

Supplement

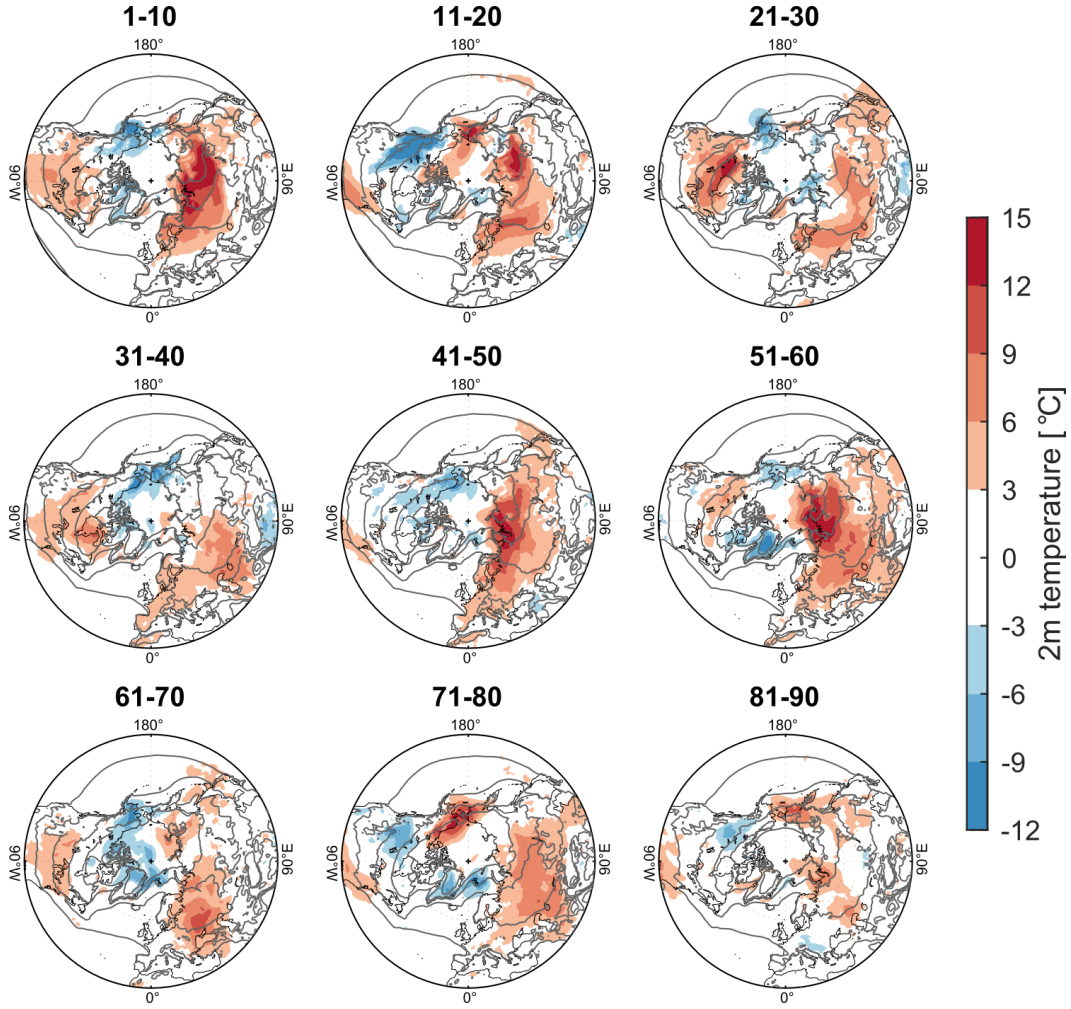


Figure S1. Evolution of 2-meter temperature climatology (contour) and corresponding anomaly (shading) calculated as 10-day-averages over the time period from January 1 to March 31. Panel headings indicate the periods that is averaged over in terms of day of the year. The contour levels for the climatology range from -30°C to $+10^{\circ}\text{C}$ in steps of 10°C .

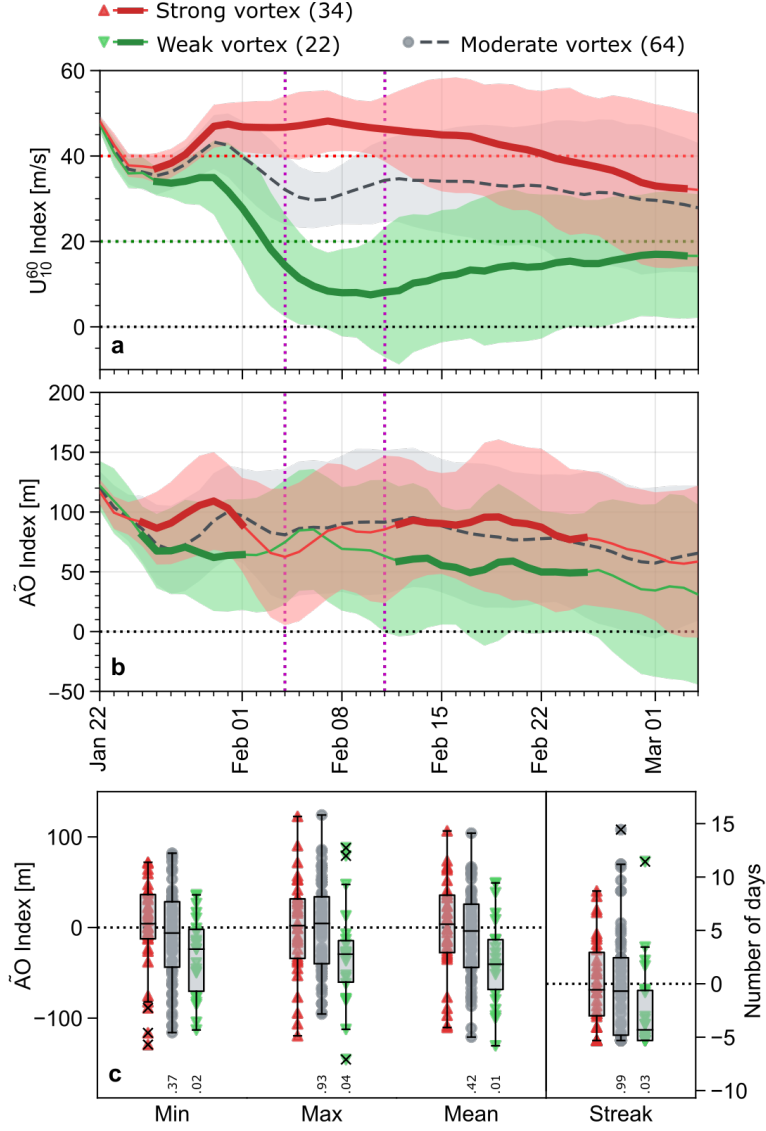


Figure S2. Evolution of U_{10}^{60} (a) and \tilde{AO} (b) indices for a lagged pre-SSD ensemble (initialized on January 16th, 19th and 22nd). Members are separated into 3 clusters based on whether their U_{10}^{60} index averaged between February 4th and 11th is above 40m/s (strong vortex), between 20m/s and 40m/s (moderate vortex) or below 20m/s (weak vortex). The curves show the cluster means, with thick line segments for the clusters with strong and weak vortex indicate statistical significant (to 95% confidence) differences between the corresponding cluster means. Vertical dashed lines indicate the time period used for the clustering and the horizontal dashed lines in subplot a indicate the thresholds of 20m/s and 35m/s and numbers in the legend represent cluster sizes. Panel c shows the normalized member distributions of the three clusters for the minimum, maximum and mean \tilde{AO} between Feb. 15th and 25th, as well as the maximum number of consecutive days with positive \tilde{AO} anomaly between Feb. 15th and Mar. 5th. All member distributions are normalized so zero corresponds to the mean of the cluster with strong vortex and numbers at the bottom show the p-values for significance tests assuming the mean of clusters with medium and weak vortex to be zero.

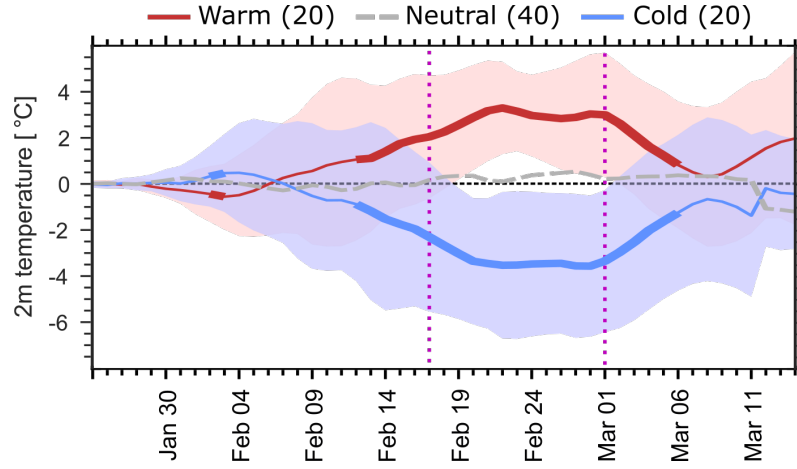


Figure S3. Evolution of 2m temperature for a lagged SSD-onset ensemble (initialized on January 25th and 28th). Members are separated into 3 clusters based on whether their 2-meter temperature (averaged over 30-120E/50-80N (see Fig. 2) and Feb. 17th to March 1st) lies below the 25th percentile (cold), between the 25th and 75th percentile (neutral) or exceeds the 75th percentile (warm). The curves show the cluster means, shadings show one standard deviation and thick line segments for the warm and cold clusters indicate statistically significant (to 95% confidence) differences between the corresponding cluster means.

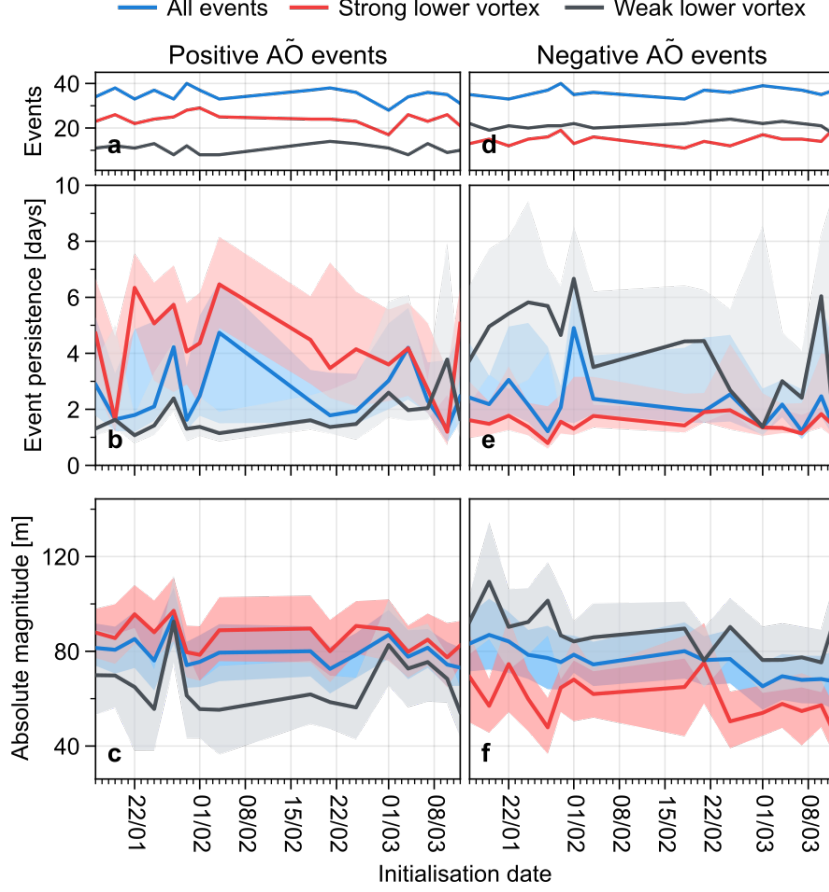


Figure S4. Comparison of different properties of clusters of positive and negative $\tilde{A}O$ extreme events as function of ensemble initialization date. Events are defined as positive events with $\tilde{A}O \geq 100\text{m}$ (left column) or negative events with $\tilde{A}O \leq 60\text{m}$ (right column) any time between 15 and 38 days after initialization as a function of lag time with respect to the event date (time of global maximum/minimum in $\tilde{A}O$ ensemble anomaly in the stated period). Events have further been clustered into two clusters with either a positive (strong stratosphere) or a negative (weak stratosphere) U_{100}^{60} ensemble anomaly (when averaged over the 10 days following the event). The top row shows the number of events in the different clusters, the middle row shows the cluster mean of e -folding decay time of events and the bottom row shows the $\tilde{A}O$ index at the event time. The shading in panels c and f corresponds to one standard deviation, while the shading in panels b and e shows the 10th and 90th confidence percentiles computed via bootstrapping.