of each cluster from 1979 to 1996 (light blue) and from 1998 to 2015 (dark blue) and the change (%). Significant changes ( $p < 0.05$ ) are indicated in red.

response is more pronounced if the stratospheric recovery is slow following a vortex disturbance (Kodera et al. 2016; Runde et al. 2016). Thus, an increase in more persistent weak polar vortex states, that is, longer-lived cluster 7 events, could potentially influence winter temperatures. In other words, the moderate changes in the mean vortex state (Fig. ES5) are much less relevant for surface conditions than the increased persistence of extremely weak states.

To study the relationship of clusters 1 and 7 events with surface weather, we create composites of (detrended) near-surface temperature (Fig. 4). As expected, strong vortex states (cluster 1) coincide with mild temperatures in the eastern United States and northern Eurasia and cold temperatures over Alaska and Greenland (Fig. 4a). In contrast, during weak vortex states (cluster 7), anomalously cold temperatures are observed in northern Eurasia whereas Canada is anomalously warm (Fig. 4b). Thus, the increased frequency in cluster 7 during recent winters might be linked to the surface cooling trends over Eurasia. To test this, we first determine different linear regression models onto mean winter (JF) near-surface temperature at each grid point and plot their  $R^2$  values (Fig. 5), indicating how much of the observed temperature variability is explained by the linear model. To account for potential biases due to trends in the regressors and the temperature time series, we detrended the variables first. Though polar cap height (PCH) variability can already explain some seasonal temperature variability (Fig. 5b), regression by cluster 7 seasonal frequency gives higher  $R^2$  values, significant over extended regions, including



**FIG. 4. Composites of detrended near-surface temperature during (a) cluster 1 and (b) cluster 7 days. Significant values ( $p < 0.05$ ) are indicated with dots.**

central Siberia, eastern Canada, and the western Atlantic sector but not the United States (Fig. 5c). The combination of ENSO (described by the mean winter Niño-3.4 index) and the seasonal frequency of cluster 7 further improves the results over the Pacific and parts of the United States (Fig. 5d), but ENSO alone has very little influence on Eurasian temperature variability (Fig. 5a). Note that the correlation between the detrended cluster 7 frequency time series and the detrended Niño-3.4 index is only 0.01, showing that they are almost completely independent.

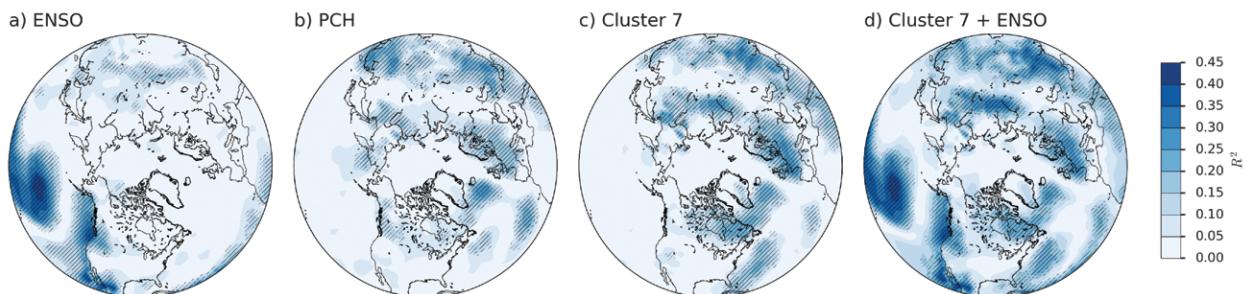
Next, we calculate the temperature trends at each grid point for each of the regression models (Figs. 6a–c). For consistency with previous studies analyzing the warm Arctic–cold continent pattern (Sun et al. 2016; Cohen et al. 2013; Cohen 2016; McCusker et al. 2016), we calculate trends over the era of Arctic amplification (Cohen et al. 2014a), that is, from 1990 onward. We apply the regression parameters from the models calculated for the detrended data from 1979 to 2015

the models including both cluster 7 and the Niño-3.4 index (Fig. 6c). Thus, although other factors certainly play a role as well, the observed cooling trends over Eurasia (Fig. 6d) are well captured by the trend toward more-persistent weak vortex states (Fig. 6b), something that can be further improved by including tropical variability (Fig. 6c).

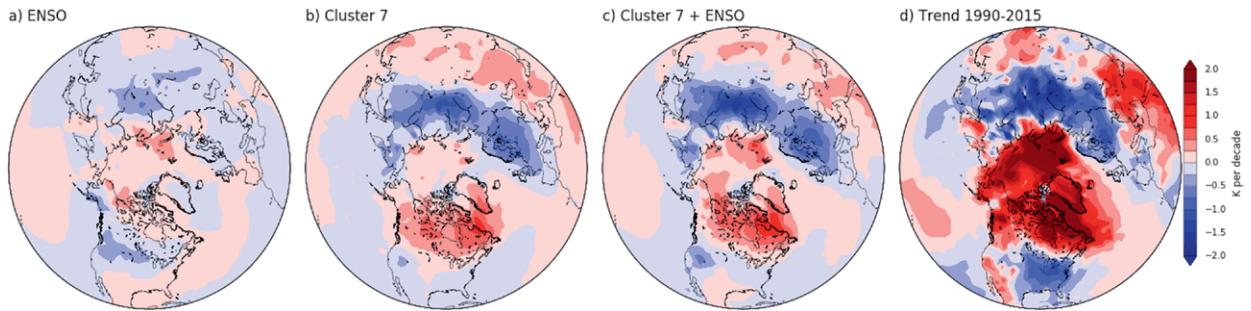
**COLD WEATHER IN EURASIA.** Several studies focused on Eurasia, as the winter cooling trend there has been more pronounced (McCusker et al. 2016; Sun et al. 2016; Li et al. 2015; Mori et al. 2014). Indeed, our analyses show that the relationship between weak polar vortex states and surface temperature is much stronger for this region, as compared to the northeastern United States (Figs. 4b, 5c, 6b).

Our predicted regression model based on cluster 7 correlates ( $r = 0.46$ ,  $R^2 = 0.21$ ) significantly ( $p < 0.01$ , according to a Student's  $t$  test) with winter temperature averaged over the Eurasian sector ( $50^\circ$ – $65^\circ$ N,

(Fig. 5) to predict temperature trends using the non-detrended regressors from 1990 to 2015. All models show a warm Arctic–cold continent pattern, with much stronger cooling over Eurasia than over North America. The explanatory power of ENSO (Fig. 6a) and PCHs at 10 hPa (not shown) is small. In contrast, regression by cluster 7 frequency (Fig. 6b) captures the observed Eurasian pattern well. The best agreement with observations (Fig. 6d) is achieved with



**FIG. 5. Explained variance ( $R^2$  values) of winter (JF)-mean temperature for regression with (a) winter-mean Niño-3.4 index, (b) winter-mean PCH, (c) cluster 7 frequency, and (d) cluster 7 frequency and the winter-mean Niño-3.4 index. Before calculating the regression models, the linear trends of the regressors and the temperature were removed. Significant ( $p < 0.05$ ) models according to  $F$  test are indicated in hatches.**



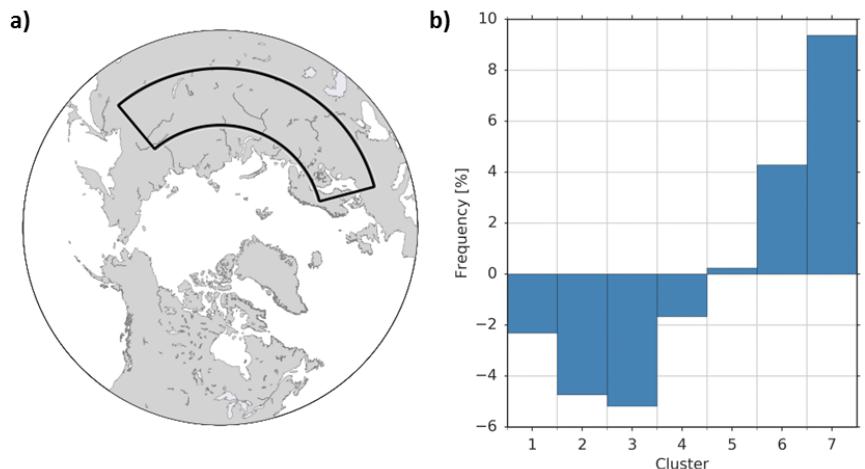
**FIG. 6.** (a)–(c) Linear trends in temperature as projected by the regression models in Figs. 5a, 5c, and 5d, respectively, and (d) observed trends for the period 1990–2015. The regression models were calculated based on detrended data from 1979 to 2015 and the projected trends are calculated for the undetrended regressors from 1990 to 2015.

15°–130°E; black box in Fig. 7a). This model performs much better than a regression model based on the PCH index at 10 hPa ( $r = 0.26$ ,  $R^2 = 0.07$ ,  $p = 0.11$ ). Thus, the seasonal frequency of weak states is a better predictor for Eurasian temperature variability than the polar cap mean. Moreover, the cluster 7-based model explains ~60% of the domain-mean Eurasian cooling trend since 1990 ( $-0.95 \text{ K decade}^{-1}$ ). For ENSO and the PCH, this is respectively only 17% and 24%. When ENSO is combined with cluster 7, the percentage of the recovered cooling trend in Eurasia jumps to 77%. This shows that the trend toward more-persistent weak polar vortex states can explain most of the winter cooling trend over northern Eurasia.

Next, we consider Eurasian cold extremes (defined as days when the temperature anomaly over the Eurasian sector is below  $-5^\circ\text{C}$ , coinciding with the 10th percentile) and calculate the relative occurrence frequency of each cluster. For the null hypothesis, that is, that stratospheric variability plays no role, one would expect for each cluster a frequency during cold extremes approximately equal to its occurrence over all winter days as displayed in Fig. 1. Though only 8.25% of all considered days were assigned to cluster 7 (Fig. 1), the likelihood of cluster 7 days roughly doubles to 17.2% if only cold days are considered (Fig. 7b), which is a significant increase ( $p < 0.01$ , according to a chi-square test). The occurrence of cluster 6 days also exceeds the expected frequency, whereas the strong vortex clusters 1–3 occur

less often than statistically expected. Similarly, only 3% of the hottest days (exceeding the 90th percentile) are cluster 7 days, which significantly ( $p < 0.01$ ) differs from the expected occurrence of ~8% (not shown).

To assess the direction of causality between weak vortex states and Eurasian cold extremes, we perform lagged coincidence analysis. In the week before the onset of cluster 7 events, most days are assigned to weak polar vortex states (51% cluster 6, 20% cluster 5), which themselves are already associated with low temperature anomalies over Eurasia. The mean Eurasian temperature anomaly preceding cluster 7 events is  $-1.2^\circ\text{C}$ , but it reaches its minimum value during cluster 7 events with an average anomaly of  $-1.9^\circ\text{C}$ . Thus, cluster 7 days represent the peak of the polar vortex disturbance as well as the peak of the cold anomalies over the northern Eurasian sector. Consistently, in the week before the onset of a cold event, the likelihood of cluster 6 is anomalously high. If we merge clusters 6 and 7, the mean Eurasian temperature during these weak vortex



**FIG. 7.** (a) Coincidence analysis for extreme cold days over the Eurasian sector ( $50^\circ\text{--}65^\circ\text{N}$ ,  $15^\circ\text{--}130^\circ\text{E}$ ). (b) The deviation from the statistically expected occurrence frequency (as displayed in Fig. 1) of each cluster is shown during cold days ( $<-5^\circ\text{C}$ ).

states is still negative ( $-1.1^{\circ}\text{C}$ ), but the temperature in the preceding week is anomalously warm at  $+0.4^{\circ}\text{C}$ . Thus, since weak vortex events (clusters 6 and 7) are preceded by positive temperature anomalies in Eurasia, we propose that the observed cooling trend in this region is more likely the consequence of the vortex weakening rather than its cause. Moreover, we found that cluster 7 Granger causes Eurasian temperature variability in winter and that the opposite is not true, which further supports this assumption (see supplemental information). This is also consistent with recent findings, showing that cold spells over Eurasia are longer lasting if accompanied by a weak polar vortex (Garfinkel et al. 2017).

### PRECURSORS AND POTENTIAL REASONS FOR WEAK POLAR VORTEX STATES.

Finally, we analyze potential reasons for the observed trends in frequency of the polar vortex clusters 1 and 7. Both observational and modeling studies have shown that strong upward wave propagation in the upper troposphere can weaken the stratospheric flow (Jaiser et al. 2013; Kretschmer et al. 2016; Kim et al. 2014; Polvani and Waugh 2004; Shaw et al. 2014) as expected on theoretical grounds (Matsuno 1971) and is often preceded by distinct sea level pressure anomalies (Baldwin and Dunkerton 1999; Cohen and Jones 2011; Kretschmer et al. 2016). Therefore, we created composites of anomalies in sea level pressure (30–10 days before the start date of cluster events) and meridional heat flux  $v^*T^*$  at 100 hPa (10 days prior to the cluster events), which is a common proxy for vertical wave propagation (Figs. 8a–d, showing only those for clusters 1 and 7). The choice of time lags was motivated by previous studies (Kretschmer et al. 2016; Kim et al. 2014; Cohen and Jones 2011), but the results are also robust for time shifts of a few days. In the month before the onset of a weak polar vortex event, sea level pressure over most of northwest Eurasia is anomalously high while sea level pressure over the Chukchi Sea, North America, and the northern Atlantic is anomalously low (Fig. 8b). This pressure dipole is followed by an anomalously strong poleward heat flux over northern Europe, central Asia, and the Chukchi and Beaufort Seas and a lower-than-normal heat flux north over the Lena River and over northern Canada (Fig. 8d). In contrast, strong polar vortex events are preceded by patterns of opposite sign of sea level pressure and heat flux anomalies but are of less amplitude (Figs. 8a,c).

Vice versa, to test if high western Siberian sea level pressure events are also followed by weak polar vortex states (in a statistically significant way), we create an

index of area-averaged sea level pressure over the Ural Mountains region ( $45^{\circ}$ – $70^{\circ}\text{N}$ ,  $40^{\circ}$ – $85^{\circ}\text{E}$ ) for December and January (Cohen et al. 2014b; Kretschmer et al. 2016). We define strong western Siberian high events when the index exceeds 1,035 hPa, which corresponds to the 93rd percentile. In the month following high sea level pressure over western Siberia in December and January, the frequency of cluster 7 events triples (from 8.25% to 26.1%,  $p < 0.01$ ), whereas that of cluster 1 events halves (from 16.12% to 7.15%,  $p < 0.01$ ; see appendix). Thus, not only are cluster 7 events preceded by high sea level pressure over the Ural Mountains, but high sea level pressure anomalies over western Siberia also strongly increase the likelihood of weak polar vortex states.

The cluster 7  $v^*T^*$  precursor anomalies (Fig. 8d) correspond to a reinforcement of the climatological poleward heat flux, which was shown to lead to a weakened polar vortex (Polvani and Waugh 2004; Dunn-Sigouin and Shaw 2015; Shaw et al. 2014). Moreover, the sea level pressure composites for cluster 7 (Fig. 8b) are consistent with different studies linking increased vertical wave propagation to tropospheric forcing (Kretschmer et al. 2016; Feldstein and Lee 2014; Cohen and Jones 2011). Constructive interference with the climatological high leads to more vertical wave activity in the upper troposphere and thus a weakening of the polar vortex (Feldstein and Lee 2014; Kretschmer et al. 2016; Cohen et al. 2014b; Smith et al. 2010). Therefore, the detected precursors of cluster 7 are in accordance with known physical mechanisms of troposphere–stratosphere coupling.

The formation of anomalous high pressure over northern Eurasia has been associated with late autumn Barents and Kara sea ice loss and enhanced Eurasian October snow cover extent (Kim et al. 2014; Kretschmer et al. 2016; Feldstein and Lee 2014; Cohen et al. 2014b). Therefore, we speculate that these processes, which have been linked to Arctic amplification (Cohen et al. 2014a; Overland et al. 2011) and which have also been reproduced by climate models (Jaiser et al. 2016; Handorf et al. 2015), contributed to the patterns that favor a weakened polar vortex represented by cluster 7 (Figs. 8b,d). Moreover, the involved time lag of approximately 3 months (Kretschmer et al. 2016) for these Arctic-driven mechanisms might explain why clustering with November and December data exhibits no trends in the frequency of the different vortex clusters (Fig. ES9). The negative sea level pressure anomalies over the North Pacific for cluster 7 events (Fig. 8b) are also similar to patterns associated with El Niño years, which are associated with a weak polar vortex (Baldwin and O’Sullivan 1995; Polvani

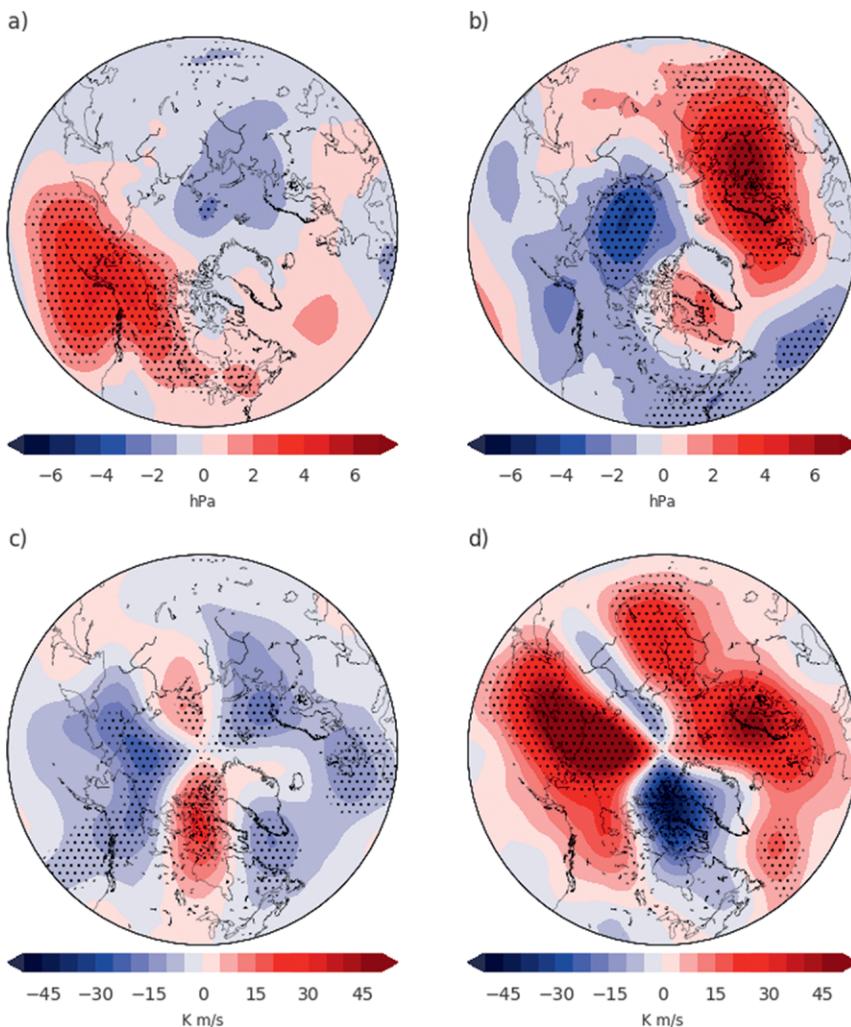
et al. 2017). However, since different ENSO indices did not show any trend over the last decades, the weakening polar vortex can probably not be explained by ENSO-related teleconnections. Nevertheless, the interplay between different tropical teleconnections (Garfinkel and Hartmann 2008), natural variability (McCusker et al. 2016), and variability in atmospheric responses to Arctic sea ice loss (Screen and Francis 2016) as well as impacts of regional differences in sea ice decline (Sun et al. 2015) might influence the stratospheric response. This interplay of possible causal drivers requires further analyses using both climate models and observations (Overland et al. 2016).

**CONCLUSIONS.** Using cluster analysis, we identified dominant patterns of the stratospheric polar vortex in boreal winter. We showed that the polar vortex weakening over the last four decades was a result of more-persistent weak polar vortex states (cluster 7) and less frequent strong polar vortex events (cluster 1) rather than an overall weakening. This shift in polar vortex states can account for most of the recent winter cooling trends over Eurasian mid-latitudes via stratosphere–troposphere coupling. The observed sea level pressure and heat flux precursors are in agreement with proposed physical mechanisms and can explain the weakening of the polar vortex via a dynamical troposphere–stratosphere coupling.

Our analysis shows that the Eurasian cooling trend in the era of Arctic amplification can largely be explained by polar vortex variability. Understanding the two-way link between stratospheric and tropospheric circulation is thus essential for understanding winter teleconnections in the Northern Hemisphere. Any improvements in wintertime seasonal forecasts are likely to depend on our comprehension of competing

drivers, including the influence of stratospheric variability (Sigmond et al. 2013; Kretschmer et al. 2016).

**ACKNOWLEDGMENTS.** We thank Peter Hitchcock and two anonymous reviewers for their useful comments and suggestions to improve the manuscript, and we thank ECMWF and GMAO for making the ERA-Interim and the MERRA-2 data available. The work was supported by the German Federal Ministry of Education and Research, Grant 01LN1304A (M.K., D.C.); National Science Foundation (NSF) Grants AGS-1303647 and PLR-1504361 (J.C.); NOAA Grant NA15OAR4310077 (J.C.); and NSF Grant AGS-1303604 (E.T.). The research project resulted from M.K. visiting J.C., and M.K. would like to thank AER and Harvard for hosting. E.T. would like to thank the Weizmann Institute for its hospitality during parts of this work.



**FIG. 8. Precursors to cluster events.** Composite of (detrended) sea level pressure anomalies 30–10 days prior to start days of (a) cluster 1 and (b) cluster 7 events. (c),(d) As in (a) and (b), respectively, but for (detrended) poleward heat flux  $v \cdot T^*$  anomalies at 100 hPa averaged 10 days before onset of cluster event. In all panels, significant values ( $p < 0.05$ ) are indicated with dots.

**APPENDIX: METHODS.** *Clustering.* The hierarchical cluster algorithm starts with  $n$  clusters (the starting vectors) and then iteratively merges two clusters until only one cluster (the mean over all vectors) exists. In each step the clusters with minimal distance are merged and their mean is calculated. Here we use Ward's metric criteria, meaning that the two clusters to be merged at each step are those that result in the minimal increase in variance in the merged cluster, over all possible unions of clusters.

While more computationally demanding, hierarchical clustering has the advantage over other clustering techniques such as  $k$ -means or self-organizing maps (SOM), that no a priori knowledge on the number of clusters is required. Each of the  $n - 1$  merging steps can be tracked back and the optimal number of clusters can thus be defined afterward. The structure of the clustering process is visualized in a dendrogram (Fig. ES1) and is used to choose the number of clusters, although that choice does require some subjective judgment (see supplemental information).

*Statistical analysis.* For the comparison of the first and second half of the studied time period (Fig. 3a), we test for significance by randomly picking blocks of 7 days of each season from the time series that contain the cluster events. The length was chosen based on the mean event length of all clusters during the whole period. The blocks are then shuffled between years and calendar slots, creating artificial time series, but the order within the blocks is maintained (preserving the intraseasonal autocorrelation of the original time series). This way, we create a new time series from which we calculate the frequency difference of the two data halves. We do this 10,000 times and calculate the percentiles of the observed frequency difference.

*Composite plots.* Before computing the temperature composites (Fig. 4), the data were detrended to prevent biases due to trends in the occurrence of the clusters. The significance of the composites is tested, creating 10,000 artificial time series by randomly picking and shuffling blocks of the original time series (with a block length of 5 days). For each newly created time series, we pick as many days as were used to form the composite, but we also keep the start days and length of the identified events from the original time series to account for a potential increase in autocorrelation during long-lasting cluster events. For the precursors we similarly composite (Fig. 8), but we neglect polar vortex data of the very first 30 days (i.e., 1–30 January 1979) since leading sea level pressure and  $v^*T^*$  values are not included in the reanalysis datasets. The

composites are then formed over the days preceding the onset of the identified cluster event.

*Coincidence analysis.* To assess the coincidence of cold events in Eurasia and weak polar vortex states, we define cold days as days when the mean temperature anomaly over the Eurasian sector is below a certain threshold, for example, below  $-5^\circ\text{C}$ . Next, we calculate the frequency of each cluster on cold days and compare to the frequency of each cluster on all days. To test significance for the observed frequency of a specific cluster  $i$ , we apply a chi-square test to the contingency table containing the cluster number (occurrence of cluster  $i$ /other than cluster  $i$ ) and the extreme event (occurrence of cold extreme/no cold extreme).

For the coincidence of anomalous sea level pressure over western Siberia and weak polar vortex states, we calculate a baseline (i.e., climatological) frequency for each cluster based on the 25–35 days following every day in December and January (neglecting December 1978, which is not included), which coincides with the absolute frequencies of the different clusters as shown in Fig. 1. We compare that to the frequency for each cluster based on the 15–35-day periods following Siberian high events. To assess the significance, we create 1,000 synthetic time series with the same number of Siberian high events as in observations, but randomly distributed in time. This way, we get a distribution of the cluster events frequencies following Siberian high events and can calculate percentiles to get the corresponding  $p$  value.

## REFERENCES

- Baldwin, M. P., and D. O'Sullivan, 1995: Stratospheric effects of ENSO-related tropospheric circulation anomalies. *J. Climate*, **8**, 649–667, [https://doi.org/10.1175/1520-0442\(1995\)008<0649:SEOERT>2.0.CO;2](https://doi.org/10.1175/1520-0442(1995)008<0649:SEOERT>2.0.CO;2).
- , and T. J. Dunkerton, 1999: Propagation of the Arctic Oscillation from the stratosphere to the troposphere. *J. Geophys. Res.*, **104**, 30937, <https://doi.org/10.1029/1999JD900445>.
- , and —, 2001: Stratospheric harbingers of anomalous weather regimes. *Science*, **294**, 581–584, <https://doi.org/10.1126/science.1063315>.
- Butler, A. H., L. M. Polvani, and C. Deser, 2014: Separating the stratospheric and tropospheric pathways of El Niño–Southern Oscillation teleconnections. *Environ. Res. Lett.*, **9**, 024014, <https://doi.org/10.1088/1748-9326/9/2/024014>.
- , D. J. Seidel, S. C. Hardiman, N. Butchart, T. Birner, and A. Match, 2015: Defining sudden stratospheric

- warmings. *Bull. Amer. Meteor. Soc.*, **96**, 1913–1928, <https://doi.org/10.1175/BAMS-D-13-00173.1>.
- Charlton, A. J., and L. M. Polvani, 2007: A new look at stratospheric sudden warmings. Part I: Climatology and modeling benchmarks. *J. Climate*, **20**, 449–469, <https://doi.org/10.1175/JCLI3996.1>.
- Cheng, X., and J. M. Wallace, 1993: Cluster analysis of the Northern Hemisphere wintertime 500-hPa height field: Spatial patterns. *J. Atmos. Sci.*, **50**, 2674–2696, [https://doi.org/10.1175/1520-0469\(1993\)050<2674:CAOTNH>2.0.CO;2](https://doi.org/10.1175/1520-0469(1993)050<2674:CAOTNH>2.0.CO;2).
- Cohen, J., 2016: An observational analysis: Tropical relative to Arctic influence on midlatitude weather in the era of Arctic amplification. *Geophys. Res. Lett.*, **43**, 5287–5294, <https://doi.org/10.1002/2016GL069102>.
- , and J. Jones, 2011: Tropospheric precursors and stratospheric warmings. *J. Climate*, **24**, 6562–6572, <https://doi.org/10.1175/2011JCLI4160.1>.
- , —, J. C. Furtado, and E. Tziperman, 2013: Warm Arctic, cold continents. *Oceanography*, **26**, 1–12, <https://doi.org/10.5670/oceanog.2013.70>.
- , and Coauthors, 2014a: Recent Arctic amplification and extreme mid-latitude weather. *Nat. Geosci.*, **7**, 627–637, <https://doi.org/10.1038/ngeo2234>.
- , J. C. Furtado, J. Jones, M. Barlow, D. Whittleston, and D. Entekhabi, 2014b: Linking Siberian snow cover to precursors of stratospheric variability. *J. Climate*, **27**, 5422–5432, <https://doi.org/10.1175/JCLI-D-13-00779.1>.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, <https://doi.org/10.1002/qj.828>.
- Dunn-Sigouin, E., and T. A. Shaw, 2015: Comparing and contrasting extreme stratospheric events, including their coupling to the tropospheric circulation. *J. Geophys. Res. Atmos.*, **120**, 1374–1390, <https://doi.org/10.1002/2014JD022116>.
- Feldstein, S. B., and S. Lee, 2014: Intraseasonal and interdecadal jet shifts in the Northern Hemisphere: The role of warm pool tropical convection and sea ice. *J. Climate*, **27**, 6497–6518, <https://doi.org/10.1175/JCLI-D-14-00057.1>.
- Garfinkel, C. I., and D. L. Hartmann, 2008: Different ENSO teleconnections and their effects on the stratospheric polar vortex. *J. Geophys. Res.*, **113**, D18114, <https://doi.org/10.1029/2008JD009920>.
- , S.-W. Son, K. Song, V. Aquila, and L. D. Oman, 2017: Stratospheric variability contributed to and sustained the recent hiatus in Eurasian winter warming. *Geophys. Res. Lett.*, **44**, 374–382, <https://doi.org/10.1002/2016GL072035>.
- Handorf, D., R. Jaiser, K. Dethloff, A. Rinke, and J. Cohen, 2015: Impacts of Arctic sea ice and continental snow cover changes on atmospheric winter teleconnections. *Geophys. Res. Lett.*, **42**, 2367–2377, <https://doi.org/10.1002/2015GL063203>.
- Hitchcock, P., and I. R. Simpson, 2014: The downward influence of stratospheric sudden warmings. *J. Atmos. Sci.*, **71**, 3856–3876, <https://doi.org/10.1175/JAS-D-14-0012.1>.
- , —, and G. L. Manney, 2013: Statistical characterization of Arctic polar-night jet oscillation events. *J. Climate*, **26**, 2096–2116, <https://doi.org/10.1175/JCLI-D-12-00202.1>.
- Horton, D. E., N. C. Johnson, D. Singh, D. L. Swain, B. Rajaratnam, and N. S. Diffenbaugh, 2015: Contribution of changes in atmospheric circulation patterns to extreme temperature trends. *Nature*, **522**, 465–469, <https://doi.org/10.1038/nature14550>.
- Jaiser, R., K. Dethloff, and D. Handorf, 2013: Stratospheric response to Arctic sea ice retreat and associated planetary wave propagation changes. *Tellus*, **65A**, 19375, <https://doi.org/10.3402/tellusa.v65i0.19375>.
- , T. Nakamura, D. Handorf, K. Dethloff, J. Ukita, and K. Yamazaki, 2016: Atmospheric winter response to Arctic sea ice changes in reanalysis data and model simulations. *J. Geophys. Res. Atmos.*, **121**, 7564–7577, <https://doi.org/10.1002/2015JD024679>.
- Kim, B.-M., S.-W. Son, S.-K. Min, J.-H. Jeong, S.-J. Kim, X. Zhang, T. Shim, and J.-H. Yoon, 2014: Weakening of the stratospheric polar vortex by Arctic sea-ice loss. *Nat. Commun.*, **5**, 4646, <https://doi.org/10.1038/ncomms5646>.
- Kodera, K., H. Mukougawa, P. Maury, M. Ueda, and C. Claud, 2016: Absorbing and reflecting sudden stratospheric warming events and their relationship with tropospheric circulation. *J. Geophys. Res. Atmos.*, **121**, 80–94, <https://doi.org/10.1002/2015JD023359>.
- Kolstad, E. W., T. Breiteig, and A. A. Scaife, 2010: The association between stratospheric weak polar vortex events and cold air outbreaks in the Northern Hemisphere. *Quart. J. Roy. Meteor. Soc.*, **136**, 886–893, <https://doi.org/10.1002/qj.620>.
- Kretschmer, M., D. Coumou, J. F. Donges, and J. Runge, 2016: Using causal effect networks to analyze different Arctic drivers of midlatitude winter circulation. *J. Climate*, **29**, 4069–4081, <https://doi.org/10.1175/JCLI-D-15-0654.1>.
- Lee, S., and S. B. Feldstein, 2013: Detecting ozone- and greenhouse gas-driven wind trends with observational data. *Science*, **339**, 563–567, <https://doi.org/10.1126/science.1225154>.
- Li, C., B. Stevens, and J. Marotzke, 2015: Eurasian winter cooling in the warming hiatus of 1998–2012. *Geophys. Res. Lett.*, **42**, 8131–8139, <https://doi.org/10.1002/2015GL065327>.

- Matsuno, T., 1971: A dynamical model of stratospheric warmings. *J. Atmos. Sci.*, **28**, 1479–1494, [https://doi.org/10.1175/1520-0469\(1971\)028<1479:ADMOTS>2.0.CO;2](https://doi.org/10.1175/1520-0469(1971)028<1479:ADMOTS>2.0.CO;2).
- McCusker, K. E., J. C. Fyfe, and M. Sigmond, 2016: Twenty-five winters of unexpected Eurasian cooling unlikely due to Arctic sea-ice loss. *Nat. Geosci.*, **9**, 838–842, <https://doi.org/10.1038/ngeo2820>.
- Molod, A., L. Takacs, M. Suarez, and J. Bacmeister, 2015: Development of the GEOS-5 atmospheric general circulation model: Evolution from MERRA to MERRA2. *Geosci. Model Dev.*, **8**, 1339–1356, <https://doi.org/10.5194/gmd-8-1339-2015>.
- Mori, M., M. Watanabe, H. Shiogama, J. Inoue, and M. Kimoto, 2014: Robust Arctic sea-ice influence on the frequent Eurasian cold winters in past decades. *Nat. Geosci.*, **7**, 869–873, <https://doi.org/10.1038/ngeo2277>.
- Overland, J. E., K. R. Wood, and M. Wang, 2011: Warm Arctic—Cold continents: Climate impacts of the newly open Arctic Sea. *Polar Res.*, **30**, 15787, <https://doi.org/10.3402/polar.v30i0.15787>.
- , and Coauthors, 2016: Nonlinear response of mid-latitude weather to the changing Arctic. *Nat. Climate Change*, **6**, 992–999, <https://doi.org/10.1038/nclimate3121>.
- Palmer, T., 2014: Record-breaking winters and global climate change. *Science*, **344**, 803–804, <https://doi.org/10.1126/science.1255147>.
- Polvani, L. M., and D. W. Waugh, 2004: Upward wave activity flux as a precursor to extreme stratospheric events and subsequent anomalous surface weather regimes. *J. Climate*, **17**, 3548–3554, [https://doi.org/10.1175/1520-0442\(2004\)017<3548:UWAFAA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<3548:UWAFAA>2.0.CO;2).
- , L. Sun, A.H. Butler, J.H. Richter, and C. Deser, 2017: Distinguishing stratospheric sudden warmings from ENSO as key drivers of wintertime climate variability over the North Atlantic and Eurasia. *J. Climate*, **30**, 1959–1969, <https://doi.org/10.1175/JCLI-D-16-0277.1>.
- Runde, T., M. Dameris, H. Garny, and D. E. Kinnison, 2016: Classification of stratospheric extreme events according to their downward propagation to the troposphere. *Geophys. Res. Lett.*, **43**, 6665–6672, <https://doi.org/10.1002/2016GL069569>.
- Scaife, A. A., and Coauthors, 2016: Seasonal winter forecasts and the stratosphere. *Atmos. Sci. Lett.*, **17**, 51–56, <https://doi.org/10.1002/asl.598>.
- Screen, J. A., and I. Simmonds, 2010: The central role of diminishing sea ice in recent Arctic temperature amplification. *Nature*, **464**, 1334–1337, <https://doi.org/10.1038/nature09051>.
- , and J. A. Francis, 2016: Contribution of sea-ice loss to Arctic amplification is regulated by Pacific Ocean decadal variability. *Nat. Climate Change*, **6**, 856–860, <https://doi.org/10.1038/nclimate3011>.
- Shaw, T. A., J. Perlwitz, and O. Weiner, 2014: Troposphere-stratosphere coupling: Links to North Atlantic weather and climate, including their representation in CMIP5 models. *J. Geophys. Res. Atmos.*, **119**, 5864–5880, <https://doi.org/10.1002/2013JD021191>.
- Shepherd, T. G., 2016: Effects of a warming Arctic. *Science*, **353**, 989–990, <https://doi.org/10.1126/science.aag2349>.
- Sigmond, M., J. F. Scinocca, V. V. Kharin, and T. G. Shepherd, 2013: Enhanced seasonal forecast skill following stratospheric sudden warmings. *Nat. Geosci.*, **6**, 98–102, <https://doi.org/10.1038/ngeo1698>.
- Smith, K. L., C. G. Fletcher, P. J. Kushner, K. L. Smith, C. G. Fletcher, and P. J. Kushner, 2010: The role of linear interference in the annular mode response to extratropical surface forcing. *J. Climate*, **23**, 6036–6050, <https://doi.org/10.1175/2010JCLI3606.1>.
- Sun, L., C. Deser, and R. A. Tomas, 2015: Mechanisms of stratospheric and tropospheric circulation response to projected Arctic sea ice loss. *J. Climate*, **28**, 7824–7845, <https://doi.org/10.1175/JCLI-D-15-0169.1>.
- , J. Perlwitz, and M. Hoerling, 2016: What caused the recent “Warm Arctic, Cold Continents” trend pattern in winter temperatures? *Geophys. Res. Lett.*, **43**, 5345–5352, <https://doi.org/10.1002/2016GL069024>.

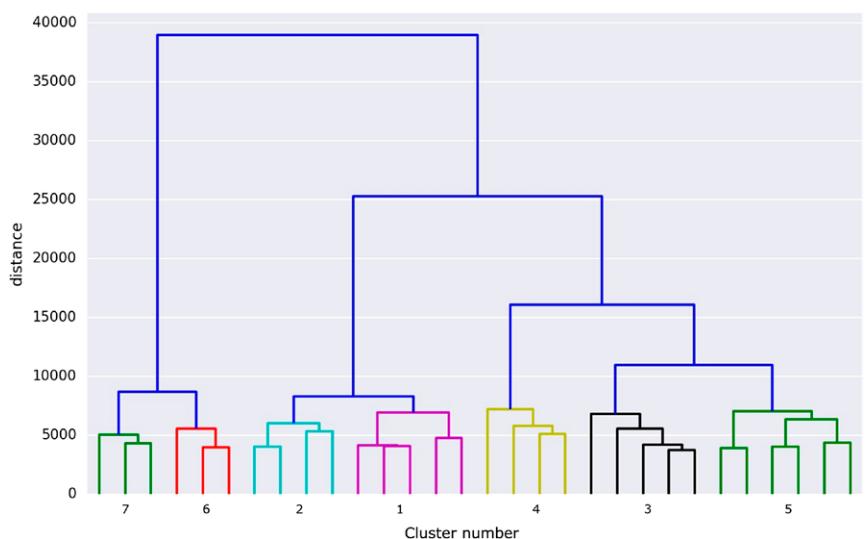
# MORE-PERSISTENT WEAK STRATOSPHERIC POLAR VORTEX STATES LINKED TO COLD EXTREMES

MARLENE KRETSCHMER, DIM COUMOU, LAURIE AGEL, MATHEW BARLOW, ELI TZIPERMAN, AND JUDAH COHEN

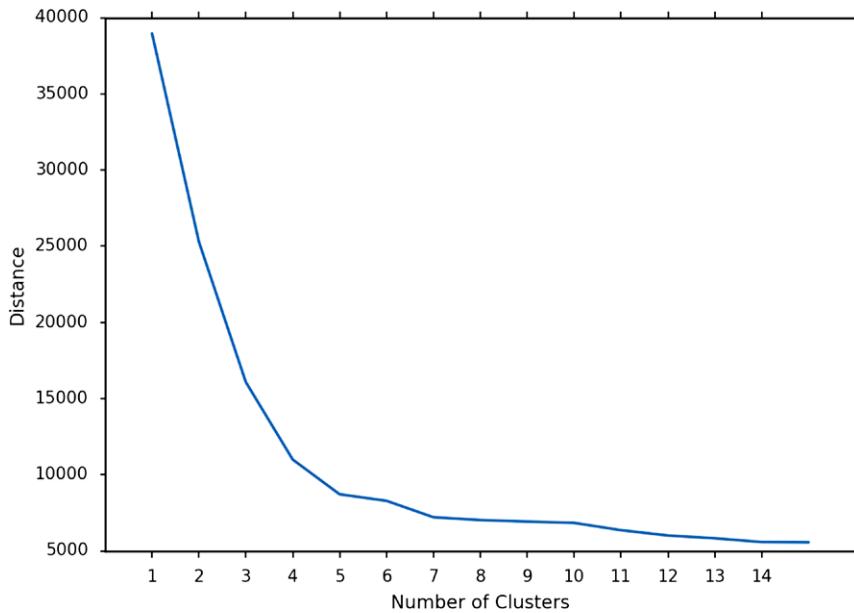
This document is a supplement to “More-Persistent Weak Stratospheric Polar Vortex States Linked to Cold Extremes” by Marlene Kretschmer, Dim Coumou, Laurie Agel, Mathew Barlow, Eli Tziperman, and Judah Cohen (*Bull. Amer. Meteor. Soc.*, **99**, 49–60) • ©2018 American Meteorological Society • Corresponding authors: Marlene Kretschmer, kretschmer@pik-potsdam.de; Judah Cohen, jcohen@aer.com • DOI:10.1175/BAMS-D-16-0259.2

**CHOICE OF NUMBER OF CLUSTERS.** The choice of number of clusters is a compromise between reducing the data dimensionality and preserving as much variability as possible. The dendrogram of a hierarchical cluster analysis visualizes the required distance in each merging step and helps to select a number of clusters. For each merging step, the required distance for merging two clusters is shown on the y axis of the dendrogram plot. Similar clusters are lined via shorter y-axis segments, and large y distances within one cluster should be avoided. In the present analysis, we used Ward’s minimum variance algorithm such that distance refers to the variance of the elements in the cluster. Therefore, large increases in this distance mean that after

merging two clusters, the within-cluster variance increased by much, whereas smaller jumps refer to the merger of similar clusters. Based on the dendrogram (Fig. ES1) of the last 30 merging steps for clustering

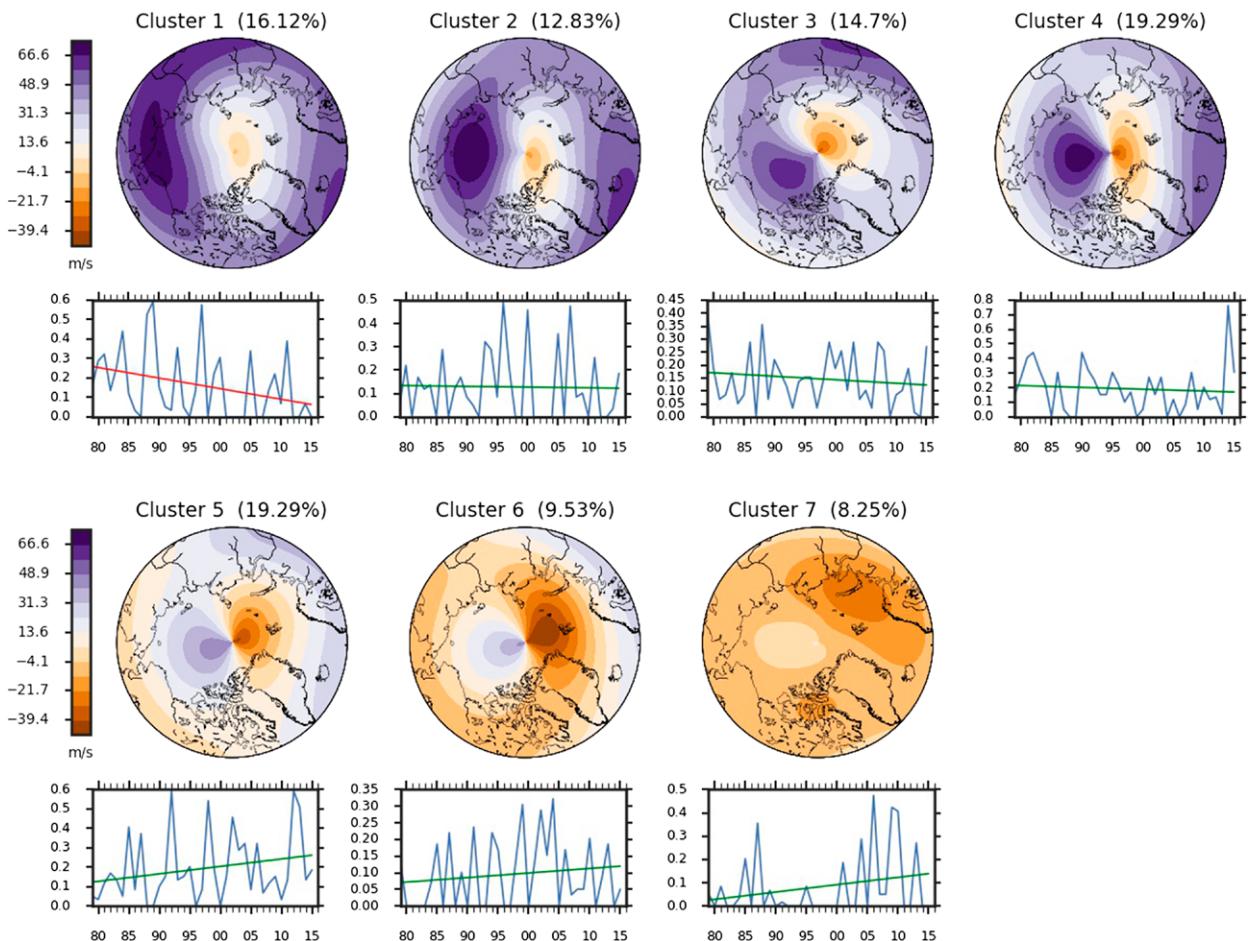


**Fig. ES1. Dendrogram of hierarchical clustering with area-weighted zonal wind velocity anomalies poleward of 60°N at 10 hPa for the last 30 steps of the clustering. Each color represents one cluster.**



**FIG. ES2. Elbow plot of hierarchical clustering with area-weighted zonal wind velocity anomalies poleward of  $60^{\circ}\text{N}$  at 10 hPa for the last 15 steps of the clustering.**

with area-weighted zonal wind anomalies in January and February, we conclude that seven is an appropriate choice for the number of clusters (i.e., the different-colored clusters) because the within-cluster distances are similar. Moreover, the so-called “elbow” plot (Fig. ES2), which shows the cluster distance as a function of the number of clusters, shows a bend for seven clusters. This means that the distance is only slightly smaller for a greater number of clusters but the distance growth accelerates for fewer clusters. As indicated by the dendrogram (Fig. ES1) only five clusters would be

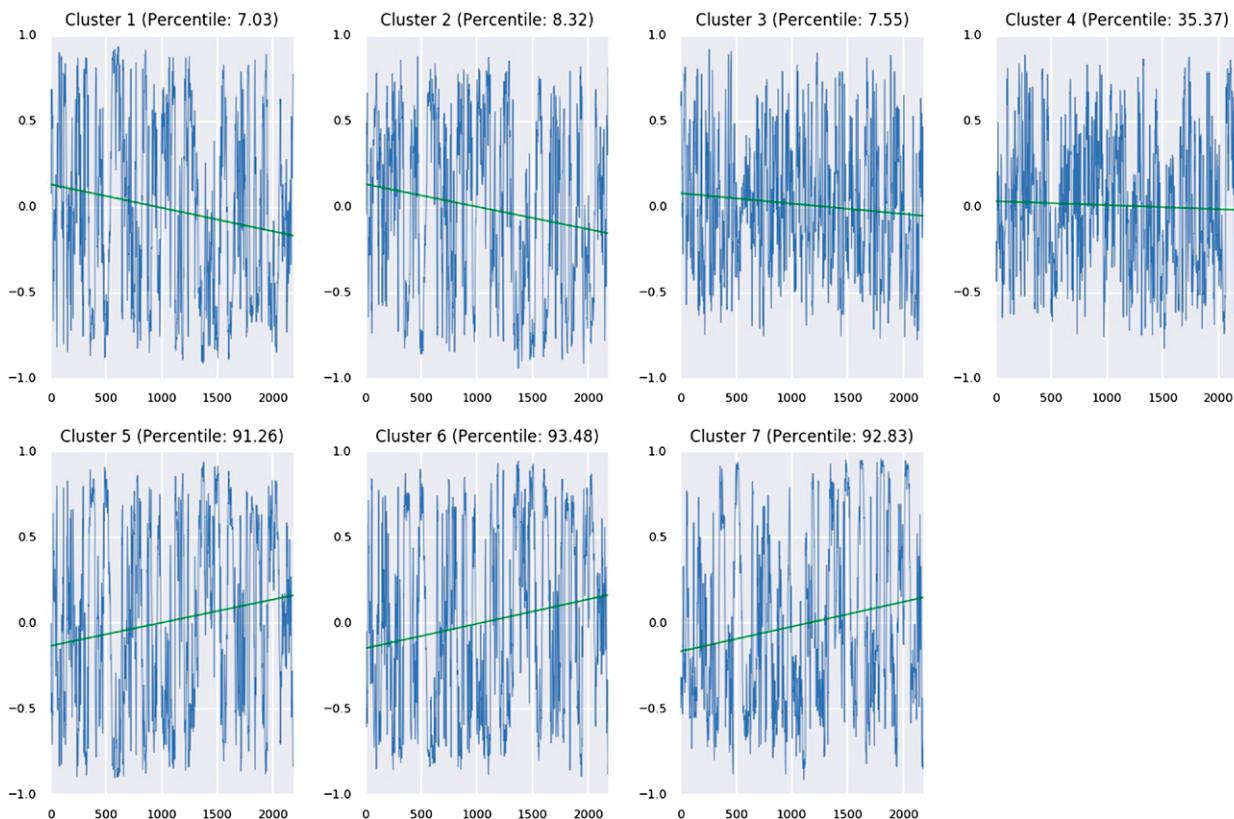


**FIG. ES3. As in Fig. 1 in the main text, but cluster representatives shown for zonal wind velocity instead of GPH.**

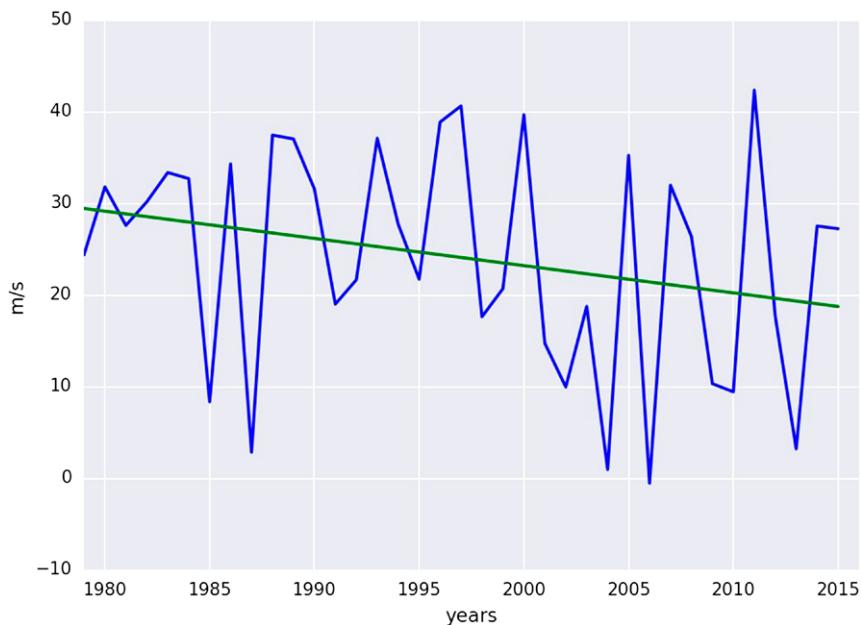
another reasonable choice, which would mean the merger of cluster 1 and cluster 2 as well as of cluster 6 and cluster 7. However, we find that our results (in particular with respect to the trends in the regression models based on cluster 7) are insensitive to this choice. Moreover, we find that for all tested numbers of clusters from 2 to 20, the strongest (weakest) vortex cluster shows a decreasing (increasing) trend in seasonal frequency as well as an increasing (decreasing) trend in daily pattern correlation (see Fig. ES4 for seven clusters).

**DAILY PATTERN CORRELATION.** To derive the trend of the daily pattern correlation time series (Fig. ES4), we perform linear regression by calculating the least squares fit trend line. Since the daily pattern correlation time series (Fig. ES4) are autocorrelated, applying a Student's *t* test to test significance of the slope would violate the assumption of independent residuals. Therefore, we apply bootstrapping on the data, preserving the intraseasonal and seasonal variability of the original time series. To assess the significance of the slopes, we perform a (one sided)

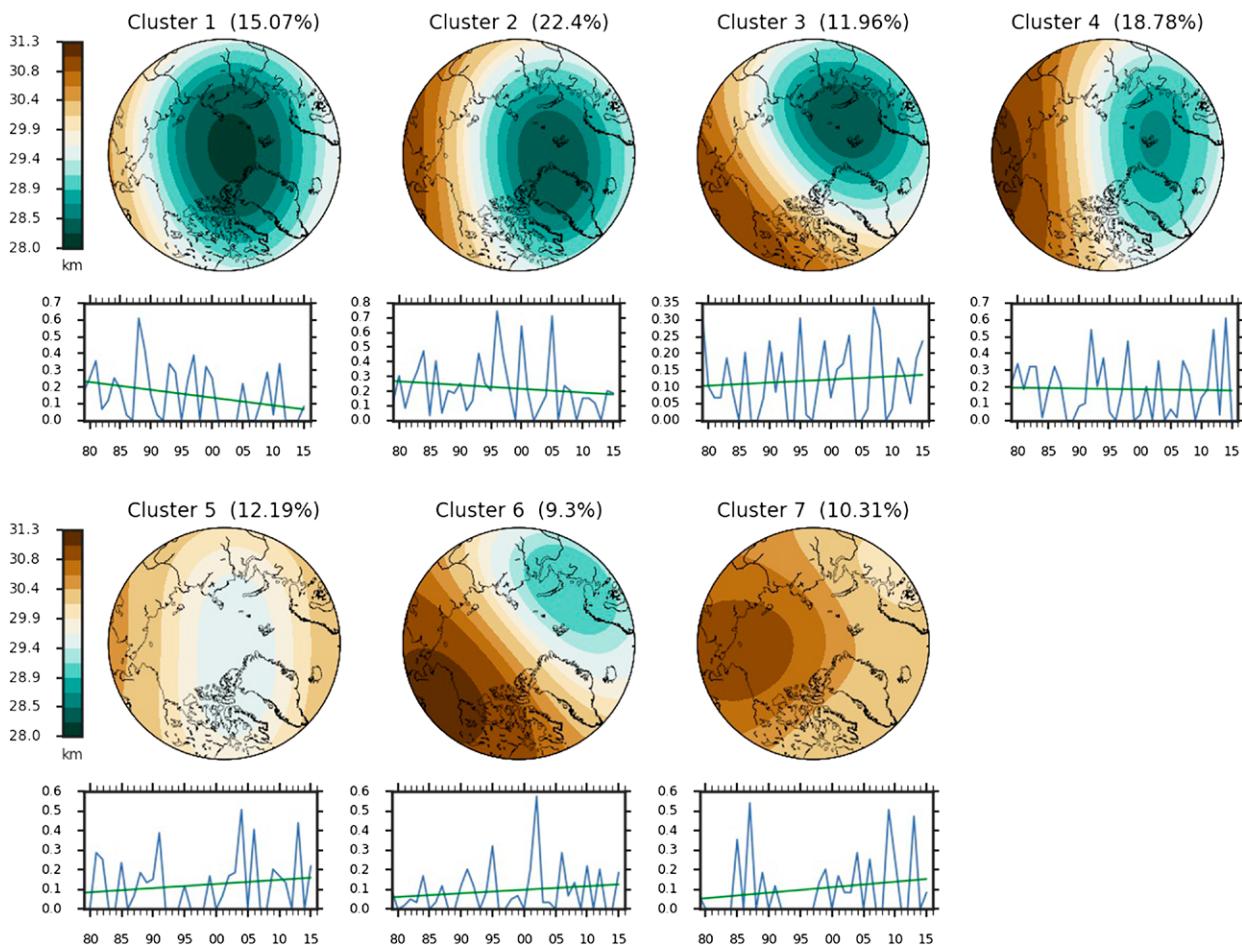
significance test with the H1 hypothesis that those clusters with a downward (upward) trend in seasonal frequency (see Fig. 1 in main text) also show a downward (upward) trend in pattern correlation. First, for each year, we determine daily anomalies for the pattern correlations by subtracting the seasonal-mean value for that year. Next, we randomly pick 15 blocks from each season, each with a length of 4 days. The block length was chosen based on the intraseasonal autocorrelation functions, which dropped below a significance threshold ( $p < 0.05$ ) after 3–4 days. The blocks are then shuffled between years and calendar slots creating artificial time series of daily anomalies, but the order within the blocks is maintained (preserving the intraseasonal autocorrelation of the original time series). Finally, we add the shuffled seasonal means of the original data to the daily values of each newly constructed year. This way, we also keep the variability between the different years. We calculate the slope of the linear least squares fit of this artificial time series and repeat the procedure 10,000 times in order to derive the (one sided) percentiles of the original slope with respect to the calculated distribution.



**FIG. ES4.** Trends in daily pattern correlation for each cluster and the percentile of the slope in brackets (see appendix in the main text).



**FIG. ES5. Mean polar cap (poleward of 60°N) wind velocity in winter (JF) from 1979 to 2015 and least squares trend line ( $p = 0.11$ ).**



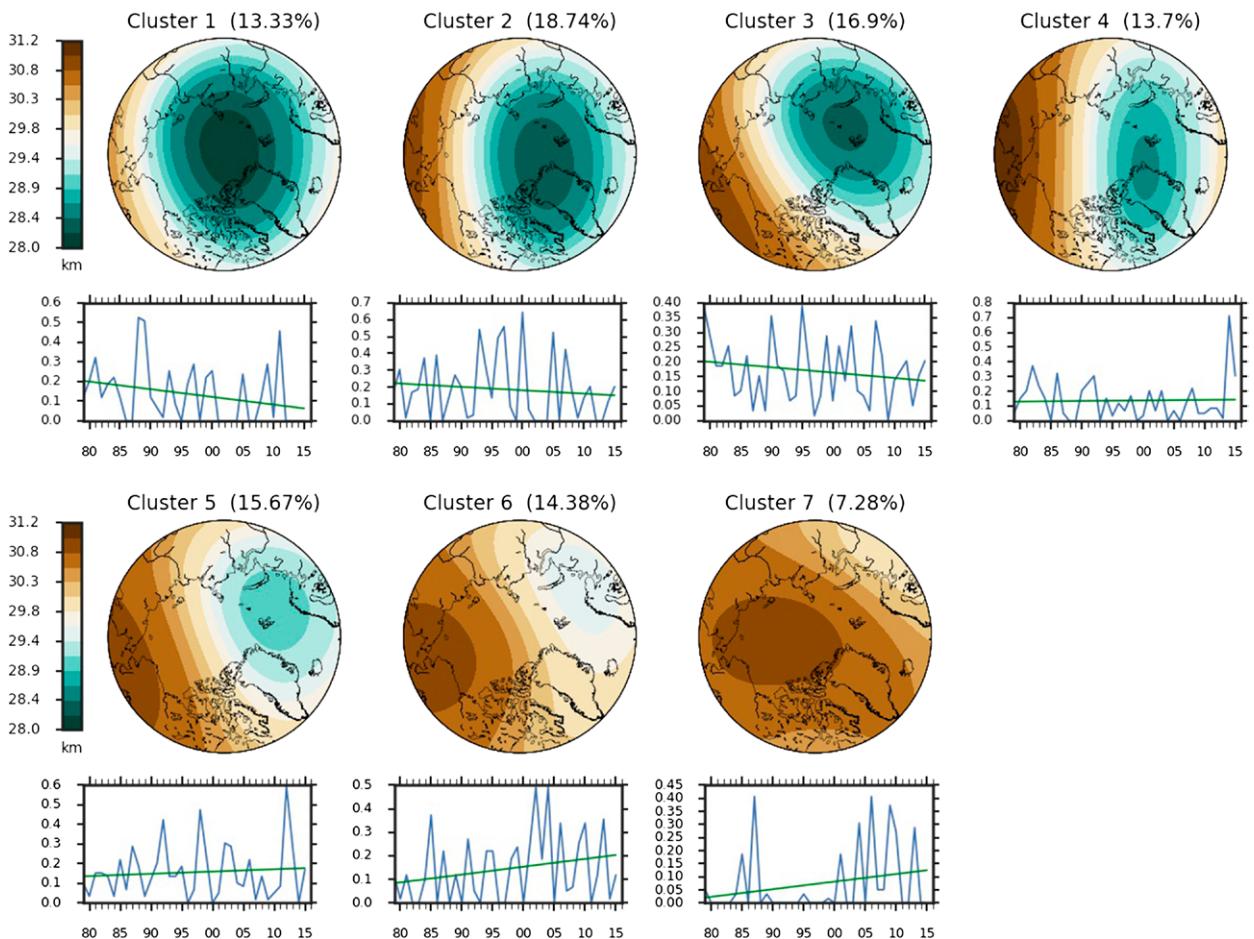
**FIG. ES6. As in Fig. 1 in the main text, but for clustering with GPH anomalies.**

**OTHER SENSITIVITY ANALYSES.** We repeat the cluster analysis with other variables that describe the stratospheric polar vortex (SPV). First, we repeat the analyses using geopotential height (GPH) data over the same polar-cap region and at same pressure level (10 hPa). Then, we repeat the analyses with zonal wind and GPH data averaged over the pressure levels at 10–50 hPa and for 100 hPa, respectively. To test sensitivity of the clustering technique, we also perform *k*-means clustering, which requires *k* starting vectors, for example, by randomly selecting *k* vectors of the input data. Then, each input vector is assigned to the starting vector it is closest to in terms of Euclidean metric and their mean is calculated. This gives *k* new clusters for which the procedure is repeated until a prechosen maximum number of iteration steps is reached. The *k*-means algorithm is computationally less expensive than hierarchical clustering. However, it has the disadvantage of being sensitive to the initialization vectors as well as to the choice of *k*. Therefore, we perform *k*-means clustering by using the centroids identified with hierarchical clustering as our starting

vectors and compare the results. Furthermore, we apply self-organizing maps (SOM) on the Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2), dataset. Thus, we are not only testing another clustering technique, but also a different reanalysis dataset. SOM requires a prechosen representation dimension (which can even be multi-dimensional). For consistency with our analysis, we choose a  $7 \times 1$  matrix.

The cluster representatives (visualized with absolute 10-hPa GPH values) are similar for both variables (*u* or GPH) and cluster algorithms (hierarchical clustering, *k*-means, SOMs); the strongest and weakest vortex clusters are especially robustly identified for all variables and cluster techniques (Figs. ES6–ES8, ES14, ES15). Moreover, we find that the signs of the slopes of the seasonal frequency time series and of the pattern correlation are consistent and significance is of similar range, confirming our finding of a weakened polar vortex in recent winters.

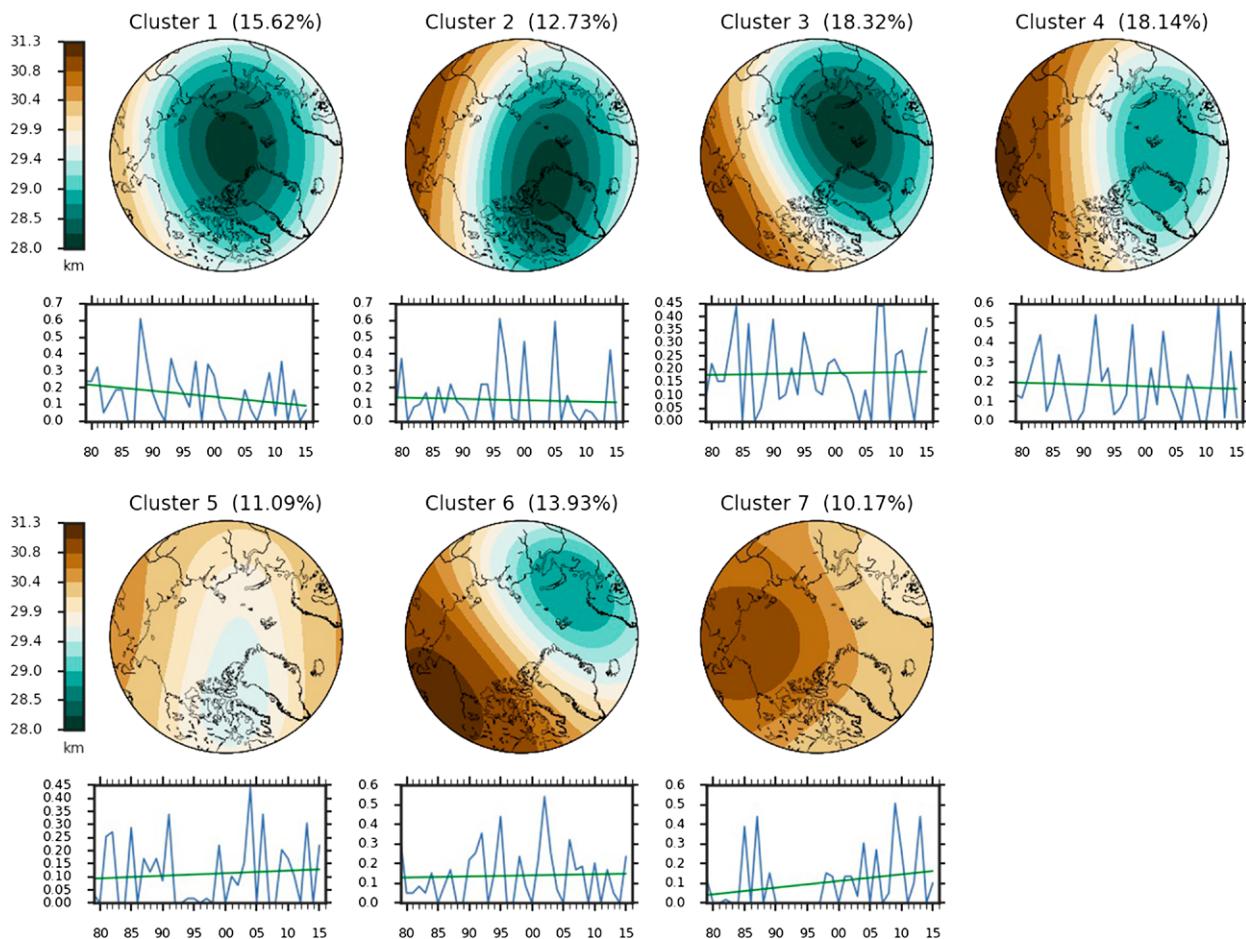
In particular the weak polar vortex states are robustly identified by the different sensitivity tests



**FIG. ES7.** As in Fig. 1 in the main text, but for *k*-means clustering.

(Figs. ES15, ES18). More precisely, 84% of cluster 7 days identified by hierarchical clustering on ERA-Interim data coincide with cluster 7 days based on self-organizing maps performed on MERRA-2 data. This number even rises to 98% if both weak polar vortex representatives cluster 6 and cluster 7 days are considered. Cluster analysis on geopotential heights instead of zonal wind velocity thus results in similar numbers. Furthermore, the results are not very sensitive to the choice of pressure level either, giving robust findings for clustering of zonal wind velocity and GPH at 100 hPa and averaged over the levels 10–50 hPa. More than 80% of the cluster 7 days at 10 hPa are also detected for clustering at lower pressure levels (100 hPa and the 10–50-hPa mean instead of 10 hPa; Fig. ES18). Consequently, the seasonal cluster 7 frequency is robustly identified by the different sensitivity tests with pairwise correlation coefficients ranging from 0.82 to 0.99 (Fig. ES15).

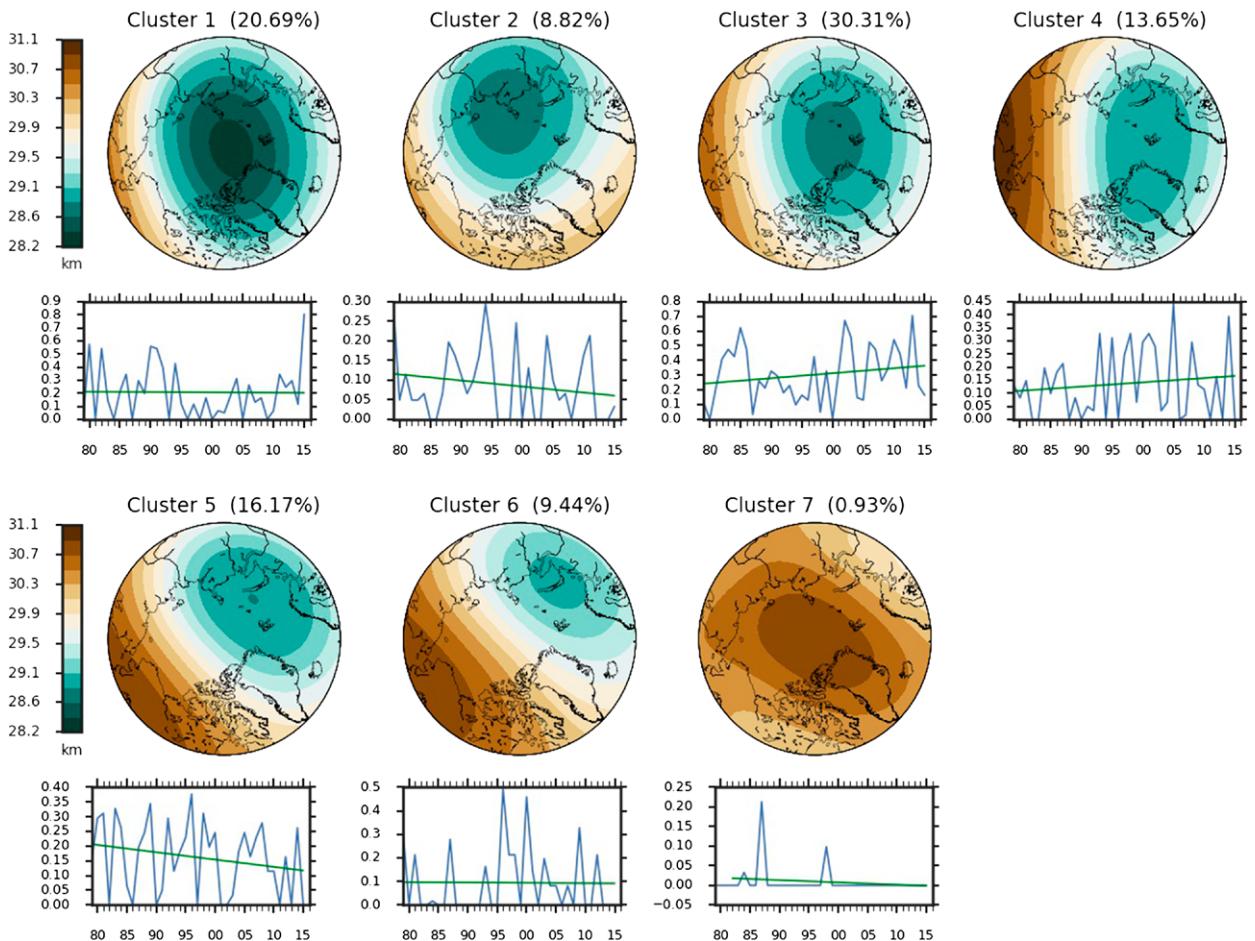
**GRANGER CAUSALITY.** To assess the direction of causality between the stratosphere and surface weather in more detail, we have applied a Granger causality test. We first projected the area-weighted daily geopotential height anomaly field poleward of 60°N onto the cluster 7 representative to construct a continuous time series. A daily domain-mean temperature index over northern Eurasia, however, might be strongly biased given strong synoptic variability in the troposphere likely diluting any potential signal coming for the stratosphere. Therefore, we have applied the Granger causality test on 5-day-mean time series. We detrended both time series, fitted the autoregressive model (using the Akaike information criterion), and finally tested for Granger causality. Using an  $F$  test, the  $H_0$  hypothesis could be rejected ( $p < 0.05$ ), meaning that the cluster 7 time series index Granger causes Eurasian temperature variability in winter. However, we found that the opposite is not true.



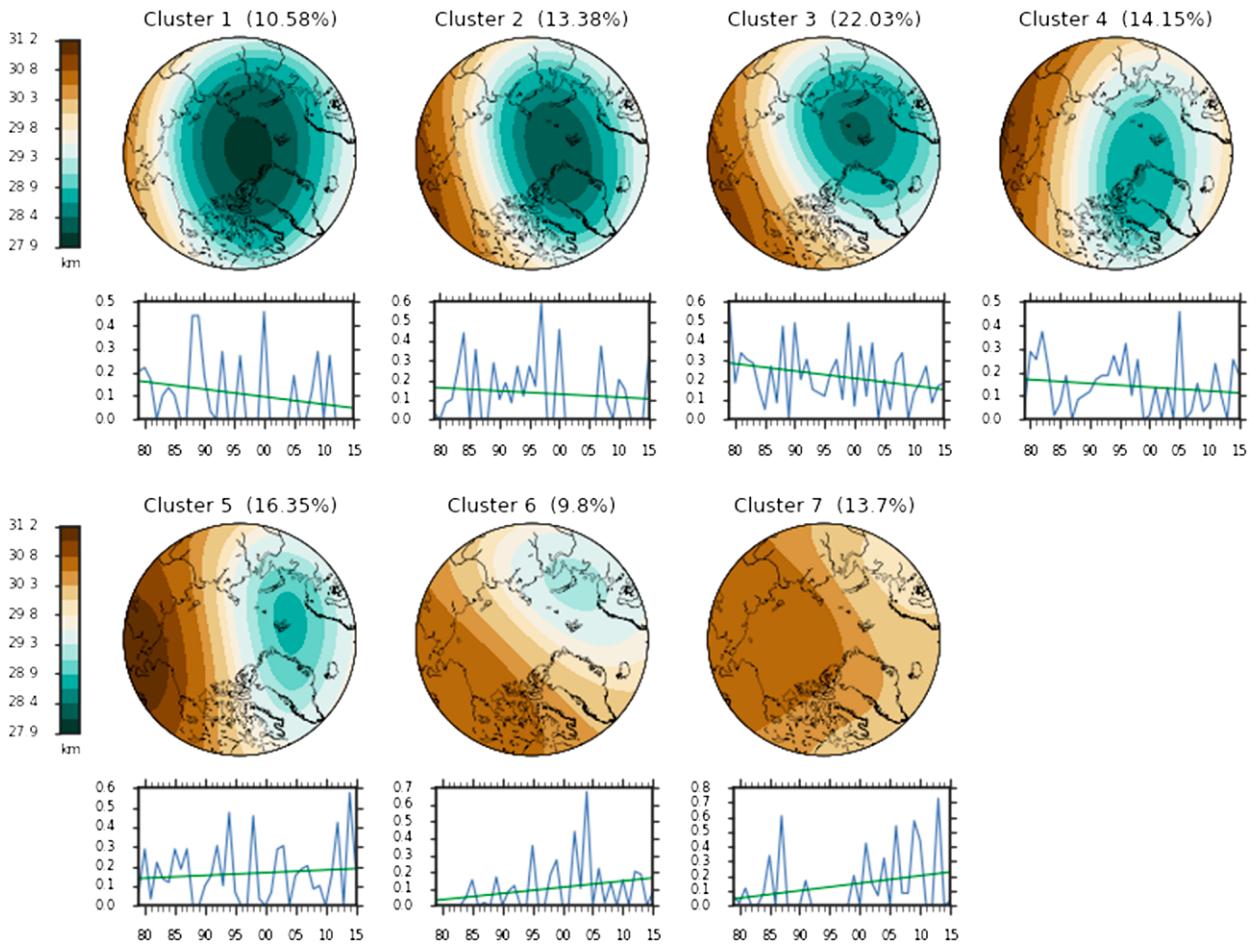
**FIG. ES8.** As in Fig. 1 in the main text, but for  $k$ -means clustering with GPH at 10 hPa.

## REFERENCE

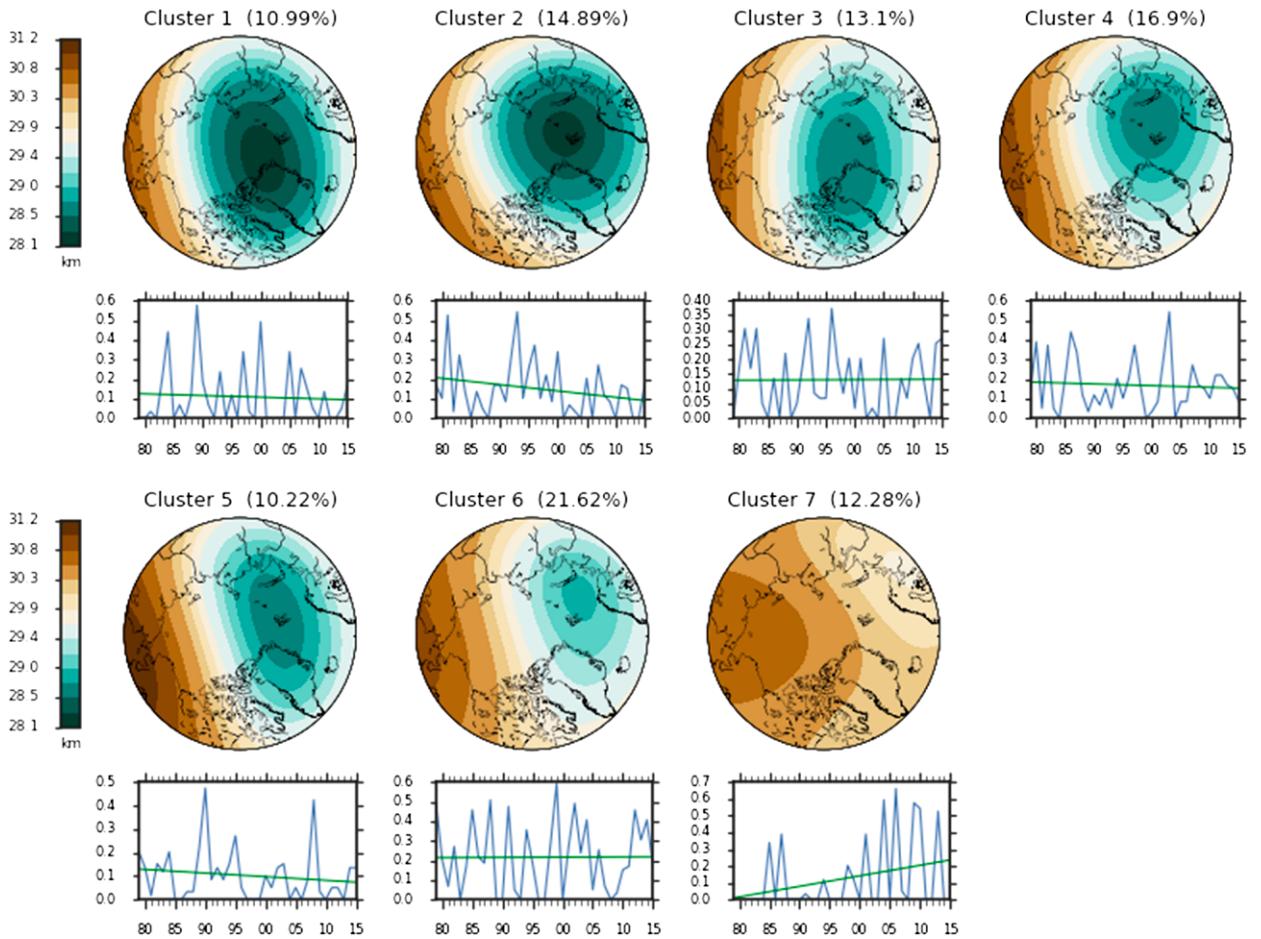
Hitchcock, P., I. R. Simpson, and G. L. Manney, 2013b: Statistical characterization of Arctic polar-night jet oscillation events. *J. Climate*, **26**, 2096–2116, <https://doi.org/10.1175/JCLI-D-12-00202.1>.



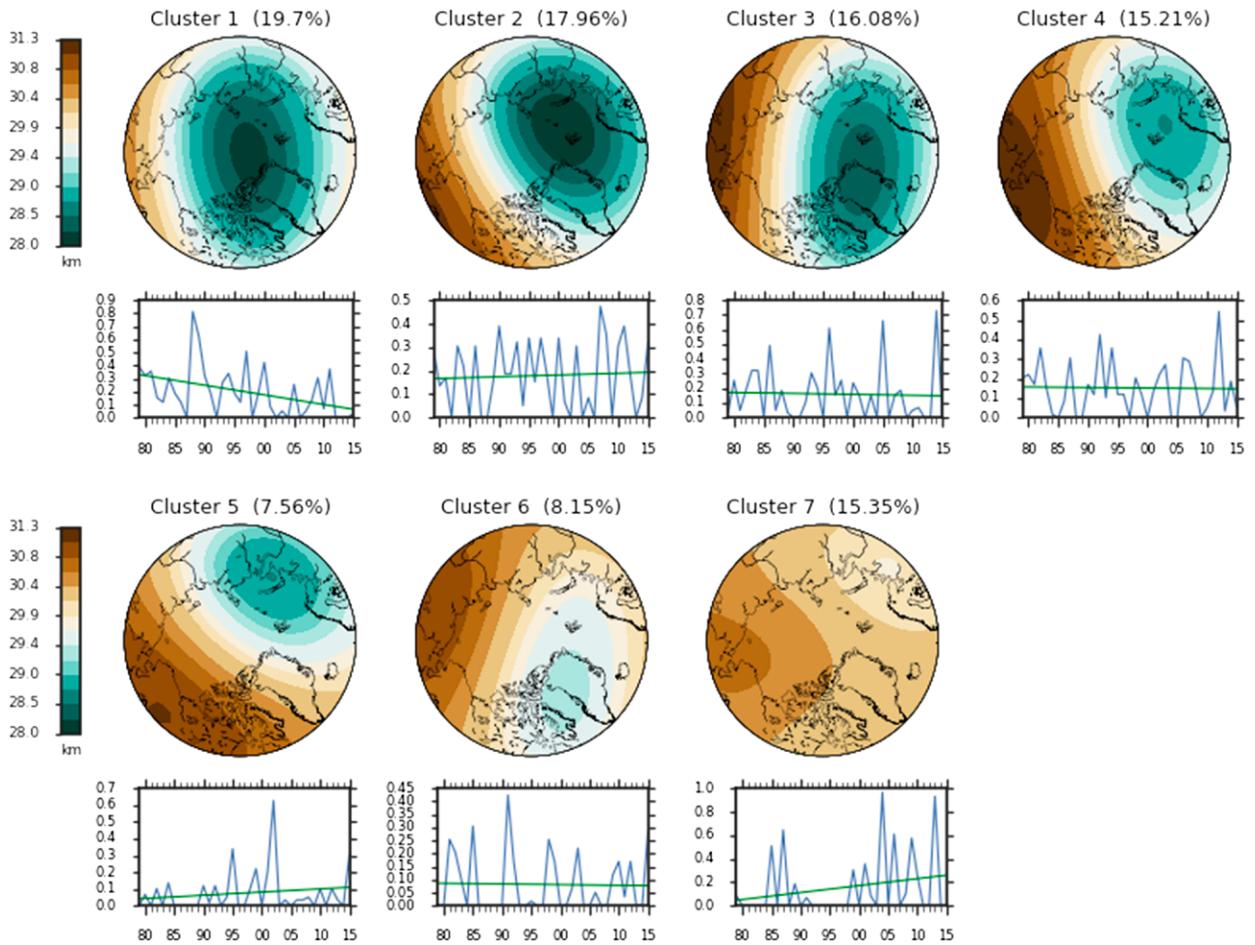
**FIG. ES9.** As in Fig. 1 in the main text, but for clustering with Nov and Dec data.



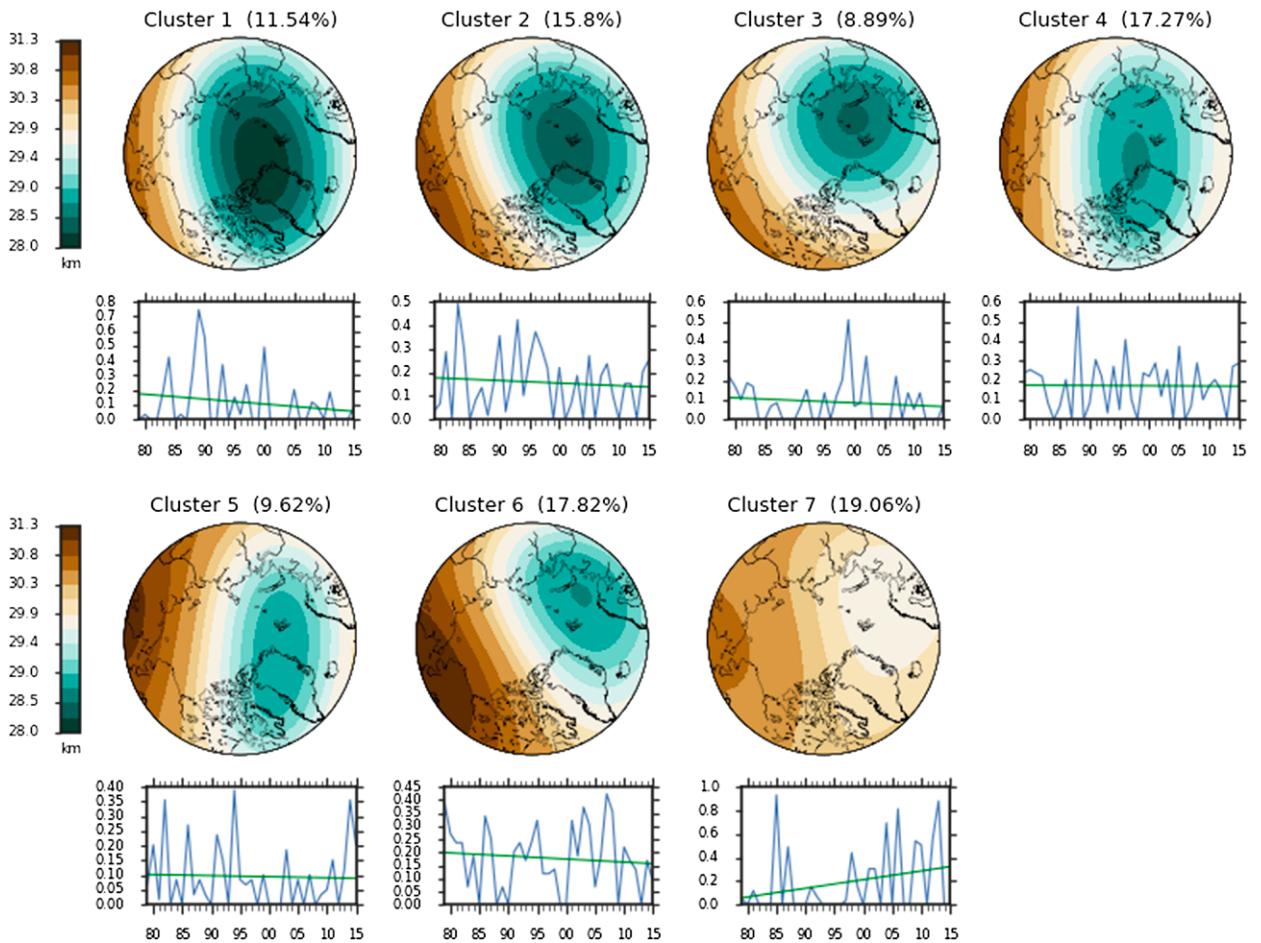
**FIG. ES10.** As in Fig. 1 in the main text, but for clustering with zonal wind velocity anomalies averaged over 10–50 hPa.



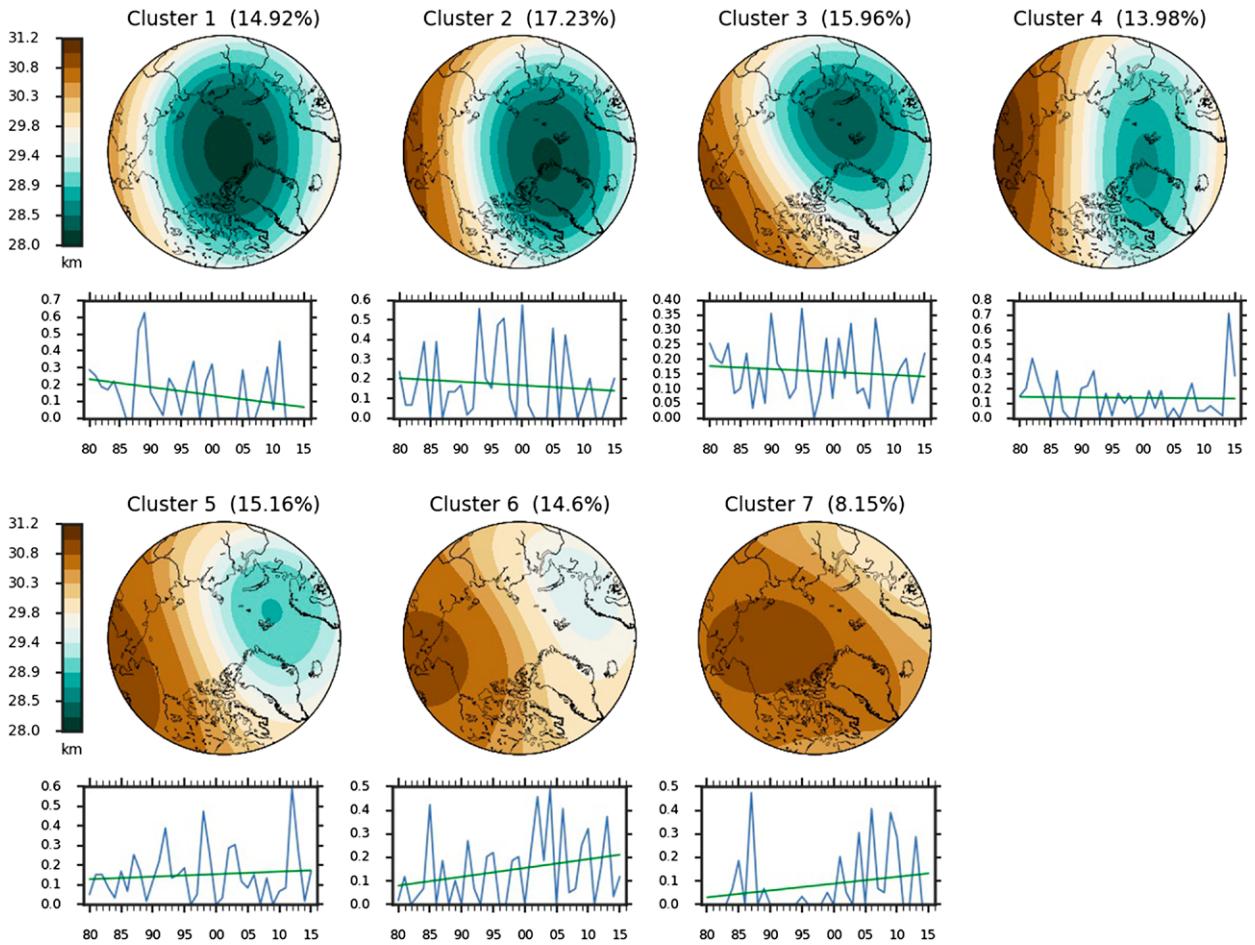
**FIG. ES11.** As in Fig. 1 in the main text, but for clustering with zonal wind velocity at 100 hPa.



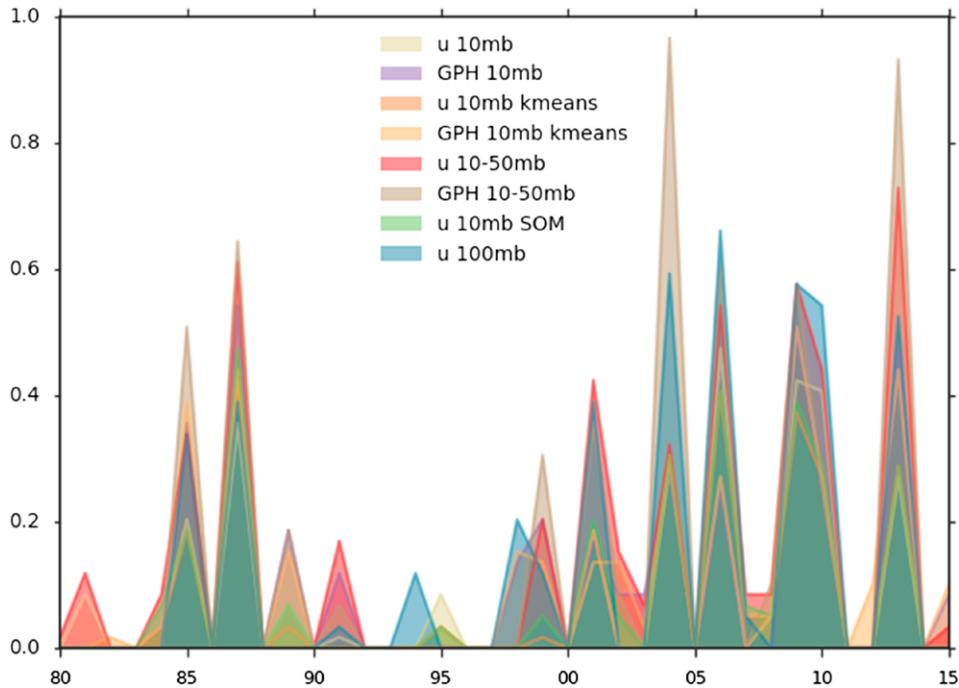
**FIG. ES12.** As in Fig. 1 in the main text, but for clustering with GPH anomalies averaged over 10–50 hPa.



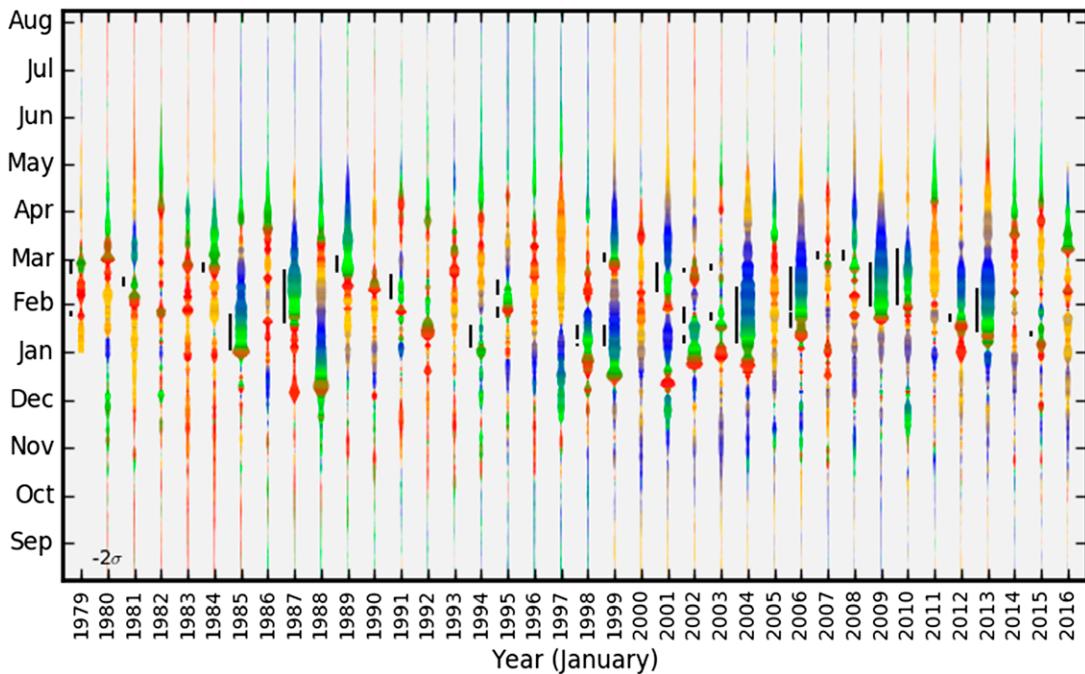
**FIG. ES13.** As in Fig. 1 in the main text, but for clustering with GPH at 100 hPa.



**FIG. ES14.** As in Fig. 1 in the main text, but for SOM clustering with MERRA-2 data from 1980 to 2015.

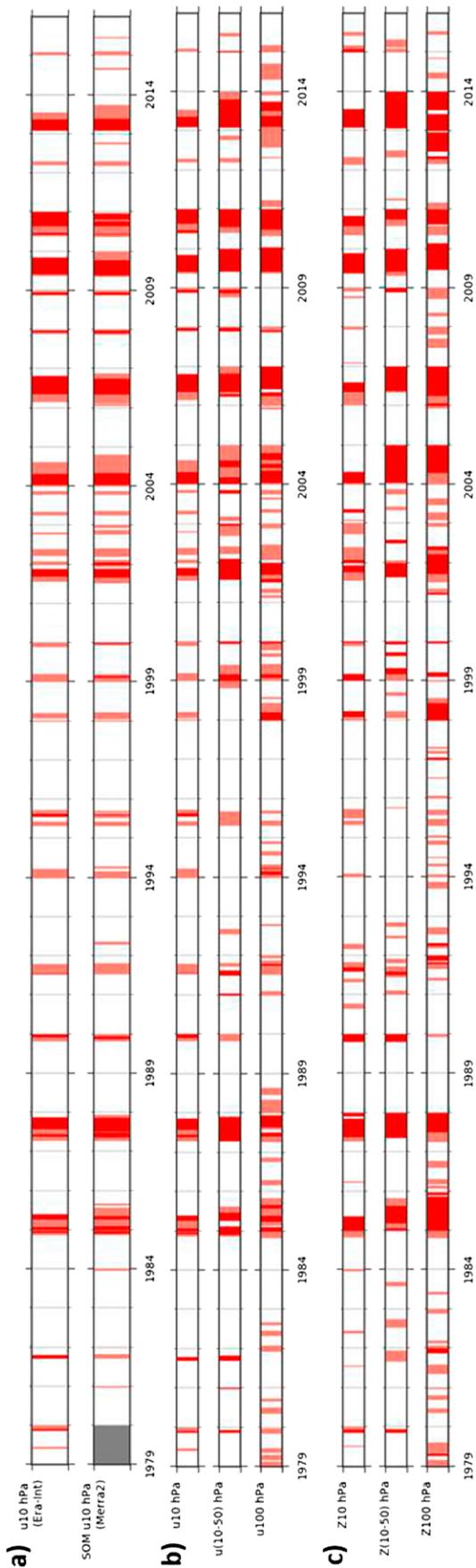


**FIG. ES15.** Time series of seasonal-mean cluster 7 frequencies for the different sensitivity analyses from 1980 to 2015.



**FIG. ES16.** “Abacus” plot of polar night jet oscillation (PJO) with cluster 6 and cluster 7 days indicated to the left of the ribbons (courtesy of Peter Hitchcock). The calculation of PJO events and the color scaling are as in Hitchcock et al. (2013a,b).





**FIG. ES18. Visualization of cluster 6 (light red) and cluster 7 (red) event evolution in each winter, marked by the gray vertical lines for the different sensitivity test. (a) Hierarchical clustering of the 10-hPa wind field with (top) ERA-Interim data and (bottom) self-organizing maps on MERRA-2 data. (b) Hierarchical clustering of (top) the 10-hPa wind field with ERA-Interim data, (middle) data averaged over 10–50 hPa, and (bottom) data at 100 hPa. (c) As in (b), but clustering based on geopotential heights.**