UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

SUBSEASONAL-TO-SEASONAL PREDICTABILITY OF COLD AIR OUTBREAKS IN THE CENTRAL UNITED STATES

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

By

OLIVER T. MILLIN

Norman, Oklahoma

2025

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/341526

SUBSEASONAL-TO-SEASONAL PREDICTABILITY OF COLD AIR OUTBREAKS IN THE CENTRAL UNITED STATES

A DISSERTATION APPROVED FOR THE
SCHOOL OF METEOROLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Jason C. Furtado (Chair)

Dr. Elinor R. Martin

Dr. Naoko Sakaeda

Dr. Joseph T. Ripberger

Dr. Carol L. Silva

© Copyright by OLIVER T. MILLIN 2025
All Rights Reserved.

Acknowledgments

Well, it has been an incredible journey over the last several years, and I could not have achieved this without the help of many people around me. First, I would like to thank Dr. Jason Furtado for being an outstanding advisor throughout graduate school. Jason, you have been a brilliant mentor and role model, helping me grow into an independent researcher. Thank you for your guidance and confidence in me. Next, I thank my committee. Dr. Elinor Martin, thank you for your support, valuable input on my research, and for being a fantastic teaching role model during our time working together. Dr. Naoko Sakaeda, thank you for your insightful feedback and steady support throughout my time at OU. Dr. Joe Ripberger and Dr. Carol Silva, thank you for your research input and your support through the National Science Foundation (NSF) Oklahoma Established Program to Stimulate Competitive Research (EPSCoR) project, you have both helped me understand the broader impacts of my work. I am deeply grateful to each of you.

I would also like to thank my co-authors, Dr. Jeffrey Basara and Dr. Chris Malloy, for your valuable input, guidance, and collaboration on our manuscripts, it was a pleasure to work with you both. A huge thank you goes to Leah Hill for your constant support and encouragment throughout my Ph.D., you have always been there for me and I cannot thank you enough. To my first office mate and great friend Ty Dickinson, thank you for your mentorship and friendship over the years. I've learned so much from you. Thanks also to my other office mates: Katie Giannakopoulos, Henry Olayiwola, Jacob Fields, Nathan Erickson, and Katie Giesa. We've worked hard together and shared many laughs along the way. To my friends Alyssa Woodward, David Nowicki, Taylor Grace, Noah Brauer, Simon Lee, Connor Bruce, Chris Downey, Justin Redemann, and Ryan Rico (and many others) – thank you for making my time in Norman so memorable.

To my friends and family in the UK: your unwavering support means the world to me. Thank you especially to my Mum, Dad, and sister, Alice. From my early fascination with meteorology to pursuing a Ph.D. thousands of miles away, you've always supported and believed in me.

Finally, I gratefully acknowledge the NSF Oklahoma EPSCoR grant for the financial support that made this dissertation possible. Now, let's take a look at this Ph.D. research. It seems only fitting to start with a football (or soccer!) reference – so, in the words of Fabrizio Romano: here we go!

Table of Contents

Acknowledgments **iv**

List Of Tables **viii**

List Of Figures **ix**

Abstract **xix**

1 Introduction **1**

2 Characteristics, Evolution, and Formation of Cold Air Outbreaks in the Great Plains of the United States **6**
2.1 Introduction . . . 6
2.2 Data and Methods . . . 8
2.2.1 Data . . . 8
2.2.2 Defining cold air outbreaks . . . 9
2.2.3 Composite analyses and weather regime identification . . . 10
2.2.4 Examining stratospheric wave reflection . . . 11
2.3 Results . . . 12
2.3.1 Near-surface temperature and atmospheric circulation signatures . . . 12
2.3.2 North American winter weather regimes and cold air outbreaks . . . 15
2.3.3 Stratospheric modes of variability . . . 20
2.3.3.1 Stratospheric wave reflection . . . 20
2.3.3.2 Stratospheric polar vortex strength . . . 26
2.3.4 CAOs and weather regime transitions . . . 29
2.4 Discussion . . . 31
2.5 Conclusions . . . 34

3 Analyzing Stratospheric and Tropical Contributions to the Subseasonal Forecasts of the December 2017 and January 2004 North American Cold Air Outbreaks **36**
3.1 Introduction . . . 36
3.2 Data and Methods . . . 38
3.2.1 Data . . . 38
3.2.2 Model description and experiments . . . 39
3.2.3 North American winter weather regimes . . . 41
3.2.4 Event choice . . . 41
3.3 Evolutions of the two central US CAO events . . . 42
3.3.1 Evolution of the extratropical circulation pattern . . . 42

3.3.2 Evolution of the stratospheric and tropical circulation patterns . . . 43
3.4 Model experiment results . . . 46
3.4.1 December 2017/January 2018 CAO analysis . . . 47
3.4.1.1 Forecasts of the reflection index . . . 47
3.4.1.2 Temperature and tropospheric circulation forecast . . . 48
3.4.1.3 North American winter weather regime forecast . . . 52
3.4.2 January 2004 CAO analysis . . . 56
3.4.2.1 Polar cap zonal wind forecasts . . . 56
3.4.2.2 Temperature and tropospheric circulation forecast . . . 57
3.4.2.3 North American winter weather regime forecast . . . 61
3.5 Discussion . . . 64
3.6 Conclusions . . . 66

4 Predictability of Central United States Cold Spells in Subseasonal Forecast Models 69
4.1 Introduction . . . 69
4.2 Data and Methods . . . 71
4.2.1 Verification data and cold spells . . . 71
4.2.2 S2S prediction systems and model cold weeks . . . 71
4.2.3 Skill assessment and metrics . . . 73
4.2.4 Windows of forecast opportunity for cold spells . . . 76
4.2.5 Weather regimes . . . 77
4.2.6 Stratospheric polar vortex forcing . . . 77
4.3 Forecast skill and reliability . . . 78
4.4 Tropospheric circulation and weather regimes in confident forecasts . . . 81
4.5 State of the SPV in confident cold spell predictions . . . 87
4.6 Confident forecasts under weak and strong polar vortex conditions . . . 90
4.6.1 Proportions of confident forecasts by SPV strength . . . 90
4.6.2 Reliability . . . 91
4.6.3 Tropospheric circulation and weather regimes . . . 93
4.6.4 Stratospheric signals for forecasts initialized in weak and strong vortex . . . 99
4.7 Summary and Conclusions . . . 102

5 Subseasonal Predictability of the February 2021 Cold Air Outbreak 105
5.1 Introduction . . . 105
5.2 Data and Methods . . . 107
5.3 Results . . . 108
5.3.1 Evolution of the February 2021 cold air outbreak . . . 108
5.3.2 Connecting Rossby wave breaking events to anomalous ridges . . . 109
5.3.3 Extreme event probability and regimes . . . 110
5.3.4 Simulation of wave breaks in subseasonal forecast models . . . 113
5.3.4.1 East Siberian Sea wave break . . . 113
5.3.4.2 Labrador Sea wave break . . . 114

5.4 Summary and Conclusions . . . 117

6 Application to Industry: Central United States Cold Air Outbreaks, Electricity Load, and Weather Regimes 120
6.1 Introduction . . . 120
6.2 Data and Methods . . . 122
6.2.1 Electricity load data and North American weather regimes . . . 122
6.2.2 Polynomial model . . . 123
6.2.3 Probability density functions and risk ratios . . . 125
6.3 Results . . . 125
6.3.1 North American winter weather regimes . . . 125
6.3.2 Wintertime 2 m temperatures and peak electricity load . . . 126
6.3.3 North American winter weather regimes and electricity load . . . 128
6.4 Summary and Conclusions . . . 134

7 Summary and Conclusions 137

Reference List 145

Appendix 158
1 Appendix A . . . 158
2 Appendix B . . . 162
3 Appendix C . . . 165
4 Appendix D . . . 168

List Of Tables

2.1 Statistics for the mean and standard deviations of the magnitude and lengths of the CAOs by North American weather regime. 19

4.1 The S2S Project Database models used in the study: (1) European Center for Medium Range Weather Forecasts (ECMWF), (2) United Kingdom Meteorological Office (UKMO), (3) Korean Meteorological Agency (KMA), (4) Bureau of Meteorology (BOM), (5) Italian National Research Council – Institute of Atmospheric Sciences and Climate (CNR-ISAC), and (6) Hydro-Meteorological Centre of Russia (HMCR). 73

4.2 The range of possible scores, the value of a perfect score, and the value of a "no skill" score for the Heidke Skill Score (HSS), Probability of Detection (POD), Miss Rate (MR), and False Alarm Ration (FAR). 75

6.1 Corresponding balancing authority regions between two sources of electricity demand data. Correspondence is based on the authors' review of mergers and acquisitions documentation and administrative reports from SPP. 124

List Of Figures

1.1 (a) A schematic depicting the qualitative estimate of forecast skill based on forecast range from short-range weather forecasts to long-range seasonal predictions, including potential sources of predictability. (b) A schematic diagram highlighting the relationship between the S2S 'extended-range' forecast range and other prediction timescales, with examples of actionable information that can enable decision-making across sectors. Adapted from White et al. (2017). 5

2.1 Lagged composites of T2M anomalies (°C, shaded) across North America for the 37 identified Great Plains CAOs since 1950, with respect to CAO onset (Day 0) for (a) Days -7 to -4, (b) Days -3 to -0, (c) Days +1 to +4 and (d) Days +5 to +8. Black contours represent significance at the 95th percentile level via a bootstrapping approach of 5000 iterations. The black box represents the region used to define the CAOs. Negative lags indicate days prior to CAO onset day (Day 0), and positive lags indicate days after CAO onset. 13

2.2 Lagged composites of the full-field GPH (dam, black contours) and only the significant ($p < 0.05$) linearly-detrended 500 hPa GPH anomalies (dam, shaded) across the Northern Hemisphere for the 37 Great Plains CAOs, with respect to CAO onset (Day 0) at 4-day intervals from (a) Days -11 to -8 through (f) Days +9 to +12. Statistical significance is calculated via a bootstrapping approach of 5000 iterations. Negative lags indicate days prior to CAO onset day (Day 0), and positive lags indicate days after CAO onset. 14

2.3 As in Fig. 2.2 but for 50 hPa full-field GPH (dam; contours) and only significant ($p < 0.05$) linearly-detrended GPH anomalies (dam; shading) 15

2.4 Composite 500 hPa linearly-detrended GPH anomalies (dam, shaded), for the five North American weather regimes for NDJFM between 1950 and 2021: (a) Alaskan Ridge (AkR), (b) Arctic High (ArH), (c) Pacific Trough (PT), (d) West Coast Ridge (WCR), and (e) Arctic Low (ArL). Percentages show the frequency of occurrence of a given regime throughout 1950-2021. 17

2.5 Onset day composites of linearly-detrended 500 hPa GPH anomalies (dam, shaded), with respect to the 1981-2010 climatology, are shown for (a) AkR-, (b) ArH- and (c) WCR-CAOs. Black contours represent statistical significance ($p < 0.05$) via 5000-iteration bootstrapping without replacement. . . 18

2.6 Magnitude (°C) of each of the 37 defined CAOs using the non-detrended T2M anomalies, categorized by weather regime: AkR (orange), ArH (red) and WCR (blue). Bold numbers below the bars represent the length (days) of each of the CAOs. 19

2.7 Composites of 100 hPa *v'T'* (K m s^{-1}, shaded) for Days -3 to 0 in (a), (c), and (e), and RI in (b), (d), and (f) for AkR-, ArH-, and WCR-CAOs, respectively. The RI is defined as in Matthias and Kretschmer (2020) but between 55-75°N, and 145-185°E and 220-260°E for Siberia and Canada, respectively (thick black lines on (a), (c), and (e)). Black contours on (a), (c), and (e) and red dots on (b), (d), and (f) show statistical significance ($p < 0.05$) via 5000-iteration bootstrapping. The two southerly dashed lines are latitudes 60°N and 70°N, as used in Fig. 2.8. Negative and positive lags indicate days before and after CAO onset day (Day 0), respectively. 22

2.8 Lag composites of the linearly-detrended standardized anomalies of GPH (shaded) and full-field WAF (vectors) at 4-day intervals from Days -11 to -8 through Days -3 to 0 for: (a) to (c) AkR-CAOs and (d) to (f) ArH-CAOs, averaged between 60-70°N. Yellow dots represent statistical significance ($p < 0.05$) of the the vertical component of the WAF via a 5000-iteration bootstrapping without replacement. Black contours represent statistical significance ($p < 0.05$) of the GPH anomalies using the same method. Negative lags indicate days prior to CAO onset day (Day 0), and positive lags indicate days after CAO onset. Scaling of the WAF vectors was achieved by dividing the vertical WAF component by a scale factor of 0.1 above 200 hPa and by a scale factor of 0.3 below and including 200 hPa, and dividing the zonal WAF component by 10 above 200 hPa and by 50 below and including 200 hPa. 24

2.9 Lag composites of the 250 hPa streamfunction anomalies (10^6 m^2 s^{-1}, shaded) and statistically significant ($p < 0.05$) standardized WAF anomalies (vectors) across the Pacific basin for 4-day intervals of Days -11 to -8 through Days -3 to 0 for (a), (c) and (e) AkR-CAOs, and (b), (d) and (f) ArH-CAOs. Black contours represent statistical significance ($p < 0.05$) of the streamfunction anomalies via a bootstrapping approach of 5000 iterations without replacement. Statistical significance ($p < 0.05$) of WAF is computed using the magnitude of the WAF anomalies and 5000-iteration one-tailed bootstrapping without replacement. Negative lags indicate days prior to CAO onset day (Day 0), and positive lags indicate days after CAO onset. 25

2.10 Lagged composites of the (a) and (b) vertical profile averaged between 70-80°N, and meridional profile at 30 hPa, (c) and (d), of zonal wind (m s^{-1}) for AkR-CAOs averaged between 200-250°E (red solid line) and in the zonal mean (blue solid line), at 4-day intervals for Days -7 to -4 and Days -3 to 0, respectively, where Day 0 is CAO onset. Red and blue dashed lines represent DJF 1981–2010 climatology of the zonal wind between 200-250°E and in the zonal mean, respectively. Grey dots indicate strong curvature (see Section 2d). Negative lags indicate days prior to CAO onset day (Day 0), and positive lags indicate days after CAO onset. 27

2.11 Composite pressure-time polar cap (60-90°N) averaged cross-sections of standardized and non-detrended zonal wind anomalies for (a) AkR- and (c) ArH-CAOs, and of linearly-detrended GPH standardized anomalies for (b) AkR- and (d) ArH-CAOs between Days -40 and +40. Dashed lines represent statistical significance at the 95th percentile via bootstrapping of 5000 iterations. Negative lags indicate days prior to CAO onset day (Day 0), and positive lags indicate days after CAO onset. 28
2.12 North American weather regime transitions for a) AkR-CAOs and b) ArH-CAOs in the US Great Plains, between Days -20 and +10. 31

3.1 Maps of MERRA-2 T2M anomalies (°C, shading) and 500 hPa GPH anomalies (dam, contours) averaged for 3-day periods from a) 23 December 2017 – 25 December 2017 through e) 4 January 2018 – 6 January 2018 for the December 2017 central US CAO event, and f) 24 January 2004 – 26 January 2004 through j) 5 February 2004 – 7 February 2004 for the January 2004 central US CAO event. Anomalies are calculated with respect to the 1999-2020 climatology, where the seasonal cycle is removed from each day. 44
3.2 The evolution of several key S2S indices for the December 2017 central US CAO event: a) the polar cap (60–90°N) averaged zonal wind standardized anomalies throughout the atmosphere (shading) from MERRA-2, b) the RI (black line) and its Siberian (red dashed line) and Canadian components (blue dashed line) from MERRA-2, and c) the MJO amplitude (black solid line) and phase (red solid line) from the BoM. Panels d)–f) show the same as a)–c), respectively, but for the January 2004 central US CAO. Black dashed lines indicate the start and end of the CAO periods. Anomalies are calculated with respect to the 1999-2020 climatology, where the seasonal cycle is removed from each day. 46
3.3 Boxplots of a)–c) the RI, d)–f) the Siberian component of the RI, and g)–i) the Canadian component of the RI all averaged between 14–25 December 2017, for subseasonal initializations of 27 November 2017, 4 December 2017, and 11 December 2017, respectively. Gray boxes denote the FREE experiments, red boxes denote the NUDGE-T experiments, and blue boxes denote the NUDGE-S experiments. The orange lines represent the ensemble median, the lower and upper bounds of the box indicate the 25th and 75th percentiles, the whiskers extend from the box to the farthest data point lying within 1.5 times the interquartile range from the box, and the circles are those points that lie outside of the whiskers. 49

3.4 The difference in T2M (°C, shading) and 500 hPa GPH (dam, contours) from the ensemble mean FREE run for a)–d) MERRA-2, e)–h) the NUDGE-S ensemble mean, and j)–m) the NUDGE-T ensemble mean, for the 27 November 2017 initialization at 3-day averaged periods from 26–28 December 2017 through 4–6 January 2018, respectively. Hatching shows statistical significance on the T2M difference field at the 95th percentile via 5000-iteration bootstrapping with replacement. The black boxes represent the area that usually has the coldest T2M response to reflection events in Messori et al. (2022) (40-55°N, 260-290°E), the purple box represents the Great Plains (30-48°N, 257-268°E), and the green box represents the northeast US (35-45°N, 270-290°E). 52
3.5 Violin plots of the ensemble forecast of area-averaged T2M anomalies (°C) for weekly CESM2 CAM6 initializations from 27 November 2017 to 18 December 2017 for a) the area most prone to cold events due to stratospheric wave reflection (40-55°N, 260-290°E) from Messori et al. (2022) averaged between 26 December 2017 to 2 January 2018, b) the Great Plains (30-48°N, 257-268°E) averaged between 26 December 2017 to 2 January 2018, and c) the northeast US (35-45°N, 270-290°E) averaged between 30 December 2017 to 6 January 2018. The blue, red, and gray distributions represent the NUDGE-S, NUDGE-T, and FREE experiments, respectively. The dashed black lines represents the MERRA-2 values. 53
3.6 The evolution of the ensemble mean WRI for the five North American winter weather regimes for the 27 November 2017 CESM2 CAM6 initialization in a) the MERRA-2 evolution, b) the FREE experiment, c) the NUDGE-S experiment, and d) the NUDGE-T experiment. The green line represents the Pacific Trough (PT), the blue line represents the West Coast Ridge (WCR), the orange line represents the Alaskan Ridge (AkR), the red line represents the Arctic High (ArH), and the purple line represents the Arctic Low (ArL). The black dashed lines show the start and end dates of the 2017/2018 CAO. The purple dashed lines represent the start and end of the period used for correlation analysis in Fig. 3.7. 55
3.7 The correlation between the MERRA-2 WRIs and the CESM2 CAM6 forecast WRI initialized on 27 November 2017 for the lead time extending from 14 days prior to the CAO to 6 January 2018 (i.e., end of averaging period for NE cold), as a function of ensemble member. 56
3.8 Boxplots of the polar cap averaged (60-90°N) zonal wind at a)–c) 10 hPa, d)–f) 50 hPa, and g)–i) 100 hPa all averaged between 12–26 January 2004, for subseasonal initializations of 29 December 2003, 5 January 2004, and 12 January 2004, respectively. Gray boxes denote the FREE experiments, red boxes denote the NUDGE-T experiments, and blue boxes denote the NUDGE-S experiments. The orange lines represent the ensemble median, the lower and upper bounds of the box indicate the 25th and 75th percentiles, the whiskers extend from the box to the farthest data point lying within 1.5 times the interquartile range from the box, and the circles are those points that lie outside of the whiskers. 58

3.9 The difference in T2M (°C, shading) and 500 hPa GPH (dam, contours) from the ensemble mean FREE run for a)–d) MERRA-2, e)–h) the NUDGE-S ensemble mean, and j)–m) the NUDGE-T ensemble mean, for the 29 December 2003 initialization at 3-day averaged periods from 27 January–29 January 2004 through 5 February–7 February 2004, respectively. Hatching shows statistical significance at the 95th percentile on the difference in the T2M via 5000-iteration bootstrapping with replacement. 60
3.10 Violin plots of the ensemble forecast of area-averaged T2M anomalies (°C) between 27 January 2004 to 31 January 2004 for weekly CESM2 CAM6 initializations from 29 December 2003 to 19 January 2004 for a) the northwest of the US (50-60°N, 230-260°E), b) the Great Plains (30-48°N, 257-268°E), and c) the northeast US (35-45°N, 270-290°E). The blue, red, and gray distributions represent the NUDGE-S, NUDGE-T, and FREE experiments, respectively. The dashed black line represents the MERRA-2 values. 61
3.11 The evolution of the ensemble mean weather regime index for the five North American winter weather regimes for the 29 December 2003 CESM2 CAM6 initialization in a) the MERRA-2 evolution, b) the FREE experiment, c) the NUDGE-S experiment, and d) the NUDGE-T experiment. The green line represents the Pacific Trough (PT), the blue line represents the West Coast Ridge (WCR), the orange line represents the Alaskan Ridge (AkR), the red line represents the Arctic High (ArH), and the purple line represents the Arctic Low (ArL). The black dashed lines show the start and end dates of the 2004 CAO during the initialization. The purple dashed line represents the start of the period used for correlation analysis in Fig. 3.12. . 63
3.12 The correlation between the MERRA-2 WRIs and the CESM2 CAM6 forecast WRI initialized on 29 December 2003 for the lead time extending from 14 days prior to the CAO to 31 January 2004 (i.e., end of the CAO), as a function of ensemble member. 64

4.1 The skill of ECMWF (blue bars), UKMO (green bars), KMA (purple bars), CNR-ISAC (orange bars), BOM (red bars), and HMCR (gray bars) models in forecasting cold weeks as a function of forecast week in terms of (a) HSS, (b) POD, (c) FAR, and (d) MR. 79
4.2 Reliability diagrams and sharpness histograms for the S2S models in forecasting (a,c) Week 3 and (b,d) Week 4 cold spells. For the reliability diagrams, the forecast probability is shown in the *x*-axis and the observed relative frequency is shown on the *y*-axis, the horizontal and vertical dashed gray lines represent the climatological frequency, the diagonal black line represents the perfect reliability 1:1 line, the diagonal dashed gray line represents the "no-skill" line, and the dashed black box is the region of confident forecasts (i.e., 50%+ forecast probability). The sharpness histograms have forecast probability on the *x*-axis and forecast ratio on the *y*-axis, where the vertical dashed black line represents the climatological frequency (i.e., the climatological expected rate of occurrence). 81

4.3 Composite maps of T2M anomalies (shading, °C) and 500 hPa GPH anomalies (contours, dam) for confident hit, false alarm and miss forecasts for (a,d,g) ERA5 initialization day, (b,e,h) ERA5 Week 3, and (c,f,i) multi-model mean Week 3, respectively. The black box shows the region used to define central US cold spells and solid contours show positive GPH anomalies while dashed contours show negative GPH anomalies. 84
4.4 Lagged North American winter weather regime frequencies with respect to forecast initialization time (lag 0) for a) the confident hit forecast initializations and c) the confident false alarm forecast initializations in ERA5. Panels b) and d) show the same as panels a) and c) but instead for the model forecasts as a function of lead time. Regime frequencies are calculated relative to the total number of samples in each group and are then smoothed with a 5-day centered filter. Bold and dashed lines represent statistically significant enhanced or suppressed frequencies relative to climatology at the 95th percentile via 5000-iteration bootstrapping from ERA5 data. Week 3 is shown in gray shading and the colored dots indicate the climatological frequency of the five winter weather regimes. 86
4.5 Composites of polar cap (60-90°N) averaged GPH anomalies (shading) as a function of lag from initialization and pressure level for Week 3 cold spell confident hit forecasts, confident false alarm forecasts, and the difference between the confident hit and false alarm forecasts in a)–c) ERA5 and d)–f) the S2S models, respectively. Circles represent statistical significance at the 95th percentile via 5000-iteration bootstrapping, the blue lines represent the Week 3 cold, and the black solid line represents initialization date. . . . 89
4.6 The total number of confident hit (blue solid bars) and confident false alarm (red solid bars) forecasts as a function of S2S model for Week 3 cold spells. The left-to-right hatching indicates the proportion of confident hit and false alarm forecasts initialized in a weak SPV state while the right-to-left hatching indicates the proportion of confident hit and false alarm forecasts initialized in a strong SPV state. The bold numbers show the respective counts of these different proportions. 91
4.7 As in Fig. 4.2 but the forecast reliability diagrams and sharpness histograms of Week 3 and Week 4 cold spells initialized in (a,b,e,f) weak SPV conditions and (c,d,g,h) strong SPV conditions. Note that here the cumulative number of forecasts used for this analysis is 695 for weak SPV forecasts and 665 for strong SPV forecasts, compared to the total number of 3454 used in Fig. 4.2. 93
4.8 As in Fig. 4.3 but for forecasts initialized in weak polar vortex states. 95
4.9 As in Fig. 4.3 but for forecasts initialized in strong polar vortex states. . . . 96
4.10 As in Fig. 4.4 but for forecasts initialized in weak polar vortex states. 98
4.11 As in Fig. 4.4 but for forecasts initialized in strong polar vortex states. . . . 99
4.12 As in Fig. 4.5 but for forecasts initialized in weak polar vortex states. 100
4.13 As in Fig. 4.5 but for forecasts initialized in strong polar vortex states. . . . 102

5.1 ERA5 500 hPa GPH anomalies (dam, contours) and T2M anomalies (°C, shaded) averaged between a) 1-4, b) 5-11, and c) 12-18 February 2021. Solid and dashed contours represent positive and negative GPH anomalies, respectively. The purple boxes represent the region defining the Great Plains (30-48°N and 256.5-268.5°E). 109
5.2 ERA5 320 K PV (PVU, shaded) for two wave breaks during the evolution of the February 2021 CAO: a) – c) an East Siberian Sea anticyclonic wave break between 2-4 February 2021 and d) – f) a Labrador Sea cyclonic wave break between 9-11 February 2021. The black contour represents the 2 PVU boundary. Black boxes in a) – c) and d) – f) represent the two wave break regions, respectively. 111
5.3 a) Probability of an extreme Great Plains T2M anomaly event between 12-18 February 2021 at the 10th percentile as a function of initialization date for ECMWF (blue line) and NCEP (red line). b) A bar plot of the ensemble distribution of 5 North American weather regimes for the 8-10 February 2021 as a function of initialization date within the ECMWF modelthe Pacific Trough (PT), West Coast Ridge (WCR), Alaskan Ridge (AkR), Arctic High (ArH), and the Arctic Low (ArL). 112
5.4 28 January 2021 initialization ECMWF 320 K PV composites (PVU, shaded) for the 5 ensemble members with the a) lowest and b) highest PV in the East Siberian Sea region (black box) averaged over 2-4 February 2021. c) The difference (PVU, shaded) between the low and high PV composites. d) – f) Composites of T2M (°C, shaded) and 500 hPa GPH (dam, line contours) anomalies for the 5 ensemble members with the lowest PV in the East Siberian Sea region. g) – i) As in d) – f) but for the 5 ensemble members with the highest PV in the East Siberian Sea region. Hatching in panels c) – f) denote statistical significance $(p < 0.1)$ in the composite difference (low - high PV) in the PV and T2M fields, computed by a 5000-iteration bootstrapping technique, without replacement. Bold dates at the top of panels d)–i) indicate the valid date for the forecasts. 115
5.5 As in Fig. 5.4 but for the Labrador Sea wave break between 9-11 February 2021. 116
5.6 28 January 2021 initialization ECMWF T2M anomaly (°C, shading) and 500 hPa geopotential height anomaly (dam, contours) composites between 14-16 February 2021 for ensemble members that a) – c) simulate both the East Siberian Sea and Labrador Sea wave breaks best (PV $<$ 25th percentile) and for ensemble members that d) – f) simulate both the East Siberian Sea and Labrador Sea wave breaks least effectively (PV $>$ 75th percentile). Solid contours represent positive anomalies and dashed contours represent negative anomalies. Hatching denotes statistical significance $(p < 0.1)$ in the composite difference (low - high PV) in the T2M field, computed by a 5000-iteration bootstrapping technique, without replacement. Bold dates at the top of panels a) – f) indicate the valid date for the forecasts. 117

6.1 As in Fig. A.13 but for ERA5 linearly detrended T2M anomalies (shading, °C). The black lines show the region of the entire Southwest Power Pool (SPP) while the purple box (31-43°N, 256-268°E) represents the region of study for electricity load data in the SPP and the area in which T2M is averaged for Fig. 6.2. A sub-domain of the entire SPP region is used for analysis as it represents the 13 divisions within the SPP that could be consistently matched between load data sources (see methods). Anomalies are with respect to the 1979-2020 climatology also used to calculate the regimes. 127
6.2 Scatter plots of peak load anomaly (MW/1000 customers) versus ERA5 (a) daily average T2M (°C) and (b) daily minimum T2M (°C), where both temperature variables have been averaged over the boxed region in Fig. 6.1. The red lines represent the second degree polynomial model for the relationship between the two parameters in each plot, with the equation of this polynomial shown on the plot, where T represents the temperature, as well as the the coefficient of determination (R^2) value for variance explained in peak load anomaly respective to the polynomial model. The blue dashed lines represent an average daily peak load between 1999-2022. 129
6.3 (a) The composite mean peak load anomaly (bars, MW/1000 customers) as a function of North American winter weather regime, where white diamonds indicate that the composite value is statistically significant at the 95th percentile via 5000-iteration bootstrapping with replacement. (b) The probability density function for peak load anomaly (lines, MW/1000 customers) as a function of North American winter weather regime via Gaussian kernel density estimation. σ, γ, and κ represent the probability density function standard deviation, skewness, and kurtosis, respectively. (c) The risk ratios of extreme peak load by North American weather regime (bars) for 90th percentile, 95th percentile, and 97.5th percentile extreme thresholds versus that of the entire distribution. Whiskers represent the 95th percentile confidence interval for each risk ratio from 5000-iteration bootstrapping with replacement. For all plots, AkR is orange, ArH is red, WCR is blue, PT is green, and ArL is purple. 132
6.4 Composite maps of ERA5 T2M anomalies (shading, °C) and 500 hPa GPH anomalies (dam, contours), respective to the 1991-2020 climatology, for days exceeding the 90th, 95th, and 97.5th percentile thresholds of all peak load anomalies for a)–c) all days, d)–f) AkR days, g)–i) ArH days, j)–l) WCR days, m)–o) PT days, and p)–r) ArL days. The sample size is given for each composite. 133

A.1 Lagged composites of the 850 hPa horizontal thermal advection (shaded, °C day^{-1}) for all 37 CAOs for a) Day -1, b) Day 0, and c) Day 1. Positive lags represent days after CAO onset day (Day 0) and negative lags represent days prior to CAO onset. The thick, black box represents the region over which the CAOs were defined. 158

A.2 Lagged composites of the 100 hPa *v'T'* (K m s^{-1}, shaded) across the Northern Hemisphere for ArH-CAOs, with respect to CAO onset (Day 0) at 4-day intervals from a) Days -23 to -20 through f) Days -3 to 0. Black contours represent significance at the 95th percentile via a bootstrapping approach of 5000 iterations. 159
A.3 Lagged composites of the 100 hPa *v'T'* (K m s^{-1}, shaded) across the Northern Hemisphere for AkR-CAOs, with respect to CAO onset (Day 0) at 4-day intervals from a) Days -23 to -20 through f) Days -3 to 0. Black contours represent significance at the 95th percentile via a bootstrapping approach of 5000 iterations. 160
A.4 Composites of the 100 hPa *v'T'* anomalies (K m s^{-1}, shaded) across the Northern Hemisphere for (a) AkR-CAOs, (b) ArH-CAOs, and (c) WCR-CAOs for Days -3 to 0. Black contours represent significance at the 95th percentile via a bootstrapping approach of 5000 iterations. 161
A.5 Lagged composite of the cumulative total 100 hPa *v'T'* (K m s^{-1}) averaged over the Siberian region (55-75°N and 145-185°E) for AkR-CAOs (orange) and ArH-CAOs (dark red) with respect to CAO onset (Day 0). Statistical significance at the 95th percentile is shown via red dots via a 5000-iteration bootstrapping approach. 161
A.6 The evolution of the ensemble mean zonal wind standardized anomaly (shading) averaged over the polar cap (60–90°N for the three subseasonal forecast initializations of 27 November 2017, 4 December 2017, and 11 December 2017 for a)–c) the FREE experiments, d)–f) the NUDGE-S experiments, and g)–i) the NUDGE-T experiments, respectively. The dashed lines show the start and end dates of the central US CAO. 162
A.7 The evolution of the ensemble mean 200 hPa streamfunction anomalies (shading, m^2 s^{-1}) in the 27 November 2017 initialization for 3-day averaging periods of 17–19 December 2017, 20–22 December 2017, 23-25 December 2017, and 26-28 December 2017 for a)–d) the FREE experiment, f)–i) the NUDGE-S experiments, and j)–m) the NUDGE-T experiments, respectively. 163
A.8 The evolution of the ensemble mean 500 hPa GPH anomalies (shading, dam) in the 27 November 2017 initialization for 3-day averaging periods of 17–19 December 2017, 20–22 December 2017, 23-25 December 2017, and 26-28 December 2017 for a)–d) the FREE experiment, f)–i) the NUDGE-S experiments, and j)–m) the NUDGE-T experiments, respectively. 163
A.9 The evolution of the zonal wind standardized anomaly (shading) averaged over the polar cap (60–90°N for the three subseasonal forecast initializations of 29 December 2003, 5 January 2004, and 12 January 2004 for a)–c) the FREE experiments, d)–f) the NUDGE-S experiments, and g)–i) the NUDGE-T experiments. The dashed lines show the start and end dates of the central US CAO. 164

A.10 ERA5 potential temperature anomalies on the 2 PVU surface (°C, shading) for two wave breaks during the evolution of the February 2021 CAO: a) – c) an East Siberian Sea anticyclonic wave break between 2-4 February 2021 and d) – f) a Labrador Sea cyclonic wave break between 9-11 February 2021. Black boxes in a) – c) and d) – f) represent the East Siberian Sea and Labrador Sea wave break regions, respectively. 165
A.11 As in Fig. 5.4 but for the 3 highest and lowest East Siberian Sea 320 K PV ensemble members in the NCEP model. 166
A.12 As in Fig. 5.5 but for the 3 highest and lowest Labrador Sea 320 K PV ensemble members in the NCEP model. 167
A.13 Composite mean maps of ERA5 linearly detrended 500 hPa GPH anomalies (shading, dam) and full-field MSLP (contours, hPa) for a) the Alaskan Ridge (AkR), b) the Arctic High (ArH), c) the Pacific Trough (PT), d) the West Coast Ridge (WCR), and e) the Arctic Low (ArL), respectively. Percentages on panels represent the frequency of days occupied by a given winter weather regime in the 1950-2023 period. Anomalies are with respect to the 1979-2020 climatology. 168
A.14 Scatter plots of peak load anomaly (MW/1000 customers) versus daily mean ERA5 T2M (°C) for (a) non-holiday weekdays and (b) weekend days and non-weekend holidays. We have selected Christmas Day, Thanksgiving Day, and New Year's Day as the holidays to select. The daily average T2M has been averaged over the boxed region in Fig. 2. The red lines represent the second degree polynomial model for the relationship between the two parameters in each plot, with the equation of this polynomial shown on the plot, where T represents the temperature, as well as the coefficient of determination (R^2) value for peak load anomaly respective to the polynomial model. The blue dashed lines represent an average daily peak load between 1999-2022. 169
A.15 Composite maps of ERA5 T2M anomalies (shading, °C), respective to the 1991-2020 climatology, for days exceeding the regime-specific 90th percentile of peak load anomalies in each weather regime of a) AkR, b) ArH, c), PT, d) WCR, e) ArL, and f) all days. The black contour represents the SPP region and the sample size is given for each composite. 170

Abstract

Wintertime cold air outbreaks (CAOs) in the central United States (US) are high-impact events characterized by the southward intrusion of polar air into the midlatitudes. These events, such as the prolific February 2021 CAO, pose serious socioeconomic, environmental, and infrastructural risks. Despite their significance, the characteristics, evolution, and predictability of CAOs in this region remain underexplored, particularly on the subseasonal-to-seasonal (S2S) timescale – i.e., two weeks to two months, where forecast skill is low. This dissertation addresses these gaps through three main objectives: (1) developing a comprehensive climatology of central US CAOs; (2) identifying and assessing key S2S sources of predictability of central US CAOs; and (3) linking extreme central US cold to electricity demand through weather regimes. The development of a central US CAO climatology and the subsequent analyses reveal that central US CAOs are strongly associated with anomalous ridging over Alaska and across the Arctic, as well as with stratospheric wave reflection and downward propagation of circulation anomalies associated with a weak stratospheric polar vortex (SPV), respectively. Furthermore, CAOs that are associated with anomalous Alaskan ridging also demonstrate a North Pacific wave train emanating from the tropical Pacific. To test the importance of the stratospheric and tropical circulations in central US CAO development, targeted nudging experiments are employed in an S2S forecast model to reveal whether improving the representation of the stratosphere and tropical variability increases S2S skill for two CAO events. For the December 2017/18 CAO event, the nudged forecast runs show significant improvement in the long-lead (Weeks 3-5) forecast for the CAO and the tropospheric circulation pattern. However, for the January 2004 CAO event, limited forecast improvement occurs in the nudging experiments, underscoring event-to-event variability in predictability and evolution of central US CAO events. We then build on our nudging experiments by evaluating S2S model skill in forecasting CAOs in general across multiple operational S2S forecast models. Results show that a persistent weak SPV and the presence of high-latitude blocking improve forecast reliability, while false alarm forecasts can be tied to regime transitions that diverge from the observed evolution. As an example of a CAO with substantial impacts, particularly on the electrical grid, that occurred following weak SPV conditions, we also assess the predictability of the February 2021 CAO using S2S forecasts from two operational models, demonstrating that the accurate simulation of two high-latitude Rossby wave breaking events, and associated

blocking patterns, contributed to improved predictability at lead times of 2–3 weeks. Finally, this work establishes a strong link between winter weather regimes, extreme cold episodes, and electricity demand across the central US: regimes with anomalous ridging over Alaska, the Arctic, and the US West Coast (i.e., those most frequently associated with extreme cold) increase the probability of extreme electricity demand by 20–100%. Together, these findings advance our understanding of the dynamical pathways that govern central US CAOs and demonstrate how precursor signals can extend the horizon of skillful forecasts. By clarifying the conditions under which predictability arises and fails, this work lays a foundation for improving S2S prediction of high-impact cold extremes.

Chapter 1

Introduction

Extreme cold produces a wide array of socioeconomic, infrastructural, and environmental impacts, both across North America and globally (Analatis et al. 2008; Smith and Sheridan 2019, 2020; Doss-Gollin et al. 2021). Specifically, spatially and temporally extended periods of extreme cold are termed cold air outbreaks (CAOs), which occur when extremely cold air from the polar regions moves into the midlatitudes (Cellitti et al. 2006; Wheeler et al. 2011; Smith and Sheridan 2018b, 2020). These CAOs can have large impacts due to their extreme and extended nature, such as power outages, electrical grid failure, water availability issues, and unfortunately loss of life. From an economic standpoint, extreme cold events and their associated winter storms can also be very costly. In the United States (US), there have been a total of 33 billion-dollar disaster freeze and winter storm events combined between 1980–2024, resulting in 1625 deaths and a total CPI-adjusted cost of $141.6bn (NCEI 2025). The central US, encompassing the Great Plains of the US, has experienced some of the strongest cold waves in history. A key example is the two-week-long extreme CAO that occurred in the Great Plains during December 1983, resulting in the freezing of water pipes and mains in Oklahoma City, Oklahoma, leaving many residents without water for days (McManus 2009). More recently, the February 2021 central and southern US CAO produced failures in natural gas and water supply lines, widespread power outages, and loss of life (Doss-Gollin et al. 2021; Millin and Furtado 2022; Bolinger et al. 2022; Albers et al. 2022). Various studies have investigated the characteristics and precursors to CAOs in the US (e.g., Konrad and Colucci 1989; Cellitti et al. 2006; Xie et al. 2017; Smith and Sheridan 2018b,a), especially those in the eastern US, but little research

has focused on extreme cold events in the central US. Understanding the precursors, evolution, and model forecast capability for these events will aid progress in improving their predictions at longer lead times, especially at the subseasonal-to-seasonal (S2S) timescale, defined by a lead time of two weeks to two months (Vitart et al. 2017).

The S2S timescale is particularly challenging as it lies between the relatively short memory of atmospheric features that occur on weather timescales, such as extratropical cyclones, and the longer memory boundary conditions contained within the ocean and the land surface for seasonal forecasts, such as the El Niño-Southern Oscillation (ENSO, Vitart and Robertson 2018). As such, the S2S timescale generally features a lower forecast skill than both weather forecasts and seasonal outlooks (Figure 1.1a). However, actionable S2S forecasts of extremes are increasingly in demand as they provide important and longer-lead information that can be used to enhance warning and impact mitigation through multiple actions such as forecast monitoring, preparedness activities, community warning updates, and preparation of disaster-response materials (Fig. 1.1b). Therefore, skillful S2S forecasts of extreme cold in the central US could help to enhance preparedness, warning, and impact mitigation across multiple sectors such as public health, energy, agriculture, transportation, water infrastructure, insurance and financial markets, and emergency management and government services. Prior studies have linked an enhancement in subseasonal predictability, also known as windows of forecast opportunity (Mariotti et al. 2020), of extreme weather events to various modes of variability, such as stratosphere-troposphere coupling (e.g., Lee et al. 2019b; Domeisen et al. 2019a,b; Domeisen and Butler 2020), atmospheric blocking and weather regimes (e.g., Lee et al. 2019b), and tropical teleconnections such as the eastward moving packet of tropical convection known as the Madden-Julian Oscillation (MJO, e.g., Madden and Julian 1971, 1972; Rodney et al. 2013; Grimm et al. 2021) and ENSO (e.g., Domeisen et al. 2015; Yan and Guo 2023). How these modes may impact S2S predictability of central US extreme cold events is still unknown.

Given the growing demand for actionable S2S forecasts of extreme events, we investigate the S2S predictability of central US CAOs, and their related modes of variability,

through the support of the National Science Foundation (NSF) Oklahoma Established Program to Stimulate Competitive Research (EPSCoR) project. The goal of this project is to find answers and solutions to weather, water, land, and infrastructure issues facing Oklahomans. Thus, as part of the S2S team of this project, the aim is to better understand (1) the S2S processes behind extreme events, (2) how to improve models and predictions of S2S processes for extreme events, and (c) how to develop data for managing water, carbon and water cycles, and infrastructure in Oklahoma. Although the forthcoming findings are relevant to Oklahoma, the large-scale nature of central US CAOs allows the results of this dissertation to have broader significance across the central US and potentially other regions of the country. This dissertation will address the project goals from the research standpoint of central US CAOs. Specifically, this work aims to:

1. Develop a climatology of CAOs in the Great Plains of the US and identify their key precursors.

2. Assess the predictability and factors affecting predictability in S2S forecasts of central US CAOs and cold spells, to identify potential windows of S2S forecast opportunity.

3. Link extreme cold events, and associated weather regimes, with electrical demand.

This dissertation will meet these project goals in the content of the following chapters. Chapter 2 will outline the process behind defining CAOs in the Great Plains of the US, before analyzing the key tropospheric and stratospheric precursors for these extreme events on S2S timescales. The work in Chapter 2 is a mostly identical version of the published work "Characteristics, Evolution, and Formation of Cold Air Outbreaks in the Great Plains of the United States" in *Journal of Climate* by Millin et al. (2022), with minor adaptations. Chapter 3 then assesses the separate roles of the stratosphere and tropical circulation on subseasonal forecasts of the January 2004 and December 2017/January 2018 CAOs, to isolate their importance in CAO subseasonal forecasts through a series of nudging experiments. Chapter 4 provides a quantitative assessment of the prediction capability of

current S2S models in simulating central US cold spells, with an analysis of the impact of weather regimes and the stratospheric polar vortex (SPV) on confident forecasts of cold spells. Chapter 5 builds on the prior work and assesses the predictability of the February 2021 CAO in two S2S models, leading to an analysis of how the simulation of high-latitude blocking provided a window of subseasonal forecast opportunity for this CAO. The material in Chapter 5 is a mostly identical version of the published work "The Role of Wave Breaking in the Development and Subseasonal Forecasts of the February 2021 Great Plains Cold Air Outbreak" in *Geophysical Research Letters* by Millin and Furtado (2022), with small adaptations. Finally, Chapter 6 will examine the impact of North American winter weather regimes on central US electricity demand, with a focus on extreme cold episodes. The work in Chapter 6 is a mostly identical version of the published work "The Impact of North American Winter Weather Regimes on Electricity Load in the Central United States" in *npj Climate and Atmospheric Science* by Millin et al. (2024), with small adaptations. The dissertation will then wrap up with the Summary and Conclusions. Overall, the work in this dissertation amounts to significant advances in knowledge as to how these extreme cold events form, what their key S2S precursors are, and how we can improve our ability to more skillfully predict extreme cold on S2S timescales.

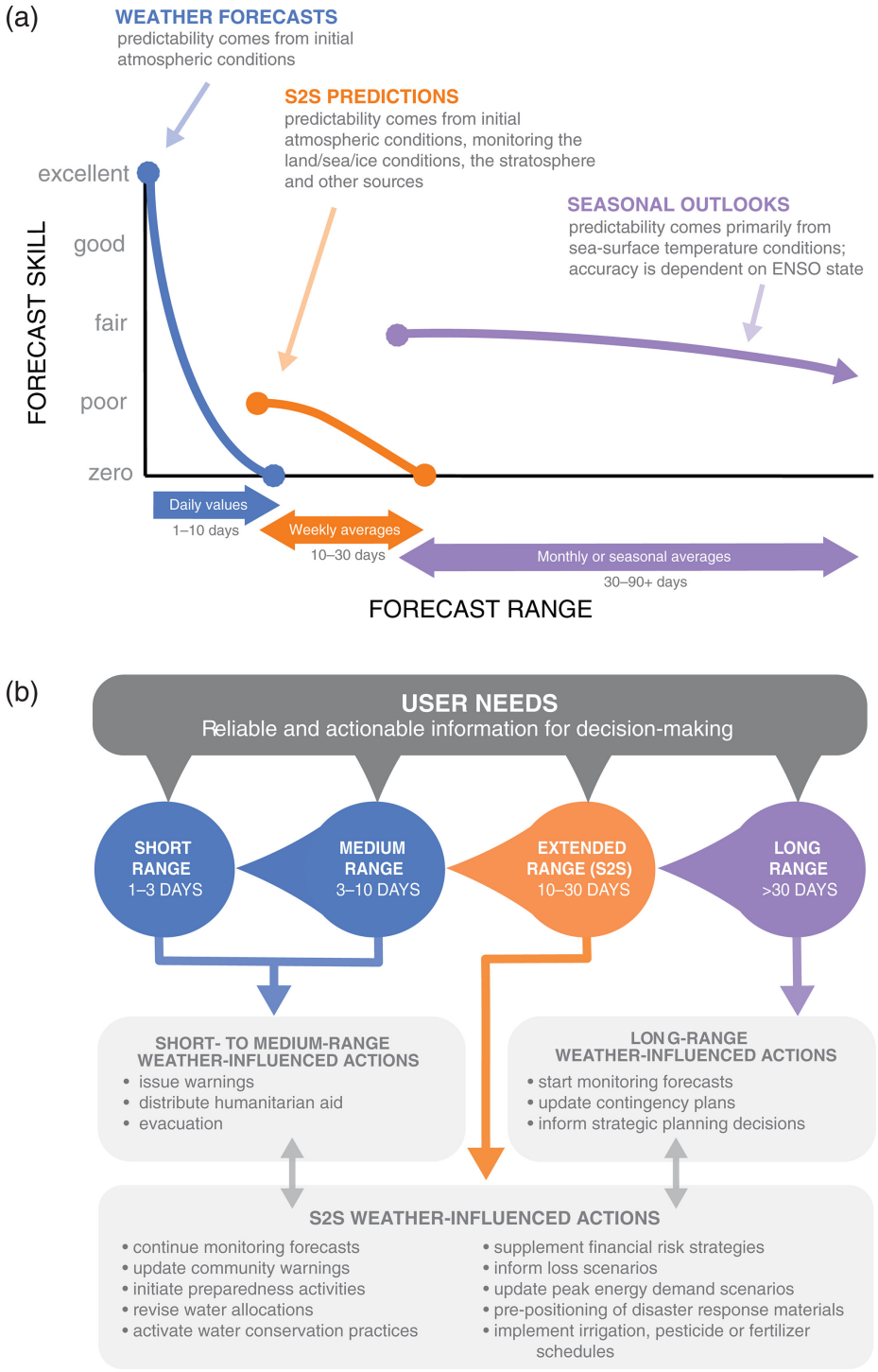


Figure 1.1: (a) A schematic depicting the qualitative estimate of forecast skill based on forecast range from short-range weather forecasts to long-range seasonal predictions, including potential sources of predictability. (b) A schematic diagram highlighting the relationship between the S2S 'extended-range' forecast range and other prediction timescales, with examples of actionable information that can enable decision-making across sectors. Adapted from White et al. (2017).

Chapter 2

Characteristics, Evolution, and Formation of Cold Air Outbreaks in the Great Plains of the United States

2.1 Introduction

The first step in determining the S2S predictability potential of CAOs in the central US is to understand the characteristics, evolution, formation, and S2S precursors of these events, which this chapter will undertake for CAOs in the Great Plains of the US. Numerous prior studies have investigated the properties and precursors of CAOs in the US, though with a strong focus on the eastern half of the US (e.g., Konrad and Colucci 1989; Cellitti et al. 2006; Xie et al. 2017; Smith and Sheridan 2018b,a). Anomalous ridging over Alaska is a recurring atmospheric signature for CAOs across North America, but mainly for northern and eastern regions (Konrad and Colucci 1989; Cellitti et al. 2006; Xie et al. 2017; Lee et al. 2019b). However, links have also been made between cold anomalies in the US Great Plains and high-latitude anomalous ridging across the Arctic (e.g., Stan and Straus 2007; Lee et al. 2019b). The formation of these large-scale patterns help to transport Arctic air masses southward into the US (Grotjahn et al. 2015), associated with overlying negative 500 hPa geopotential height (GPH) anomalies and upstream blocking (Loikith and Broccoli 2012; Westby and Black 2015). Furthermore, Cellitti et al. (2006) found that a weak tropospheric polar vortex, with increased Arctic ridging, is favored prior to CAO onset. Smith and Sheridan (2018a) found results of a similar conclusion but also found that the SPV is persistently weak prior to eastern US CAOs, associated with the negative phase

of the Northern Annular Mode (NAM, Thompson and Wallace 2000) and North Atlantic Oscillation (NAO, Wallace and Gutzler 1981).

SPV variability can have substantial impacts on surface weather (e.g., Baldwin and Dunkerton 2001; Limpasuvan et al. 2004, 2005; Domeisen et al. 2013; Afargan-Gerstman and Domeisen 2020), especially following *major* sudden stratospheric warmings (SSWs). During these events, the dominant westerly winds around the SPV reverse to easterlies in response to strong upward and poleward wave driving, transporting westward angular momentum and anomalous heat, destroying the vortex (Baldwin and Dunkerton 2001; Nakagawa and Yamazaki 2006; Charlton and Polvani 2007). Weak SPV conditions can then downward propagate through the stratosphere and into the troposphere, resulting in the negative phase of the NAM in the troposphere (e.g., Thompson and Wallace 2000; Baldwin and Dunkerton 2001; Hitchcock and Simpson 2014; White et al. 2019) and an equatorward shift of the polar jet stream (e.g., Limpasuvan et al. 2004; Maycock et al. 2020). This stratosphere-troposphere coupling of the NAM in Eurasia and North America increases the probability of CAOs there (Kolstad et al. 2010), with North American CAOs favored under displaced SPV events (Mitchell et al. 2013).

However, the link between SPV variability and North American CAOs is less straightforward than that for Eurasian CAOs. Alternatively, downward coupling via stratospheric wave reflection has been proposed to explain an alternate pathway of the stratospheric influence on CAO development in North America. Rather than planetary-scale wave absorption during weak SPV events, wave activity from the Siberian troposphere is reflected downward by the stratosphere over Canada, inducing tropospheric circulation changes (e.g., Perlwitz and Harnik 2003, 2004; Kodera et al. 2008; Harnik 2009; Shaw et al. 2010; Kretschmer et al. 2018a; Matthias and Kretschmer 2020). This mechanism requires two key conditions: (1) a high-latitude vertical reflecting surface in the upper stratosphere marked by negative wind shear and increased curvature of the zonal wind profile; and (2) a meridional

waveguide in the middle-to-lower extratropical stratosphere, marked by a zonal wind maximum with increased curvature of the zonal wind with latitude (Perlwitz and Harnik 2003; Matthias and Kretschmer 2020).

Prompted by the impactful February 2021 Great Plains CAO and previous literature, this chapter investigates the general characteristics of CAOs in the Great Plains, what their dominant onset weather regimes are, the role that the stratosphere has in their development, and implications for S2S predictability. Our results further emphasize the importance of stratosphere-troposphere dynamical coupling in wintertime extreme weather events and for skillful S2S forecasts.

2.2 Data and Methods

2.2.1 Data

This chapter utilizes daily-averaged November to March (NDJFM) European Centre for Medium Range Weather Forecasts (ECMWF) Reanalysis 5th (ERA5) data (Hersbach et al. 2020), from 1 January 1950 to 31 March 2021 at a 0.5° horizontal resolution with 37 vertical levels from 1000 hPa to 1 hPa. We use this dataset since it is the latest ECMWF reanalysis extending back to 1950. We analyze the following variables: 2 m temperature (T2M), geopotential height (GPH), zonal wind (u), meridional wind (v), and air temperature (T). From these, derived quantities are calculated. The first of these is the Plumb wave activity flux (WAF), a measure of quasi-stationary wave propagation, as defined by Plumb (1985):

$$\mathbf{F}_s = p\cos\phi \begin{pmatrix} v'^2 - \dfrac{1}{2\Omega a \sin 2\phi}\dfrac{\partial(v'\Phi')}{\partial\lambda} \\ -u'v' + \dfrac{1}{2\Omega a \sin 2\phi}\dfrac{\partial(u'\Phi')}{\partial\lambda} \\ \dfrac{2\Omega\sin\phi}{S}\left[v'T' - \dfrac{1}{2\Omega a\sin 2\phi}\dfrac{\partial(T'\Phi')}{\partial\lambda}\right] \end{pmatrix} \tag{2.1}$$

where p is pressure/1000 mb, a is the radius of Earth, Φ is geopotential, λ is longitude, ϕ is latitude, and S is the static stability, defined as:

$$S = \frac{\partial \hat{T}}{\partial z} + \frac{\kappa \hat{T}}{H}, \tag{2.2}$$

where z is height, κ is the ratio of the gas constant for dry air to the specific heat at constant pressure, and H = 8300 m is the scale height approximation for the atmosphere. Carats (ˆ) denote the areal average north of 20°N, and primes $\left(' \right)$ indicate departures from the zonal mean. Since planetary-scale waves are dominant in forcing stratospheric variability, the WAF is filtered for wavenumbers 1-3 to capture their impact. Secondly, we calculate the 100 hPa meridional eddy heat flux (*v'T'*), used for capturing potential regions of enhanced vertical wave propagation within the stratosphere (note that the vertical component of $\mathbf{F}_s$ is proportional to *v'T'* in Equation 2.1). Finally, the quasi-geostrophic streamfunction (ψ) is calculated by:

$$\psi = \frac{Zg}{f}, \tag{2.3}$$

where Z is the GPH, g = 9.81 m s^{-2} is gravitational acceleration, and $f = 2\Omega\sin\phi$ is the Coriolis parameter, with Ω being the angular velocity of Earth. Unless stated otherwise, all anomalies are calculated with respect to the daily 1981-2010 climatology in this chapter, where we remove the seasonal cycle from each day.

2.2.2 Defining cold air outbreaks

Great Plains CAOs are defined by using the December, January, and February (DJF) linearly-detrended and area-averaged T2M anomalies over the region of 92-103°W and 30-48°N (i.e., the region defining the Great Plains). This region is outlined by the black box in Figure 2.1. To be labeled a CAO, there are both magnitude and temporal criteria to meet:

1. Five or more consecutive days must be below the 10th percentile of the linearly-detrended and area-averaged DJF T2M anomalies.

2. Each CAO is separated from the end of a previous CAO by at least four days, as a measure of synoptic independence.

The start of a CAO (onset day) is the first day in which the above conditions are met. A relaxation to criterion (1) is also applied, in which a maximum of 1 day during a CAO is allowed to miss the magnitude threshold. This day cannot be the start or end day of a CAO, and the end of the CAO is the last day the magnitude threshold is met before two or more consecutive days miss the threshold. This slight relaxation of criterion (1) captures events which may otherwise miss the criteria above, but are still likely to be impactful.

2.2.3 Composite analyses and weather regime identification

To diagnose the characteristics and driving mechanisms behind the CAOs, we produce a series of time-lagged composite analyses using the variables mentioned earlier. While compositing is a powerful tool, a disadvantage is that averaging all the events masks event-to-event variability. Thus, to complement the composites, we expand the *k*-means clustering method of Lee et al. (2019b) for calculating weather regimes over North America (180-330°E and 20-80°N) back to 1950. This method is a form of unsupervised machine learning that groups patterns together that show similarities in a meteorological field. An empirical orthogonal function (EOF) analysis is performed on the linearly-detrended ND-JFM 500 hPa GPH anomalies, with respect to the 1979-2020 climatological mean to remain consistent with Lee et al. (2019b). The data are weighted by the square root of the cosine of latitude prior to eigenanalysis. The 12 leading principal component (PC) time series are retained, accounting for approximately 80% of the explained variance. Five regimes are then defined, by using *k*-means clustering on the 12 PC time series via the python package *sci-kit learn* (Pedragose et al. 2011). We use five clusters, instead of four as in Lee et al. (2019b), to account for any shifts on the decadal or longer timescale in the North Pacific over this longer period. Each day in the record is assigned a distinct regime based on the minimum Euclidean distance to the cluster centroid. The CAO events are then subsetted

by the regime on Day 0 (i.e., the first day of the CAO). Composites of relevant atmospheric variables are subsequently computed to assess the dynamics and evolution of the CAOs associated with different regimes. Statistical significance for all composite maps is computed by using bootstrap resampling of 5000 iterations, without replacement.

2.2.4 Examining stratospheric wave reflection

We also examine the role of wave propagation and stratospheric wave reflection on Great Plains CAOs. For wave reflection, we use the reflection index (RI) defined by Matthias and Kretschmer (2020). This RI describes the difference between area-averaged standardized anomalies of the 100 hPa *v'T'* over Siberia and Canada (defined between latitudes 45-75°N and longitudes of 120-185°E and 225-250°E, respectively), as per equation (2) in Matthias and Kretschmer (2020):

$$RI = (v'T')^*_{Sib} - (v'T')^*_{Can}, \tag{2.4}$$

where the asterisk (*) denotes standardized quantities and other variables are as defined earlier. A positive RI indicates more positive anomalous *v'T'* in Siberia than in Canada, which is more indicative of a wave reflection mechanism, especially if there are positive anomalies in Siberia and negative anomalies in Canada (indicating anomalously downward propagating/possibly reflected waves).

To identify reflective surfaces in the stratosphere, the curvature (σ) of the vertical and meridional profiles of the zonal wind are calculated via equation (1) in Matthias and Kretschmer (2020):

$$\sigma = \frac{\dfrac{\partial^2 u}{\partial \chi^2}}{\left[1 + \left(\dfrac{\partial u}{\partial \chi}\right)^2\right]^{\frac{3}{2}}}, \tag{2.5}$$

where χ represents either height (z) or meridional distance (y). Increased curvature is defined by exceeding the climatological curvature.

2.3 Results

2.3.1 Near-surface temperature and atmospheric circulation signatures

Our definition for the Great Plains CAOs yielded 37 events between 1950 and 2021. Fig. 2.1 shows the lag composites for the T2M anomalies across North America, where Day 0 corresponds to the onset day of the CAOs. During Days -7 to -4, there is a lack of significant ($p < 0.05$) T2M anomalies across the Great Plains (Fig. 2.1a). During Days -3 to 0 (Fig. 2.1b), significant negative T2M anomalies appear across the Great Plains, with the coldest anomalies across the northern portion of the domain. The extent and magnitude of cold anomalies significantly increase across the region of study after Day 0 (Fig. 2.1c), with widespread negative T2M anomalies ranging from -6°C in Texas to around -13°C in the northern Great Plains. Even colder T2M anomalies exist to the northwest of the Great Plains, indicating the dominant northwest to southeast propagation. Between Days +5 and +8, the magnitude of T2M anomalies decreases and becomes more uniform in the Great Plains, ranging from approximately -4°C to -7°C (Fig. 2.1d). The presence of these significant, but weaker, T2M anomalies in Fig. 2.1d demonstrates the long duration of some of these events.

The T2M composites in Fig. 2.1 show negative anomalies building across the Great Plains from the northwest, which is likely associated with large-scale circulation patterns. Figures 2.2 and 2.3 show the 500 hPa and 50 hPa lagged composites, respectively, of statistically significant linearly-detrended GPH anomalies (shaded) and the full GPH field (contours) across the Northern Hemisphere for all CAO events. Days -11 through -8 show the development of a dipole of 500 hPa GPH anomalies in the North Atlantic (Fig. 2.2a), resembling the negative phase of the NAO, and significant, positive 500 hPa GPH anomalies develop over Alaska during Days -7 to -4 (Fig. 2.2b). From Days -3 to 0, a region of positive 500 hPa GPH anomalies exists across Greenland, but a stronger, anomalous

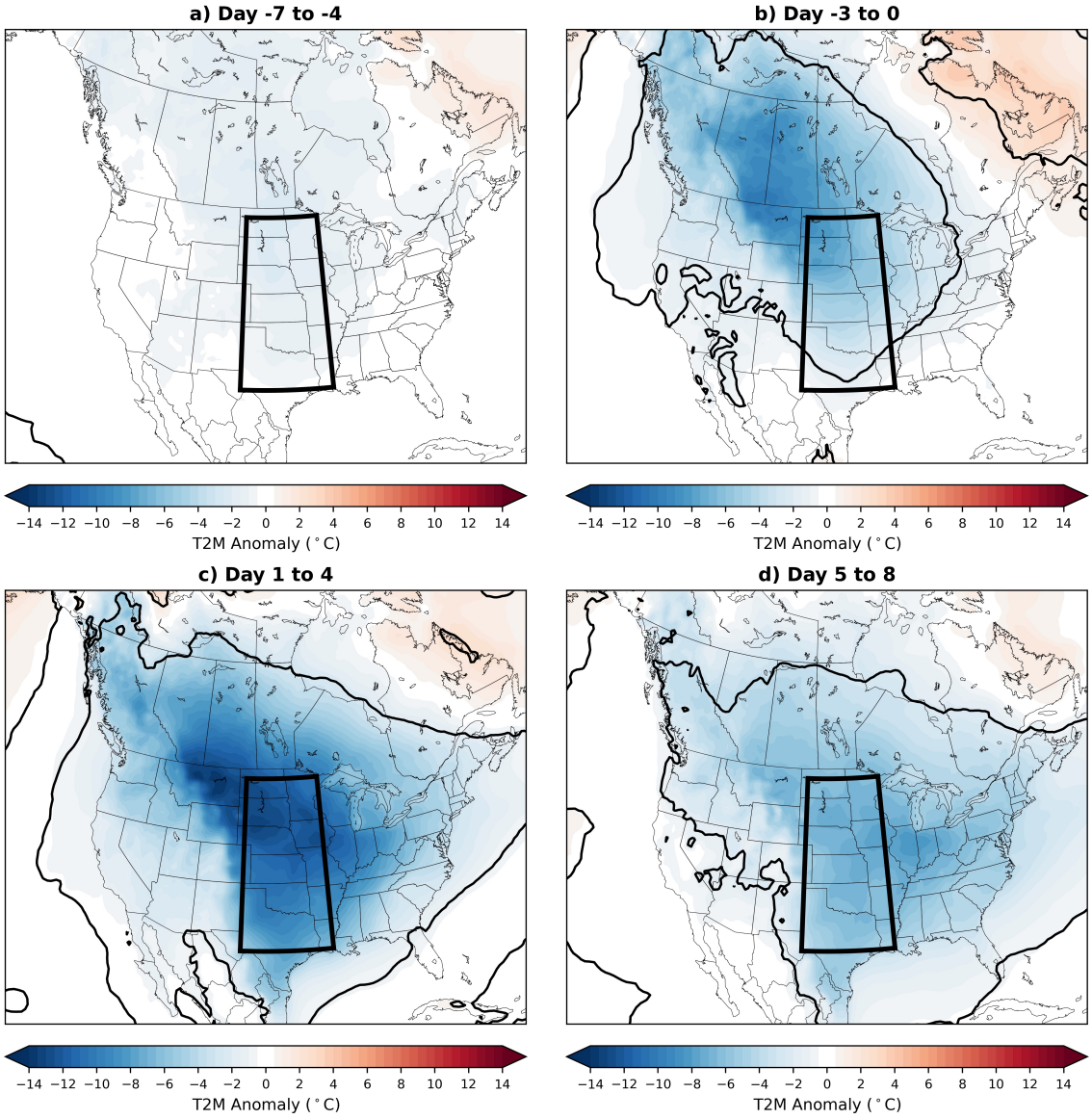


Figure 2.1: Lagged composites of T2M anomalies (°C, shaded) across North America for the 37 identified Great Plains CAOs since 1950, with respect to CAO onset (Day 0) for (a) Days -7 to -4, (b) Days -3 to -0, (c) Days +1 to +4 and (d) Days +5 to +8. Black contours represent significance at the 95th percentile level via a bootstrapping approach of 5000 iterations. The black box represents the region used to define the CAOs. Negative lags indicate days prior to CAO onset day (Day 0), and positive lags indicate days after CAO onset.

ridge rapidly develops over Alaska (Fig. 2.2c), which is a well-known pattern favorable for US CAOs (e.g., Cellitti et al. 2006; Xie et al. 2017; Lee et al. 2019b). At the same time, significant anomalous troughing develops over North America, accompanying cold air advection across the Great Plains via a north to northwesterly flow (Figure A.1). The

anomalous ridging over Alaska persists through to Days +5 to +8 (Figs. 2.2d and 2.2e), while the anomalous troughing over North America weakens and pushes eastward (Figs. 2.2e and 2.2f).

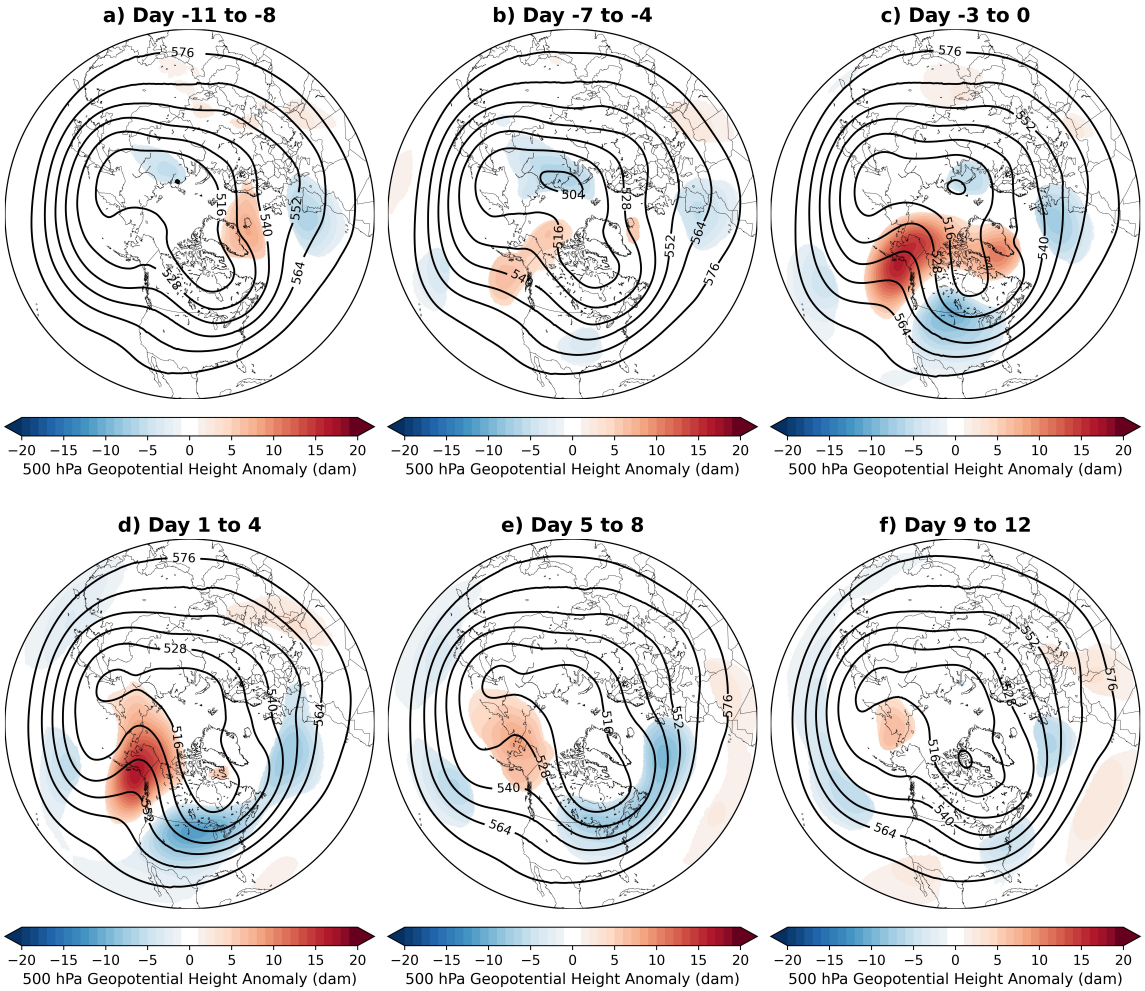


Figure 2.2: Lagged composites of the full-field GPH (dam, black contours) and only the significant ($p < 0.05$) linearly-detrended 500 hPa GPH anomalies (dam, shaded) across the Northern Hemisphere for the 37 Great Plains CAOs, with respect to CAO onset (Day 0) at 4-day intervals from (a) Days -11 to -8 through (f) Days +9 to +12. Statistical significance is calculated via a bootstrapping approach of 5000 iterations. Negative lags indicate days prior to CAO onset day (Day 0), and positive lags indicate days after CAO onset.

In the stratosphere, the 50 hPa GPH anomaly composites show a reasonably equivalent-barotropic pattern relative to the 500 hPa GPH anomalies, especially in the days leading to CAO onset (Figs. 2.3a through 2.3c). However, significant negative GPH anomalies over North America in the stratosphere persist longer than those in the troposphere. Positive 50

hPa GPH anomalies are also present over Eurasia, alongside significant anomalous troughing over North America, indicative of an elongated SPV over North America (Figs. 2.3d through 2.3f), whereby the shape of the vortex is more elliptical than circular in nature (Cohen et al. 2021).

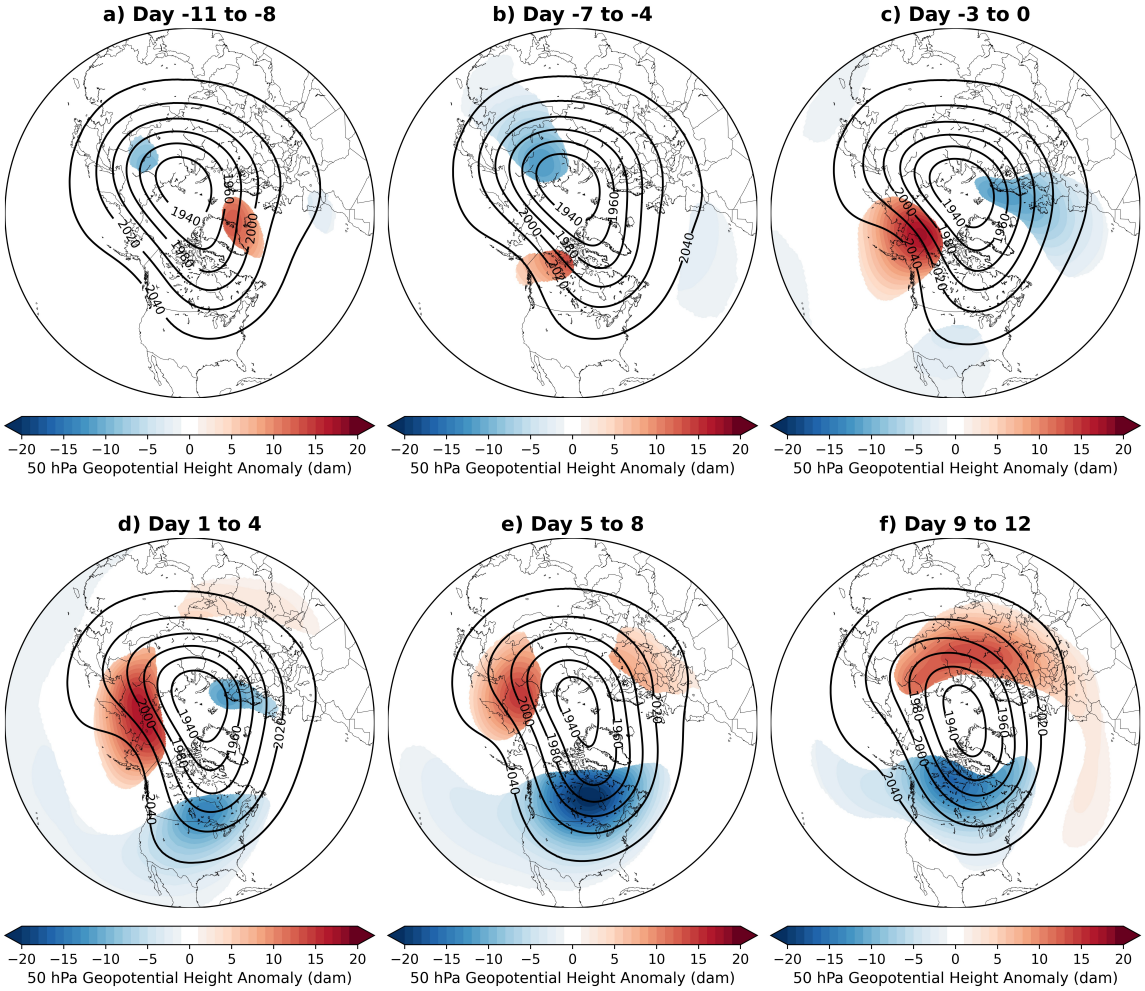


Figure 2.3: As in Fig. 2.2 but for 50 hPa full-field GPH (dam; contours) and only significant ($p < 0.05$) linearly-detrended GPH anomalies (dam; shading) .

2.3.2 North American winter weather regimes and cold air outbreaks

In the previous section, we found that there is anomalous tropospheric ridging in two key areas for Great Plains CAOs on and before CAO onset: Alaska and Greenland. Thus, we next investigate which CAOs are associated with either (or both) of these dominant

patterns, to assess potential differences in dynamics and predictability. Our method of choice is to conduct regime analysis, like Lee et al. (2019b) but for the extended period of 1950-2021. The regime calculation generated five North American weather regimes: the Alaskan Ridge (AkR), the Arctic High (ArH), the Pacific Trough (PT), the West Coast Ridge (WCR), and the Arctic Low (ArL). These regimes are shown in Figures 2.4a through 2.4e, respectively. We find the same AkR, ArH, PT, and ArL regimes as those generated in Lee et al. (2019b) but with a differing order of frequency, with the WCR as the additional regime. Our analysis has the following regimes from most to least frequent: the ArL (Fig. 2.4e, 24.0%) with anomalous North Pacific ridging and anomalous troughing in the Arctic regions, the WCR (Fig. 2.4d, 21.6%) which features a 500 hPa GPH anomaly wave train in the Pacific-North American (PNA) region, the PT (Fig. 2.4c, 20.1%) which shows an enhanced Aleutian low, the ArH (Fig. 2.4b, 17.3%) with anomalous ridging over Greenland and northern Canada, and the AkR (Fig. 2.4a, 17.0%) with anomalous ridging over Alaska.

After assigning each day in NDJFM from 1 January 1950 to 31 March 2021 a weather regime classification, we next subset the 37 Great Plains CAOs based on the regime on the CAO onset day. Of the 37 Great Plains CAOs: 12 are AkR-CAOs, 18 are ArH-CAOs, and 7 are WCR-CAOs. Unsurprisingly, no CAOs are defined as PT- or ArL-CAOs, as these regimes are not conducive for cold air advection into the Great Plains (Figs. 2.4c and 2.4e). Figure 2.5 shows CAO onset day composites of 500 hPa linearly-detrended GPH anomalies for the AkR- (Fig. 2.5a), ArH- (Fig. 2.5b) and WCR-CAOs (Fig. 2.5c). Qualitatively, the AkR- and ArH-CAO composites match to their characteristic patterns in Figs. 2.4a and 2.4b exceptionally well, but with larger anomalies, suggesting that these CAOs develop from amplified tropospheric patterns. The large sample size used to generate the characteristic regime composites could also result in comparatively smaller magnitude GPH anomalies to that of the CAOs (Fig. 2.5). Quantitatively, these patterns also match well with their corresponding regime patterns across the North American domain; the AkR regime (Fig. 2.4a) correlates with the AkR-CAO composite (Fig. 2.5a) with a value of r = 0.86 ($p < 0.05$), and the ArH regime (Fig. 2.4b) correlates with the ArH-CAO composite

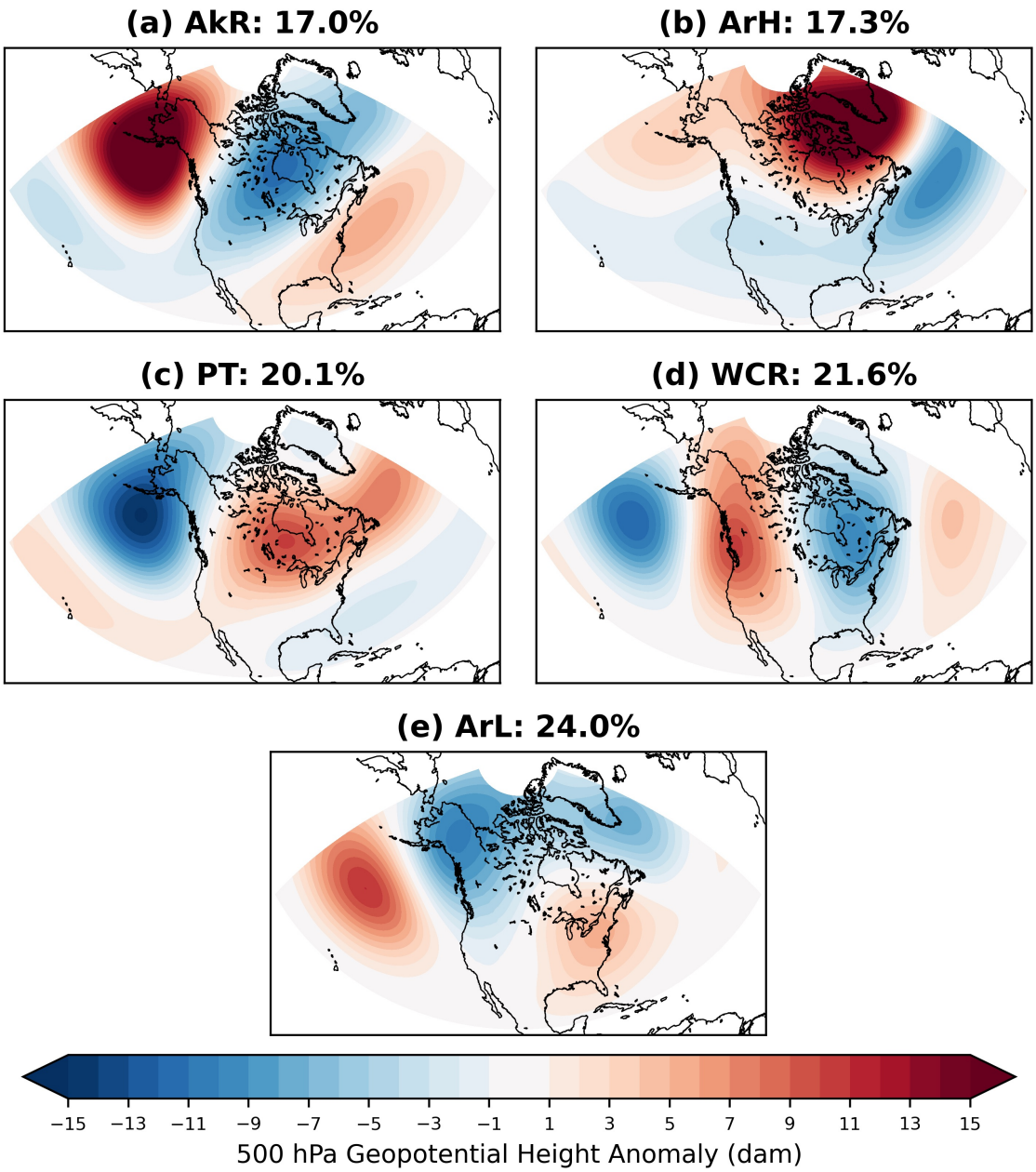


Figure 2.4: Composite 500 hPa linearly-detrended GPH anomalies (dam, shaded), for the five North American weather regimes for NDJFM between 1950 and 2021: (a) Alaskan Ridge (AkR), (b) Arctic High (ArH), (c) Pacific Trough (PT), (d) West Coast Ridge (WCR), and (e) Arctic Low (ArL). Percentages show the frequency of occurrence of a given regime throughout 1950-2021.

(Fig. 2.5b) with a value of $r = 0.89$ ($p < 0.05$). The WCR-CAO composite pattern (Fig. 2.5c) correlates less strongly to its characteristic pattern (Fig. 2.4d), however it is still statistically significant ($r = 0.70$; $p < 0.05$). This weaker pattern correlation arises because

"WCR-CAOs" do not feature anomalous ridging across the entire western US (Fig. 2.5c). Our regime-based composites (Figs. 2.5a and 2.5b) demonstrate that the composites for all CAOs (Fig. 2.2c) mask the relative importance of the AkR and ArH regimes which may involve very different characteristics and evolution, which we next explore.

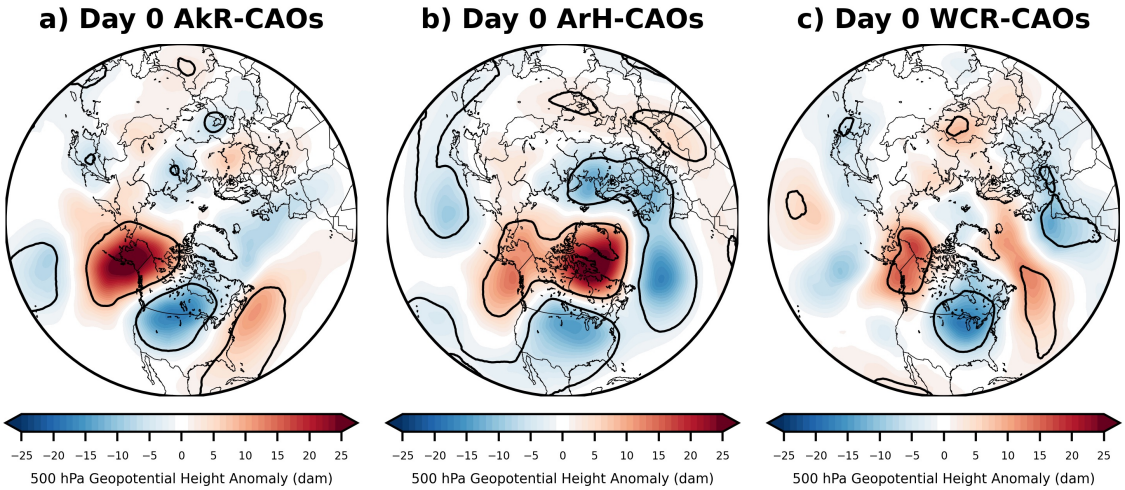


Figure 2.5: Onset day composites of linearly-detrended 500 hPa GPH anomalies (dam, shaded), with respect to the 1981-2010 climatology, are shown for (a) AkR-, (b) ArH- and (c) WCR-CAOs. Black contours represent statistical significance ($p < 0.05$) via 5000-iteration bootstrapping without replacement.

Figure 2.6 shows the magnitude of the CAO events, found by taking the mean domain-averaged T2M anomalies over the duration of the event (noted in bold in Fig. 2.6). The average duration of a CAO is 7.5 days, the average magnitude is -10.4°C, and a CAO occurs approximately once every two winter seasons. Exceptions are two CAOs in the winter seasons of 1972/73, 1978/79 and 1983/1984, and four CAOs in 1977/1978. The intense and persistent AkR-CAO on 16 December 1983 stands out, with an average T2M anomaly of -13°C over 15 days. This event was an exceptional, and prolonged, CAO across much of North America (e.g., McManus 2009). We also computed a "CAO severity index"; i.e., the product of the magnitude and duration of each event and found that the three most severe CAOs based on this index are: (1) 16 December 1983 (AkR, -195°C·day), (2) 7 February 2021 (ArH, -177°C·day), and (3) 21 January 1966 (ArH, -148°C·day). Table 2.1 shows mean statistics for Great Plains CAOs by regime type. ArH-CAOs are slightly longer

on average, the magnitude of the AkR-CAOs is slightly greater, and the WCR-CAOs are shorter and less intense than the ArH- and AkR-CAOs (Fig. 2.6 and Table 2.1), but these differences are not statistically significant. Given these three regime-types have anomalous ridges in the PNA and North Atlantic sectors, their related CAO events may have different wave forcing, stratospheric influences, and developmental timescales.

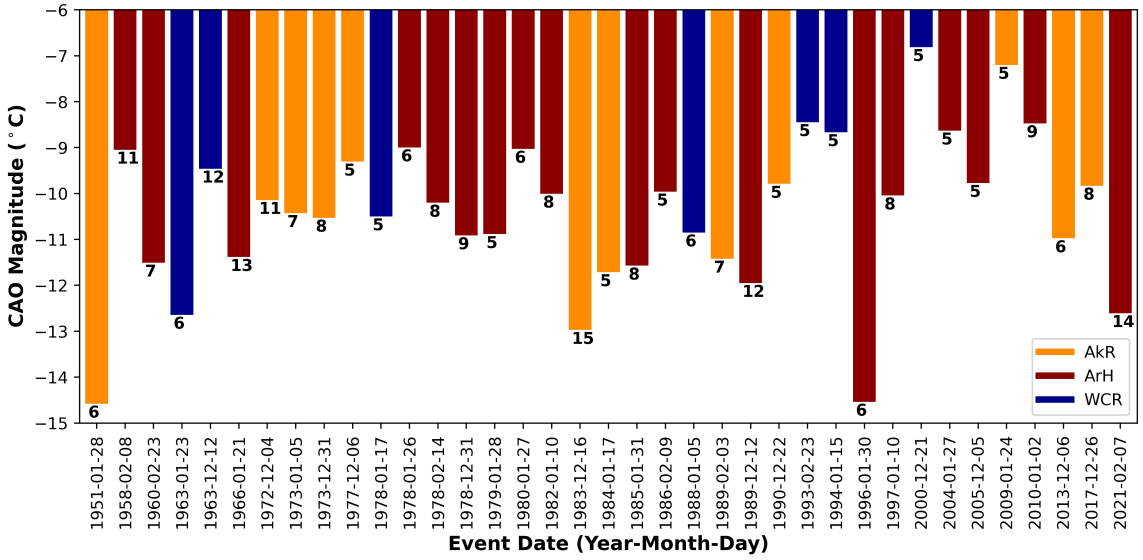


Figure 2.6: Magnitude (°C) of each of the 37 defined CAOs using the non-detrended T2M anomalies, categorized by weather regime: AkR (orange), ArH (red) and WCR (blue). Bold numbers below the bars represent the length (days) of each of the CAOs.

Table 2.1: Statistics for the mean and standard deviations of the magnitude and lengths of the CAOs by North American weather regime.

CAO Mean Magnitude and Lengths				
CAO Regime Type	Magnitude (°C)	Magnitude Standard Deviation (°C)	Length (days)	Length Standard Deviation (days)
AkR (n = 12)	-10.7	±1.78	7.3	±2.87
ArH (n = 18)	-10.5	±1.52	8.1	±2.76
WCR (n = 7)	-9.6	±1.75	6.3	±2.37

2.3.3 Stratospheric modes of variability

2.3.3.1 Stratospheric wave reflection

In this section, we investigate the patterns of *v'T'* and WAF anomalies to diagnose differences in large-scale wave forcing between AkR-, ArH-, and WCR-CAOs. We first focus on *v'T'* at 100 hPa, as this variable is a proxy for vertically propagating waves into the stratosphere. Figures 2.7a, 2.7c, and 2.7e show the 100 hPa *v'T'* for the AkR-, ArH-, and WCR-CAOs, respectively, averaged over Days -3 to 0. For AkR- and ArH-CAOs, *v'T'* is positive and significant over eastern Siberia. ArH-CAOs feature a slightly greater positive heat flux/vertical wave propagation signature over this region than the AkR-CAOs, with another region of significant positive *v'T'* east of Greenland (Fig. 2.7c). Statistically significant positive *v'T'* is actually present over Siberia during ArH-CAOs up to seven days before CAO onset (Figure A.2). Conversely, the strong positive *v'T'* over the Siberian region is significant but only seen immediately prior to AkR-CAO onset (Fig. 2.7a and Figure A.3). Another notable feature is the negative 100 hPa *v'T'* over northwest North America (Figs. 2.7a and 2.7c); this feature is much stronger for AkR-CAOs than ArH-CAOs. Together with the Siberian positive *v'T'*, this heat flux dipole is a hallmark signature for stratospheric wave reflection events over the PNA region (Matthias and Kretschmer 2020), indicating the presence of downward propagating/possibly reflected waves over North America. Finally, the WCR-CAOs do not feature a dipolar signature, with a lack of statistically significant signals over both Siberia and northwest North America (Fig. 2.7e).

Based on the zonal *v'T'* dipole found in ArH- and AkR-CAOs, we modified the original RI from Matthias and Kretschmer (2020) by adjusting the boxes to focus on the regions of significant v'T' over Siberia and Canada (black boxes in Figs. 2.7a, 2.7c, and 2.7e). This index uses the standardized anomalies of the 100 hPa *v'T'* over Canada subtracted from that over Siberia (see Figure A.4 for Day -3 to 0 v'T' anomalies only). The composites of the time-lagged modified RI are shown in Figs. 2.7b, 2.7d, and 2.7f for AkR-, ArH-, and WCR-CAOs, respectively. AkR-CAOs feature a rapid rise in the RI starting at Day -8, peaking

at a value of 2 one day prior to CAO onset (Fig. 2.7b). Analogous to the conclusions from the *v'T'* composites (Fig. A.2), the RI for the ArH-CAOs exceeds zero about a week prior to CAO onset, reaching a statistically significant value of 1.2 on Day -2 (Fig. 2.7d). The ArH-CAO RI features a more gradual increase toward its peak, alongside smaller negative *v'T'* in northwest North America (Fig. 2.7c) than in AkR-CAOs. This gradual increase in RI indicates more constant and prolonged vertical wave propagation from Siberia in ArH-CAOs. As such, ArH-CAOs have approximately 800 K m s^{-1} more cumulative total *v'T'* at 100 hPa over Siberia than AkR-CAOs by Day 0 (Figure A.5). This difference in cumulative *v'T'* suggests a greater weakening of the SPV during ArH-CAOs versus AkR-CAOs, which is later verified via composites. The WCR-CAOs feature no statistically significant signal prior to CAO onset due to the lack of significant *v'T'* in the RI dipole region (Figs. 2.7e and 2.7f). As such, the stratosphere has a minimal influence on WCR-CAOs. Since AkR- and ArH-CAOs appear to have more connections to the polar stratosphere and that the WCR-CAOs make up only $\sim$18% of the total CAO sample size, we focus on the characteristics and formation of the dominant AkR- and ArH-CAOs for the remainder of this chapter.

The statistically significant peaks in the RI for AkR- and ArH-CAOs (Figs. 2.7b and 2.7d) emphasizes the question of whether stratospheric wave reflection plays a role in their development. Matthias and Kretschmer (2020) defined wave reflection events when the RI value exceeds 1.5 for 10 consecutive days. However, for our shorter-lived CAO events, the RI peaks immediately prior to CAO onset with notable asymmetry in the magnitudes of each pole of the index by regime type, especially for AkR-CAOs. The magnitude of positive *v'T'* (i.e., upward wave propagation) from Siberia is slightly larger for ArH-CAOs than for AkR-CAOs. Also, the magnitude of negative *v'T'* (i.e., downward propagating/possibly reflected waves) in northwest North America is much greater for AkR-CAOs than ArH-CAOs. These two observations are also shown in the anomalous *v'T'* fields in Fig. A.4. Since the RI measures the *difference* between the 100 hPa *v'T'* standardized anomalies in these two areas, the index alone cannot determine if a reflective event is occurring. To clarify the high-latitude mechanism occurring in the days leading to Great Plains CAOs,

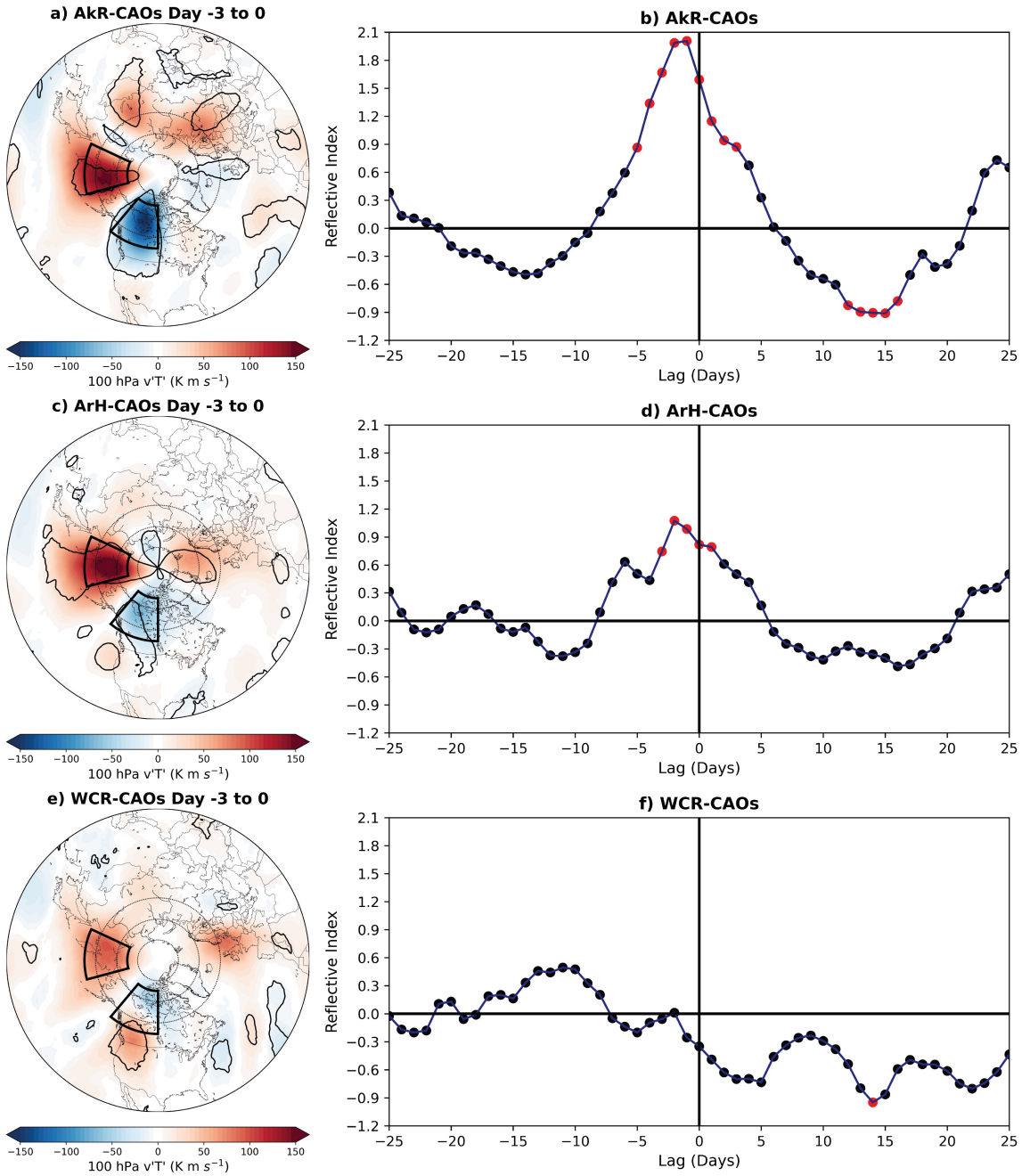


Figure 2.7: Composites of 100 hPa *v'T'* (K m s^{-1}, shaded) for Days -3 to 0 in (a), (c), and (e), and RI in (b), (d), and (f) for AkR-, ArH-, and WCR-CAOs, respectively. The RI is defined as in Matthias and Kretschmer (2020) but between 55-75°N, and 145-185°E and 220-260°E for Siberia and Canada, respectively (thick black lines on (a), (c), and (e)). Black contours on (a), (c), and (e) and red dots on (b), (d), and (f) show statistical significance ($p < 0.05$) via 5000-iteration bootstrapping. The two southerly dashed lines are latitudes 60°N and 70°N, as used in Fig. 2.8. Negative and positive lags indicate days before and after CAO onset day (Day 0), respectively.

Figure 2.8 shows the lagged composites of the averaged 60-70°N linearly-detrended GPH standardized anomalies (shaded contours) and full-field WAF (vertical and zonal components; vectors) for AkR- and ArH-CAOs. The AkR-CAOs feature persistent and significant anomalous stratospheric troughing over Eurasia through to Day -4 (Figs. 2.8a and 2.8b). Simultaneously, there is anomalous stratospheric and tropospheric ridging over Alaska, alongside upward WAF propagating into the stratosphere over eastern Eurasia and Siberia (100-160°E). Fig. 2.8b indicates downward propagating WAF from the stratosphere over Alaska and into northwest Canada, with significant upward WAF from the surface to the mid-troposphere. Finally, Fig. 2.8c shows amplification of the anomalous ridging over Alaska and the wave fluxes seen in the previous lags, with notable stratospheric wave reflection occurring between Days -3 and 0. The wave reflection is denoted by statistically significant WAF propagating from the troposphere to the stratosphere over Siberia, alongside statistically significant downward WAF from the stratosphere to the troposphere over Canada. This panel corresponds to the peak in the AkR-CAO RI (Fig. 2.7b). The evident mid-tropospheric WAF convergence results in troughing over Canada, decelerating the westerly flow, and amplifying the mid-tropospheric wave pattern (Perlwitz and Harnik 2003).

The ArH-CAOs show a different overall evolution of wave activity. Significant anomalous tropospheric troughing develops over Eurasia between Days -11 and -4, along with anomalous stratospheric troughing between Days -11 and -8, and upward WAF from the upper troposphere into the stratosphere over Siberia (Figs. 2.8d and 2.8e). Significant anomalous tropospheric ridging already exists near 300°E during Days -11 to -8 (Fig. 2.8d), earlier than the appearance of the anomalous ridging for AkR-CAOs (Figs. 2.8a-2.8c). This anomalous ridging intensifies toward CAO onset and retrogrades westward (Fig. 2.8f). Importantly, the canonical wave reflection pattern is not seen during ArH-CAOs. Therefore, while the RI indicates a statistically significant peak immediately before both ArH- and AkR-CAOs (Fig. 2.7), stratospheric wave reflection only occurs during AkR-CAOs (Fig. 2.8c).

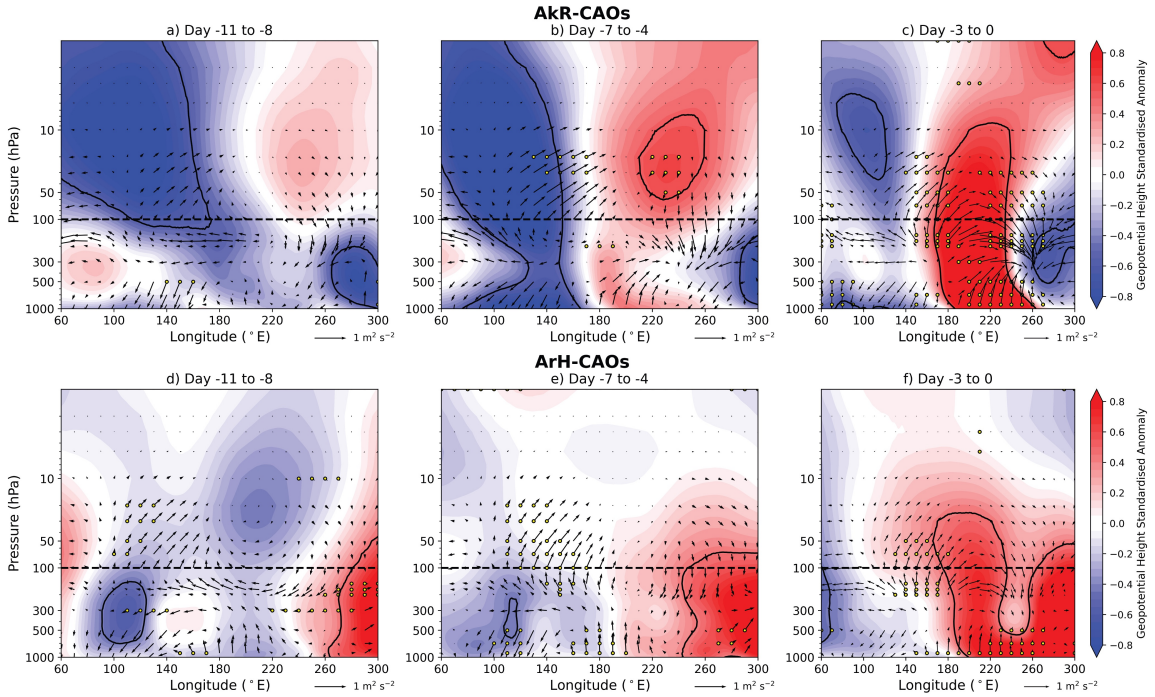


Figure 2.8: Lag composites of the linearly-detrended standardized anomalies of GPH (shaded) and full-field WAF (vectors) at 4-day intervals from Days -11 to -8 through Days -3 to 0 for: (a) to (c) AkR-CAOs and (d) to (f) ArH-CAOs, averaged between 60-70°N. Yellow dots represent statistical significance ($p < 0.05$) of the the vertical component of the WAF via a 5000-iteration bootstrapping without replacement. Black contours represent statistical significance ($p < 0.05$) of the GPH anomalies using the same method. Negative lags indicate days prior to CAO onset day (Day 0), and positive lags indicate days after CAO onset. Scaling of the WAF vectors was achieved by dividing the vertical WAF component by a scale factor of 0.1 above 200 hPa and by a scale factor of 0.3 below and including 200 hPa, and dividing the zonal WAF component by 10 above 200 hPa and by 50 below and including 200 hPa.

Horizontal wave propagation differences between the AkR- and ArH-CAOs are also apparent across the Pacific domain. Figure 2.9 shows the lag-composited 250 hPa streamfunction anomalies and only the statistically significantly large ($p < 0.05$) standardized WAF (meridional and zonal) anomalies in this region prior to and including CAO onset. During AkR-CAOs (Figs. 2.9a, 2.9c and 2.9e), a Pacific wave train develops from the subtropical Pacific to North America and intensifies toward CAO onset. This wave train is not seen for ArH-CAOs (Figs. 2.9b, 2.9d and 2.9f). Instead, there are significant streamfunction and WAF anomalies in the North Atlantic associated with ArH-CAOs (Figs. 2.9b, 2.9d and 2.9f). These findings suggest that precursor patterns leading to AkR-CAOs in Figs. 2.8

and 2.9 may initiate through remote tropical forcing. Indeed, previous studies have linked certain phases of the MJO, and ENSO, to wave breaking and high-latitude blocking in the Gulf of Alaska, setting up an AkR regime-like pattern (e.g., Renwick and Wallace 1996; Henderson and Maloney 2018). By contrast, ArH-CAOs feature pan-Arctic connections, likely tied to the weakened SPV and its downward influence on the NAM (e.g., Baldwin and Dunkerton 2001; Hitchcock and Simpson 2014).

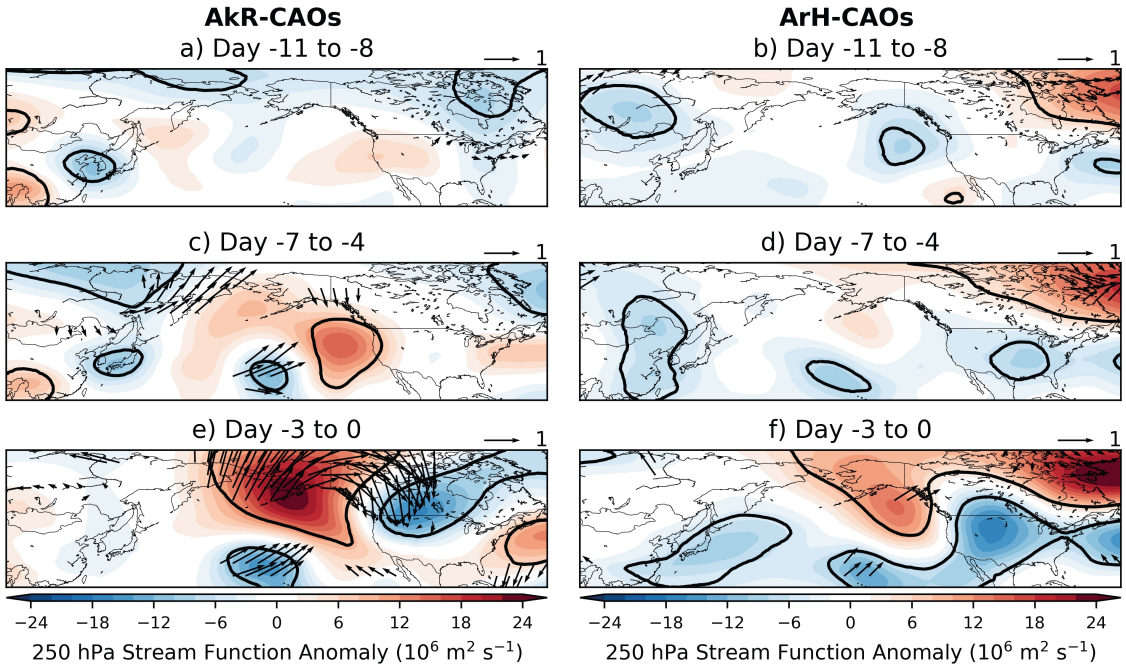


Figure 2.9: Lag composites of the 250 hPa streamfunction anomalies (10^6 m^2 s^{-1}, shaded) and statistically significant ($p < 0.05$) standardized WAF anomalies (vectors) across the Pacific basin for 4-day intervals of Days -11 to -8 through Days -3 to 0 for (a), (c) and (e) AkR-CAOs, and (b), (d) and (f) ArH-CAOs. Black contours represent statistical significance ($p < 0.05$) of the streamfunction anomalies via a bootstrapping approach of 5000 iterations without replacement. Statistical significance ($p < 0.05$) of WAF is computed using the magnitude of the WAF anomalies and 5000-iteration one-tailed bootstrapping without replacement. Negative lags indicate days prior to CAO onset day (Day 0), and positive lags indicate days after CAO onset.

We now investigate the specific conditions that constitute wave reflection, as outlined in previous studies (e.g., Perlwitz and Harnik 2003, 2004; Matthias and Kretschmer 2020). Figures 2.10a and 2.10b show the vertical profile of area-averaged zonal wind between 70-80°N and 200-250°E (i.e., the area of downward reflection in Fig. 2.8c), and Figs. 2.10c

and 2.10d show the meridional profile of the zonal wind averaged over the same longitudes at 30 hPa, analogous to Matthias and Kretschmer (2020). Here we composite Days -7 to -4 and Days -3 to 0 to represent the two panels in Fig. 2.8 where the reflection mechanism appears. Figs. 2.10a and 2.10b show an anomalously strong zonal wind between 200-250°E throughout the stratosphere, and negative vertical wind shear above 10 hPa approaching CAO onset (solid red line). The zonal mean zonal wind profile also demonstrates slightly stronger zonal winds in the stratosphere (Figs. 2.10a and 2.10b, solid blue line). Increased curvature in the regional vertical profile is consistent in the mid-to-upper stratosphere leading up to CAO onset (Figs. 2.10a and 2.10b; solid red line and grey dots), with the zonal mean zonal wind vertical profile also showing increased curvature at the same lags (Figs. 2.10a and 2.10b; solid blue line and grey dots). The meridional profiles of the zonal wind (Figs. 2.10c and 2.10d) demonstrate increased curvature (grey dots) of the regional (solid red line) and zonal mean (solid blue line) zonal wind in both the mid-latitudes and high-latitudes approaching CAO onset, indicating a meridional waveguide between approximately 45-80°N. The regionally averaged meridional profile (Fig. 2.10c, solid red line) is most striking, with a well observed anomalous peak in zonal winds in the high-latitude stratosphere with increased curvature, a condition required for downward coupling from wave reflection (Matthias and Kretschmer 2020). Hence, this analysis suggests that the AkR-CAOs satisfy the wave reflection conditions.

2.3.3.2 Stratospheric polar vortex strength

Given the strong *v'T'* and WAF signature over Siberia for ArH-CAOs (Figs. 2.7c, 2.7d, 2.8e and 2.8f), we next investigate the role that SPV variability may have on the development of ArH-CAOs. SPV variability is of importance since strong or weak SPV states have implications for S2S predictability - e.g., the downward propagation of NAM anomalies (e.g., Baldwin and Dunkerton 2001; Tripathi et al. 2015; White et al. 2019). Figure 2.11 shows pressure-time lag composite cross sections of the standardized anomalies of the

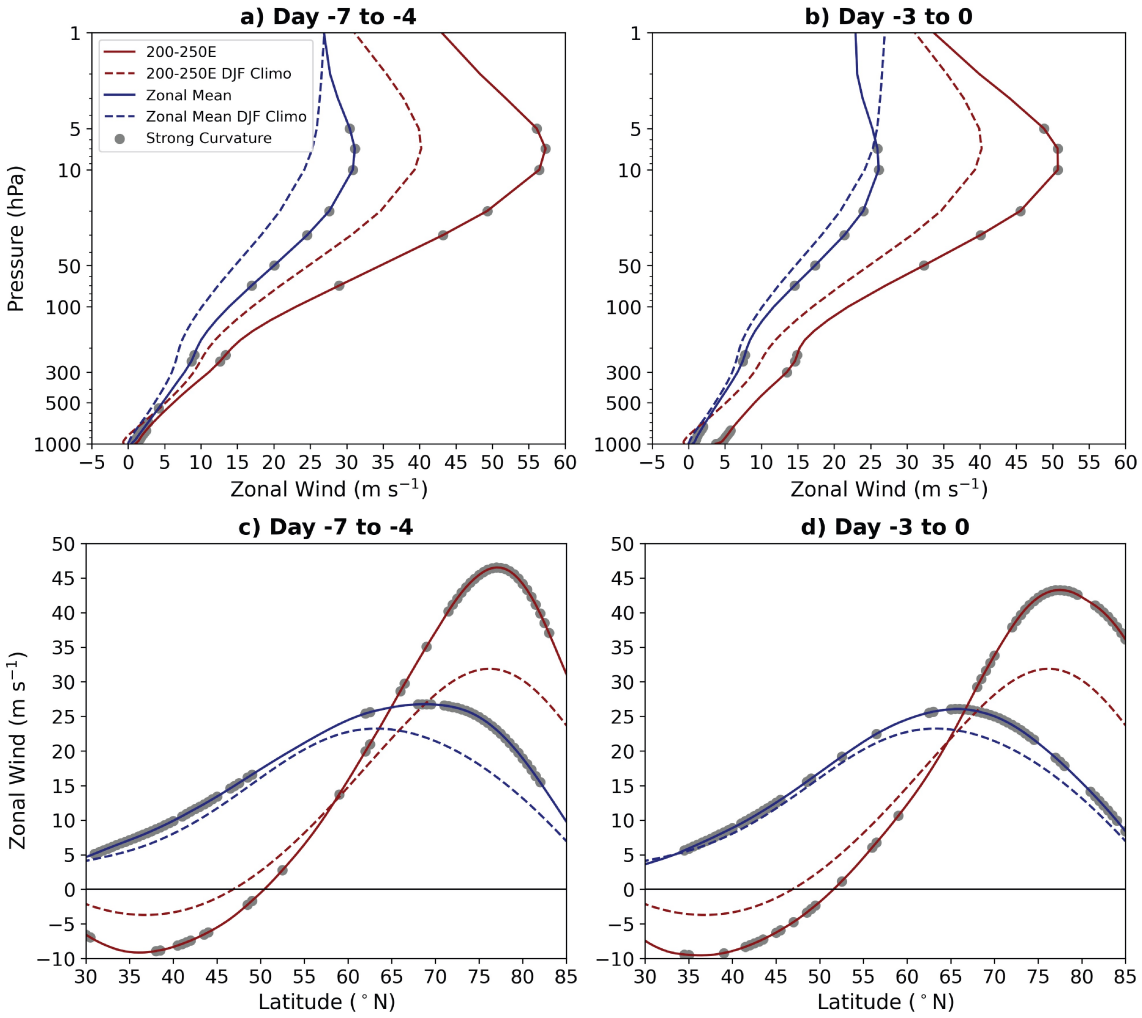


Figure 2.10: Lagged composites of the (a) and (b) vertical profile averaged between 70-80°N, and meridional profile at 30 hPa, (c) and (d), of zonal wind (m s^{-1}) for AkR-CAOs averaged between 200-250°E (red solid line) and in the zonal mean (blue solid line), at 4-day intervals for Days -7 to -4 and Days -3 to 0, respectively, where Day 0 is CAO onset. Red and blue dashed lines represent DJF 1981–2010 climatology of the zonal wind between 200-250°E and in the zonal mean, respectively. Grey dots indicate strong curvature (see Section 2d). Negative lags indicate days prior to CAO onset day (Day 0), and positive lags indicate days after CAO onset.

zonal wind (Figs. 2.11a and 2.11c) and linearly-detrended GPH standardized anomalies (Figs. 2.11b and 2.11d) for AkR- and ArH-CAOs, both area-averaged over the polar cap (60-90°N). Here, we perform lag compositing from Days -40 to +40 (Day 0 is the CAO onset date) since the polar stratosphere has significantly longer timescales for its evolution. Fig. 2.11a illustrates that AkR-CAOs are preceded by statistically significantly ($p < 0.05$)

stronger than average polar cap zonal winds 1-3 weeks before CAO onset throughout much of the troposphere and stratosphere up to 30 hPa. These anomalously strong winds are accompanied by anomalously negative GPH anomalies throughout the stratosphere and upper troposphere (Fig. 2.11b). Interestingly, there is a very sharp transition to weaker than average zonal winds and more positive GPH anomalies just before Day 0, supporting the result that AkR-CAOs undergo a rapid transition after wave reflection, associated with stronger SPV conditions (Fig. 2.8c, Fig. 2.11b and Kretschmer et al. 2018a).

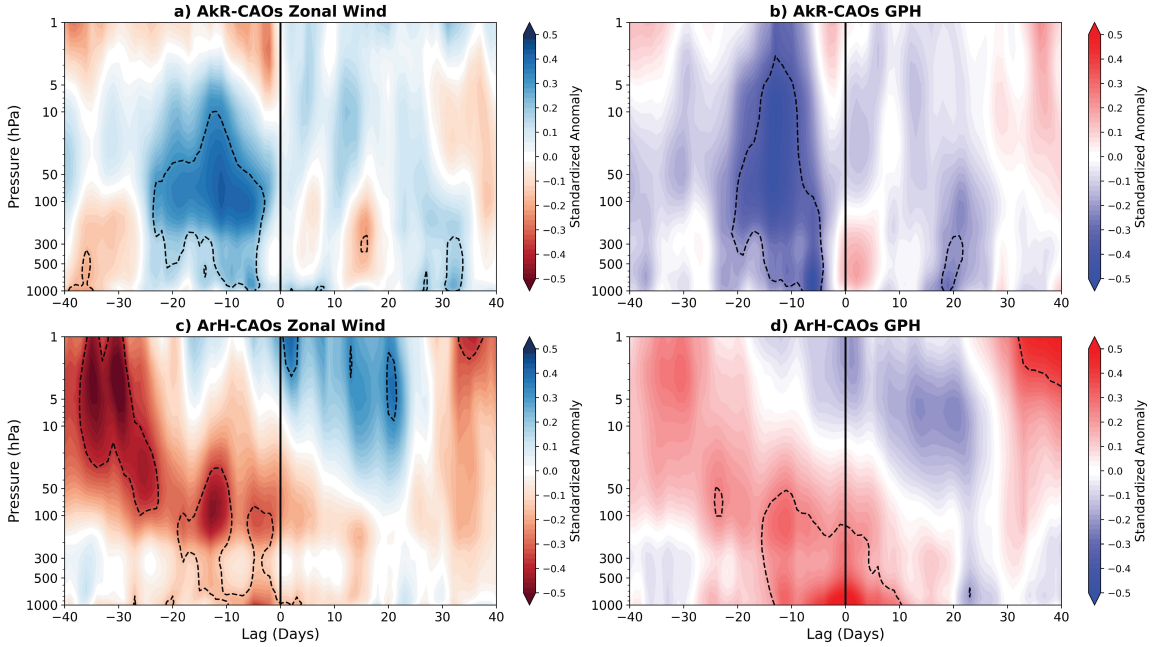


Figure 2.11: Composite pressure-time polar cap (60-90°N) averaged cross-sections of standardized and non-detrended zonal wind anomalies for (a) AkR- and (c) ArH-CAOs, and of linearly-detrended GPH standardized anomalies for (b) AkR- and (d) ArH-CAOs between Days -40 and +40. Dashed lines represent statistical significance at the 95th percentile via bootstrapping of 5000 iterations. Negative lags indicate days prior to CAO onset day (Day 0), and positive lags indicate days after CAO onset.

The patterns of zonal wind and GPH anomalies seen in AkR-CAOs are in stark contrast to those for ArH-CAOs (Figs. 2.11c and 2.11d). The lag composites of both the zonal wind anomalies and GPH anomalies for ArH-CAOs present a classic 'dripping paint' signature, i.e., downward propagation of negative zonal wind anomalies and positive GPH anomalies from the stratosphere into the troposphere (Baldwin and Dunkerton 2001). Fig. 2.11c shows significantly weaker than normal zonal winds in the upper stratosphere between

Days -30 and -40. This signal propagates downward into the mid-to-lower stratosphere between Days -30 to -10, before reaching the surface near CAO onset. Downward propagation appears in the GPH anomalies for the ArH-CAOs as well (Fig. 2.11d), with statistically significant downward propagation of positive GPH anomalies to the surface from around 50 hPa in the two weeks preceding these events. Although not statistically significant, the downward propagation of these GPH anomalies can be seen higher in the stratosphere on longer timescales, consistent with negative zonal wind anomalies (Fig. 2.11c). A possible explanation is that while a weak SPV is a favorable precursor to Great Plains CAOs, a *major* SSW may not be necessary. In support of this explanation, only 7 of the ArH-CAOs were preceded by a major SSW in the period of up to 6 weeks prior to CAO onset (see Butler et al. (2017) and Afargan-Gerstman and Domeisen (2020) for a list of dates), including the *major* SSW in January 2021 (Lee 2021). No major SSWs preceded AkR- and WCR-CAOs.

2.3.4 CAOs and weather regime transitions

The results of Fig. 2.11 clearly demonstrate that the state of the SPV influences the formation of CAOs in the Great Plains, but its exact dynamical influence depends on the regime-type at CAO onset. Such an observation now begs the question of whether the influence of the SPV on CAO formation can be diagnosed in advance, to determine whether a CAO will occur with a given onset day regime. To assess this prospect, Figure 2.12 shows North American weather regime transitions between Day -20 and Day +10 for both AkR- and ArH-CAOs. Firstly, AkR-CAOs mainly show a transition to the AkR regime on synoptic timescales (i.e., a few days), from persistent WCR, PT, and/or ArL states (Fig. 2.12a). PT and ArL regimes are associated with anomalously negative GPH anomalies in the Arctic stratosphere, indicating a stronger SPV (Lee et al. 2019b). Therefore, AkR-CAOs preferentially originate from regimes associated with stronger SPV conditions, aligning with the results in Fig. 2.11a. The WCR regime has a closely located anomalous ridging node to the

AkR, also making it a preferential precursor regime (Figs. 2.4a and 2.4d). Thus, persistent periods of WCR, PT, and/or ArL regimes could increase the probability of an AkR-CAO in the US Great Plains several days to weeks later.

The persistence of the ArH regime prior to ArH-CAO onset is visualised in Fig. 2.12b. In most cases, the ArH regime is the dominant regime for approximately a week or longer before CAO onset. ArH regime days are associated with anomalously positive GPH anomalies in the Arctic stratosphere (Lee et al. 2019b), yielding an anomalously weak SPV. The persistent signal of the ArH regime before ArH-CAO onset relates to the results of Figs. 2.11c and 2.11d. The downward propagation of negative zonal wind anomalies and positive GPH anomalies eventually yields a tropospheric ridging signal in the Arctic troposphere within the 20 days preceding CAO onset. However, is it possible that such a scenario could happen without a CAO? To answer this, we investigated events with 5 or more consecutive ArH days that were preceded by 14-day averaged standardized zonal wind anomalies poleward of 60°N of less than -0.3 at 100 hPa, and/or less than -0.4 at 10 hPa 25-35 days prior to the last ArH day (see Fig. 2.11c). Out of a total of $n = 43$ such events, eight of the ArH-CAOs fall into this category. The other 35 identified events do not meet our CAO criteria but indeed feature T2M anomaly patterns analogous to Fig. 2.1 with cold anomalies propagating from the northwest, but with weaker magnitudes of $\sim$-5°C in the Great Plains. Thus, a weak SPV followed by persistent ArH regimes increases the risk of cold periods for the Great Plains.

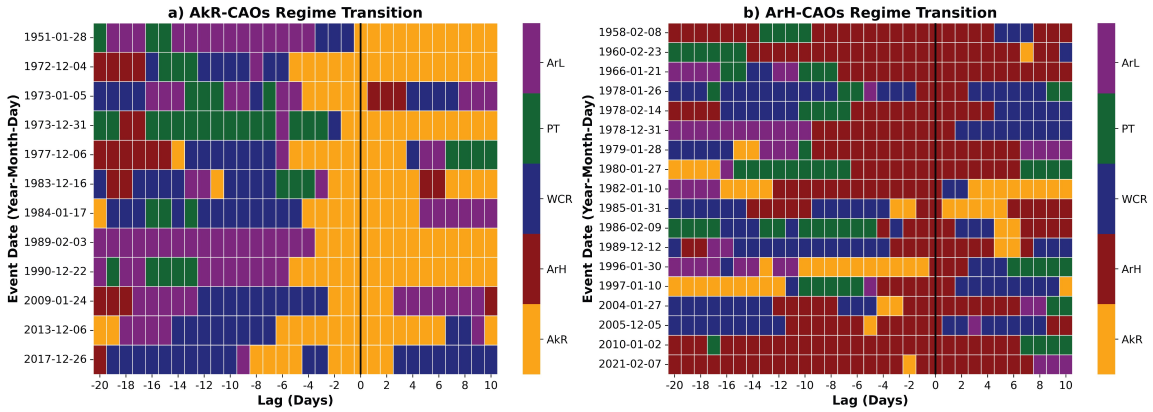


Figure 2.12: North American weather regime transitions for a) AkR-CAOs and b) ArH-CAOs in the US Great Plains, between Days -20 and +10.

2.4 Discussion

Wintertime CAOs in the US Great Plains are high-impact events, though few studies currently exist about their formation and potential S2S predictability. This chapter fills part of this gap. Consistent with previous studies, we show that CAOs in the Great Plains are associated with two dominant weather regimes; anomalous ridging over Alaska or anomalous ridging over Greenland and northern Canada (Figs. 2.5a and 2.5b). Subsequently, the dynamics leading up to these two CAO types are unique and act on different timescales. The differing dynamical characteristics found between different regime-type CAOs demonstrate that using a regime-based approach to classify the CAOs is an efficient, computationally inexpensive, and powerful tool in distinguishing between key differences in dynamical forcings for extreme events.

The AkR-CAOs feature a rapid development of anomalous ridging over Alaska with a wave train across the North Pacific and anomalous horizontal WAF from the central Pacific, associated with a short timescale wave pulse from Siberia leading to stratospheric wave reflection (Figs. 2.8a through 2.8c and Fig. 2.9). Thus, Great Plains AkR-CAOs are sensitive to changes in upstream GPH patterns across the North Pacific, which themselves have strong teleconnections to MJO activity (e.g., Renwick and Wallace 1996; Henderson and Maloney 2018; Green and Furtado 2019). The rapid development of the AkR-CAOs (i.e.,

in a few days) aligns with previous studies which note that stratospheric wave reflection generally occurs in association with rapidly developed wave activity into the stratosphere (Harnik 2009; Kodera et al. 2016; Kretschmer et al. 2018a; Matthias and Kretschmer 2020).

In contrast, ArH-CAOs feature large-scale weak SPV conditions propagating downward over subseasonal timescales (Figs. 2.11c and 2.11d), with both weak upper and lower SPV conditions favored. These results complement past studies on the downward influence of the stratosphere on the troposphere via projecting onto the negative phase of the NAM, which is associated with the southward migration of the polar jet stream and an enhanced risk of Northern Hemisphere CAOs (e.g., Baldwin and Dunkerton 2001; Limpasuvan et al. 2004; Hitchcock and Simpson 2014; White et al. 2019; Maycock et al. 2020). While previous literature suggests that SPV variability is not a robust precursor *on its own* for US CAOs (e.g., Cellitti et al. 2006; Smith and Sheridan 2018a), our study suggests that the precursor potential of SPV variability to Great Plains CAOs relates to the North American winter weather regime in place in the troposphere. Weak SPV conditions and a persistent ArH regime prior to CAO onset relate to ArH-CAOs, while strong SPV conditions and persistent stronger weather regimes relate to the AkR-CAOs (Figs. 2.11a, 2.11c, and 2.12). Moreover, ArH-CAOs feature a longer timescale downward propagation of negative zonal wind anomalies up to around 5 weeks before CAO onset. However, the stronger SPV before AkR-CAOs is seen up to 3 weeks before CAO onset. Thus, the ArH-CAOs feature a longer precursor timescale, potentially resulting in long-lead S2S predictability.

Interestingly, our modified RI identifies significant peaks for AkR- and ArH-CAOs immediately before onset, but with the AkR-CAOs showing evolution over synoptic timescales (e.g., 5 days or less, Figs. 2.7b and 2.7d). However, ArH-CAOs did not demonstrate the reflective trait but instead had intense Siberian positive 100 hPa *v'T'* anomalies which lacked the focused dipole in AkR-CAOs, in association with deeper negative anomalies over Canada. The disparity between RI values and the lack of wave reflection for ArH-CAOs shows that one should examine spatial maps of the 100 hPa *v'T'* alongside the RI.

Using both tools shows that the SPV can absorb the intense Siberian wave activity in ArH-CAOs, aiding the weak vortex regime, but the stronger SPV preceding AkR-CAOs more readily facilitates wave reflection (Figs. 2.8 and 2.11).

The relationship of the CAOs with SPV variability also has climate change implications. That is, the frequency and intensity of these events in a future warmer world could be sensitive to climate change-induced effects on the strength and position of the SPV, as well as on the tropospheric circulation patterns. There has been some uncertainty surrounding the exact impact of anthropogenic climate change on the SPV, with future climate simulations showing no consensus on *major* SSW frequency, and even a decrease in weak SPV downward coupling signals (Ayarzaguena et al. 2020; Hamouda et al. 2021). However, recent studies have shown that the frequency of weak SPV states has increased in the Arctic amplification era, especially the occurrence of "SPV stretching" events (Kretschmer et al. 2018b; Cohen et al. 2021). These "SPV stretching" events are often associated with Alaskan ridging and a dipole in WAF analogous to that of the 100 hPa *v'T'* in Fig. 2.7a. Therefore, an increase in these events could increase AkR-CAO risk in a warming climate. Future work will investigate the frequency changes of AkR and ArH regimes under future climate change, their relationship to the SPV, and how these trends affect CAOs across North America, particularly the Great Plains.

Finally, although the formation of the reflective mechanism associated with AkR-CAOs is fast, there may be potential for anomalous Alaskan ridging to be linked with the tropics. The MJO and ENSO have already been linked with the development of North Pacific blocking, with previous studies showing (1) an increase in North Pacific wintertime blocking during El Niño when followed by MJO phase 7, and (2) increased sensitivity of blocking frequency in the Bering Strait to the ENSO cycle (Renwick and Wallace 1996; Henderson and Maloney 2018). Furthermore, Green and Furtado (2019) showed that MJO phases 7 and 8, alongside a strong SPV, aids the development of anomalous tropospheric ridging over Alaska. Conversely, a weak SPV alongside several MJO phases correspond to anomalous Arctic ridging. Further work should assess MJO phase dependence and ENSO

modulation for Great Plains CAOs, and their relationships with the SPV and high-latitude blocking. These lines of research are important for assessing whether these modes of variability could also act as key S2S predictors for the fast timescale processes involved in AkR-CAOs and other extreme weather events.

2.5 Conclusions

In this chapter, 37 US Great Plains CAOs were defined and analyzed between 1950-2021. The main conclusions were as follows:

- The CAOs had an average duration of 7.5 days and an average T2M anomaly of -10.4°C across the domain. These events occurred approximately once every two winter seasons on average.
- The dominant onset day regimes for these CAOs were the AkR (i.e., Alaskan Ridge, Figs. 2.4a and 2.5a) and the ArH (i.e., Arctic High, Figs. 2.4b and 2.5b). ArH-CAOs are slightly longer and AkR-CAOs slightly colder, although these differences are not significant (Table 2.1).
- AkR-CAOs involve stratospheric wave reflection, consisting of upward wave activity from Siberia alongside downward wave activity over Canada, and a stronger than average SPV preceding the event (Figs. 2.8c, 2.11a and 2.11b).
- CAOs that start with an ArH regime feature a longer timescale build-up of positive heat flux from Siberia (Figs. 2.7c and A.2), accompanied by a weak upper and lower SPV, featuring downward propagation of zonal wind and GPH anomalies (Figs. 2.11c and 2.11d).

A caveat to this work is that using domain-averaged T2M anomalies to define CAOs may mean that extremely cold conditions do not always occur throughout the entire region. Future work should consider splitting the study area into the southern and northern Great

Plains and studying each separately. Another way to define these events may involve implementing a spatial extent threshold to identify more regional Great Plains CAOs, as in Smith and Sheridan (2020). The differences in the timescale and evolution between these two dominant types of CAOs (i.e., AkR- and ArH-CAOs) invites future research. One particular caveat that should be addressed is the further investigation of the fraction of times in regimes and weak/strong SPV conditions in which CAOs do not occur, to further enable application to S2S prediction. Additionally, in the light of recent work by Lillo et al. (2021) and Biernat et al. (2021), future work should investigate any potential link with tropopause polar vortices migrating within the large-scale flow.

Another avenue of future work is identifying potential S2S predictability windows of opportunity for Great Plains CAOs involving tropical to extra-tropical interactions of the MJO and ENSO, which we investigate through tropical nudging in the next chapter. The rapid Alaskan ridge development could be linked to Rossby wave breaking from convection in the tropics, potentially modulated by ENSO (Renwick and Wallace 1996; Moore et al. 2010; Seo et al. 2016; Seo and Lee 2017; Henderson and Maloney 2018). The investigation of how predictable these CAOs are in S2S model hindcasts is shown later in Chapter 4. Additionally, studying the formation, evolution, and dynamics of Greenland and Alaskan blocking may provide further sources of S2S predictability for Great Plains CAOs. Understanding these high-latitude blocking areas is important as they could portend CAOs in North America, especially since weather models inherently struggle representing their frequency and longevity (Quinting and Vitart 2019). Investigating these concepts carries great potential for improving the predictability of these CAOs on the S2S timescale, and for enhancing impact mitigation.

Chapter 3

Analyzing Stratospheric and Tropical Contributions to the Subseasonal Forecasts of the December 2017 and January 2004 North American Cold Air Outbreaks

In the previous chapter we demonstrated that central US CAOs are associated with distinct stratospheric and potentially tropical subseasonal precursors. Building on these findings, we now analyze two CAO case studies from the Great Plains database. These events were characterized by differing stratospheric conditions but occurred during periods of active tropical convection. The MJO (Madden and Julian 1971, 1972) is a well-known source of subseasonal prediction skill for the Northern Hemisphere extratropics, owing to its influence on the polar jet stream via Rossby wave trains (e.g., Johnson et al. 2014; Garfinkel et al. 2014; Henderson et al. 2016; Green and Furtado 2019; Quinting et al. 2023). Therefore, in this chapter, we investigate how both stratospheric variability and tropical convection, in isolation and collectively, affect the subseasonal forecasts of the two selected CAO events, using targeted nudging experiments.

3.1 Introduction

As discussed in Chapter 2, North American cold extremes on subseasonal timescales can follow SPV weakening and are often linked to specific North American winter weather regimes (Lee et al. 2019b; Millin et al. 2022). For instance, the ArH regime is associated with anomalously weak SPV conditions and coincides with 18 of the 37 Great Plains CAOs from 1950–2021. Meanwhile, 12 of the 37 events are linked to stratospheric wave reflection

and the AkR regime, facilitating southward Arctic air transport (Millin et al. 2022). Notably stratospheric wave reflection contributed to the development of the AkR-type CAO during winter 2017/18 (Matthias and Kretschmer 2020). Chapter 2 suggested that such AkR-type events may be influenced by tropical variability, as their development is tied to wave trains originating in the tropical Pacific (Millin et al. 2022).

The tropical-extratropical connection is one pathway by which the MJO influences North American weather. In particular, MJO phases 6–8 have been associated with increased blocking over the North Pacific and North Atlantic due to Rossby wave excitation from enhanced tropical Pacific convection (Henderson et al. 2016; Henderson and Maloney 2018). Active MJO phases have also been linked to stronger upward wave driving into the stratosphere over the North Pacific, increasing the likelihood of *major* SSWs and modulating surface temperature responses 10–14 days after SSW onset in North America (Garfinkel et al. 2012; Green and Furtado 2019). These findings underscore the importance of considering tropical and stratospheric variability when assessing subseasonal predictability of CAOs.

Although it is understood that both the Arctic stratosphere and MJO contribute to subseasonal skill for North American weather patterns, their relative impacts remain to be quantified. Most existing studies, including those discussed in earlier chapters, rely on reanalysis or operational model output, which inherently reflect the combined influence of both. As such, these data do not allow clear attribution of forecast skill to individual sources. To address this gap, we conduct targeted nudging experiments with a subseasonal forecasting model for our two CAO case studies. Nudging modifies the model's physics tendencies, typically for prognostic variabiles (i.e., zonal wind, meridional wind, and air temperature), to steer the model toward a specific target state, often reanalysis. This approach has been used to isolate the role of the stratosphere in European cold events following *major* SSWs and has produced anomalous cold responses in North American forecasts (Kautz et al. 2019; Huang et al. 2022; Ding et al. 2023). It has also been applied to tropical regions, where improving the representation of tropical dynamics has been shown to

reduce forecast error both in the stratosphere and for MJO-related teleconnections (e.g., Karpechko et al. 2023). We apply this technique here to separately assess the influence of tropical convection and stratospheric variability on the subseasonal evolution of our two CAO events.

3.2 Data and Methods

3.2.1 Data

To analyze the evolution of two selected central US CAOs, we use daily averaged data from the Modern-Era Retrospective Analysis for Research and Applications Version 2 (MERRA-2) reanalysis (Gelaro et al. 2017). We use this dataset between 1 January 1981 to 31 December 2022, with a 0.5°x0.625° horizontal resolution and 42 vertical pressure levels distributed unequally from 1000 hPa to 0.1 hPa. MERRA-2 is selected for event analysis to ensure consistency with the nudging experiment input files (see below). We obtain several key variables associated with CAO evolutions: T2M, GPH, meridional wind (v), zonal wind (u), and air temperature (T). Additionally, we also calculate the RI to assess the role of stratospheric wave reflection in these events – an important mechanism in inducing CAOs in the central US (Matthias and Kretschmer 2020; Millin et al. 2022). The RI is now derived in a slightly modified way compared to Chapter 2, following the method of Messori et al. (2022) and is defined as the difference between the standardized anomalies of the area-averaged 100 hPa meridional eddy heat flux $(v'T')$ over Siberia (45-75°N, 140-200°E) and Canada (45-75°N, 230-280°E):

$$RI = (v'T')^*_{Sib} - (v'T')^*_{Can}, \tag{3.1}$$

where asterisks indicate standardized quantities and primes denote deviations from the zonal mean. A positive RI suggests an anomalously higher meridional eddy heat flux over

Siberia than Canada, potentially indicating stratospheric wave reflection. This is particularly evident when the Siberian component is positive (i.e., upward wave propagation) and the Canadian component is negative (i.e., downward wave propagation). Additionally, we obtain phase and amplitude data for the MJO from the Australian Bureau of Meteorology (BoM 2025) to evaluate the intensity and position of tropical convective activity leading up to the events.

3.2.2 Model description and experiments

Nudging in S2S forecasts holds an important role in quantifying the impact of the modes of variability on surface predictability, extreme event attribution, and furthering the understanding of stratosphere-troposphere coupling mechanisms, as outlined in the Stratospheric Nudging And Predictable Surface Impacts (SNAPSI) protocol (Hitchcock et al. 2022). Hence we use the Community Earth System Model version 2 (CESM2) subseasonal prediction system (Richter et al. 2022), which includes the Community Atmosphere Model Version 6 (CAM6) as its atmospheric component, to perform targeted nudging experiments for our cold events. CAM6 operates at approximately 1° horizontal resolution in latitude and longitude and includes 32 vertical levels from the surface to 2 hPa ($\sim$40km). CESM2 CAM6 subseasonal forecasts span a 45-day lead time, with 21 ensemble members initialized weekly each Monday. CESM2 CAM6 is chosen for its high prediction skill in forecasting variables such as T2M, the MJO, and the NAO, and its the ability to nudge specific regions of the atmosphere toward predefined evolutions (Richter et al. 2022). To calculate anomalies in the subseasonal forecast experiments, we use CESM2 CAM6's 11-member historical initialized hindcasts from 1999 to 2020 to develop a climatology that accounts for lead time and initialization dependencies. Because CESM2 CAM6 hindcasts are initialized weekly, each calendar day within the hindcast period lacks initializations for every year, unlike other models in the S2S model database (Vitart et al. 2017). To address this irregularity in initialization dates, we adapt the method of Pegion et al. (2019)

to generate our model climatology. First, we compute the mean across ensemble members and initializations for each calendar day initialization. Then, we apply a rolling triangular smoother with a 31-day window ($\pm$15 days).

To isolate the influence of stratospheric variability vs. tropical atmospheric variability on the evolution of central US CAOs, we conduct three types of experiments that nudge air temperature, zonal wind, and meridional wind toward the MERRA-2 reanalysis evolution across specific vertical and horizontal domains. The first experiment, called NUDGE-S, nudges the stratosphere horizontally everywhere, with full nudging at and above $\sim$100 hPa. A hyperbolic transition centered at $\sim$150 hPa tapers the nudging to zero at $\sim$200 hPa. This nudging approach is slightly modified compared to previous studies in the way it nudges lower in the atmosphere, but ensures full nudging of the lower stratosphere, capturing potential stratosphere-troposphere coupling mechanisms like wave reflection and downward propagation from SSWs. In the second experiment, called NUDGE-T, nudging is applied throughout the atmospheric column at all heights and across all longitudes but only between 10°S–10°N, with a hyperbolic transition tapering to zero at around 30°S/N. No nudging occurs poleward of these latitudes. These bounds are similar to those used in previous studies (e.g., Jung et al. 2010; Dias et al. 2021; Karpechko et al. 2023; Vitart and Balmaseda 2024). Our third experiment, called FREE, is a control simulation where the atmosphere is allowed to freely evolve without any nudging, for baseline comparison.

Each nudging experiment uses 3-hour MERRA-2 input files (UCAR 2024) for the 32-level CAM6 configuration, with nudging performed on a 12-hour timescale, corresponding to a nudging coefficient of 0.25 for the 3-hour meteorology files. Furthermore, the nudging setup allows the model to linearly interpolate target values to the current model time, with a constant time scale based on the time interval of the target values. By comparing these nudging experiments, we can assess how much the stratospheric or tropical circulation influences the subseasonal forecasts of central US CAOs. We will evaluate the impact of the nudging experiments through a series of ensemble-mean analyses highlighting the

difference in T2M and 500 hPa GPH versus that of the FREE experiment, where the statistical significance on the difference between these evolutions will be calculated via 5000-iteration bootstrapping. Furthermore, we also evaluate the impact of the ensemble-mean 500 hPa GPH response in terms of North American winter weather regimes.

3.2.3 North American winter weather regimes

One way to examine circulation differences during CAOs in each forecast experiment is to look at the evolution of North American winter weather regimes (Millin et al. 2022). We apply the same regime definition framework as in the earlier chapters to the MERRA-2 reanalysis data over the period of November to March for 1981-2022. Specifically, 500 hPa GPH anomalies in the North American domain (20-80°N, 180-330°E) during November to March are used to define the regimes. In this case, the analysis begins with an EOF analysis applied to the linearly detrended 500 hPa GPH anomalies, but now with respect to the 1981–2022 climatology, where the seasonal cycle is removed from each day. To further evaluate the strength of each regime in both MERRA-2 and the CESM2 CAM6 forecast experiments, we calculate the weather regime index (WRI) by projecting each day's 500 hPa GPH anomalies onto the cluster mean patterns (e.g., Michel and Rivière 2011). The projections are then standardized across the 1981-2022 period for MERRA-2 and using the WRI long-term mean and long-term standard deviations for the CESM2 CAM6 reforecasts between 1999-2020.

3.2.4 Event choice

To investigate the impact of stratospheric and tropical influences on the evolution of central US CAOs and their associated circulation patterns, we select two events from the Millin et al. (2022) Great Plains CAO database in Chapter 2. Our first event is the central US CAO from 26 December 2017–2 January 2018, characterized by stratospheric wave reflection (Matthias and Kretschmer 2020) and an active MJO, occurring under an AkR weather

regime (Millin et al. 2022). The second event is a central US ArH-type CAO from 27 January 2004–31 January 2004, which was preceded by a major SSW on 5 January 2004 (Butler et al. 2017), and similarly featured active MJO conditions. These two cases offer a unique opportunity to explore the role of stratospheric representation in subseasonal forecasting of central US CAOs, each with distinct stratospheric mechanisms but influenced by concurrent tropical convection.

3.3 Evolutions of the two central US CAO events

3.3.1 Evolution of the extratropical circulation pattern

We begin our analysis by examining the synoptic evolution of our selected CAO events. Figure 3.1 illustrates the evolution of T2M and 500 hPa GPH anomalies across the Northern Hemisphere for both CAOs in MERRA-2. The December 2017/January 2018 CAO is marked by a strong, anomalous ridge over Alaska ($\sim$+30 dam) leading up to and during the first three days of the event. This ridge coincided with anomalous troughing over Canada and the northern US and was associated with anomalously negative T2M anomalies from Canada into the central and eastern US (Figs. 3.1a and 3.1b). From 29 December to 31 December 2017, the anomalous ridge over the US West Coast intensified, shifting more negative T2M anomalies into the central and eastern US from northwest North America. Meanwhile, the downstream troughing pattern shifted eastward, extending into the North Atlantic (Fig. 3.1c). Between 1 January and 6 January 2018, anomalous ridging expanded along the US West Coast with downstream troughing across the eastern half of North America. This pattern produces above-normal T2M in the western half of North America and below-normal T2M in the eastern half. (Figs. 3.1d and 3.1e). Although the CAO primarily spanned 26 December 2017 to 2 January 2018 within the central US, strong negative T2M anomalies persisted over the eastern US beyond this period.

The synoptic evolution of the 27–31 January 2004 central US CAO contrasts substantially with that of the December 2017/January 2018 event (Figs. 3.1f-j). In the days leading up to this CAO, a strong anomalous ridge existed over southern Greenland (exceeding +30 dam), flanked by anomalous troughing to the south over the North Atlantic (Fig. 3.1f). Additionally, anomalous ridging existed over the East Siberian Sea, while troughing extended from Canada into the central US, generating negative T2M anomalies across northwest and northeast North America (Fig. 3.1f). In the first three days of the CAO event, these negative T2M anomalies advanced from northwest Canada into the central US, particularly in the northern US Great Plains (Fig. 3.1g). The Greenland/northern Canada anomalous ridge persisted until 4 February 2004, with a deep trough to the south, forming a robust negative NAO-like pattern, particularly pronounced between 27 January–1 February 2004 (Figs. 3.1g and 3.1h). As the CAO event subsided, moderately negative T2M anomalies lingered across the US and North America (Figs. 3.1h–3.1j), though the intensity of the cold lessened compared to the peak values seen in Figs. 3.1f and 3.1g. Taken together, our two events reveal distinct differences in the duration of negative temperature anomalies and in the dominant high-latitude blocking features (Alaskan ridging vs. Greenland/northern Canada ridging) that shape each CAO's evolution. While these key blocking areas are known to facilitate CAO development in the central and eastern US, they also exhibit unique relationships with stratospheric variability (Lee et al. 2019b; Millin et al. 2022). Consequently, we now turn to an analysis of the stratospheric and tropical forcings associated with these events.

3.3.2 Evolution of the stratospheric and tropical circulation patterns

Figure 3.2 illustrates the lagged evolution of the polar cap (60-90°N) averaged zonal wind standardized anomalies, RI, and MJO phase and amplitude for the two selected CAO events. Approximately one month before the December 2017/January 2018 CAO, the polar cap averaged zonal wind standardized anomalies featured anomalously positive values

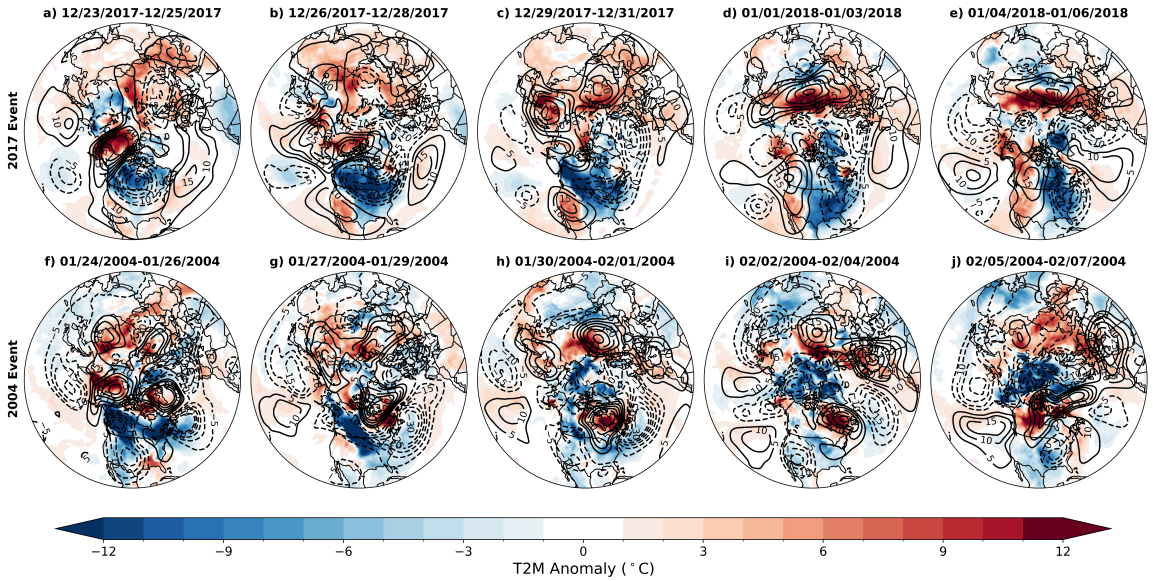


Figure 3.1: Maps of MERRA-2 T2M anomalies (°C, shading) and 500 hPa GPH anomalies (dam, contours) averaged for 3-day periods from a) 23 December 2017 – 25 December 2017 through e) 4 January 2018 – 6 January 2018 for the December 2017 central US CAO event, and f) 24 January 2004 – 26 January 2004 through j) 5 February 2004 – 7 February 2004 for the January 2004 central US CAO event. Anomalies are calculated with respect to the 1999-2020 climatology, where the seasonal cycle is removed from each day.

in the upper stratosphere, signaling a stronger than normal polar vortex circulation (Fig. 3.2a). About 3 to 4 weeks before the CAO, there were anomalously negative polar zonal wind anomalies in the troposphere, later accompanied by anomalously negative zonal wind anomalies in the mid-to-upper stratosphere 2-3 weeks before the event (around 6 December 2017). These anomalies persisted until the start of the CAO event before shifting to a stronger vortex state (Fig. 3.2a). Concurrently, as the mid-to-upper stratospheric polar zonal wind anomalies become more negative, the RI increased from +3σ to +5σ (Fig. 3.2b, black line) until the onset of the CAO (Fig. 3.2b). These high RI values occurred alongside multiple anomalously positive Siberian meridional eddy heat flux pulses (Fig. 3.2b, red dashed line) followed by anomalously negative Canadian meridional eddy heat flux pulses at 100 hPa (Fig. 3.2b, blue dashed line). Therefore, the December 2017/January 2018 CAO was preceded by several stratospheric wave reflection events. Additionally, two active periods of the MJO coincided with or slightly lagged the Siberian heat flux pulses. These active periods featured the MJO in phases 6–8, which are phases linked to North

Pacific and North Atlantic blocking (Fig. 3.2c, Henderson et al. 2016; Henderson and Maloney 2018). The RI and MJO timing suggest that active tropical convection may be a driver of the extratropical wave pattern associated with AkR-type CAOs in the central US, as proposed in Chapter 2 (Millin et al. 2022).

The evolution of the stratospheric and tropical circulations for the January 2004 CAO differs notably from that of December 2017/January 2018. The January 2004 event followed a major SSW on 5 January 2004, indicated by strong negative anomalies of the polar cap averaged zonal wind in the mid-to-upper stratosphere (Fig. 3.2d). These anomalies gradually propagated to the troposphere, with two surface 'drips'– one around 15 January 2004 and another at the CAO onset on 27 January 2004. These two "drips" were accompanied by strong, anomalous ridging over Greenland/northern Canada (Fig. 3.1). Hence, there is some evidence that the major SSW may have influenced the January 2004 CAO event. Unlike the December 2017/January 2018 CAO, the RI evolution shows a large peak ($+3\sigma$, Fig. 3.2e) caused by a large, brief meridional eddy heat flux pulse over Siberia around 29 December 2003 (Fig. 3.2e, red dashed line). This pulse likely weakened the SPV. Notably, this RI peak is smaller than in Fig. 3.2b and lacks a subsequent strong negative Canadian heat flux response, indicating that stratospheric wave reflection did not occur in this event. Despite the absence of wave reflection, the January 2004 CAO was preceded by a highly active MJO (amplitude > 2) in phases 6–8, which shifted into Phases 1, 2, and 3 a week before the CAO began (Fig. 3.2f).

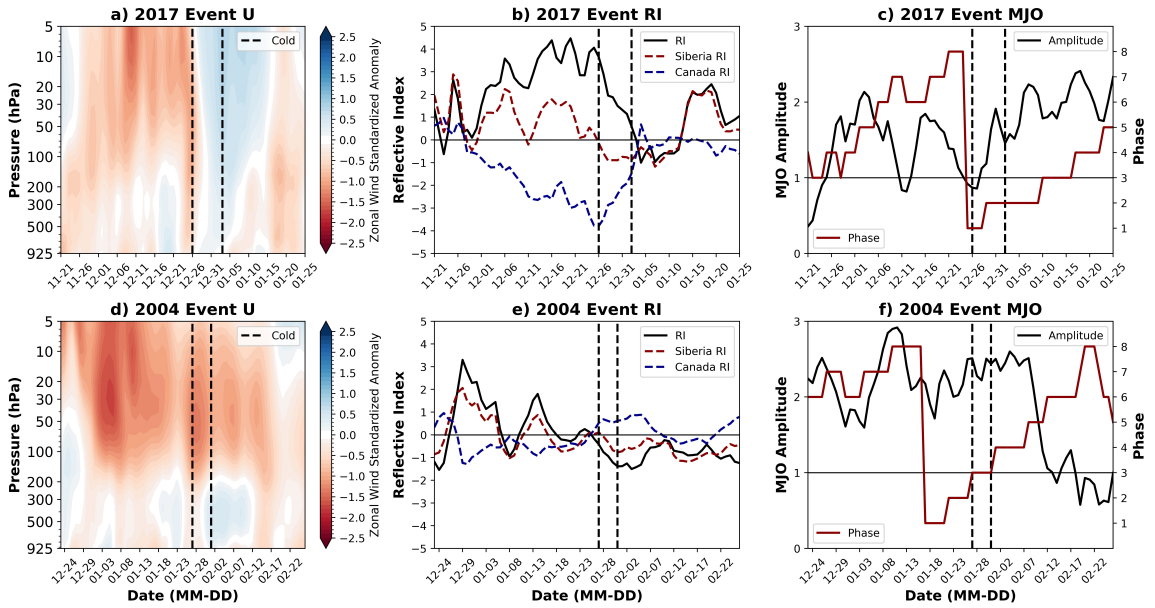


Figure 3.2: The evolution of several key S2S indices for the December 2017 central US CAO event: a) the polar cap (60–90°N) averaged zonal wind standardized anomalies throughout the atmosphere (shading) from MERRA-2, b) the RI (black line) and its Siberian (red dashed line) and Canadian components (blue dashed line) from MERRA-2, and c) the MJO amplitude (black solid line) and phase (red solid line) from the BoM. Panels d)–f) show the same as a)–c), respectively, but for the January 2004 central US CAO. Black dashed lines indicate the start and end of the CAO periods. Anomalies are calculated with respect to the 1999-2020 climatology, where the seasonal cycle is removed from each day.

3.4 Model experiment results

Our analyses of the evolution of the atmospheric circulation in each CAO event reveal that each event features different tropospheric weather regimes, stratospheric mechanisms, and extratropical-tropical interactions. We next examine the role of the stratosphere and tropics in subseasonal forecasts of these CAOs through the nudging experiments, to investigate their impacts collectively and in isolation.

3.4.1 December 2017/January 2018 CAO analysis

3.4.1.1 Forecasts of the reflection index

For the December 2017/January 2018 CAO, we now examine the RI in our forecast experiments. Figure 3.3 illustrates the forecast spread of the RI and its two components for the target period 14-25 December 2017. This target period corresponds to the period of peak RI and the strongest downward wave response prior to the CAO (Fig. 3.2b). We show the results for three subseasonal CAO forecast lead times for each experiment (FREE, NUDGE-T, and NUDGE-S). For the 27 November 2017 initialization ($\sim$4 weeks before the CAO onset), the forecast RI in the FREE run is centered around 0 (i.e., no reflection), with the maximum of the distribution reaching about +3σ, lower than the MERRA-2 observed value of $\sim$+4σ (Fig. 3.3a, gray). In contrast, the NUDGE-T experiment shifts the forecast RI distribution toward higher values, reaching a maximum closer to +4σ (Fig. 3.3a, red). This finding suggests that, even though the extratropical stratosphere is not directly nudged in the NUDGE-T experiment, the tropical atmosphere contributes to a better forecast of the RI for this event. The NUDGE-S experiment performs best for this initialization, as may be expected, showing a much narrower ensemble spread and values close to, and even exceeding, the observed evolution (Fig. 3.3a, blue). Indeed, in the NUDGE-S experiment, there is a closely similar evolution of the polar cap zonal wind standardized anomalies in the stratosphere to that of MERRA-2 (Figs. A.6 and 3.2a). A similar pattern of ensemble spread appears in the 4 December 2017 initialization forecast RI ($\sim$3 weeks before the CAO, Fig. 3.3b), though the FREE experiment exhibits an increase in the value of the forecast RI. By the 11 December 2017 initialization ($\sim$2 weeks before the CAO and 3 days before the RI forecast period), all experiments begin to converge, which is expected as nudging has less impact at shorter lead times to the RI averaging period, due to the ability of the FREE run to better capture the shorter-term variability (Fig. 3.3c).

Examining the forecasts for the two components of the RI separately reveals similar conclusions to that of the total RI (Figs. 3.3d–3.3f for the Siberian component; Figs. 3.3g–3.3i for the Canadian component). Ultimately, across initializations, the NUDGE-S consistently outperforms the other experiments in simulating both components of the RI, with narrower spreads as well. Interestingly, the NUDGE-T exhibits the highest Siberian RI forecast among the experiments in the 27 November 2017 and 4 December 2017 initializations (Figs. 3.3d and 3.3e), indicating a potential strong association between tropical forcing and upward wave fluxes over Siberia . Closer to the event, all experiments converge to the observed values in both components. Thus, given the large experimental differences on the longer subseasonal timescales, the representation of the stratosphere and tropics may have a substantial impact on the subseasonal forecast of reflection-induced cold for this event.

3.4.1.2 Temperature and tropospheric circulation forecast

Next, we investigate the evolution of T2M and 500 hPa GPH anomalies for this CAO event. For these comparisons amongst the model experiments, we choose to investigate the 27 November 2017 initialization run only. To compare the ensemble-mean tropospheric circulation and T2M forecasts, we use the FREE run as a baseline and subtract its forecast values for the fields from MERRA-2 (to see the errors between the FREE run and the observed evolution) and the two nudged experiments forecast values. We focus on forecast fields during the CAO event (26 December 2017 to 6 January 2018). Figure 3.4 shows the differences in full-field T2M (shading, °C) and 500 hPa GPH (contours, dam) for MERRA-2, NUDGE-S, and NUDGE-T relative to the FREE experiment. The FREE run largely underestimates 500 hPa GPH across the North Pacific, Alaska, the southwestern US, and the Arctic Ocean during the first three days of the CAO, where strong ridging is observed (Fig. 3.1), while overestimating GPH over Canada and the northern US, where a deep trough is present (Fig. 3.4a). As a result, the FREE run is not nearly as cold as MERRA-2

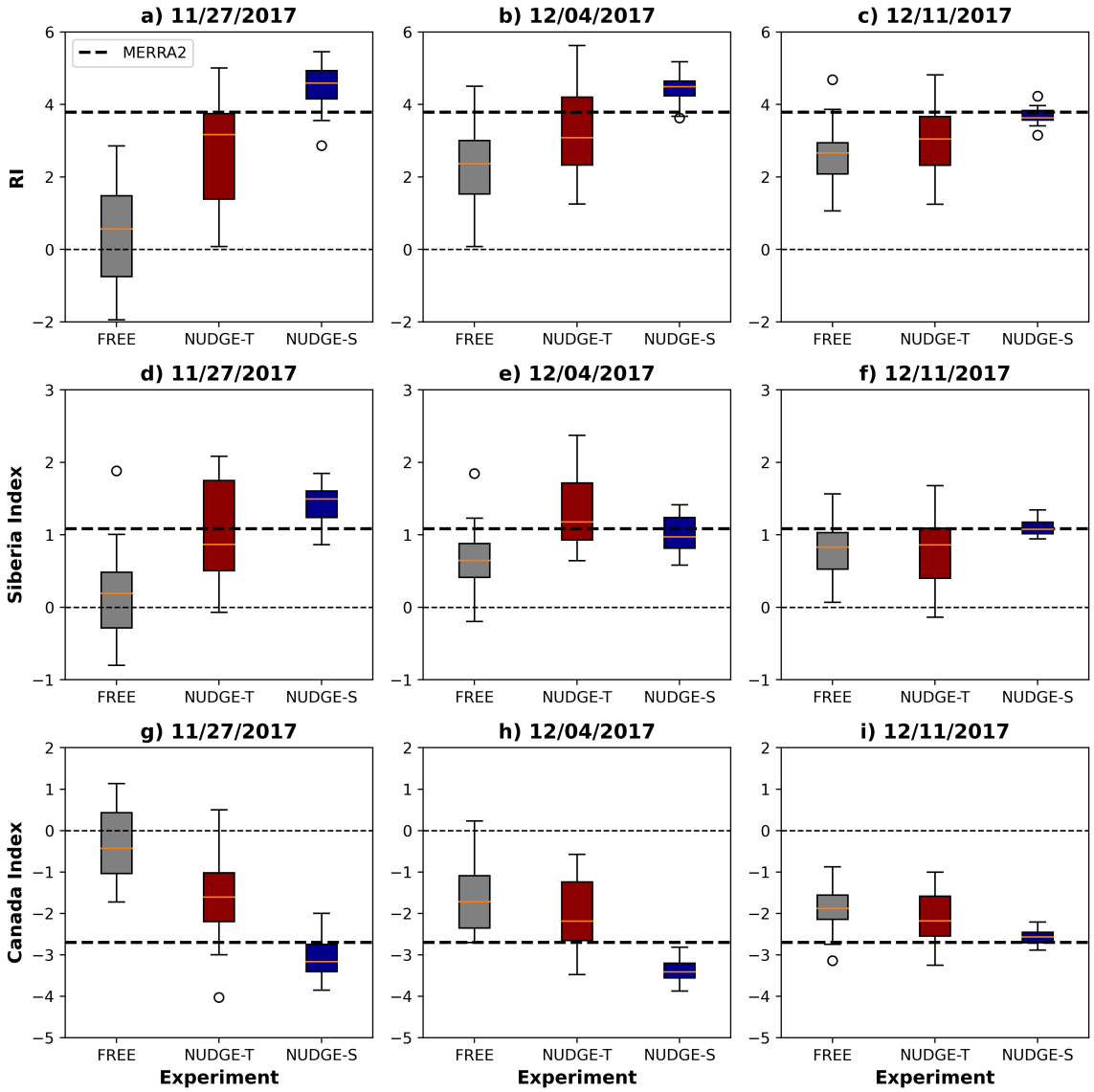


Figure 3.3: Boxplots of a)–c) the RI, d)–f) the Siberian component of the RI, and g)–i) the Canadian component of the RI all averaged between 14–25 December 2017, for subseasonal initializations of 27 November 2017, 4 December 2017, and 11 December 2017, respectively. Gray boxes denote the FREE experiments, red boxes denote the NUDGE-T experiments, and blue boxes denote the NUDGE-S experiments. The orange lines represent the ensemble median, the lower and upper bounds of the box indicate the 25th and 75th percentiles, the whiskers extend from the box to the farthest data point lying within 1.5 times the interquartile range from the box, and the circles are those points that lie outside of the whiskers.

across much of Canada, the eastern US, and the northern US (Fig. 3.4a, negative values in shaded contours). From 29–31 December 2017, similar differences persist, but the FREE run better captures the GPH across the Arctic Ocean (Fig. 3.4b). Throughout 1–6 January

2018, the FREE run fails to accurately simulate the west/east dipole in ridging/troughing, respectively. The difference between MERRA-2 and the FREE run resembles the pattern in Fig. 3.1, with anomalous ridging across the US West Coast and anomalous troughing over eastern North America. This results in lower T2M values in MERRA-2 extending farther south across the central US and, particularly from 4–6 January, the northeastern US (Figs. 3.4c and 3.4d). The underestimation of extremely high or low 500 hPa GPH or T2M values is also more likely on subseasonal timescales versus reanalysis, as S2S models often struggle to maintain the magnitude of extremes at longer lead times and thus explains why often probabilistic forecasts are used (Vitart and Robertson 2018).

Figs. 3.4e–3.4h show the differences in T2M and 500 hPa GPH between the NUDGE-S and FREE experiments. From 26–31 December 2017, the NUDGE-S experiment simulates higher GPH over Alaska and lower GPH over northeast Canada and the northern US compared to the FREE experiment (Figs. 3.4e and 3.4f). Additionally, the NUDGE-S experiment produces colder T2M over northeastern Canada, the eastern US, and the northern US during these periods. Notably, the NUDGE-S run is significantly colder across regions sensitive to extreme cold following a stratospheric reflection event (Messori et al. 2022), including the northeastern US (black and green boxes in Fig. 3.4, respectively). The difference pattern closely resembles that of MERRA-2 (Figs. 3.4a and 3.4b), though of smaller magnitude, suggesting that nudging the stratosphere improves the simulation of tropospheric circulation and near-surface temperature on subseasonal timescales for this CAO event. From 1–6 January 2018, the NUDGE-S experiment continues to enhance ridging over Alaska, extending into the US West Coast, similar to the MERRA-2 difference maps (Figs. 3.4g and 3.4h). The NUDGE-S run also produces lower GPH across the eastern half of the US than the FREE experiment, accompanied by significantly negative T2M differences across Canada and the northeast US, which extend further south from 4–6 January 2018 (Fig. 3.4h). Despite the overall improvement in subseasonal forecasts with stratospheric nudging, the US Great Plains shows little temperature response, suggesting

that this region is less sensitive to improved stratospheric representation with wave reflection, at least for this particular event (purple boxes in Fig. 3.4).

The NUDGE-T forecast consistently exhibits a stronger signal for higher GPH over Alaska than both the FREE and NUDGE-S forecasts, suggesting that improved representation of the tropics in the CESM2 subseasonal model may enhance anomalous ridge development over Alaska and the North Pacific (Figs. 3.4j–3.4m). The stronger ridging in NUDGE-T vs. the FREE run also supports the RI experimental differences in Siberia (Fig. 3.3d), where enhanced ridging likely amplifies the upward heat flux response from that region. However, while the Siberian component of the RI aligns with this explanation, the downward wave response into Canada is weaker than in the NUDGE-S (Fig. 3.3g). This weaker signal is evident in Figs. 3.4j–3.4m; although the ridging response in NUDGE-T is stronger across Alaska than in the other experiments, it fails to capture the magnitude of the anomalous trough over Canada and the northern US. By extension, improvements to T2M forecasts across the central and eastern US remain modest, particularly between 26–28 December 2017 (Fig. 3.4j). In fact, between 1–3 January 2018, the NUDGE-T experiment even produces *higher* T2M than the FREE experiment in the eastern US (Fig. 3.4l).

Fig. 3.5 presents violin plots of ensemble T2M anomalies for three regions: the wave reflection-sensitive area (26 December 2017–2 January 2018), the Great Plains (26 December 2017–2 January 2018), and the northeast US (30 December 2017–6 January 2018). Each panel compares forecasts from the FREE (gray), NUDGE-T (red), and NUDGE-S (blue) experiments. In both the reflection-sensitive area (Fig. 3.5a) and the northeast US (Fig. 3.5c), the NUDGE-S experiment consistently produces the coldest temperature response, with most of the distribution falling below average, especially in the northeast US. While none of the experiments fully capture the magnitude of the MERRA-2 T2M anomalies (black dashed line), the NUDGE-T run is generally the next best performer, with the FREE run performing the worst, particularly at longer lead times (e.g., 27 November and 4 December 2017 initializations). This ranking is especially pronounced in the reflection-sensitive area, underscoring the influence of subseasonal stratospheric representation on

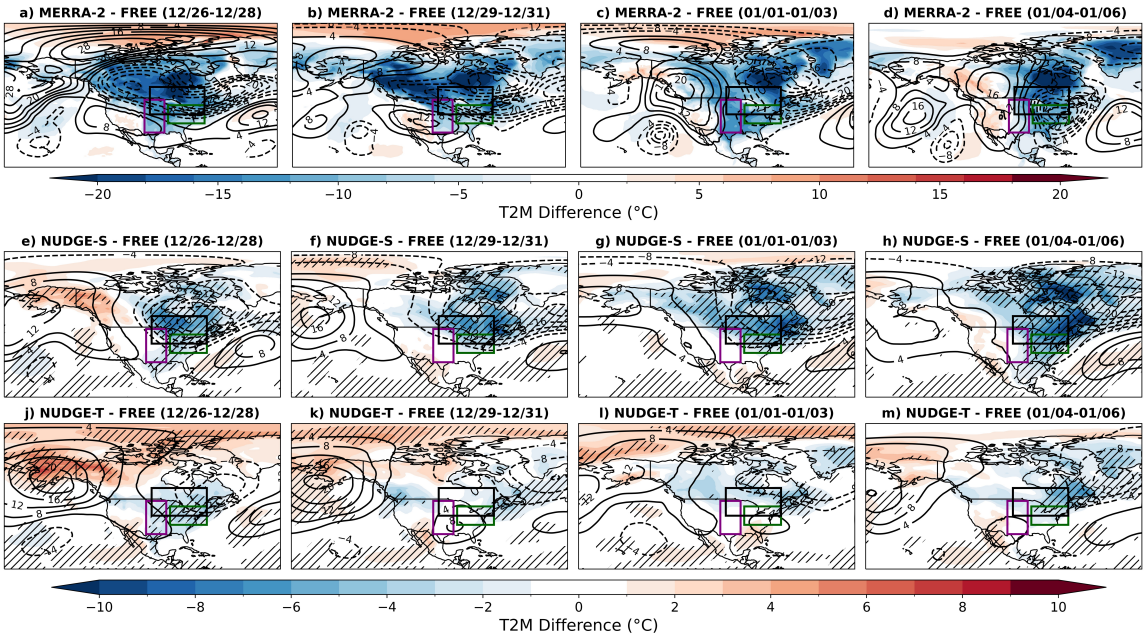


Figure 3.4: The difference in T2M (°C, shading) and 500 hPa GPH (dam, contours) from the ensemble mean FREE run for a)–d) MERRA-2, e)–h) the NUDGE-S ensemble mean, and j)–m) the NUDGE-T ensemble mean, for the 27 November 2017 initialization at 3-day averaged periods from 26–28 December 2017 through 4–6 January 2018, respectively. Hatching shows statistical significance on the T2M difference field at the 95th percentile via 5000-iteration bootstrapping with replacement. The black boxes represent the area that usually has the coldest T2M response to reflection events in Messori et al. (2022) (40-55°N, 260-290°E), the purple box represents the Great Plains (30-48°N, 257-268°E), and the green box represents the northeast US (35-45°N, 270-290°E).

extreme cold temperature forecasts in North America (Fig. 3.5a). Across all experiments, the largest differences occur for initializations on the subseasonal lead times. Closer to the CAO event, even the FREE run gains skill, reducing the impact of the nudging. As noted earlier, T2M anomalies in the Great Plains remain relatively unaffected by experiment type, suggesting that reflection-induced cold events in this region depend on dynamically more than just improving the representation of the stratosphere (Fig. 3.5b).

3.4.1.3 North American winter weather regime forecast

To further examine differences in tropospheric evolution across our experiments, we analyze the ensemble mean evolution of our five North American winter weather regimes. Note that the characteristic patterns of these regimes remain virtually unchanged from those in

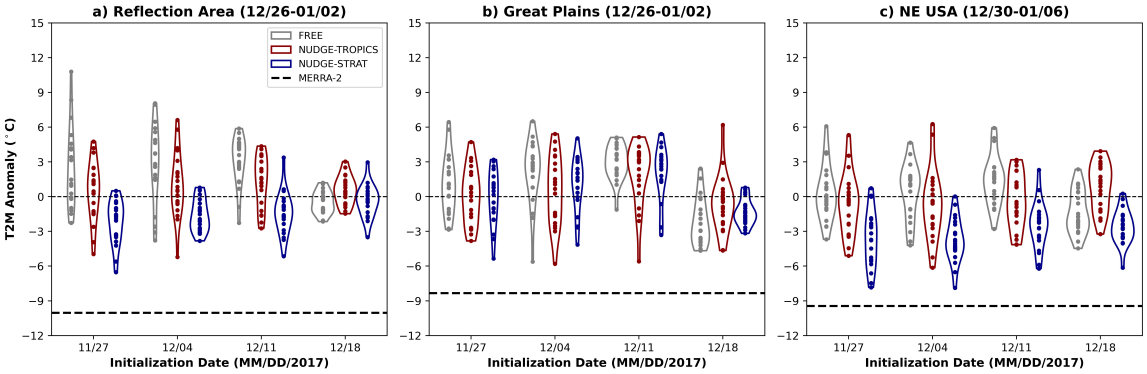


Figure 3.5: Violin plots of the ensemble forecast of area-averaged T2M anomalies (°C) for weekly CESM2 CAM6 initializations from 27 November 2017 to 18 December 2017 for a) the area most prone to cold events due to stratospheric wave reflection (40-55°N, 260-290°E) from Messori et al. (2022) averaged between 26 December 2017 to 2 January 2018, b) the Great Plains (30-48°N, 257-268°E) averaged between 26 December 2017 to 2 January 2018, and c) the northeast US (35-45°N, 270-290°E) averaged between 30 December 2017 to 6 January 2018. The blue, red, and gray distributions represent the NUDGE-S, NUDGE-T, and FREE experiments, respectively. The dashed black lines represents the MERRA-2 values.

Chapter 2, except the order of occupation frequency is ArH, AkR, WCR, PT, and ArL, from least to most frequent. Figure 3.6 illustrates the ensemble mean evolution of the WRI for these five regimes in MERRA-2 and the three experiments, initialized on 27 November 2017. In MERRA-2, the period begins in a PT regime (Fig. 3.6a, green line) before rapidly transitioning to a dominant WCR regime from approximately 7–17 December 2017 (Fig. 3.6a, blue line), with WRI values increasing to near +3σ. Concurrently, the ArL regime (Fig. 3.6a, purple line) weakens, with its WRI falling to -2σ, reflecting its strong spatial anticorrelation with the WCR pattern (Fig. 2.4). In the two weeks preceding the onset of the central US CAO, the WRIs for the AkR (Fig. 3.6a, orange line) and WCR (Fig. 3.6a, blue line) regimes signals strengthen, with the AkR regime emerging as the dominant regime at CAO onset. As the CAO progresses, the AkR regime weakens, and the WCR regime once again becomes the dominant pattern, consistent with Fig. 3.1.

The FREE experiment does not accurately simulate the evolution of winter weather regimes compared to MERRA-2 (Fig. 3.6b). Throughout most of the forecast, the dominant regime remains the ArH regime (WRI values of +1 to +2σ; Fig. 3.6b, red line),

followed closely by the WCR (Fig. 3.6b, blue line) and PT (Fig. 3.6b, green line) regimes. However, the FREE run fails to establish the WCR regime as one of the dominant regimes during the CAO period as seen in MERRA-2 (Fig. 3.6a). By contrast, the NUDGE-S experiment shows a substantial improvement in the WRI forecast, closely resembling the MERRA-2 evolution (Fig. 3.6c). Although it does not fully capture the reduction and subsequent rebound of the WCR regime before the CAO, it maintains the WCR regime as the strongest regime throughout the event. These results further reinforce our earlier findings; i.e., nudging the stratosphere toward the observed state on subseasonal timescales improves the representation of North American tropospheric circulation, leading to a more accurate CAO forecast. Finally, the NUDGE-T ensemble mean WRI evolution also shows an increase in the WCR WRI and a decrease in the ArL WRI between 7–27 December 2017 (Fig. 3.6d). However, it maintains higher AkR and ArH WRI values compared to the other experiments until the CAO begins, with the AkR regime becoming the dominant signal around 29 December 2017 (Fig. 3.6d). This forecast evolution aligns with prior research illustrating that active MJO convection in phases 6–8 enhances ridging (and blocking) across the North Pacific. Indeed, ensemble mean maps of the 200 hPa streamfunction anomalies leading up to, and including, the CAO onset show that the NUDGE-T simulation features a stronger wave pattern originating from the central tropical Pacific, aligning with strong ridging in the 500 hPa GPH anomaly field across Alaska and northwest Canada (Figures A.7 and A.8).

To further assess the impact of nudging on WRI evolution in the forecast, we evaluate how individual ensemble members simulate the WRIs on subseasonal timescales by displaying the correlation between MERRA-2 and ensemble forecast WRI between 12 December 2017 to 6 January 2018 (Figure 3.7). The free experiment exhibits inconsistent WRI performance, with frequent negative correlations relative to MERRA-2, except for modest correlations in the AkR regime (Fig. 3.7a). In contrast, the NUDGE-S shows improved WRI skill, particularly for the WCR regime, which dominated during the CAO (Figs. 3.7b and 3.6). Correlations in NUDGE-S reach as high as 0.87 for the WCR, with

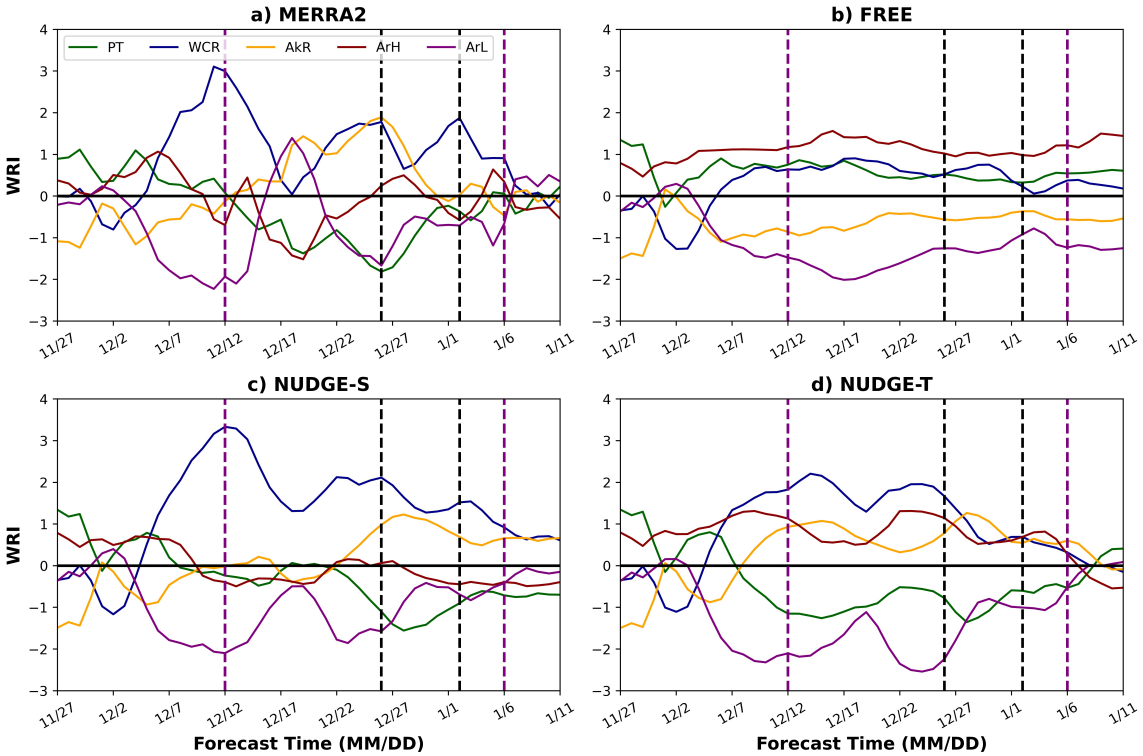


Figure 3.6: The evolution of the ensemble mean WRI for the five North American winter weather regimes for the 27 November 2017 CESM2 CAM6 initialization in a) the MERRA-2 evolution, b) the FREE experiment, c) the NUDGE-S experiment, and d) the NUDGE-T experiment. The green line represents the Pacific Trough (PT), the blue line represents the West Coast Ridge (WCR), the orange line represents the Alaskan Ridge (AkR), the red line represents the Arctic High (ArH), and the purple line represents the Arctic Low (ArL). The black dashed lines show the start and end dates of the 2017/2018 CAO. The purple dashed lines represent the start and end of the period used for correlation analysis in Fig. 3.7.

similarly strong performance for the ArL regime due to their inverse relationship. NUDGE-T also shows improved correlations compared to the FREE, but to a lesser extent than NUDGE-S (Fig. 3.7c). The results from Fig. 3.7, combined with that from the prior figures, suggest that stratospheric nudging substantially enhances the model's ability to simulate stratospheric wave reflection and its downstream impact over North America. While tropical nudging contributes, it alone cannot reproduce the full surface response. Optimal forecasting of stratosphere-driven cold events like this CAO likely requires accurate representation of both stratospheric and tropical circulations, with stratospheric dynamics playing the dominant role in this case. To further assess the impacts of tropical and stratospheric circulations on CAO forecasts in the CESM2 CAM6 subseasonal predictions, we

now extend our analysis to a case study of the January 2004 CAO, which occurred following a major SSW and in the midst of active tropical convection.

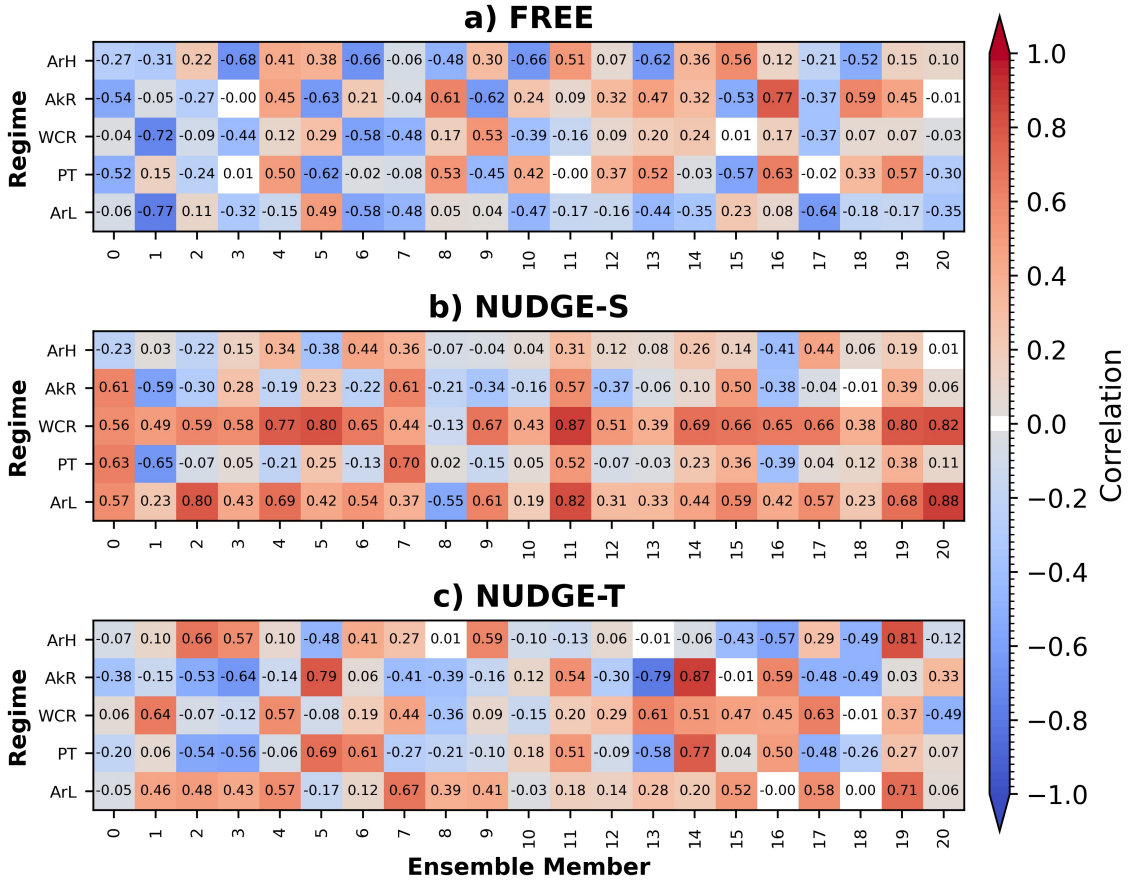


Figure 3.7: The correlation between the MERRA-2 WRIs and the CESM2 CAM6 forecast WRI initialized on 27 November 2017 for the lead time extending from 14 days prior to the CAO to 6 January 2018 (i.e., end of averaging period for NE cold), as a function of ensemble member.

3.4.2 January 2004 CAO analysis

3.4.2.1 Polar cap zonal wind forecasts

Since the January 2004 CAO occurred following a major SSW on 5 January 2004, Figure 3.8 presents the ensemble forecast evolution of the polar cap averaged zonal wind at 10 hPa (U10; Figs. 3.8a-c), 50 hPa (U50; Figs. 3.8d-f), and 100 hPa (U100; Figs. 3.8g-i) for the target period 12-26 January 2004. For U10, the NUDGE-S experiment maintains

a narrow distribution centered around the observed MERRA-2 values (∼-1 m/s, dashed line) across all three initializations, consistent with the experimental design (Figs. 3.8a–3.8c). However, for the 29 December 2003 initialization, the NUDGE-T and FREE experiments exhibit greater uncertainty in the forecast of the winds prior to the SSW, with the NUDGE-T experiment showing the highest variability (Figs. 3.8a,d,g, red). On the 5 January 2004 initialization (the central date of the major SSW), the FREE and NUDGE-T ensembles predominantly predict weaker-than-observed zonal winds in the two weeks preceding the CAO, whereas the NUDGE-S experiment remains constrained to the observed evolution (Figs. 3.8b,e,h). These differences suggest that, without a nudged extratropical stratosphere, the FREE and NUDGE-T experiments tend to push the upper stratospheric polar zonal wind toward substantially stronger easterlies, highlighting the uncertainty in subseasonal stratospheric evolution. By the 12 January 2004 initialization (Figs. 3.8c,f,i), both the NUDGE-T and FREE experiments converge toward the observed value, displaying similar ensemble distributions as forecast lead time decreases and skill improves.

3.4.2.2 Temperature and tropospheric circulation forecast

Given the larger ensemble spread in the forecasts for U10, U50, and U100 in the NUDGE-T and FREE experiments for the 29 December initialization and that this initialization is prior to the major SSW, we use this initialization to explore differences in the tropospheric evolution across the experiments when forecasting the January 2004 CAO event. Figure 3.9 is similar to Fig. 3.4 but using the 29 December 2003 initialization to evaluate the forecasts for 27 January to 7 February 2004. During the first three days of the CAO event (27-29 January 2004), the FREE run substantially underestimates the magnitude of the GPH across Greenland, northern Canada, and the Arctic Ocean compared to MERRA-2, which are all areas of anomalous ridging in the true evolution (compare Figs. 3.1f and 3.9a). Additionally, the FREE run does not capture the magnitude of the troughing south of Greenland/northern Canada and near the Aleutian Islands, leading to a warmer bias across

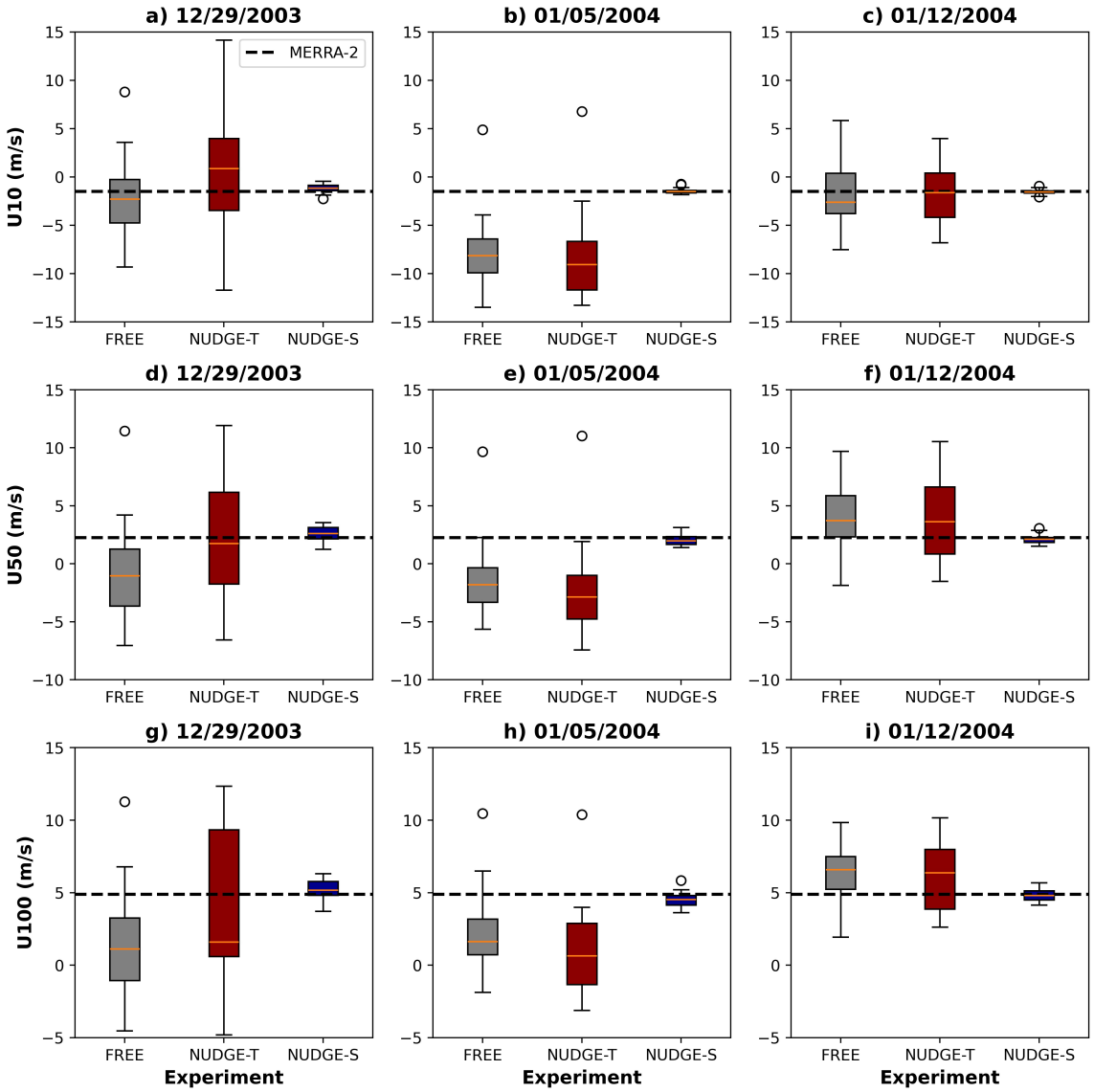


Figure 3.8: Boxplots of the polar cap averaged (60-90°N) zonal wind at a)–c) 10 hPa, d)–f) 50 hPa, and g)–i) 100 hPa all averaged between 12–26 January 2004, for subseasonal initializations of 29 December 2003, 5 January 2004, and 12 January 2004, respectively. Gray boxes denote the FREE experiments, red boxes denote the NUDGE-T experiments, and blue boxes denote the NUDGE-S experiments. The orange lines represent the ensemble median, the lower and upper bounds of the box indicate the 25th and 75th percentiles, the whiskers extend from the box to the farthest data point lying within 1.5 times the interquartile range from the box, and the circles are those points that lie outside of the whiskers.

western Canada (black box, Fig. 3.9), the northern US, and the Great Plains (purple box, Fig. 3.9). Between 30 January and 1 February, many of the same errors persist, (Fig. 3.9b), resulting in an overall warm bias across Canada and the central/eastern US compared to

MERRA-2. Although the CAO has finished by 2 February 2004, the FREE run continues to underestimate the negative T2M anomalies in the central US (Figs. 3.9c and 3.9d). Additionally, the FREE experiment struggles to simulate both the anomalous ridging over northeast Canada and the Rossby wave train that develops across the US, which plays a key role in sustaining more modest cold across the central US after the end of the CAO by making the GPH pattern less progressive.

The NUDGE-S experiment features higher GPH than the FREE run across Greenland between 27 January 2004 and 1 February 2004, highlighting the ability of the model to simulate ridging in the region that often experiences anomalous ridging after stratospheric weakenings (Figs. 3.9e and 3.9f, Millin et al. 2022). Over the same period, the NUDGE-S experiment also better represents the lower GPH across the East Coast of the US than the FREE run. Interestingly, the NUDGE-S experiment shows significantly lower T2M across much of the eastern half of Canada, associated with lower GPH there than in the FREE run. However, the NUDGE-S experiment produces significantly higher T2M than the FREE run across western Canada (black box) and the Great Plains (purple box) (Figs. 3.1g,h and 3.9e,f). Between 2–7 February 2004, higher T2M than the FREE run persists across western Canada and the Great Plains in the NUDGE-S experiment, with the negative T2M departures weakening across much of the rest of Canada, and some modest negative T2M departures from the FREE run in the northeast US (Figs. 3.9g and 3.9h). Thus, by nudging the stratosphere, the model is gaining some benefit across the eastern half of Canada, but does not simulate well the lower T2M in western Canada and the central US later.

By contrast, the NUDGE-T experiment consistently produces lower T2M than the FREE run across much of western and central Canada and into the northern US throughout the entire period (Figs. 3.9j–3.9m). The difference is most notable across western Canada (black box) throughout 27 January 2004 to 1 February 2004, where the NUDGE-T experiment produces significantly lower T2M across this key region where the extremely negative T2M anomalies were in the observations (Fig. 3.1). Indeed, the NUDGE-T run consistently exhibits higher GPH than the FREE run to the west of Alaska and just off

the US West Coast (Figs. 3.9j–3.9m). This area of higher GPH aligns with that shown in the difference between MERRA-2 and the FREE run, where anomalous ridging develops in the observed evolution (Fig. 3.1). The simulation of more ridging in the North Pacific likely makes the overall tropospheric pattern less progressive and leads to the better simulation of lower T2M across Canada and the northern US, despite the relatively weaker simulation of strong ridging across Greenland and northern Canada versus the NUDGE-S experiment. Regardless, the simulation of lower T2M in the Great Plains and northeast US is weak throughout, although not with positive T2M departure as in the NUDGE-S experiment (Figs. 3.9j–3.9m). Thus, the NUDGE-T experiment tends to perform better for this case study for Canada and the northern US, despite the NUDGE-S simulation showing the most accurate polar cap zonal wind evolution in the stratosphere following the major SSW (Fig. A.9).

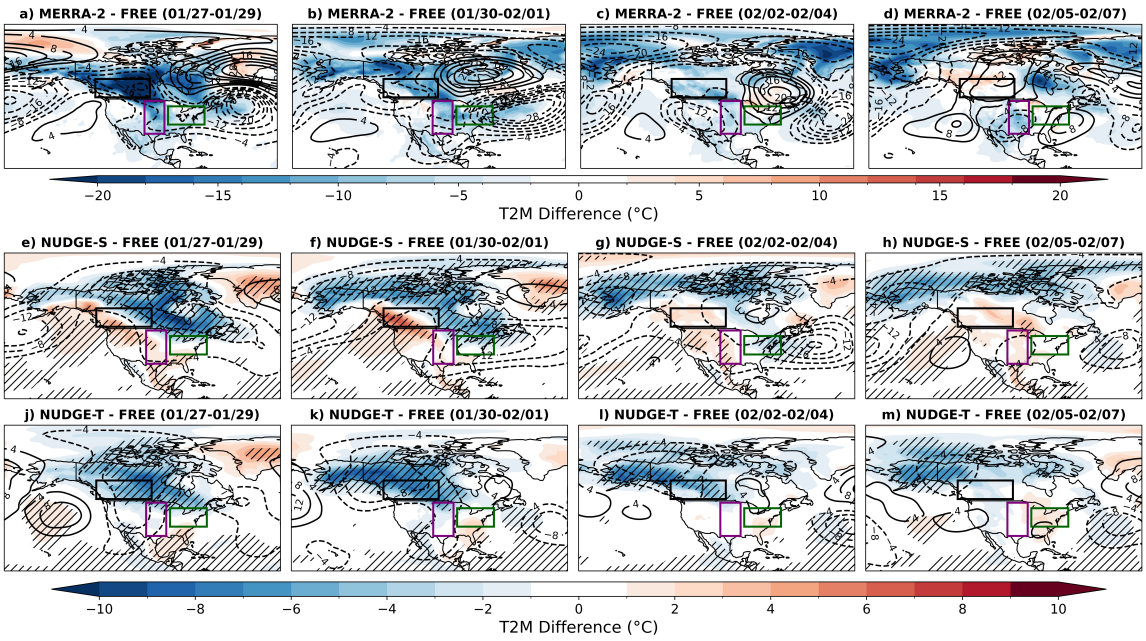


Figure 3.9: The difference in T2M (°C, shading) and 500 hPa GPH (dam, contours) from the ensemble mean FREE run for a)–d) MERRA-2, e)–h) the NUDGE-S ensemble mean, and j)–m) the NUDGE-T ensemble mean, for the 29 December 2003 initialization at 3-day averaged periods from 27 January–29 January 2004 through 5 February–7 February 2004, respectively. Hatching shows statistical significance at the 95th percentile on the difference in the T2M via 5000-iteration bootstrapping with replacement.

Figure 3.10 shows the violin plots of ensemble T2M anomalies averaged between 27–31 January 2004. For the 29 December 2003 initialization in western Canada, the NUDGE-T ensemble distribution of T2M anomalies shows a clear shift toward more negative anomalies when compared to the NUDGE-S and FREE run, with much of the NUDGE-S experiment ensemble above 3°C (Fig. 3.10a), aligning with Fig. 3.9. Furthermore, the NUDGE-T experiment has some of the ensemble members with the most negative T2M anomalies in the Great Plains, with the NUDGE-S again showing the warmest ensemble forecast. In the northeast US, there is very little difference between experiments during the CAO (Fig. 3.10c). Similar results hold for the 5 January and 12 January 2004 initializations across all regions, before the experiments converge toward more negative T2M anomalies on the 19 January 2004 initialization (Fig. 3.10). Overall, the difference between experiments for this event is modest in these three regions versus the 2017/18 CAO event, highlighting the complex relationship between SSWs, tropical variability, and CAOs in North America.

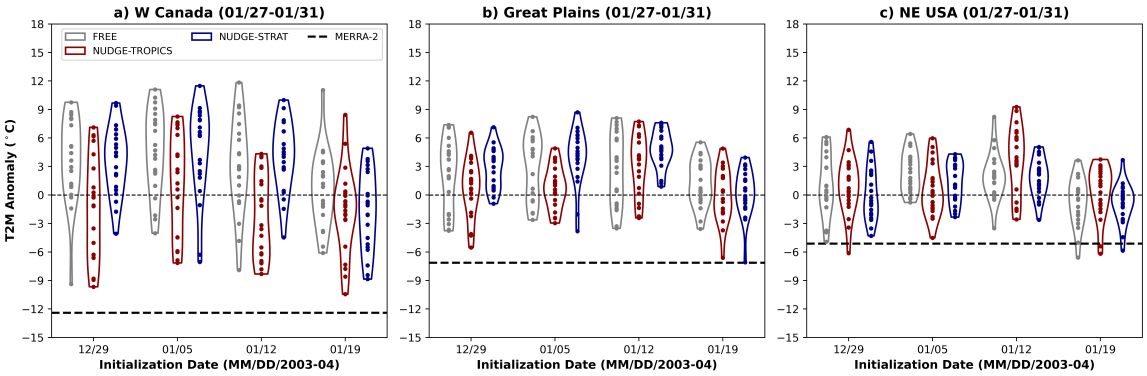


Figure 3.10: Violin plots of the ensemble forecast of area-averaged T2M anomalies (°C) between 27 January 2004 to 31 January 2004 for weekly CESM2 CAM6 initializations from 29 December 2003 to 19 January 2004 for a) the northwest of the US (50-60°N, 230-260°E), b) the Great Plains (30-48°N, 257-268°E), and c) the northeast US (35-45°N, 270-290°E). The blue, red, and gray distributions represent the NUDGE-S, NUDGE-T, and FREE experiments, respectively. The dashed black line represents the MERRA-2 values.

3.4.2.3 North American winter weather regime forecast

Figure 3.11 shows the North American winter WRIs for the three experiments and in MERRA-2 for the 29 December 2003 initialization. The MERRA-2 evolution of the WRI

shows a dominant AkR regime during the first ten days of the forecast before transitioning to a WCR between 6–13 January 2004. The North American weather pattern then transitions to a dominant ArH regime around 18 January 2004 (Fig. 3.11a). During the first two weeks of the forecast, all experiments produce a WRI evolution very similar to reanalysis, with a dominant AkR regime at the beginning of January (Figs. 3.11b–3.11d). After the first two weeks, the FREE run maintains a dominant ArH regime (Fig. 3.11b, red line), while the NUDGE-S experiment better captures the emergence of the WCR regime before the ArH regime emerges in the days leading to the CAO and during the CAO (Fig. 3.11c). The NUDGE-T experiment also captures the initial emergence of the WCR regime around 13 January 2004 (Fig. 3.11d), then the ArH emerges as the dominant regime 7 to 10 days prior to the CAO. Therefore, even though each experiment was unable to capture the extreme magnitude of the observed 500 hPa GPH and T2M on longer subseasonal timescales, the model forecasts somewhat accurately capture the regime evolution for this CAO event. Interestingly, the NUDGE-T experiment is the only experiment that captures the increased AkR WRI before the CAO start, likely corresponding to the higher GPH across the west of Alaska and the North Pacific relative to the FREE run (Figs. 3.9j–3.9m).

To further investigate the behaviour of the weather regimes, Figure 3.12 presents the correlation between WRIs from the forecasts and MERRA-2 during the two weeks preceding the 2004 CAO through the end of the event. Overall, the FREE experiment shows mixed performance in reproducing WRI correlations with MERRA-2 across ensemble members (Fig. 3.12a). The ArL and WCR are reasonably well represented in the FREE run, exhibiting consistently positive correlations across ensemble members, reaching values as high as 0.91 and 0.86 for ArL, and 0.76 and 0.75 for WCR. In contrast, other regimes, particularly AkR and ArH, display weaker or inconsistent correlations.

A key outcome of this analysis is the marked improvement in WRI forecast correlation when the stratosphere is nudged, as evidenced by the generally positive correlations during the CAO period in the NUDGE-S experiment (Fig. 3.12b). This finding shows the benefit of accurately representing the stratosphere and its coupling with the troposphere post SSW,

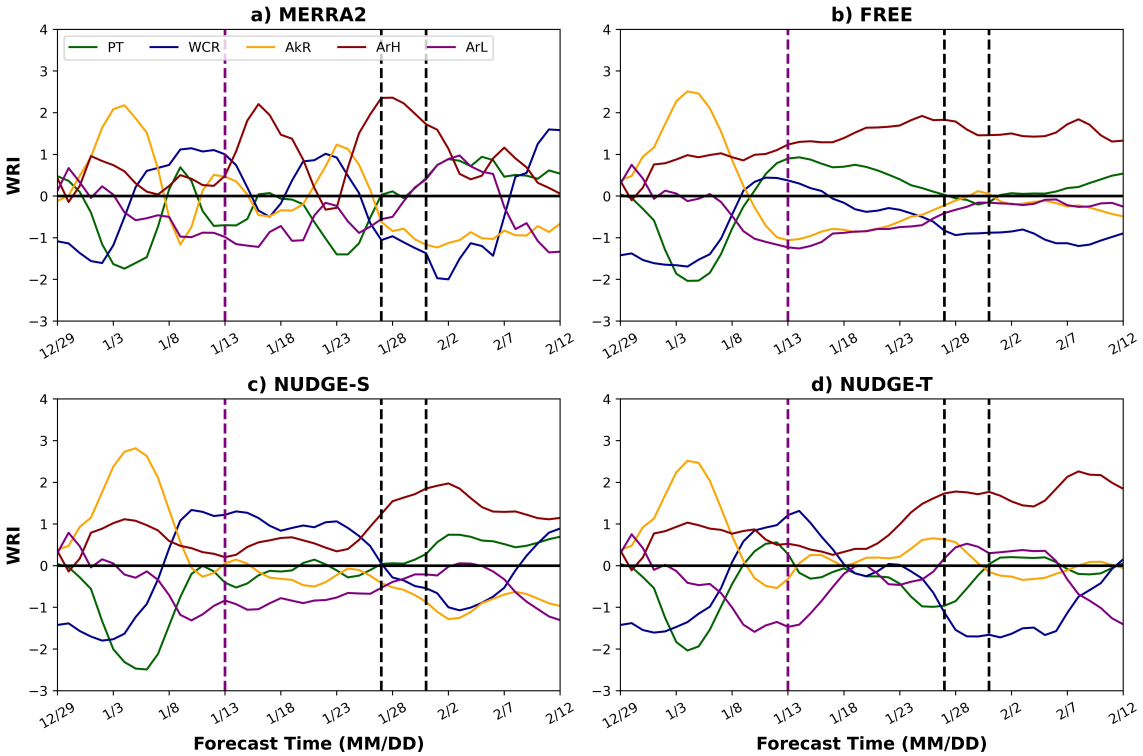


Figure 3.11: The evolution of the ensemble mean weather regime index for the five North American winter weather regimes for the 29 December 2003 CESM2 CAM6 initialization in a) the MERRA-2 evolution, b) the FREE experiment, c) the NUDGE-S experiment, and d) the NUDGE-T experiment. The green line represents the Pacific Trough (PT), the blue line represents the West Coast Ridge (WCR), the orange line represents the Alaskan Ridge (AkR), the red line represents the Arctic High (ArH), and the purple line represents the Arctic Low (ArL). The black dashed lines show the start and end dates of the 2004 CAO during the initialization. The purple dashed line represents the start of the period used for correlation analysis in Fig. 3.12.

particularly for regimes commonly associated with CAOs in the central US (i.e., Akr, ArH, and WCR). The NUDGE-T experiment further supports this result, showing the highest ensemble member correlations for the WCR and ArL regimes and predominantly positive correlations for ArH and AkR (Fig. 3.12c). As such, the nudging of the tropical circulation during active periods of the MJO may additionally help the representation of surface weather regimes post-SSW, especially those with the Pacific blocking highs (e.g, the AkR and WCR regimes).

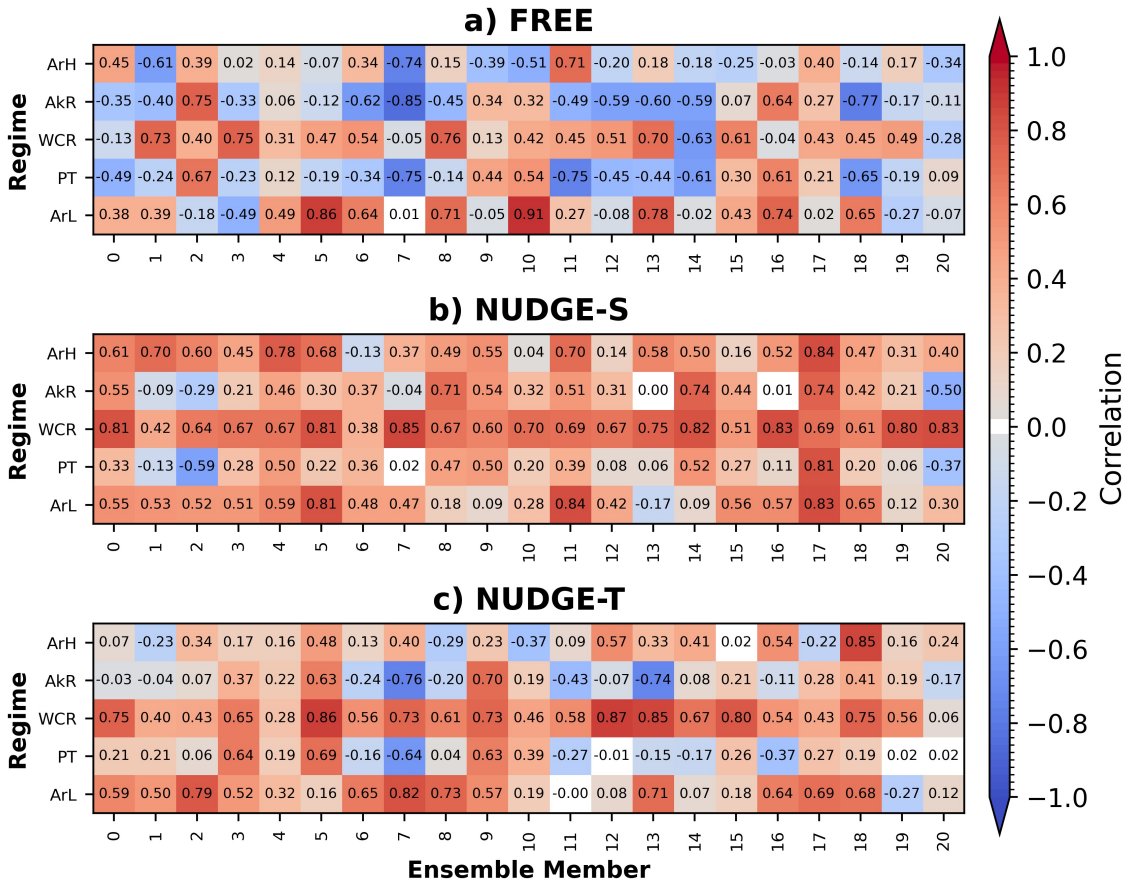


Figure 3.12: The correlation between the MERRA-2 WRIs and the CESM2 CAM6 forecast WRI initialized on 29 December 2003 for the lead time extending from 14 days prior to the CAO to 31 January 2004 (i.e., end of the CAO), as a function of ensemble member.

3.5 Discussion

Wintertime CAOs in the central US are extreme events that have significant socioeconomic, environmental, and infrastructural impacts that often reach into other highly-populated regions, such as the eastern US. Past works have established significant statistical relationships between modes of climate variability and North American CAOs (e.g., Matthias and Kretschmer 2020; Lee et al. 2019b; Millin et al. 2022). This chapter borrows from parallel studies on European CAO events (e.g., Kautz et al. 2019; Huang et al. 2022) and addresses the gap in the literature on the mechanisms of how these features may impact the subseasonal prediction of North American CAOs. We employ nudging experiments in the CESM2 CAM6 subseasonal model simulations of two dynamically different central

US CAO events to compare and contrast the contributions of tropical versus stratospheric circulation variability on the subseasonal forecasts of these CAOs.

In the case of the December 2017/January 2018 CAO, the NUDGE-S experiment results in robust forecasts of cold temperatures throughout Canada and the eastern US on subseasonal timescales (Figs. 3.4 and 3.5). Additionally, the NUDGE-S experiment also improves the forecasts of the North American winter weather regimes (proxies for sensible weather patterns) in the two weeks leading up to and including the CAO period weeks ahead (Figs. 3.6 and 3.7). Thus, better representations of the stratosphere, including the wave reflection mechanism, results in an improved subseasonal forecast for the 2017/2018 CAO versus the FREE run. Despite this improvement, the T2M response in the central US was relatively modest, highlighting that other processes may have played a role in the cold development there. The NUDGE-T experiment also improved the simulation of the wave reflection during December 2017, particularly the increase in upward Siberian flux (Fig. 3.3). This experiment also corresponded to a better representation of the Alaskan ridge and a modest improvement in the T2M forecast across the eastern US (Figs. 3.4 and 3.5). Indeed, this finding agrees with prior work showing that active MJO convection in the western and central tropical Pacific is associated with increased ridging near Alaska (Henderson et al. 2016; Henderson and Maloney 2018; Millin et al. 2022). As such, there may be importance to both tropical and stratospheric processes in the evolution of this CAO event

For the January 2004 CAO event, we found that both the NUDGE-S and NUDGE-T produced little improvement in T2M forecasts across the US when compared to the FREE run (Figs. 3.9 and 3.10). In fact, the T2M response in the Great Plains in the NUDGE-S experiment was *warmer* than the FREE run during the CAO period. This finding is somewhat contradictory to canonical relationships between major SSW events and mid-latitude CAOs, though the connection between SSWs to North American cold is likely less robust than that for Europe (Domeisen et al. 2019a; Davis et al. 2022; Millin et al. 2024). Despite less improvement in the T2M response than in the December 2017/2018 nudged runs, the

NUDGE-S and NUDGE-T experiments improved forecasts of the North American winter weather regimes compared to the FREE run (Fig. 3.11 and 3.12). One possible explanation for this apparent contradiction in skill between temperature and weather regime forecasts is timing of CAO events. CAOs have large-scale forcing elements but their timing and magnitude rely on smaller scale features including the synoptic timing of troughs, thermodynamics, tropopause polar vortices, and synoptic-scale wave breaks (Cavallo and Hakim 2009, 2010, 2013; Millin and Furtado 2022). Furthermore, not every major SSW propagates downward to yield surface impacts (e.g., Karpechko et al. 2017). The January 2019 (e.g., Rao et al. 2019) and January 2021 (e.g., Davis et al. 2022) major SSWs are two recent examples of events not directly connected to North American winter weather. Ultimately, the subseasonal prediction problem linking major SSW activity to CAOs in North America may be two-pronged. Major SSWs may increase the probability of CAOs and thus, from a probabilistic perspective, there is increased forecast confidence for such cold events weeks post-SSW (Baldwin and Dunkerton 2001; Domeisen et al. 2019a; Furtado et al. 2021). However, *timing* of the CAO event and its magnitude likely lies outside the subseasonal window and is a deterministic forecasting problem.

3.6 Conclusions

In this chapter, we conducted targeted nudging experiments with the CESM2 CAM6 subseasonal forecasts for two central US CAO events from the database in Chapter 2 (Millin et al. 2022). The December 2017/January 2018 CAO featured robust stratospheric wave reflection, was preceded by an active MJO in Phases 6–8, and evolved during AkR and WCR weather regimes (Figs. 3.1 and 3.2). The January 2004 CAO was dynamically different and was preceded by both a major SSW (Fig. 3.2d), active MJO conditions in phases 6–8 (Fig. 3.2f), and high-latitude blocking (Fig. 3.1f–3.1j). The nudging experiments involved separately nudging the stratospheric circulation and the tropical circulation to evaluate the

relative contribution of each to the Week 3-5 forecasts of these two central US CAO events. The main findings were as follows:

- For the December 2017/January 2018 CAO event, nudging the stratosphere produced a significantly more robust negative T2M anomaly response across Canada and the northeastern US during the event, in response to a stronger downward wave response from stratospheric wave reflection and a better representation of tropospheric weather regimes compared to the control experiment.

- Nudging the tropical circulation for the December 2017/January 2018 CAO resulted in some smaller, but significant, cold T2M responses across the eastern US. However, the largest improvement in the forecast was the amplification of the anomalous ridging across Alaska compared to the control experiment.

- The impact of both stratospheric and tropical nudging on the January 2004 CAO was modest compared to the control run, with both experiments preferring to load colder T2M into Canada rather than the US, despite an overall improvement of the North American winter weather regime forecast post-SSW.

A caveat of this work is that we only studied two events, limiting the generalization of our findings on the contributions of the stratosphere and tropics on the evolution of North American CAOs. As such, we recommend reproducing our experiment design for more CAO events. An additional caveat is that we selected events only from a central US CAO database, whereas events from other domains within the US may also be fruitful. Additionally, we provided nudged experiments for the stratosphere and tropics separately to isolate their impacts, but the nudging of *both* the stratosphere and tropics together would be interesting to consider. Furthermore, we did not examine which variables were most strongly nudged or the geographic regions where nudging was most pronounced. Future work should address these questions to better identify areas of poor model performance or determine whether the standard forecast was already adequate.

Future work should extend this analysis to look at the impact of the stratosphere and tropics on blocking frequency in subseasonal forecasts, which are hallmark features for many extreme weather events during the cold season such as extreme precipitation and even winter warm spells (Grace et al. 2024). Further understanding the stratospheric and tropical circulations will be crucial to S2S predictability of surface extremes and their potential impacts. One forecast metric may be the winter weather regimes themselves, as their large-scale nature may make them "easier" targets for subseasonal forecasts, both in dynamical forecast systems and data-driven models. Indeed, future work should also consider employing nudging on the full reforecast suite to assess how the representation of these features affects the overall skill of the model during the winter season.

Chapter 4

Predictability of Central United States Cold Spells in Subseasonal Forecast Models

4.1 Introduction

We have shown that several key modes of variability influence the formation and predictability of CAOs in the central US: the SPV, the MJO, and North American winter weather regimes. However, these analyses are from either reanalysis or individual event subseasonal forecasts. We now quantify the predictability of weekly cold spells in various S2S model reforecasts, while tying the results to SPV variability and North American weather regimes.

A way of enhancing predictability of such cold events can be through making use of windows of forecast opportunity, which are periods where the skill of an S2S forecast is temporarily higher than normal, usually due to favorable conditions within features of the climate system (Mariotti et al. 2020). One such mode of variability is the SPV, where prior work has shown that the downward propagation of anomalously weak zonal winds (including both major and minor SSWs) into the troposphere can enable the development of cold across Eurasia (Baldwin and Dunkerton 2001; Wang and Chen 2010; Butler et al. 2017; Lee et al. 2019b; Afargan-Gerstman and Domeisen 2020; Liu et al. 2022; Millin et al. 2022), and also in North America as shown in prior chapters. In contrast, strong SPV conditions, developing from a lack of upward and poleward wave activity, are typically associated with anomalous midlatitude warmth, especially in Eurasia (Tripathi et al.

2015). However, we showed that Great Plains CAOs may develop after strong SPV conditions through stratospheric wave reflection associated with anomalous ridging over Alaska (Chapter 2 and Millin et al. 2022). SPV strength, and associations with North American winter weather regimes, may be important precursors for successful cold predictions in the central US.

The strength of the SPV may also modulate North American weather regime transitions. Lee et al. (2022) demonstrated that under weak SPV conditions, the dominant transition is nearly always toward the ArH. However, under strong vortex conditions the regime transition is dominantly toward the PT or ArL regimes (Figs. 2.4c and 2.4e). Thus, regimes and the SPV have individual and collective potential impacts on subseasonal predictability of central US cold. Overall, while important precursors to central US anomalous-to-extreme cold have been identified, the quantification of skill and predictability features for central US cold spells in models is still unknown. In this chapter, we quantify the predictability of CAOs and subseasonal cold spells in the central US using various skill metrics in S2S operational forecast models, and assess the weather regime and stratospheric characteristics associated with accurate and inaccurate forecasts. We predominantly follow the methods of Erner and Karpechko (2024), who conducted a similar study but for Eurasian cold spells. Their findings showed that confident cold spell predictions for northern Eurasia have anomalous cold already in the initial conditions in conjunction with either a negative NAO or Scandinavian block. Additionally, they find the S2S models overpredict the frequency of cold spells, with false alarms occurring more often in a weak SPV. We produce some of the same analyses for our central US cold spells to explore the SPV-regime relationships in S2S forecasts, but also quantify the prediction skill of the subseasonal forecast systems.

4.2 Data and Methods

4.2.1 Verification data and cold spells

This work uses daily-averaged November through April (NDJFMA) ERA5 reanalysis from 1 November 1993 to 30 April 2021 for our verification dataset (Hersbach et al. 2020). This reanalysis product is retrieved on a 0.5° horizontal resolution with 37 vertical levels from 1000 hPa to 1 hPa. We obtain both T2M and GPH for our analyses. Anomalies in ERA5 are calculated with respect to the 1993–2020 climatology (year represents season start), where the seasonal cycle is removed from each day. This climatological window is chosen because it matches the period in which model forecasts are obtained.

To define *cold spells*, we follow the method of Erner and Karpechko (2024) but for the central US. We begin with area-averaging the ERA5 T2M anomalies across the central US and compute weekly averages by employing a forward-looking overlapping rolling mean starting on 1 November in each season. For example, the weekly average for week commencing 1 November 1993 will be the average of 1–7 November 1993, and the weekly average for week commencing 2 November 1993 will be the average of 2–8 November 1993, and so forth. Cold spells are defined when the weekly-averaged T2M anomalies across the central US fall below the lower tercile (33rd percentile). Finally, for verification scoring, we assign binary values of 1 for a cold week and 0 for a non-cold week to the weekly-averaged T2M anomalies.

4.2.2 S2S prediction systems and model cold weeks

In this study, we use reforecasts from six S2S prediction systems in the S2S Project Database (Table 4.1, Vitart et al. 2017). We have chosen these specific models for several reasons:

1. They provide a minimum of 19 seasons of reforecasts to allow for a robust statistical analysis.

2. They have 7 or more ensemble members to allow for more representative cold spell probability calculations.

3. Each model is initialized on the same calendar day each year, which allows for a robust climatology calculation for each model.

The models chosen for this study are ECMWF, the United Kingdom Meteorological Office (UKMO), the Korean Meteorological Agency (KMA), the Bureau of Meteorology (BOM), the National Research Council of Italy – Institute of Atmospheric Sciences and Climate (CNR-ISAC), and the Hydro-Meteorological Centre of Russia (HMCR). We select the reforecasts that are initialized in November through March only, between 1993–2020. Each model's reforecast span may differ within this time range (Table 4.1). Collectively, these models yield a total of 3454 reforecasts for analysis.

From these S2S models, we retrieve T2M and GPH fields. Note that in the S2S models, the T2M data is daily averaged and the GPH data is taken at 00 UTC once per day. Furthermore, the S2S model data is retrieved at a 1.5° horizontal resolution with 10 vertical levels from 10 hPa to 1000 hPa. Since the BOM model is provided in its native grid, we bilinearly interpolate the data to the same 1.5° horizontal resolution as the other models. Anomalies are calculated with respect to a lead- and initialization-dependent model climatology with a leave-one-out method. That is, the climatology for each reforecast initialization date is the mean of all reforecasts across year and ensemble member space in each model's reforecast period that are initialized on the same calendar day except for the reforecast on that particular date.

To calculate model forecast cold weeks/spells, we firstly calculate the area-averaged T2M anomalies over the central US before weekly averaging the forecast data. Forecast anomalies are averaged between leads 0–6 days (Week 1), 7–13 days (Week 2), 14–20 days (Week 3), and 21–27 days (Week 4). For each initialization date and week, we compute the 33rd percentile across year and ensemble member space in each model's reforecast period on the same calendar day, except for the reforecast on that particular date. This lead- and

initialization-dependent lower tercile threshold in T2M anomalies is then used to classify each week as a non-cold (0) or cold (1) week. Forecast probabilities of a cold week are then computed by taking the mean of the binary cold week forecast across ensemble members and compared with ERA5. While forecast skill is assessed for Weeks 1–4 initially, we focus most of the analysis on Weeks 3 and 4 as we want to gauge the subseasonal prediction capability and drivers in these models. We will now outline the skill metrics to be used.

Table 4.1: The S2S Project Database models used in the study: (1) European Center for Medium Range Weather Forecasts (ECMWF), (2) United Kingdom Meteorological Office (UKMO), (3) Korean Meteorological Agency (KMA), (4) Bureau of Meteorology (BOM), (5) Italian National Research Council – Institute of Atmospheric Sciences and Climate (CNR-ISAC), and (6) Hydro-Meteorological Centre of Russia (HMCR).

Model	Model Version	Time Range (days)	Time Period Used	Reforecast Size	Runs Per NDJFM	Total Runs Used
ECMWF	CY47R3	0–46	2001–2021	11	43	880
UKMO	GloSea6	0–60	1993–2016	7	20	460
KMA	GloSea6–GC3.2	0–60	1993–2016	7	20	460
BOM	POAMA P24	0–62	1993–2013	33	30	600
CNR-ISAC	GLOBO	0–35	2001–2020	8	30	570
HMCR	RUMS	0–46	1993–2015	11	22	484

4.2.3 Skill assessment and metrics

To assess the skill of the subseasonal prediction systems in forecasting cold weeks, we use several widely-used skill scores (CAWCR 2015), which we compute using Python's *"climpred"* package (Brady and Spring 2021). The first of these is the Heidke Skill Score (HSS), which is a generalized skill score that measures the proportion of correct forecasts after removing those that would be correct due to random chance and is given by the following equation:

$$HSS = \frac{(a+d) - (n_{exp})_{random}}{N - (n_{exp})_{random}} \tag{4.1}$$

where the number of expected correct forecasts by random chance, $(n_{exp})_{random}$, is:

$$(n_{exp})_{random} = \frac{(a+c)(a+b) + (d+c)(d+b)}{N} \tag{4.2}$$

where a, b, c, and d represent the forecast hits, false alarms, misses, and correct rejections (i.e., contingency table components), respectively, and N is the total number of forecasts. Furthermore, we calculate the probability of detection (POD), which describes the proportion of correct forecasts, and the miss rate (MR), which describes the proportion of events not detected by the models:

$$POD = \frac{a}{a+c} \tag{4.3}$$

$$MR = 1 - POD \tag{4.4}$$

The next skill score we use is the false alarm ratio (FAR), which describes the fraction of predicted events that did not occur:

$$FAR = \frac{b}{a+b} \tag{4.5}$$

Table 4.2 shows the range of possible scores and the scores equating to the perfect and "no skill" outcomes. In addition to skill scores, we also generate reliability diagrams for Weeks 3 and 4 cold week forecasts to assess how model forecast probabilities compare to the observed frequencies (Met Office 2025). To get the reliability, we use the *"xskillscore"* Python package, which bins the forecast probabilities and then compares them to the relative frequencies in ERA5. Several lines comprise a reliability diagram: (1) the diagonal "perfect reliability" 1:1 line representing a perfect forecast where forecast probabilities are equal to observed relative frequencies, (2) the climatological event frequency lines which

Table 4.2: The range of possible scores, the value of a perfect score, and the value of a "no skill" score for the Heidke Skill Score (HSS), Probability of Detection (POD), Miss Rate (MR), and False Alarm Ration (FAR).

Skill Score	Range	Perfect Score	No Skill Score
Heidke Skill Score	-1 to 1	1	0
Probability of Detection	0 to 1	1	0
Miss Rate	0 to 1	0	1
False Alarm Ratio	0 to 1	0	1

are vertical and horizontal lines placed at 0.33 (i.e., 33rd percentile) in our case, and (3) the "no skill" line which lies halfway between the "perfect reliability" line and the climatological event frequency. If the reliability curve has a slope that is smaller than the "no skill" line, then the prediction system is said to be overconfident (i.e., cold weeks are over-predicted), whereas above this line and the forecast is overall adding value.

We complement the reliability diagrams by computing sharpness histograms, which show the relative proportion of forecasts falling into each forecast probability bin (Met Office 2025). As noted in Erner and Karpechko (2024), the expectation in this analysis is that the frequency peak is to occur at around 0.33 forecast probability, since we used the lower tercile threshold in weekly-averaged T2M anomaly to define cold weeks. Finally, we use probability bins at 0.1 spacing between 0 and 1 for models that have 10 or more ensemble members (i.e., ECMWF, BOM, and HMCR), but the reliability is calculated using bins with 1/M spacing for those with less than 10 ensemble members (i.e., UKMO, KMA, and CNR-ISAC) for appearance purposes, where M is the number of ensemble members. Thus, sharpness histograms are occasionally missing counts for UKMO, KMA, and CNR-ISAC as they do not contain forecast probabilities that fall into every bin spaced at even 0.1 intervals.

4.2.4 Windows of forecast opportunity for cold spells

While skill scores and reliability metrics provide a broad assessment of model performance, more targeted insights can be gained by focusing on forecasts with higher ensemble agreement (i.e., windows of forecast opportunity). Bloomfield et al. (2021) found that there is an increase in extended-range forecast skill over Europe when at least 50% of ensemble members agree on the flow pattern, and Erner and Karpechko (2024) successfully used this definition in an analysis of the predictability of European cold spells. As such, we also focus on central US cold spells that are predicted by at least 50% of ensemble members, here termed *confident forecasts*, to assess the key signatures of enhanced predictability for central US cold spells on subseasonal timescales. Specifically, we define a *hit confident forecast* as a forecast that correctly predicts a cold spell with a probability of at least 50%. Subsequently, *a false alarm confident forecast* is defined as a forecast that incorrectly forecasts a cold spell with a probability of at least 50%. Finally, a *miss forecast* is when a forecast does not predict an observed Week 3 cold spell with at least 50% probability. Most importantly, tracking the hit and false alarm confident forecasts allows us to conclude what conditions may yield enhanced predictability subseasonally vs. overconfidence. For each category of forecast, we pool the initializations in that group together to represent the multi-model mean (MMM). For model field composites and other statistics, all initializations are kept including initialization dates that match between models as each forecast is considered an independent sample. However, for composites of the ERA5 fields for these different forecast categories, we remove repeating initialization dates to ensure we are not averaging the same map from observations each time. Statistical significance on all analyses are determined via a bootstrapping approach with 5000 iterations with replacement for significance testing on regime frequencies and model composites, but without replacement for ERA5 composites.

4.2.5 Weather regimes

North American winter weather regimes have a substantial influence on the development of central US cold spells due to their quasi-persistent nature (Lee et al. 2019b; Millin et al. 2022). We define North American winter weather regimes using the same method as in prior chapters (Lee et al. 2019b; Millin et al. 2022). Instead of NDJFM, we use NDJFMA linearly-detrended 500 hPa GPH anomalies between 1993–2020, with respect to the 1993–2020 climatology, in ERA5 to define the regimes across the North American region (180–330°E, 19.5–79.5°N), to account for end of March forecasts that go into April. We note that the regimes' characteristic patterns do not change after the alteration in month and year range compared to Fig. 2.4, with the exception of the PT and WCR switching frequency positions. To investigate the evolution of North American winter weather regimes in confident forecasts of subseasonal cold spells, we also employ the ERA5 k-means clustering model on the ensemble mean 500 hPa GPH anomalies in the S2S models to get model forecast regimes.

4.2.6 Stratospheric polar vortex forcing

SPV strength has been shown to be an important precursor to central US cold in prior studies and chapters, despite its somewhat more complicated relationship with extreme cold in North America when compared to Europe (Millin et al. 2022). As such, we assess how the strength of the SPV impacts the predicted probability of cold spells and the subsequent regime evolution post weak and strong vortex conditions. To do so, we use the 50 hPa GPH anomalies from ERA5, with respect to the 1993–2020 climatology, at the time of initialization to determine the SPV strength, as in Erner and Karpechko (2024). Specifically, weak and strong SPV conditions are defined by the polar cap averaged (60-90°N) 50 hPa GPH anomalies being above or below the 80th and 20th percentiles of the NDJFMA 1993–2020 period, respectively.

4.3 Forecast skill and reliability

Figure 4.1 shows the skill scores for all six S2S models in forecasting cold weeks, stratified by forecast week. In Week 1, five models (KMA, UKMO, ECMWF, HMCR, and CNR-ISAC) exhibit strong performance, with HSS values between 0.7 and 0.9 (Fig. 4.1a) and POD above 0.8 (Fig. 4.1b). The FAR and MR are generally below 0.2 (Figs. 4.1c and 4.1d). The BOM model lags behind, with a HSS near 0.5 and higher FAR and MR (both around 0.3). In Week 2, skill declines across all metrics. ECMWF maintains the greatest skill (e.g., HSS around 0.4), followed by UKMO, KMA, CNR-ISAC, and HMCR. BOM again performs noticeably worse, primarily due to elevated FAR and MR (Figs. 4.1c and 4.1d).

By Week 3, model performance declines notably, with HSS values falling below 0.2 across all models and BOM becoming more comparable to the other prediction systems (Fig. 4.1a). POD values hover around 0.4, while FAR and MR increase to approximately 0.6 (Figs. 4.1b–4.1d). At Week 4, HSS drops further to below 0.1 for all models (Fig. 4.1a), although POD and FAR remain similar to those in Week 3 (Figs. 4.1b–4.1c). This suggests that rising MR and FAR are primarily responsible for the continued decline in HSS at longer lead times. Nevertheless, all six models maintain a positive HSS and a POD of around 0.4 at Weeks 3 and 4, indicating that they retain meaningful skill in forecasting cold weeks at subseasonal lead times, albeit with some overconfidence. We also tested whether skill holds for predicting CAO days instead of cold weeks, using the NDJFM 20th percentile threshold over three or more consecutive days (per Chapter 2) to ensure sample stability. Week 1 and 2 skill remains high and comparable to cold weeks for most models (except BOM), but drops sharply beyond 14 days (not shown).

We further evaluate the S2S models through a reliability assessment. Figure 4.2 presents reliability diagrams and sharpness histograms for Week 3 and Week 4 cold spells. For Week 3, most models exhibit underconfidence at low forecast probabilities (0–0.3), with reliability curves lying above the perfect reliability and no-skill lines, indicating too many false negatives (Fig. 4.2a). As forecast probabilities increase, the curves typically fall below the

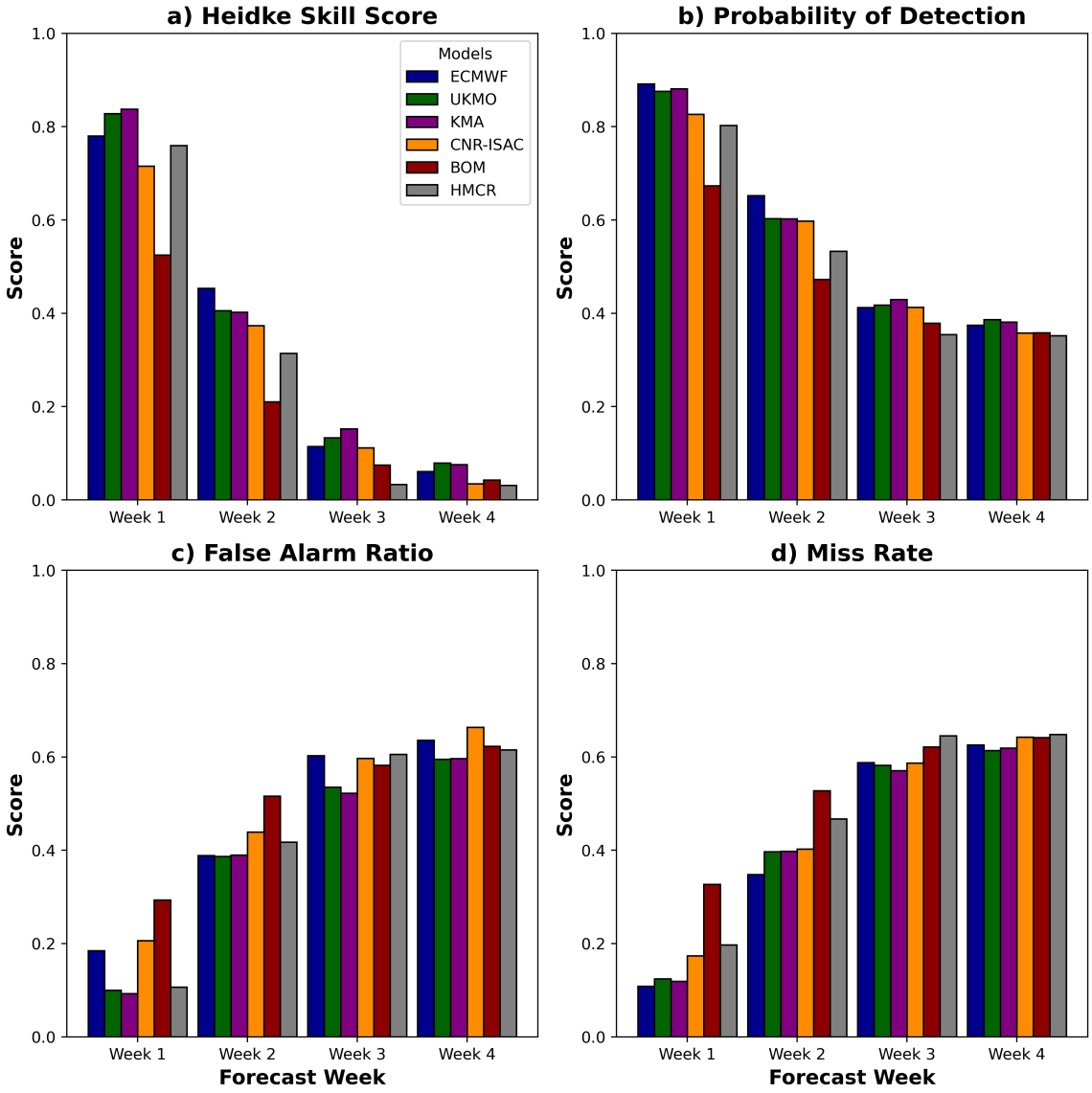


Figure 4.1: The skill of ECMWF (blue bars), UKMO (green bars), KMA (purple bars), CNR-ISAC (orange bars), BOM (red bars), and HMCR (gray bars) models in forecasting cold weeks as a function of forecast week in terms of (a) HSS, (b) POD, (c) FAR, and (d) MR.

no-skill line, reflecting overconfidence and an excess of false positives. However, ECMWF (blue), KMA (purple), and BOM (red) show some skill at higher probabilities (>0.4), with curves that lie between the no-skill and perfect reliability lines. In particular, ECMWF and KMA display higher observed frequencies at higher forecast probabilities, indicating the presence of some skillful and confident forecasts. Despite biases at both low and high forecast probabilities, the overall reliability curves show a rising pattern from bottom left

to top right, suggesting that the models retain some resolution in Week 3 forecasts. The accompanying sharpness histograms reveal that most models issue the majority of forecasts in the lower probability bins, except BOM, which peaks in the 0.3–0.4 range, likely due to the scarcity of confident forecasts at such long leads (Fig. 4.2c). Across models, the number of forecasts generally declines with increasing forecast probability, indicating that while the models can produce high probability forecasts for cold spells at Week 3, such forecasts are relatively rare. Overall, the S2S models demonstrate sharpness by issuing forecasts with probabilities distinct from the climatological frequency (0.33), though BOM exhibits limited sharpness, with a peak near the climatological value.

For Week 4 cold spells, the reliability curves generally fall below the skillful region, particularly at forecast probabilities above 0.4, indicating that the models tend to be over-confident at this lead time (Fig. 4.2b). Notable exceptions include BOM (red line) for forecast probabilities between 0.4–0.7 and HMCR (gray line) between 0.8–0.9, both of which show some skillful confident forecasts. The sharpness histogram further reveals that most forecasts cluster around the climatological frequency (dashed black line), especially in the 0.2–0.5 probability range (Fig. 4.2d). This peak near the climatological value suggests that most forecasts offer limited deviation from the base rate and thus exhibit little sharpness at Week 4. Given that Week 3 forecasts demonstrate greater reliability, sharpness, and overall skill compared to Week 4 (Figs. 4.1 and 4.2), the remainder of our analysis focuses primarily on Week 3. We begin by examining the associated tropospheric circulation and near-surface temperature characteristics of confident forecasts.

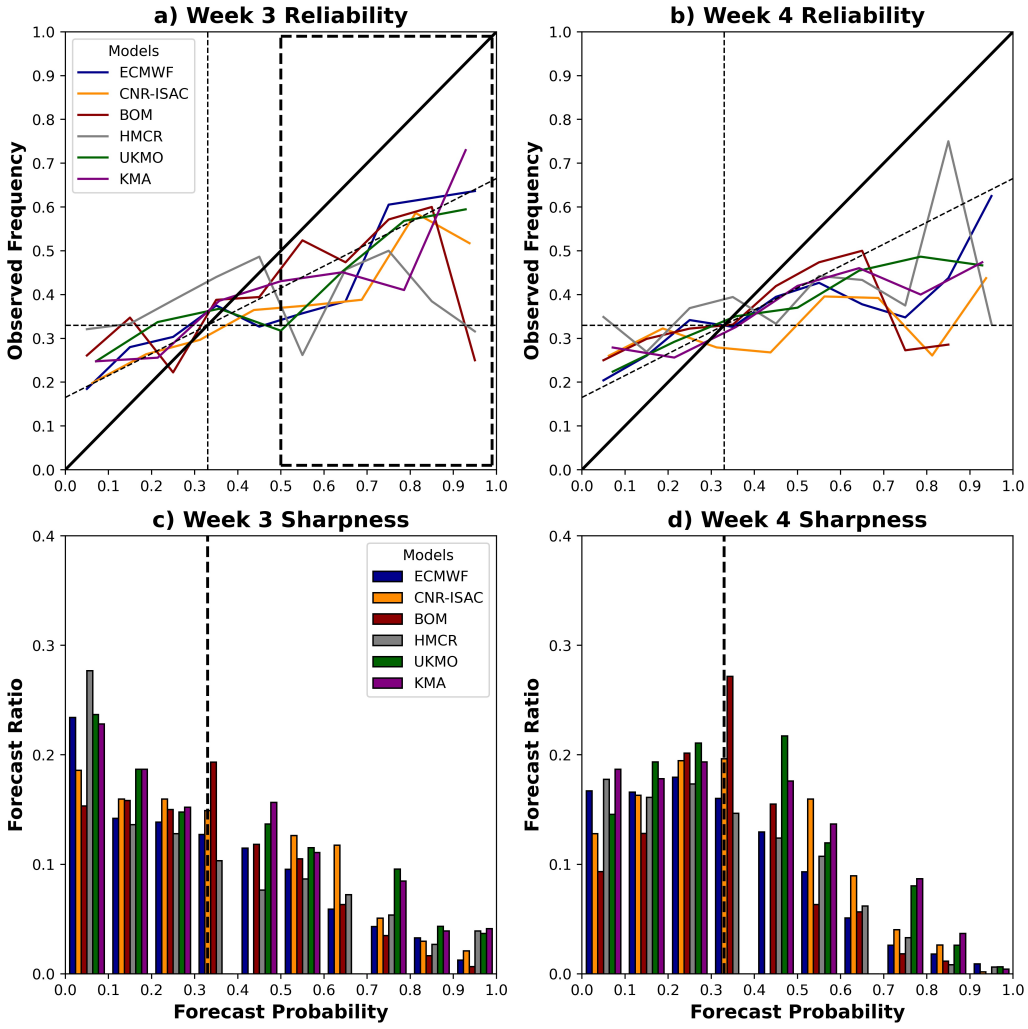


Figure 4.2: Reliability diagrams and sharpness histograms for the S2S models in forecasting (a,c) Week 3 and (b,d) Week 4 cold spells. For the reliability diagrams, the forecast probability is shown in the x-axis and the observed relative frequency is shown on the y-axis, the horizontal and vertical dashed gray lines represent the climatological frequency, the diagonal black line represents the perfect reliability 1:1 line, the diagonal dashed gray line represents the "no-skill" line, and the dashed black box is the region of confident forecasts (i.e., 50%+ forecast probability). The sharpness histograms have forecast probability on the x-axis and forecast ratio on the y-axis, where the vertical dashed black line represents the climatological frequency (i.e., the climatological expected rate of occurrence).

4.4 Tropospheric circulation and weather regimes in confident forecasts

To better understand the skill and limitations of Week 3 cold spell forecasts, we analyze confident forecasts across the six S2S models. Specifically, we examine composites of 500

hPa GPH anomalies and T2M anomalies for different outcomes (Figure 4.3). In ERA5, the initialization day composite for confident hits shows pronounced anomalous ridging (up to +3 dam) over Greenland and northeastern Canada, accompanied by troughing to the south over the North Atlantic, resembling the negative phase of the NAO and the ArH regime (Fig. 4.3a). This supports the idea that confident hit forecasts may be linked to ArH regimes at initialization, consistent with previous studies and potential associations with weak SPV variability on subseasonal timescales (e.g., Lee et al. 2019b; Millin et al. 2022, 2024). In contrast, confident false alarms show much weaker circulation anomalies at initialization, with only faint ridging over the North Atlantic and Arctic Ocean (Fig. 4.3d), suggesting a lack of strong pre-existing signals. Interestingly, the miss forecasts exhibit a different signal entirely, with evidence of anomalous ridging over Siberia at initialization (Fig. 4.3g), hinting at a distinct circulation pattern not conducive to accurate cold spell prediction over the central US.

For Week 3 confident hit forecasts in ERA5, the verifying composite shows strong anomalous ridging (up to +6 dam) over Alaska and the US West Coast, flanked by deep troughing over the North Pacific, central/eastern US, and the North Atlantic. This pattern drives widespread negative T2M anomalies across the central and eastern US (Fig. 4.3b). Additional ridging over Greenland and northern Canada further suggests that a persistent ArH regime may play a key role in enabling skillful subseasonal forecasts of central US cold spells. The overall circulation bears resemblance to previously identified AkR-, ArH-, and WCR-like patterns associated with CAOs in the central US in Chapter 2 (Fig. 2.4, Millin et al. 2022). In contrast, ERA5 composites for confident false alarms at Week 3 reveal a wave train pattern with anomalous troughing over Alaska and anomalous ridging over the central and northern US, leading to positive T2M anomalies in those regions, the opposite of the forecast (Fig. 4.3e). These forecasts are not only confidently wrong, but they also misrepresent the sign of temperature anomalies, highlighting a fundamental circulation mismatch.

ERA5 patterns for the miss group closely resemble those of the confident hits, with anomalous ridging over Alaska and a downstream anomalous trough. However, they lack the ArH signature over Greenland and northern Canada (Fig. 4.3h). This absence suggests that the ArH regime may serve as an important distinguishing feature between successful and missed forecasts of central US cold, potentially in conjunction with AkR and WCR patterns. The MMM forecast composites support these interpretations. For confident hits, the MMM forecasts a circulation very similar to ERA5, with anomalous ridging over Alaska, the western US, and Greenland/northern Canada, along with downstream troughing (Fig. 4.3c). In contrast, confident false alarm MMM forecasts feature anomalous ridging over Alaska, but lack the ArH regime signature, instead displaying a more progressive wave pattern (Fig. 4.3f). This may indicate that confident false alarms are more likely when the forecast features a pattern resembling AkR or WCR regimes with a lack of an anomalous Greenland ridge. The lack of ArH also hints at a potential connection to stratospheric preconditioning (e.g., SPV variability) for skillful Week 3 cold forecasts. Finally, the MMM composites for Week 3 misses show weak and incoherent circulation patterns, consistent with the poor performance, and accompanied by weak positive T2M anomalies across the US (Fig. 4.3i).

Since the composite maps for confident hit and false alarm forecasts of Week 3 cold spells reveal distinct North American winter weather regime patterns, we now examine the frequency evolution of the five regimes in Figure 4.4, relative to forecast initialization and lead time in ERA5. Frequencies are computed at each lag day with respect to Day 0 (initialization), normalized by the number of samples in each forecast group. For confident hit forecasts, the PT regime (green line) dominates the 2–4 weeks preceding initialization, with significantly enhanced frequency reaching around 30% (Fig. 4.4a). In the final two weeks before initialization, the ArH regime (red line) rapidly becomes dominant. This aligns well with the observed anomalous ridging over the ArH region seen in Fig. 4.3a, reinforcing the importance of the ArH as a precursor signal. The ArH regime maintains a strong presence throughout Week 3, reaching over 30% frequency in ERA5. As the event nears, however,

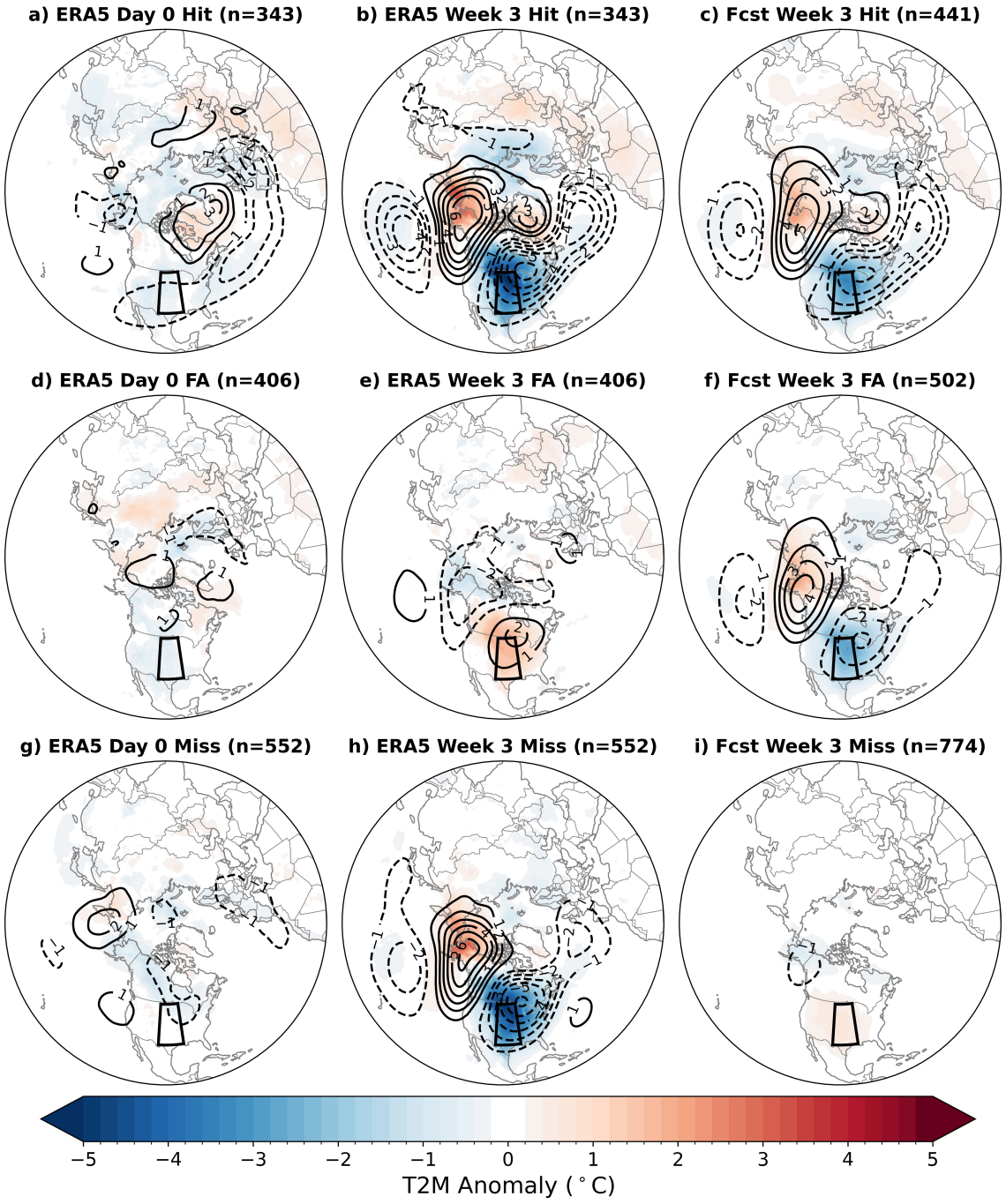


Figure 4.3: Composite maps of T2M anomalies (shading, °C) and 500 hPa GPH anomalies (contours, dam) for confident hit, false alarm and miss forecasts for (a,d,g) ERA5 initialization day, (b,e,h) ERA5 Week 3, and (c,f,i) multi-model mean Week 3, respectively. The black box shows the region used to define central US cold spells and solid contours show positive GPH anomalies while dashed contours show negative GPH anomalies.

the WCR (blue line) emerges as the dominant regime during Week 3, accompanied by a sharp uptick in AkR (orange line, Fig. 4.4a). Simultaneously, the PT and ArL (purple line)

regimes decline to less than 10% frequency, consistent with their unfavorable conditions for central US cold spells (Millin et al. 2022, 2024). The strong WCR signal in Week 3 is somewhat unexpected given the spatial prominence of the anomalous Alaskan ridging in Fig. 4.3b. However, this likely reflects overlapping features in WCR-like events, including deep North Pacific troughing and possible strong AkR contributions to certain CAOs (see Figs. 2.4d and 4.3b). Importantly, the model forecast frequencies largely mirror those observed in ERA5 for all five regimes in confident hit forecasts (Fig. 4.4b), highlighting the models' ability to reproduce the key regime evolutions that lead to successful cold spell forecasts. In particular, the significantly elevated ArH frequency prior to and at initialization suggests a critical preconditioning role for this regime in enabling skillful Week 3 cold spell forecasts (Millin et al. 2022).

The regime frequency evolution for the confident false alarm forecasts provides insight into how the models may become overconfident in predicting Week 3 cold spells. In ERA5, no regime stands out 2–4 weeks before initialization; frequencies for all five regimes hover near 20% (Fig. 4.4c). In the two weeks leading up to initialization, the WCR regime becomes significantly enhanced, followed by a brief increase in ArH frequency starting during the week before initialization. However, as Week 3 approaches, both ArH and AkR decline, the latter becoming significantly suppressed. Strikingly, during Week 3 itself, the PT and ArL regimes become significantly enhanced in ERA5, reaching frequencies of 25–30%. Since these regimes are generally associated with anomalous warmth over the central US, this suggests that false alarm forecasts are not only wrong about cold occurrence but also fail to systematically predict circulation patterns that drive warm anomalies instead (see Fig. 4.3e).

Model forecasts for these false alarm events reflect some aspects of the observed evolution. They capture the initial rise in ArH frequency during the first lead week, peaking near 20% (Fig. 4.4d). However, unlike ERA5, the models continue to increase the ArH frequency toward Week 3, maintaining a significantly enhanced signal. Similarly, the AkR regime is significantly overestimated throughout the forecast period, despite its decline in

the verifying state. Perhaps most tellingly, the WCR regime dominates the false alarm model forecasts for Week 3, with peak frequencies reaching around 35% during the target week (Fig. 4.4d). This pattern suggests that false alarm forecasts are often linked to a more progressive, wave train-like circulation across North America, consistent with Fig. 4.3f. Prior work has shown that WCR-associated CAOs tend to be shorter and less intense, often lacking strong stratospheric percursors (Millin et al. 2022). This may help explain the relatively muted ArH and AkR signals the model forecasts for false alarms, in contrast to the stronger SPV-linked signals seen in confident hit forecasts (Figs. 4.4b and 4.4d). These distinctions highlight the potential role of stratospheric variability in differentiating between skillful and overconfident cold forecasts, which we now examine.

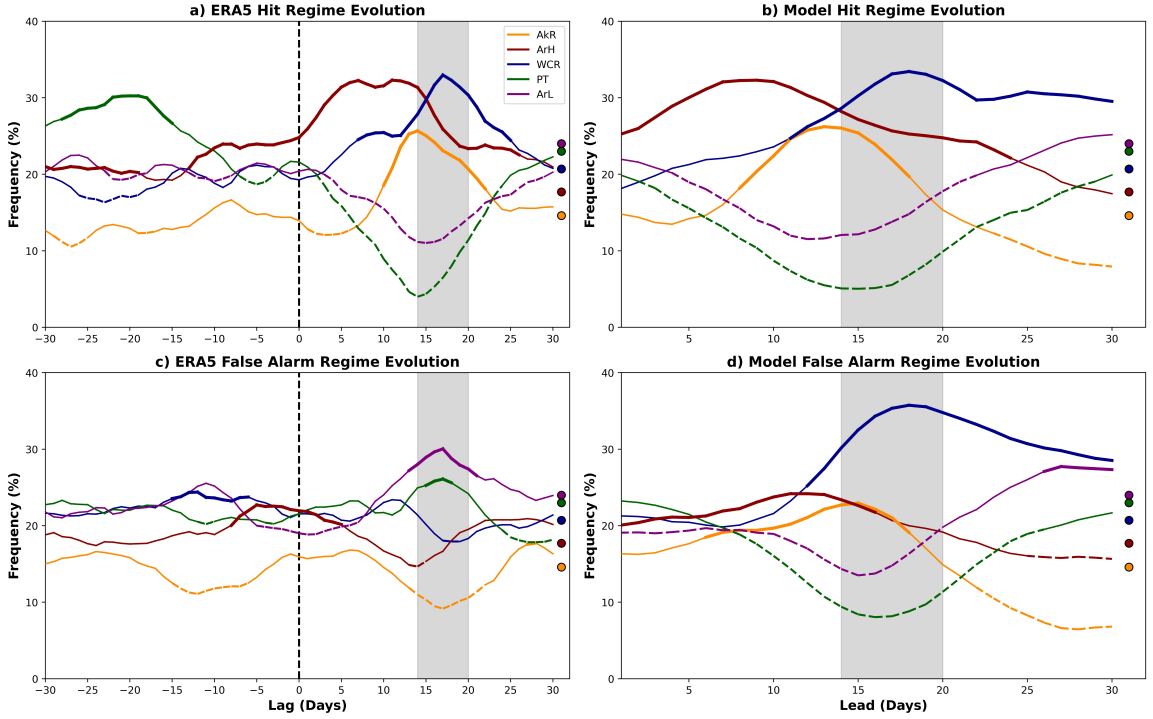


Figure 4.4: Lagged North American winter weather regime frequencies with respect to forecast initialization time (lag 0) for a) the confident hit forecast initializations and c) the confident false alarm forecast initializations in ERA5. Panels b) and d) show the same as panels a) and c) but instead for the model forecasts as a function of lead time. Regime frequencies are calculated relative to the total number of samples in each group and are then smoothed with a 5-day centered filter. Bold and dashed lines represent statistically significant enhanced or suppressed frequencies relative to climatology at the 95th percentile via 5000-iteration bootstrapping from ERA5 data. Week 3 is shown in gray shading and the colored dots indicate the climatological frequency of the five winter weather regimes.

4.5 State of the SPV in confident cold spell predictions

Figure 4.5 presents the composite standardized GPH anomalies averaged over the polar cap (60–90°N) for confident hit and false alarm forecasts of Week 3 cold spells, along with their differences, in both ERA5 and the S2S models. For confident hit forecasts, ERA5 shows significantly positive GPH anomalies in the mid-to-upper stratosphere beginning around four weeks prior to initialization (Fig. 4.5a), indicating a weaker-than-normal SPV. These positive anomalies gradually descend through the stratosphere, reaching the lower levels near the initialization date and projecting onto the troposphere as anomalous ridging, likely linked to the ArH signal seen in ERA5 500 hPa GPH anomalies on Day 0 (Figs. 4.3a, 4.4a, and 4.5a). The weakened SPV signal persists into the lower stratosphere through the Week 3 forecast period, with the most pronounced and statistically significant downward propagation into the troposphere coinciding with the cold spell. A recovery of the SPV, characterized by negative GPH anomalies in the upper stratosphere, emerges roughly 20–30 days after initialization, suggesting that weak vortex events, including major SSWs, play a key role in subseasonal predictability of central US cold spells (Fig. 4.5a). Importantly the S2S models show a comparable evolution. The forecast composite for confident hit cases also features significantly positive GPH anomalies in the mid-to-lower stratosphere, persisting from initialization through Week 3 (Fig. 4.5d). These anomalies exhibit a "dripping" structure indicative of downward propagation into the troposphere, aligning well with the observed timing of the cold spell. This agreement between model and reanalysis signals reinforces the notion that weak SPV conditions are a crucial preconditioning mechanism for skillful Week 3 forecasts of central US cold spells.

Analyzing the stratospheric conditions associated with confident false alarm forecasts of Week 3 cold spells provides further insight into the absence, or misrepresentation, of key stratospheric signals that may contribute to erroneous predictions. In ERA5, these false alarm cases exhibit little to no significant stratospheric anomalies prior to initialization

(Fig. 4.5b). Following initialization, the observed evolution in ERA5 shows a strengthening of the SPV, indicated by negative GPH anomalies in the upper stratosphere that descend toward the troposphere through Week 3. This downward propagation of a stronger vortex signal is associated with the development of warmer weather regimes (Fig. 4.4c), consistent with anomalous warmth observed during these falsely predicted cold spells. In contrast, the S2S model composite shows a divergent evolution from the reanalysis. Although the models do simulate a stronger SPV aloft, with persistent negative GPH anomalies in the upper stratosphere, it simultaneously forecasts a rapidly developing and significant tropospheric ridging signal, confined below 100 hPa. This low-level ridging emerges a few days before Week 3 and persists through the forecast cold spell period (Fig. 4.5e). This disconnect between the model's stratospheric and tropospheric evolution may explain the overconfidence in predicting cold spells despite the lack of conducive stratospheric preconditioning, highlighting a potential source of false alarm error in subseasonal forecasts.

Now, we present the composite differences in polar cap GPH standardized anomalies between confident hit and false alarm forecasts in both ERA5 and the subseasonal prediction systems (Figs. 4.5c and 4.5f). These differences highlight the polar circulation features that help distinguish correct from incorrect forecasts when a Week 3 cold spell is predicted. In ERA5, a notably weaker SPV, manifested as significant positive GPH anomalies in the mid-to-upper stratosphere, emerges as a key precursor that separates confident hit forecasts from false alarms prior to initialization (Fig. 4.5c). In the models, the same distinction holds: confident hit forecasts are characterized by persistent mid-to-upper stratospheric positive GPH anomalies, whereas false alarm forecasts lack this signal entirely (Fig. 4.5f). These results suggest that confident subseasonal cold spell forecasts are more likely to be correct when preceded by a substantial weakening of the SPV. Conversely, confident false alarms appear to arise when forecasts are made in the absence of strong stratospheric forcing, pointing to a more tropospheric origin of the predicted circulation pattern. Specifically, false alarms tend to be associated with the simulation of WCR- and AkR-like patterns,

which may be driven by tropospheric dynamics or potentially shallower stratospheric processes, such as wave reflection (Messori et al. 2022; Millin et al. 2022; Schutte et al. 2025). This highlights the importance of stratospheric preconditioning in enhancing the reliability of Week 3 cold spells.

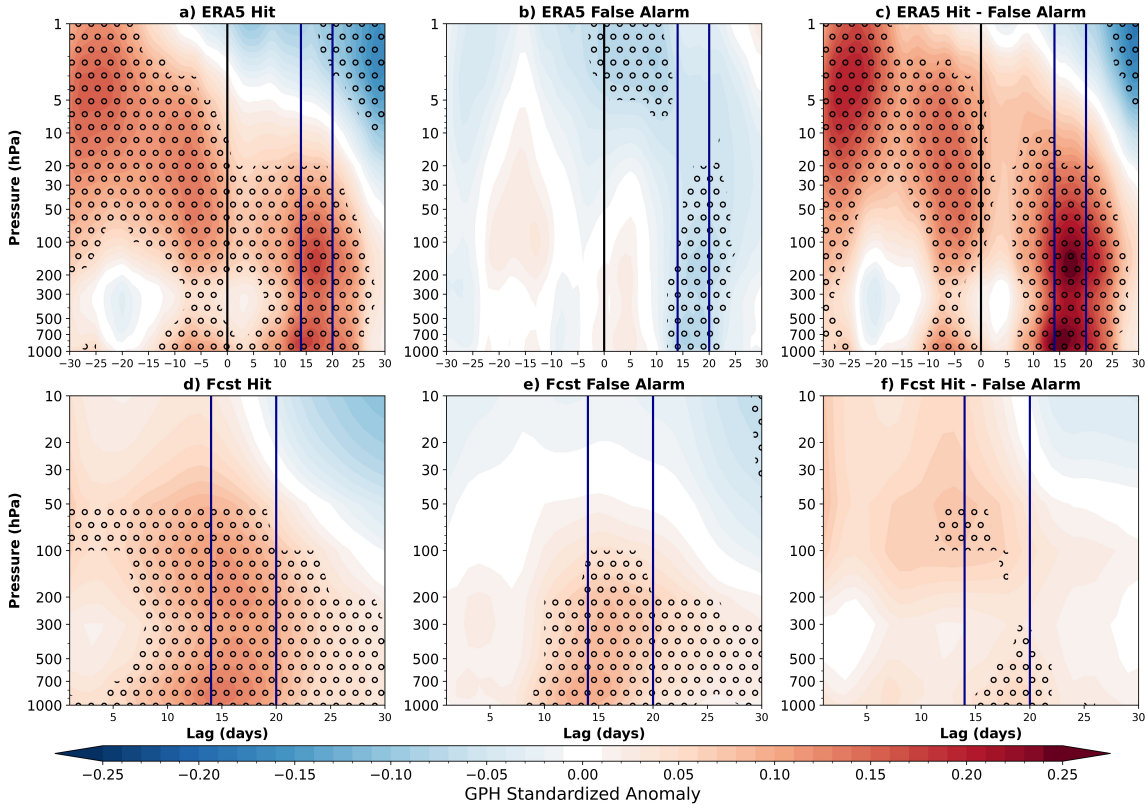


Figure 4.5: Composites of polar cap (60-90°N) averaged GPH anomalies (shading) as a function of lag from initialization and pressure level for Week 3 cold spell confident hit forecasts, confident false alarm forecasts, and the difference between the confident hit and false alarm forecasts in a)–c) ERA5 and d)–f) the S2S models, respectively. Circles represent statistical significance at the 95th percentile via 5000-iteration bootstrapping, the blue lines represent the Week 3 cold, and the black solid line represents initialization date.

4.6 Confident forecasts under weak and strong polar vortex conditions

4.6.1 Proportions of confident forecasts by SPV strength

Before evaluating the reliability of subseasonal cold spell forecasts initialized in weak and strong SPV states, we examine how confident forecasts are distributed by SPV strength, particularly given Fig. 4.5, which suggests SPV strength may serve as a key preconditioning factor for enhanced subseasonal predictability of central US cold spells. Figure 4.6 presents the number of confident hit forecasts (blue bars), confident false alarm forecasts (red bars), and the proportion of each that were initialized in either strong or weak SPV states (indicated by hatching). Among the models, ECMWF produces the highest number of both confident hits (98) and false alarms (116), likely due to its larger sample size (Table 4.1). While CNR-ISAC has fewer forecasts overall than both ECMWF and BOM, its relatively high number of false alarms is consistent with its tendency toward overconfidence shown in Fig. 4.2a. BOM, despite having 600 forecasts, produces substantially fewer confident hits and false alarms compared to ECMWF and CNR-ISAC. UKMO, KMA, and HMCR exhibit similar number of confident hits and false alarms, although HMCR tends to produce more false alarms overall.

Importantly, Fig. 4.6 illustrates that all models show more confident hit forecasts when initialized during weak SPV conditions than during strong SPV conditions. For instance, ECMWF generates 26 confident hits under weak vortex conditions compared to just 12 under strong vortex conditions. In contrast, the number of confident false alarms under weak and strong vortex states is more balanced across models. In fact, for UKMO, KMA, and ISAC-CNR, there are more false alarms initiated under strong vortex conditions than weak ones. These results suggest that weak SPV conditions may enhance forecast skill for Week 3 central US cold spells (see also Fig. 4.5). This connection is likely due to the robust link between a weak SPV, an equatorward-shifted jet stream, and the negative phase

of the tropospheric NAM. These conditions are conducive for enhanced blocking over the North Atlantic and North Pacific, which are often associated with cold-related weather regimes and thus an increased likelihood of extreme cold at subseasonal timescales (Kodera et al. 2016; Charlton-Perez et al. 2018; Lee et al. 2019b). Although weak SPV states can still lead to false alarms, strong SPV conditions appear more likely to be associated with overconfident, and ultimately, incorrect forecasts in these subseasonal prediction systems.

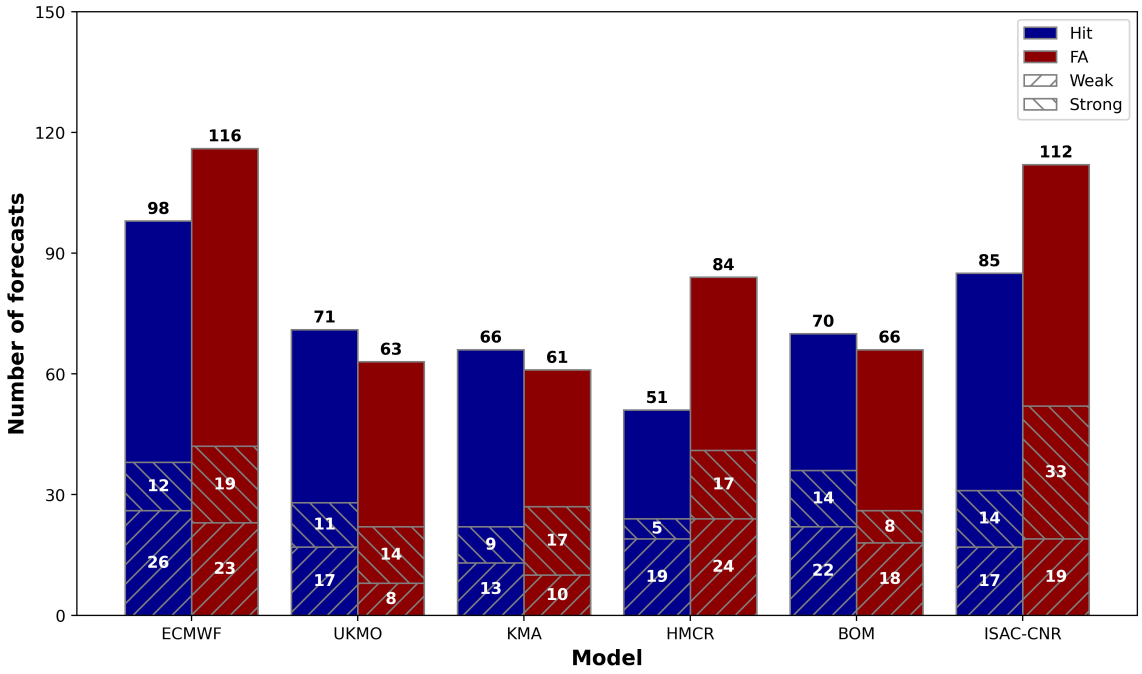


Figure 4.6: The total number of confident hit (blue solid bars) and confident false alarm (red solid bars) forecasts as a function of S2S model for Week 3 cold spells. The left-to-right hatching indicates the proportion of confident hit and false alarm forecasts initialized in a weak SPV state while the right-to-left hatching indicates the proportion of confident hit and false alarm forecasts initialized in a strong SPV state. The bold numbers show the respective counts of these different proportions.

4.6.2 Reliability

Figure 4.7 presents reliability diagrams and sharpness histograms for Week 3 and Week 4 cold spell forecasts from the six chosen S2S models, stratified by whether forecasts were initialized under strong or weak SPV conditions. The Week 3 reliability curves for forecasts initialized in a weak vortex (Fig. 4.7a) show a noticeable shift toward the perfect

reliability line compared to the reliability for all Week 3 forecasts (Fig. 4.2a). These curves are also more consistently sloped from bottom left to upper right, particularly for forecast probabilities above 0.5, indicating that a weak SPV at initialization improves both the confidence and accuracy of cold spell predictions. The corresponding sharpness histograms for Week 3 cold spells initialized in weak SPV states (Fig. 4.7e) show the highest concentration of forecasts in the lowest probability bins ($<$0.3), but also display a more even distribution across the full range of forecast probabilities compared to all forecasts (Fig. 4.2c). This broader spread, including a higher frequency of high-probability forecasts, reflects improved sharpness when forecasts are initialized in weak SPV conditions.

For Week 4 cold spells initialized in a weak SPV (Fig. 4.7b), the reliability curves similarly show improvement relative to all forecasts (Fig. 4.2b), particularly at higher forecast probabilities. However, these curves are generally flatter than their Week 3 counterparts, indicating more limited skill at longer lead times. The Week 4 sharpness histograms (Fig. 4.7f) mirror those of Week 3 in shape but to a lesser degree. The distributions are flatter, with more forecasts clustered near the climatological frequency (0.33) and some increase in high-probability forecasts, indicating modest gains in sharpness relative to Fig. 4.2d. Overall, while initializing in a weak SPV enhances both skill and sharpness for Week 4 cold spell forecasts, the improvements are smaller than those seen in Week 3.

The reliability of Week 3 cold spell forecasts initialized in strong SPV conditions shows an opposite pattern to that seen for weak SPV forecasts (Fig. 4.7c). These forecasts are generally overconfident, with reliability curves shifted toward the lower right of the diagram compared to both Figs. 4.7a and 4.2a. Some models, such as ECMWF, KMA, and BOM, display more reliable behaviour at higher forecast probabilities, but this is not consistent across all models. The sharpness histograms for strong SPV Week 3 forecasts (Fig. 4.7g) indicates some degree of sharpness, with a broad spectrum of forecast probabilities. However, the number of forecasts exceeding 0.7 probability is notably lower than in the weak SPV sharpness distribution (Fig. 4.7e), suggesting that strong SPV conditions reduce forecast confidence for Week 3 cold spells. Additionally, strong SPV forecasts show a

higher frequency of low-probability predictions, further reflecting reduced confidence and skill. Week 4 cold spell forecasts initialized in strong SPV conditions show exhibit similar underconfidence patterns to Week 3 (Fig. 4.7d), with even more clustering around the climatological frequency in the sharpness histograms, with forecasts rarely reaching the highest probability bins compared to the weak SPV forecasts (Fig. 4.7h). This implies a reduced ability to generate confident cold spell forecasts at longer lead times under strong SPV conditions.

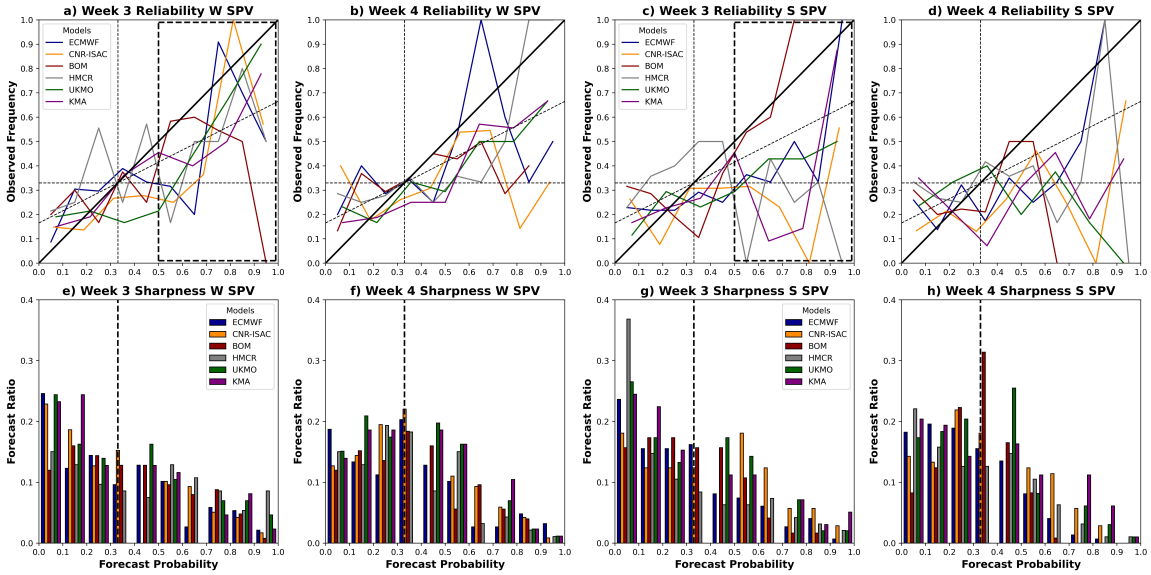


Figure 4.7: As in Fig. 4.2 but the forecast reliability diagrams and sharpness histograms of Week 3 and Week 4 cold spells initialized in (a,b,e,f) weak SPV conditions and (c,d,g,h) strong SPV conditions. Note that here the cumulative number of forecasts used for this analysis is 695 for weak SPV forecasts and 665 for strong SPV forecasts, compared to the total number of 3454 used in Fig. 4.2.

4.6.3 Tropospheric circulation and weather regimes

We now repeat the analysis of Fig. 4.3, focusing on confident hit and confident false alarm forecasts that were initialized in either weak or strong SPV conditions. Figure 4.8 shows the composite 500 hPa GPH anomalies (contours) and T2M anomalies (shading) for confident forecasts initialized during weak SPV states. For both confident hits and false alarms, the ERA5 mean pattern at initialization is similar, with a characteristic weak SPV signature;

anomalous ridging over the Arctic (Figs. 4.8a and 4.8d). In both cases, the central US features negative T2M anomalies and anomalous troughing, and negative T2M anomalies are also evident across northern Eurasia, highlighting the canonical weak vortex-related cold response in that region (Butler et al. 2017). A key difference between the confident hit and false alarm forecasts in weak SPV states is the magnitude of Arctic GPH anomalies at initialization, which is stronger in false alarms ($\sim$+8 dam) than in the hits ($\sim$+6 dam). This suggests that more pronounced Arctic ridging in the initialization state may bias the model toward forecasting central US cold spells later in the run (Figs. 4.8a and 4.8d). Week 3 verification fields show negative T2M anomalies and anomalous troughing in the central US, alongside strong anomalous ridging over northern Canada and Greenland ($\sim$8 dam), connecting with anomalous ridging over Alaska. This pattern is indicative of the ArH regime and a negative NAO-like structure, associated with a southward displaced Atlantic storm track in weak vortex confident hit forecasts (Fig. 4.8b).

In contrast, the verifying mean state for confident false alarm forecasts reveals anomalous warmth across North America, and the anomalous ridging signal over northern Canada and Greenland is weaker and less coherent. The lack of an AkR-like signal in this case suggests that both AkR and ArH regimes may be important downstream features following a weak SPV, consistent with prior findings in Chapter 2 (Fig. 4.8e, Millin et al. 2022). The corresponding Week 3 model confident hit forecasts closely resemble the observed ERA5 pattern (Fig. 4.8c). In contrast, the forecast pattern for false alarms features weaker and more high latitude-confined ridging, implying that simulating a stronger and more spatially extensive Arctic ridging enhances forecast accuracy for Week 3 cold spells (Fig. 4.8f). Finally, a notable feature across all panels in Fig. 4.8 is the consistent presence of negative T2M anomalies over Northern Eurasia. This suggests the potential for compound cold events affecting both Eurasia and the central US following weak SPV conditions, reinforcing the broader climatological significance of the weak vortex state.

Figure 4.9 is the same as Fig. 4.8 but for strong SPV initial conditions. The first notable feature is that the number of confident hit forecasts under strong SPV conditions is about

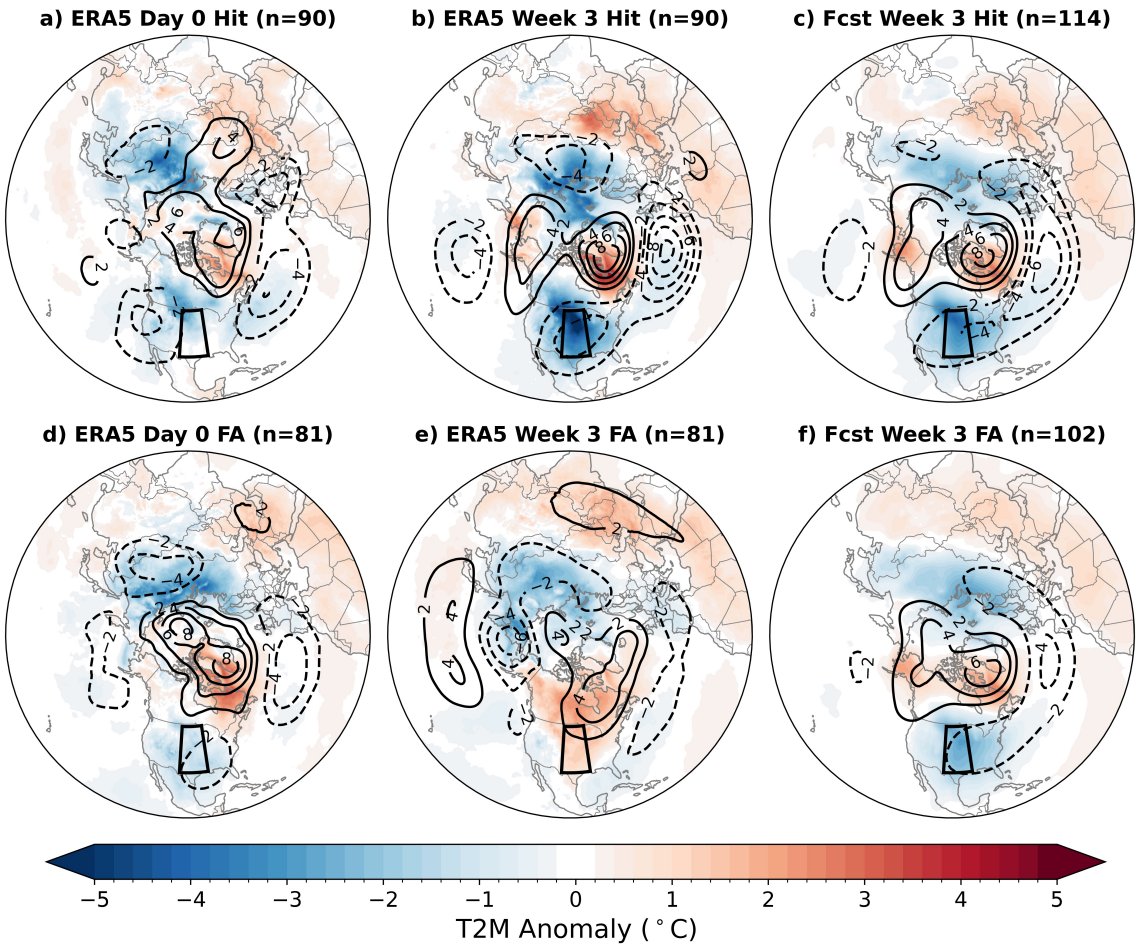


Figure 4.8: As in Fig. 4.3 but for forecasts initialized in weak polar vortex states.

half of that of weak SPV cases, suggesting less favorable conditions for skillful long-lead forecasts of central US cold spells. At initialization in ERA5, both confident hit and false alarm forecasts exhibit strong anomalous troughing over the Arctic (<-6 dam), a hallmark of strong SPV states (Figs. 4.9a and 4.9d). The confident hit forecasts show the most intense negative GPH anomalies across the Arctic at this stage. In the verifying Week 3 ERA5 fields for confident hit forecasts, strong negative T2M anomalies develop over the central US, co-occurring with a well-defined wave train pattern: anomalous ridging over Alaska and Eurasia flanked by deep anomalous troughing in the North Pacific and central/eastern US (Fig. 4.9b). The model forecast captures much of this structure, though ridging is more isolated in a pure AkR-like configuration, lacking the strong North Pacific

troughing associated with the WCR pattern (Fig. 4.9c). This configuration resembles that seen in AkR-CAOs over the Great Plains in Chapter 2 (Millin et al. 2022).

In contrast, confident false alarm forecasts initialized under strong SPV conditions generally feature a weaker wave pattern, appearing as a broad dipole over North America, with anomalous ridging over Alaska and anomalous troughing over Canada (Fig. 4.9f). The verifying ERA5 state for these false alarms maintains a pattern similar to that at initialization, with a tightly confined Arctic trough and positive T2M anomalies across the central US (Fig. 4.9e). These findings suggest that the strength and structure of the simulated wave pattern, and potentially the configuration of the SPV, play an important role in whether S2S models can confidently and correctly predict Week 3 cold spells following strong SPV conditions.

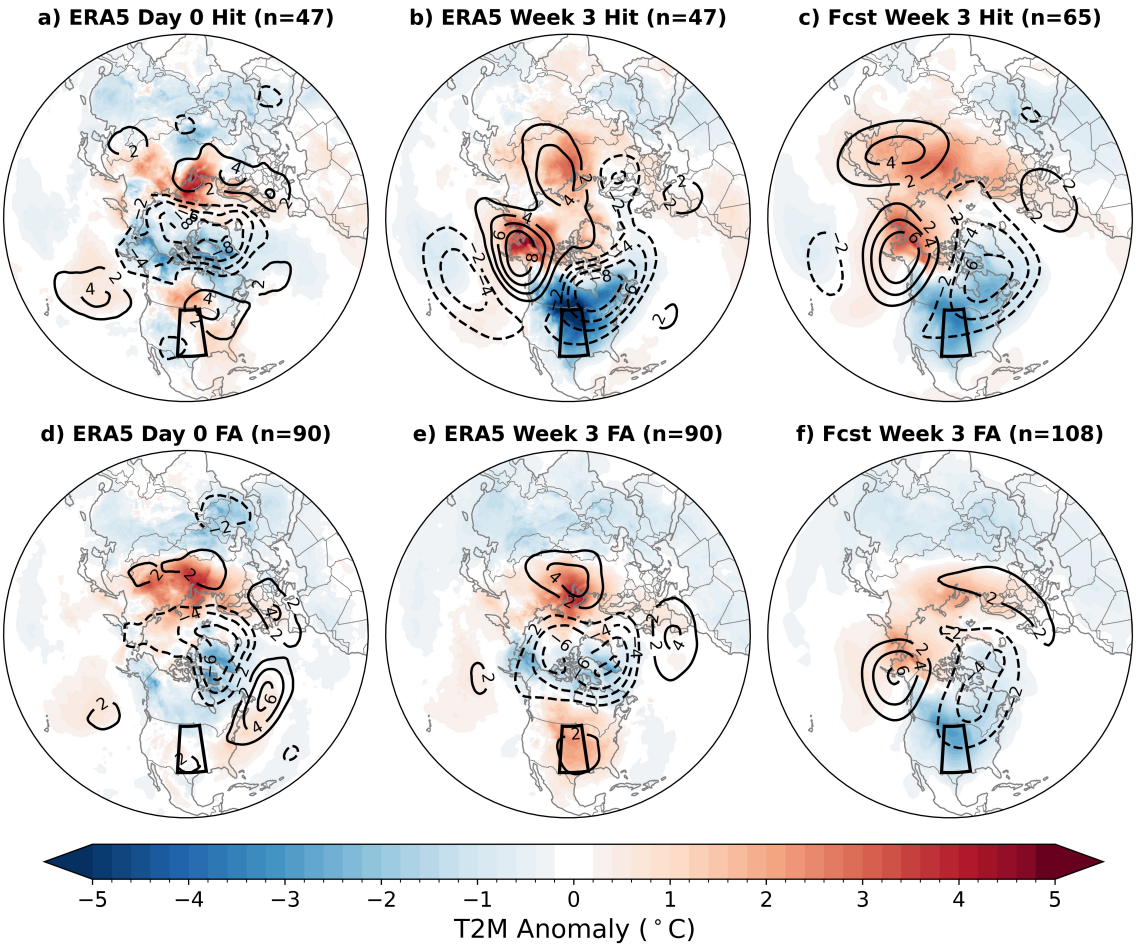


Figure 4.9: As in Fig. 4.3 but for forecasts initialized in strong polar vortex states.

The weather regime analyses for confident hit and false alarm forecasts of Week 3 cold spells, stratified by SPV strength, reinforce the circulation-based results discussed previously (Figs. 4.10 and 4.11). In ERA5, confident hit forecasts initialized in weak SPV conditions exhibit a markedly enhanced frequency of the ArH regime, which becomes apparent as early as two weeks prior to initialization and increases to nearly 50% in the days leading up to the cold spell (Fig. 4.10a). During this period, the remaining weather regimes show little or no enhancement, underscoring the dominant role of the ArH in the lead-up to Week 3 confident hit cold spells following weak SPV conditions. Consistently, the model forecasts for this same group also simulate significantly elevated ArH regime frequencies (Fig. 4.10b). This alignment between the observed and simulated regime evolution highlights the importance of capturing the emergence and persistence of the ArH regime on subseasonal timescales. Indeed, it is probabilistically unlikely to move away from the ArH regime under weak SPV conditions, explaining the dominance of the ArH in this forecast category (Lee et al. 2022).

For the false alarm confident forecasts initialized with a weak SPV, the ArH regime is consistently observed at significantly elevated frequencies of around 30% for approximately four weeks before initialization (Fig. 4.10c). However, after initialization, the ArH regime diminishes in frequency, allowing the PT regime to dominate the cold week period in ERA5, which is not conducive to extreme cold in the central US and helps explain why anomalous warmth across the central US tends to occur in the verification (Fig. 4.8e). Interestingly, in the model forecast, the confident false alarm forecasts start with a relatively lower ArH frequency than the confident hit forecasts. However, this frequency increases as the cold week approaches, eventually exceeding the ArH frequency seen in confident hit forecasts (Fig. 4.10d). This suggests that the persistence of the ArH regime prior to Week 3 cold spells is a key differentiator between confident hit and confident false alarm forecasts initialized with a weak SPV.

The evolution of regime frequencies in ERA5 for confident hit forecasts initialized with strong SPV conditions reveals that, prior to initialization, the ArL regime is a favorable

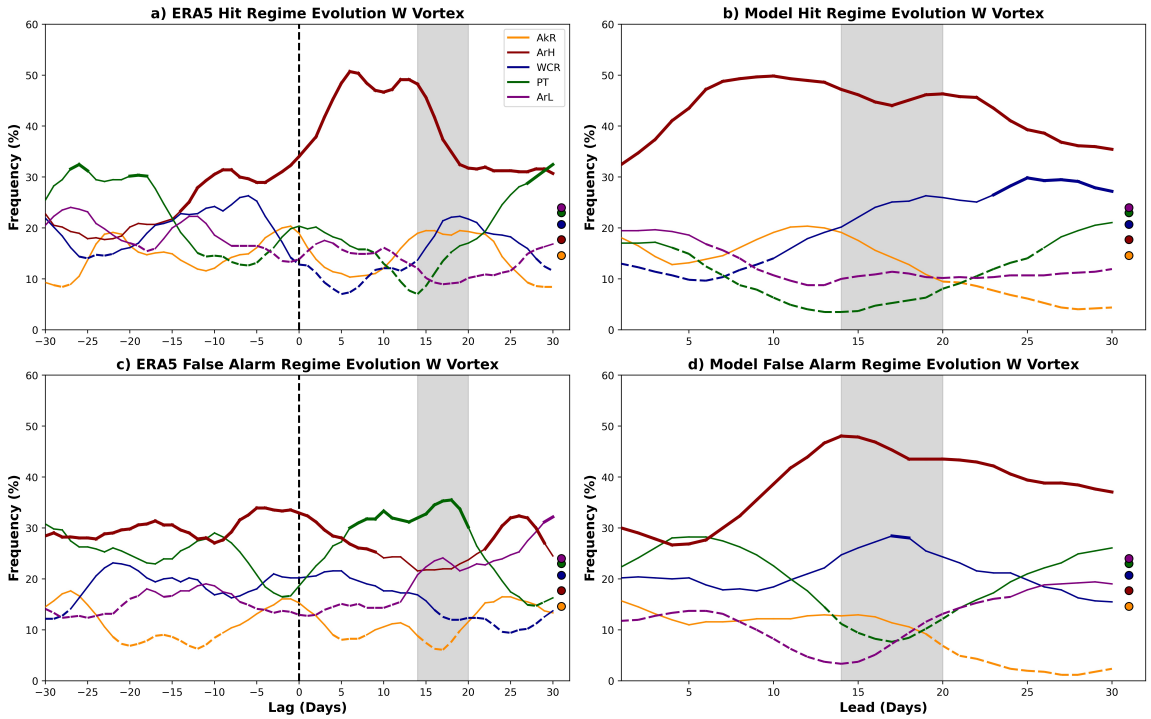


Figure 4.10: As in Fig. 4.4 but for forecasts initialized in weak polar vortex states.

weather regime (Fig. 4.11a). In the days leading up to the Week 3 cold spell in ERA5, the regime frequencies shift, with the WCR and AkR regimes becoming dominant and significantly enhanced, reaching frequencies between 30–40%. The model forecast shows a similar rapid increase in the dominance of the AkR and WCR regimes. However, the model forecasts tends to simulate AkR conditions more frequently in these forecasts (Fig. 4.11b). This difference in the dominant regime frequency is likely attributable to the positioning of the GPH anomalies shown in Figs. 4.9b and 4.9c. Nevertheless, the rapid but significant transition to WCR and AkR regimes in the strong SPV confident hit forecasts mirrors the synoptic-scale development of these features and also highlights potential associations with stratospheric wave reflection (Millin et al. 2022).

The confident false alarm forecasts reveal that the ArL regime is significantly enhanced up to approximately five days prior to initialization in ERA5, after which the PT regime frequency becomes enhanced around initialization (Fig. 4.11c). In contrast to the confident hit forecasts, the ArL regime dominates in ERA5 during the cold week. Meanwhile, the

model forecasts show significantly enhanced AkR and WCR regimes during the cold week (Fig. 4.11d). Interestingly, the model forecast regime frequencies show an evolution that results in a transition from the PT to the AkR and WCR regimes, which is a very unlikely transition under strong SPV conditions (Lee et al. 2022).

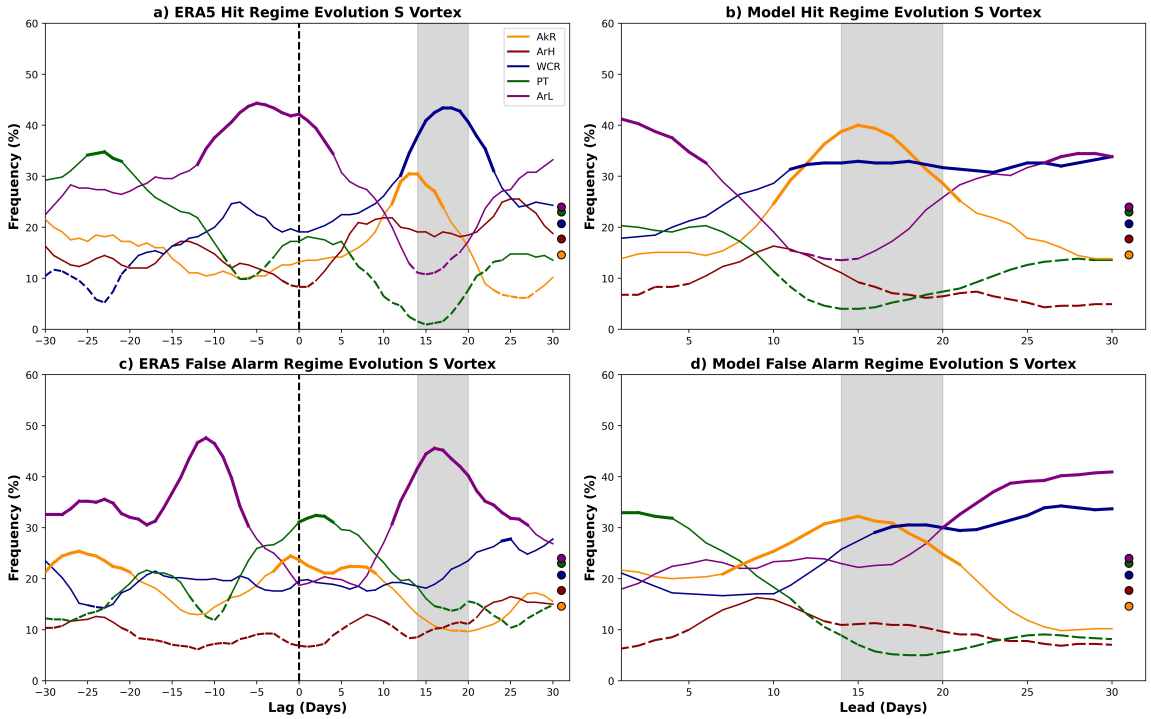


Figure 4.11: As in Fig. 4.4 but for forecasts initialized in strong polar vortex states.

4.6.4 Stratospheric signals for forecasts initialized in weak and strong vortex

To further explore the differences between these forecast groups, Figure 4.12 shows composites of polar cap-averaged standardized GPH anomalies in both ERA5 and the model forecast for confident hit and false alarm Week 3 cold spell forecasts initialized under weak SPV conditions. In ERA5, the weak SPV hit group exhibits strong and statistically significant positive GPH anomalies throughout the mid-to-upper stratosphere at initialization, particularly between 10–50 hPa (Fig. 4.12a). Several "drips" of significant positive anomalies descend into the troposphere prior to initialization and become increasingly persistent

approaching the cold spell, consistent with the sustained ArH signal seen in Fig. 4.10a. The corresponding ERA5 composite for the weak SPV false alarm group shows a broadly similar stratospheric evolution (Fig. 4.12b); however, it notably lacks positive GPH anomalies in the troposphere during the Week 3 forecast period. The most pronounced difference between the ERA5 composites emerges in the upper stratosphere approximately 3–4 weeks prior to initialization, where the hit forecasts exhibit substantially stronger positive GPH anomalies between 1–10 hPa (Fig. 4.12c). This suggests that a key distinguishing factor between the two groups is the earlier onset and greater intensity of upper stratospheric weakening in the hit forecasts. In contrast, the difference in the model forecast evolution between the two groups in minimal (Figs. 4.12d–4.12f), apart from the hit group showing a slightly larger near-surface positive GPH anomaly beyond Week 3. This suggests that false alarm forecasts under weak SPV conditions tend to overestimate the extent of downward coupling.

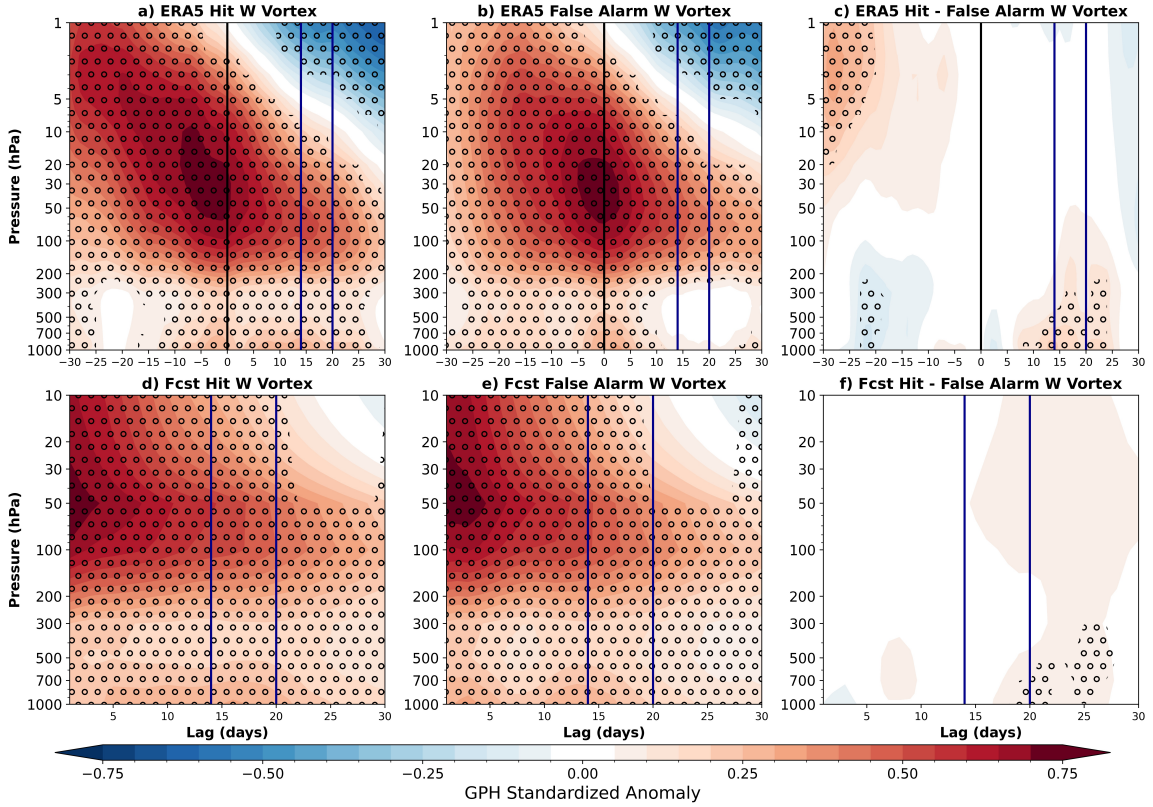


Figure 4.12: As in Fig. 4.5 but for forecasts initialized in weak polar vortex states.

Figure 4.13 presents composites of polar cap-averaged standardized GPH anomalies in both ERA5 and the model forecast for confident hit and false alarm Week 3 cold spell forecasts initialized under strong SPV conditions. In ERA5, the hit forecast group shows a more isolated negative GPH anomaly confined to the mid-to-lower stratosphere, emerging approximately 10 days prior to initialization and persisting through to the Week 3 cold period (Fig. 4.13a). This stratospheric signal, along with the rapid development of a tropospheric positive GPH anomaly, resembles the wave reflection-driven SPV signature associated with AkR-type central US CAOs described in Fig. 2.11. In contrast, the ERA5 composite for the false alarm group displays a stronger and more vertically extensive SPV signal throughout the stratosphere, which is also more persistent prior to initialization. However, this does not translate into a positive tropospheric GPH anomaly during the Week 3 period in ERA5 (Fig. 4.13b), resulting in a statistically significant positive anomaly difference between the two groups in the troposphere (Fig. 4.13c). The model forecast fields for both groups are similar, exhibiting strong SPV conditions at initialization and a more positive GPH anomaly in the troposphere during Week 3 (Figs. 4.13d–4.13f). This indicates that the model forecasts a similar tropospheric response for both cases, despite the absence of such a response in the ERA5 false alarm group composite. Taken together, these results suggest that the structure and timing of the SPV preconditioning may be critical: a more shallow and isolated stratospheric anomaly, potentially linked to wave reflection near initialization, may distinguish hit forecasts. Conversely, the false alarm cases may represent situations where the model attempts to predict a weather regime transition away from the PT toward the AkR in response to the strong SPV, which in reality is unlikely to occur and difficult to forecast accurately (Lee et al. 2022).

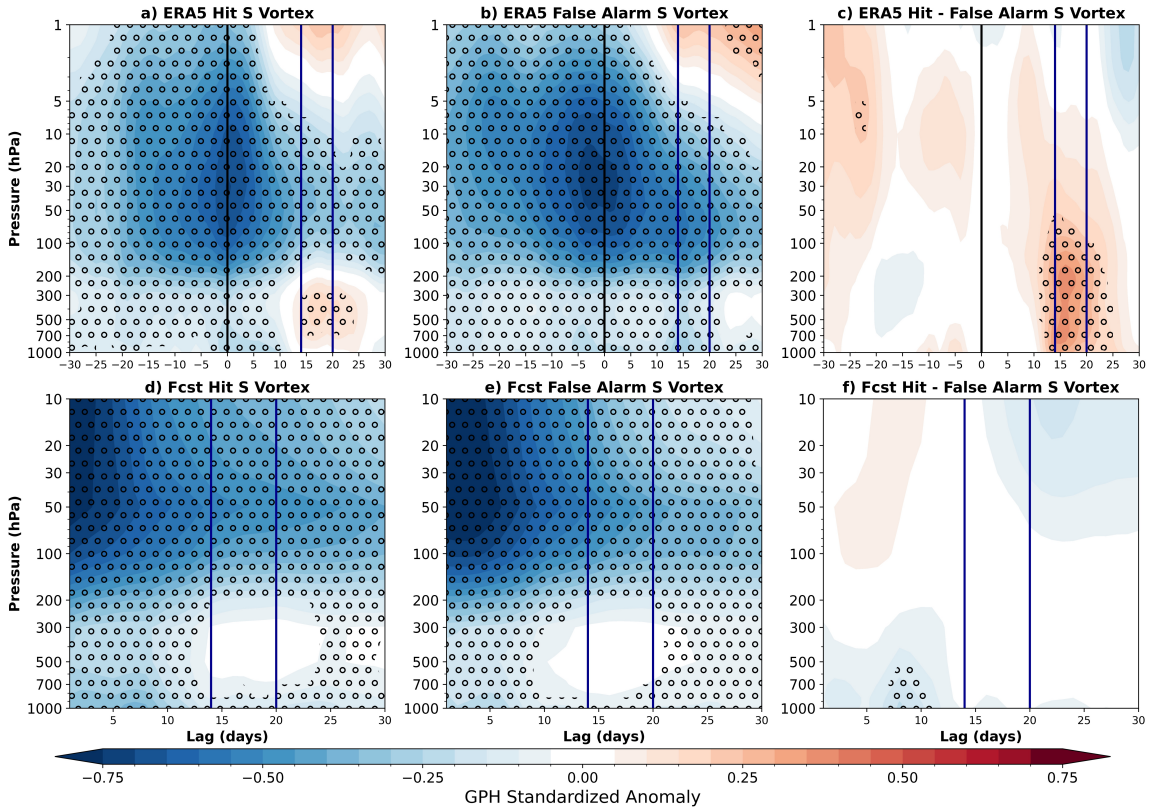


Figure 4.13: As in Fig. 4.5 but for forecasts initialized in strong polar vortex states.

4.7 Summary and Conclusions

This study provides a comprehensive assessment of the predictability of cold spells in six S2S models over 1993–2020. We find that cold spell forecasts, defined by weekly T2M anomalies below the lower tercile, exhibit a rapid degradation in HSS, POD, FAR, and MR beyond Week 2 (Fig. 4.1). The BOM model generally exhibits the lowest skill, perhaps associated with its lower model lid and coarser resolution. Nevertheless, forecasts at these leads still retain some resolution and sharpness (Fig. 4.2), allowing for meaningful analysis of confident hits and false alarms.

Our analysis reveals that confident hit Week 3 cold spell forecasts are characterized by a consistent dynamical evolution: (1) enhanced ArH regime frequency prior to onset (Figs. 4.3a, 4.4a, and 4.4b), (2) persistence of ArH, AkR, and WCR regimes during the cold spell (Figs. 4.4a and 4.4b), and (3) a gradually weakening SPV signal propagating downward

subseasonally (Figs. 4.5a and 4.5d). These findings align with prior work, including Chapter 2, linking ArH-dominated regimes and stratospheric weakening to CAO development (Lee et al. 2019b; Millin et al. 2022). In contrast, confident false alarms typically lack a coherent stratospheric precursor and are instead associated with shallow, tropospheric ridging (Figs. 4.5b and 4.5e), underscoring the role of stratosphere–troposphere coupling in enabling skillful subseasonal cold spell forecasts.

Stratifying forecasts by SPV state at initialization further reveals that weak SPV conditions are associated with improved reliability in Weeks 3–4, whereas strong SPV conditions tend to degrade forecast quality (Fig. 4.7). This echoes earlier findings over Europe (e.g., Erner et al. 2022) and extends them to the central US. Specifically, weak SPV confident hits are distinguished by earlier and more persistent SPV weakening and sustained ArH occupation, whereas weak SPV false alarms lack this long-lived precursor signal (Figs. 4.10 and 4.12). Meanwhile, strong SPV confident hits resemble wave-reflection-driven CAOs, showing an isolated strong SPV signal near initialization and forecast transitions to AkR or WCR. In contrast, strong SPV false alarms show an unrealistic regime transition from PT to AkR, highlighting model deficiencies in representing regime evolution under strong vortex conditions (Figs. 4.11 and 4.13).

These findings emphasize the potential of the SPV state as a key predictor of subseasonal forecast confidence. In particular, a weakened SPV may offer a "window of opportunity" for skillful Week 3–4 cold spell forecasts, especially when paired with a persistent ArH. Future research should explore how wave reflection and regime transition probabilities influence forecast evolution. Operationally, incorporating SPV state and regime tendencies into probabilistic forecast interpretation could enhance the reliability of extended-range cold spell guidance in North America. Indeed, we will next show in Chapter 5 that accurate representation of anomalous ridging over Alaska and northern Canada/Greenland, associated with the AkR and ArH regimes, enhanced the subseasonal forecast of the February 2021 CAO, highlighting the critical role these regimes play in predicting cold spells on subseasonal timescales.

While our results suggest a possible role for wave reflection in strong SPV cases, we did not explicitly quantify this mechanism. Future work should compute RI composites (e.g., Messori et al. 2022) to assess its role more directly. Additionally, incorporating more models from the full S2S Project database could improve robustness of the results, with the caveat of differing ensemble sizes and hindcast periods. Fig. 4.8 also suggests the potential for compound cold spells across North America and Eurasia; the role of a weak SPV in such events merits further investigation. Finally, while we focused on regime and stratospheric influences, future studies should examine the role of other modes of variability, particularly the MJO, in modulating cold spell predictability. For example, Büeler et al. (2021) demonstrated that European weather regime forecast skill improves following MJO phases 4 and 7, but is reduced after phase 2. This shows the value of examining MJO influences not only on central US cold spell prediction, but also on the associated North American weather regimes.

Chapter 5

Subseasonal Predictability of the February 2021 Cold Air Outbreak

5.1 Introduction

In earlier chapters, we identified 37 wintertime CAOs in the Great Plains from 1950 to 2021. Through a series of composite analyses, nudging experiments, and a comprehensive S2S model evaluation of cold spells, we found that stratosphere–troposphere coupling, weather regimes, and tropical teleconnections each play a notable role in both the development and skillful prediction of anomalous cold in the central United States. In particular, a key finding of Chapter 2 was that the recent February 2021 CAO was found to be a two-week event that was the second most severe CAO in this period based on magnitude and longevity combined. This CAO led to extreme cold across the central US, causing power failures, traffic incidents, widespread winter storm warnings, and various cold-related fatalities (NWS 2021a; BBC News 2021). Texas was impacted particularly hard due to failures of power generation, natural gas supply, and water infrastructure (Doss-Gollin et al. 2021). Given the high degree of severity of this event, there is great importance in understanding potential sources of atmospheric predictability in both observations and models that could improve forecasts of such events on the S2S timescale (Vitart and Robertson 2018).

Significant attention has been given to the role of the SPV in the February 2021 CAO, given that a *major* SSW occurred around one month prior (Lee 2021). However, recent studies concluded that the stratosphere had a minimal direct role in the February

2021 CAO; rather, tropospheric variability may have been more important in its evolution (Bolinger et al. 2022; Davis et al. 2022; Zhang et al. 2022). One tropospheric feature that is often associated with extreme cold is strong, anomalous mid-to-high latitude ridging, termed "blocking" when quasi-stationary and persistent (Schwierz et al. 2004; Nakamura and Huang 2018). Anomalous pan-Arctic or Alaskan ridging are example of such anticyclonic blocking patterns (e.g., Cellitti et al. 2006; Konrad and Colucci 1989; Lee et al. 2019b; Stan and Straus 2007), both of which we find in later sections also impacted the February 2021 CAO. Early papers found that the development of atmospheric blocking dominantly initiates in the Atlantic and Pacific regions, has a stable impact on temperature and precipitation patterns, and is linked with a persistent, quasi-stationary disruption of the zonal flow (Rex 1950a,b, 1951). Blocking anticyclones are often associated with Rossby wave breaking, whereby planetary-scale waves induce potential vorticity (PV) mixing, subsequently altering the large-scale flow (McIntyre and Palmer 1983; Huang and Nakamura 2013).

Rossby wave breaking is identified by the intrusion of low PV air into otherwise high PV environments and/or using potential temperature on constant PV surfaces (Weijenborg et al. 2012; Thornton et al. 2016; Lee et al. 2019a), and can have either anticyclonic or cyclonic configurations (Masato et al. 2011). There are two main characteristics resulting from Rossby wave breaking-induced blocking: (a) a dipole pattern with the anticyclone on the poleward side and a cyclone on the equatorward side, or (b) an amplified ridge with upstream and downstream troughing (Pelly and Hoskins 2003). Despite this knowledge, S2S forecast models struggle to capture well both the intensity and timing of blocking events (Quinting and Vitart 2019). Therefore, this chapter investigates the ability of two S2S models to capture the simulation of blocking regions in the forecast of the February 2021 CAO, with the goal of assessing the potential impact of blocking simulation on enhanced S2S predictability of CAOs.

5.2 Data and Methods

This chapter uses daily-averaged ERA5 data for January and February 2021 at a 0.5° horizontal resolution. Variables analyzed include: T2M, 500 hPa GPH, PV on the 320 K isentropic surface, and potential temperature on the 2 PV units, known as PVU (1 PVU = 10^{-6} m^2 s^{-1} K kg^{-1}), surface. The first three of these variables are also acquired (at a 1.5° horizontal resolution) from the ECMWF and National Centers for Environmental Prediction (NCEP) S2S real-time models (Vitart and Robertson 2018), where the Climate Forecast System Model Version 2 (CFSv2) is used by NCEP (ECMWF 2021). These models have 51 and 16 ensemble members, and 46 and 44 day lead times, respectively. Forecast initializations for both models are chosen to match the ECMWF initialization dates, from 4 January 2021 to 11 February 2021 initialized every Monday and Thursday. Lead-time dependent climatologies are developed to calculate model anomalies by using the ECMWF and NCEP models' 11-member and 4-member hindcast ensembles between 2001-2020 and 1999-2010, respectively.

We investigate the models' ability to detect the signal for extreme Great Plains cold during mid-February 2021 as follows. The extreme event threshold is calculated by taking the 10th percentile of the T2M anomaly from the climatology of hindcast initializations, which yields a lead-dependent extreme event threshold for a given model initialization and grid point, similar to Kautz et al. (2019). The choice of the 10th percentile is to remain in accordance with the threshold used to define Great Plains CAOs in Millin et al. (2022). To complement the extreme event probability analysis, we apply EOF and *k*-means clustering (Pedragose et al. 2011) analyses, as in Lee et al. (2019b), to the 500 hPa GPH anomalies over the North American domain (180-330°E, 19.5-79.5°N) for November through March between 1950-2021, but regridded to 1.5° horizontal resolution to match the S2S models. We then apply the clustering prediction to the S2S forecasts to enable the prediction of the five key weather regimes (Fig. 2.4) for the February 2021 CAO. Finally, we isolate ensemble members in ECMWF and NCEP that simulate two Rossby wave breaks more

successfully versus those that do not, to study the composite differences on the forecast of the CAO. More specifically, we average the 320 K PV over these two wave break regions, and separate those ensemble members that forecast the lowest and highest PV, where the lowest PV members are those that simulate the wave breaking more successfully than other ensemble members. Likewise, the high PV members are those that feature incorrect simulation of the wave break event.

5.3 Results

5.3.1 Evolution of the February 2021 cold air outbreak

The synoptic evolution of the February 2021 CAO shows signals for persistent blocking anticyclones with wave break signatures. Figure 5.1 presents T2M and 500 hPa GPH anomalies during February 2021 from ERA5 data (Hersbach et al. 2020). Anomalous ridging initiates in the North Pacific and extends into the East Siberian Sea between 1–4 February, resulting in the flow of cold T2M anomalies into northwest North America (Fig. 5.1a). This amplified ridge configuration, bookended by anomalous troughs, is similar to that associated with North Pacific wave breaking (Pelly and Hoskins 2003). Between 5-11 February, anomalous ridging intensifies over Greenland/northern Canada alongside anomalous North Pacific and Alaskan ridging (Fig. 5.1b), isolating negative temperature and height departures across Canada. Finally, exceptionally negative T2M anomalies infiltrate the Great Plains between 12-18 February, as the anomalous ridging area of Greenland/northern Canada pushes southward (Fig. 5.1c).

Based on the tropospheric evolution in Fig. 5.1, we investigate how the simulation of two Rossby wave breaks impact the S2S forecast of the February 2021 CAO. We hypothesize that the correct simulation of these features will aid the forecast of the CAO across the Great Plains on S2S timescales, especially in regard to its association with the evolution of high-latitude atmospheric ridging. Given that S2S models may struggle in simulating

the characteristics of these persistent ridging features (e.g., Quinting and Vitart 2019), this work has the potential to place such prior results in the perspective of an extreme cold event while isolating the need for enhanced S2S prediction of ridging in order to better forecast cold extremes.

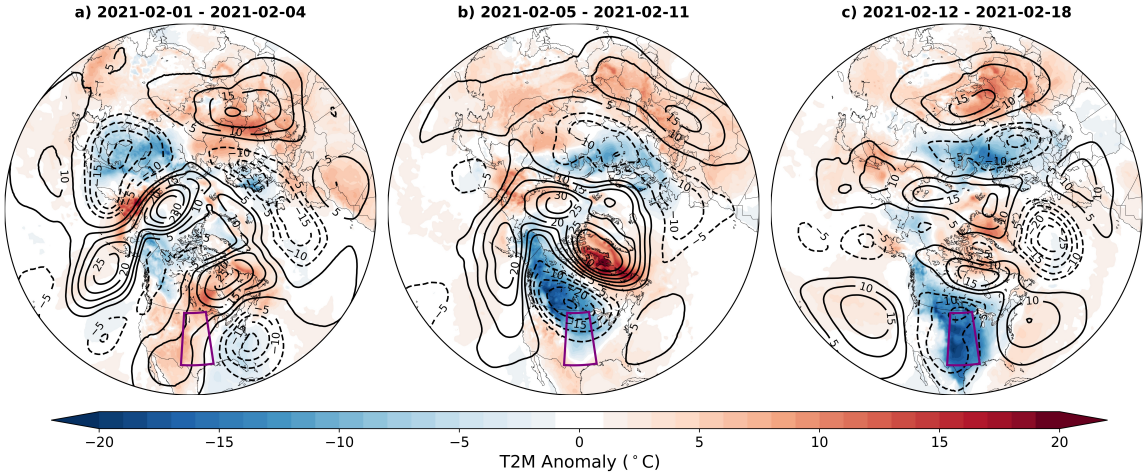


Figure 5.1: ERA5 500 hPa GPH anomalies (dam, contours) and T2M anomalies (°C, shaded) averaged between a) 1-4, b) 5-11, and c) 12-18 February 2021. Solid and dashed contours represent positive and negative GPH anomalies, respectively. The purple boxes represent the region defining the Great Plains (30-48°N and 256.5-268.5°E).

5.3.2 Connecting Rossby wave breaking events to anomalous ridges

Figure 5.2 shows the 320 K PV evolution of two crucial wave breaks in ERA5: 1) over the East Siberian Sea between 2-4 February 2021 (Figs. 5.2a – 5.2c) and 2) over the Labrador Sea between 9-11 February 2021 (Figs. 5.2d – 5.2f). The first wave break begins with an intrusion of low PV air extending from the North Pacific into the East Siberian Sea (Fig. 5.2a). Subsequently, this area of low PV air cuts off and overturns in the East Siberian Sea between 3-4 February (Figs. 5.2b and 5.2c), completing the wave break. Subsequently, an area of anomalous ridging forms (Fig. 5.1a), directing the flow of cold air into northwest North America. Analysis of this wave break in terms of potential temperature anomalies on the 2 PVU surface reveal it to be an anticyclonic wave break (Figures A.10a–A.10c, Masato et al. 2011).

The second wave break shows low PV air moving from the North Atlantic to the Labrador Sea on 9 February (Fig. 5.2d), on the northward side of a strong cyclone which previously was a northeast US winter storm (NWS 2021b; WPC 2022). Between 10-11 February, the low PV air is cut off and injected into northern Canada (Figs. 5.2e and 5.2f), subsequently intensifying the anomalous ridging across Greenland/northern Canada and translating it southward (Fig. 5.1c). This mechanism allows the previously loaded cold air in northwest North America to move into the Great Plains (Fig. 5.1c). Further analysis of this wave break in terms of potential temperature on the 2 PVU surface reveals it to be a cyclonic wave break (Figures A.10d – A.10f, Masato et al. 2011).

The evolution of the 320 K PV shows that these wave breaking features have an impact on the strong, anomalous areas of ridging that influence downstream temperature anomalies. We next investigate the extreme event probability and weather regime tendency of this CAO event to make connections to wave breaking events within the S2S models.

5.3.3 Extreme event probability and regimes

To isolate forecast features that may be sources for long-range predictability of the CAO, namely wave breaking, we first investigate the extreme event probability for the February 2021 CAO. Figure 5.3a shows the extreme event probability for T2M anomalies as a function of initialization date for both ECMWF (blue line) and NCEP (red line), averaged over 12-18 February 2021 and the Great Plains region (purple box Fig. 5.1). Extreme event probability is negligible at long lead times (Kautz et al. 2019). However, for the 28 January initialization, there is an increase in probability to $\sim$5% and $\sim$20% in ECMWF and NCEP, respectively. The extreme event probability decreases again for the 1 February initialization before increasing rapidly closer to the event (Fig. 5.3a). The increase in extreme event probability on the 28 January initialization raises the possibility that the models may be simulating important subseasonal predictors $\sim$2-3 weeks out.

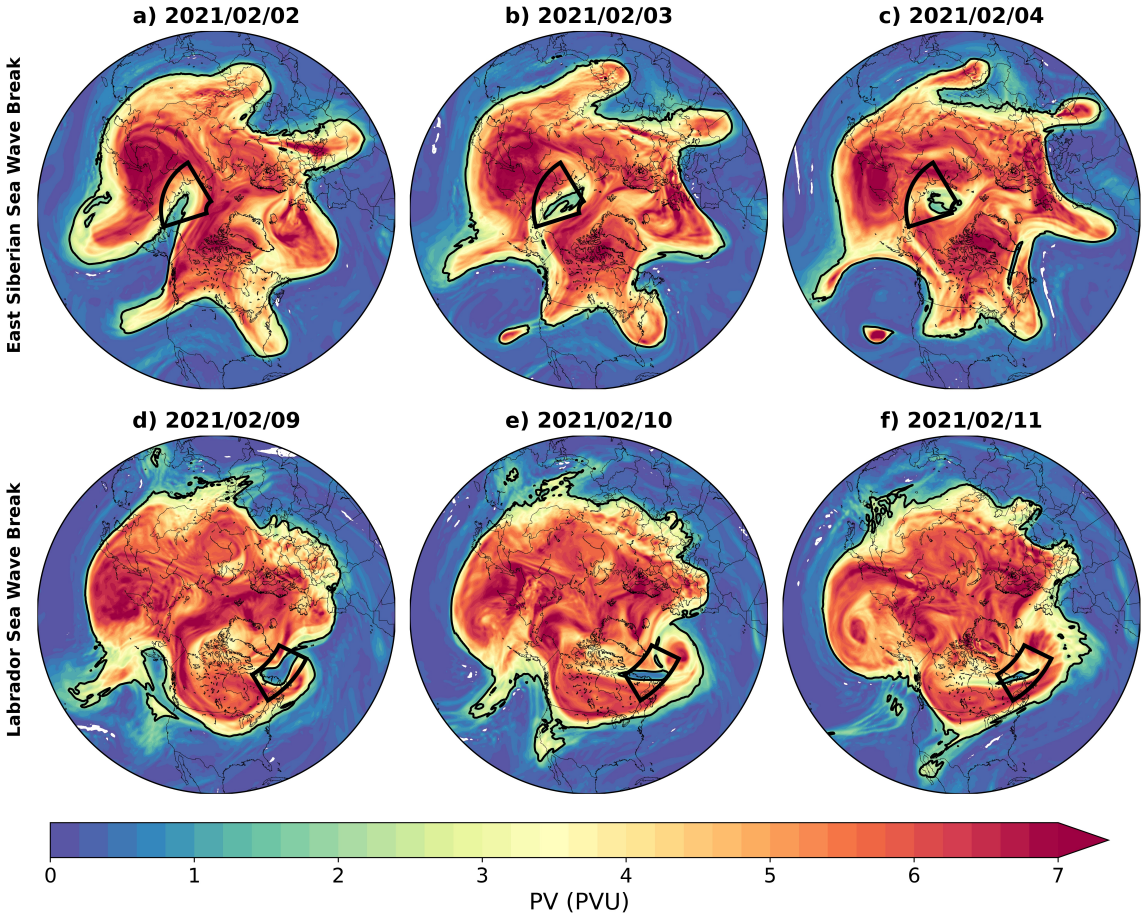


Figure 5.2: ERA5 320 K PV (PVU, shaded) for two wave breaks during the evolution of the February 2021 CAO: a) – c) an East Siberian Sea anticyclonic wave break between 2-4 February 2021 and d) – f) a Labrador Sea cyclonic wave break between 9-11 February 2021. The black contour represents the 2 PVU boundary. Black boxes in a) – c) and d) – f) represent the two wave break regions, respectively.

To give further insight into the predicted tropospheric flow patterns, Fig. 5.3b shows the averaged North American weather regime frequency (Lee et al. 2019b; Millin et al. 2022) as a function of initialization date, for the target period of 8-10 February 2021 (i.e., the period of strongest Greenland/northern Canada ridging). Most January initializations indicate a dominant ArL regime, which is not a preferential indicator for extreme cold across the US (Lee et al. 2019b; Millin et al. 2022). However, there is a shift toward the ArH regime, which features anomalous ridging across northern Canada/Greenland, for the 28 January initialization. This weather regime *is* a preferential pattern for extreme cold in the US and specifically the Great Plains (Lee et al. 2019b; Millin et al. 2022). This

regime shift begins on the same initialization that the extreme event probability first shows an increase, indicating that the 28 January 2021 initialization is a key lead time to study the simulation of tropospheric precursors to the CAO.

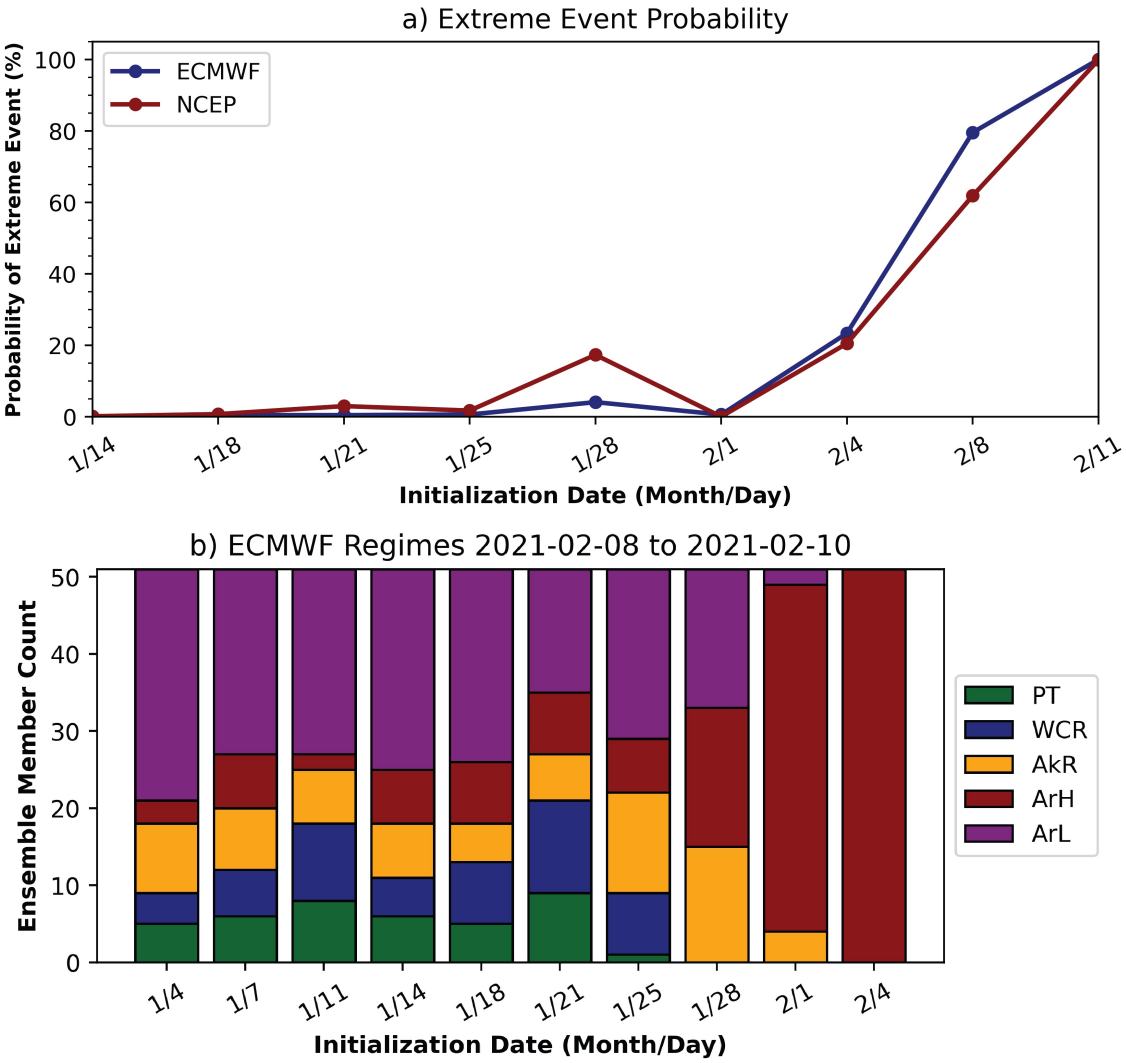


Figure 5.3: a) Probability of an extreme Great Plains T2M anomaly event between 12-18 February 2021 at the 10th percentile as a function of initialization date for ECMWF (blue line) and NCEP (red line). b) A bar plot of the ensemble distribution of 5 North American weather regimes for the 8-10 February 2021 as a function of initialization date within the ECMWF modelthe Pacific Trough (PT), West Coast Ridge (WCR), Alaskan Ridge (AkR), Arctic High (ArH), and the Arctic Low (ArL).

5.3.4 Simulation of wave breaks in subseasonal forecast models

5.3.4.1 East Siberian Sea wave break

To assess the impact of the first anticyclonic wave break on the forecast of this CAO, 320 K PV is first averaged over the East Siberian Sea region (black boxes in Figs. 5.4a – 5.4c) between 2-4 February 2021 for the 28 January initialization in both ECMWF and NCEP S2S models. The $\sim$10% of ensemble members with the lowest and highest PV in that box are then separated, corresponding to those ensemble members that simulate the anticyclonic wave break more successfully versus those that do not, respectively. This procedure yields 5 ensemble members in each category for ECMWF (3 members per category were used for NCEP, which has one third the number of ensemble members). As seen in Figs. 5.4a to 5.4c, the 5 members with the lowest PV show a cut-off area of low PV ($<$ 2 PVU) in the East Siberian Sea, simulating this wave break well 5-7 days out (Fig. 5.4a). Conversely, the 5 ensemble members with the highest PV in the East Siberian Sea feature no indication of an anticyclonic wave break (Fig. 5.4b). Subsequently, the composite difference between low and high PV members shows a statistically significant ($p < 0.1$) negative difference in the East Siberian Sea. PV maps for NCEP show very similar results (Figs. A.11a through A.11c).

Figures 5.4d – 5.4f and 5.4g – 5.4i show the composite T2M (shading) and 500 hPa GPH (line contours) anomalies between 14-16 February 2021 for the low PV (i.e., best East Siberian Sea wave break simulation) and high PV (i.e., poorest East Siberian Sea wave break simulation) ensemble members in ECMWF, respectively, for the 28 January initialization. On 14 February, the low PV members exhibit negative T2M anomalies from northwest Canada into the US Great Plains, accompanied by anomalous troughing in northern Canada and anomalous ridging in northeast North America, both of which persist through to 16 February (Figs. 5.4d – 5.4f). These negative temperature anomalies are significantly different ($p < 0.1$) from the high PV members across the Great Plains, which instead show anomalous ridging and positive T2M anomalies across the US through the period (Figs.

5.4g – 5.4i). Therefore, the members which produced the wave break in the East Siberian Sea facilitate the downstream pattern that is more favorable for the evolution of the CAO into the US Great Plains. Results from the NCEP model are qualitatively similar, though the negative T2M anomalies for the low PV members spread further eastward (Figs. A.11d – A.11i).

5.3.4.2 Labrador Sea wave break

We repeat the analyses from Fig. 5.4 for the wave break on 9-11 February 2021 over the Labrador Sea (Figs. 5.5a – 5.5c). The low PV composite shows a tongue of low PV air through the Labrador Sea, north of high PV air likely associated with cyclone development (Fig. 5.5a; Pelly and Hoskins 2003), as in Fig. 5.2d. This composite is in contrast to that of the high PV members, which features no wave break in the Labrador Sea (Fig. 5.5b), resulting in a statistically significant $(p < 0.1)$ negative composite difference (Fig. 5.5c). Comparable results are found with the NCEP ensemble members (Figures A.12a – A.12c).

Turning to the T2M and 500 hPa GPH anomalies (Figs. 5.5d – 5.5i), the lowest PV members for the Labrador Sea wave break region replicate well the strong southward push of negative T2M anomalies into the Great Plains, with anomalous ridging across much of Canada and troughing in western North America (Figs. 5.5d – 5.5f). The T2M anomalies are especially similar to the event evolution (Fig. 5.1). These results are mirrored in NCEP (Figs. A.12d–A.12f). In contrast, the high PV ensemble members produce a similar anomalous ridging pattern over the southeast US to that from the high PV ensemble members in the East Siberian Sea wave break composites (compare Figs. 5.5g – 5.5i with Figs. 5.4g – 5.4i). Indeed, between 14-15 February, *positive* T2M anomalies exist over the southern Great Plains for the high PV ensemble members, with the negative T2M and GPH anomalies confined to northwest North America (Figs. 5.5g and 5.5h). The equatorward surge of cold air is not widely seen in the high PV ensemble members for the NCEP model either (Figs. A.12g–A.12i). The lack of a focused, anomalous ridge across the Greenland/northern

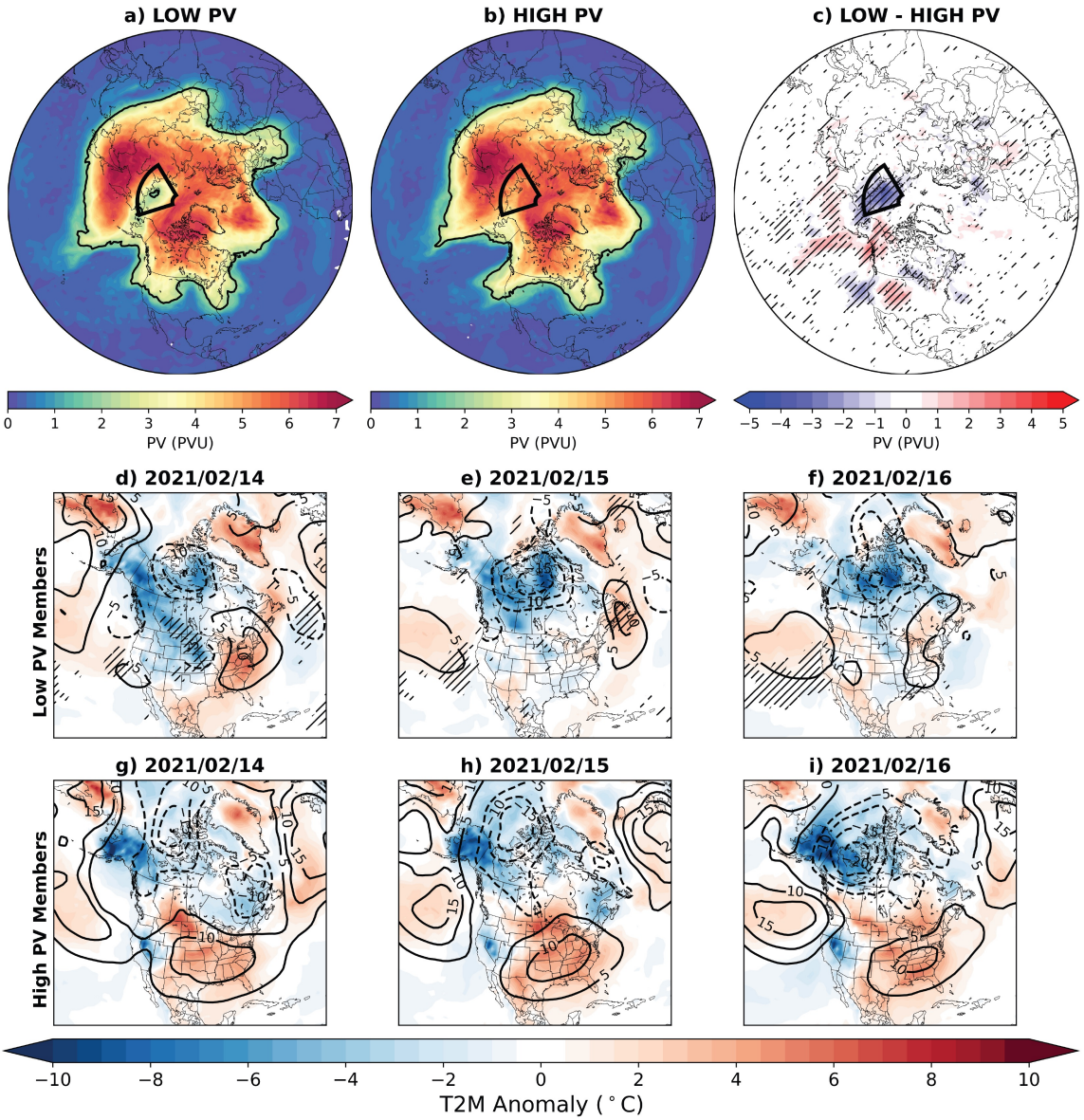


Figure 5.4: 28 January 2021 initialization ECMWF 320 K PV composites (PVU, shaded) for the 5 ensemble members with the a) lowest and b) highest PV in the East Siberian Sea region (black box) averaged over 2-4 February 2021. c) The difference (PVU, shaded) between the low and high PV composites. d) – f) Composites of T2M (°C, shaded) and 500 hPa GPH (dam, line contours) anomalies for the 5 ensemble members with the lowest PV in the East Siberian Sea region. g) – i) As in d) – f) but for the 5 ensemble members with the highest PV in the East Siberian Sea region. Hatching in panels c) – f) denote statistical significance ($p < 0.1$) in the composite difference (low - high PV) in the PV and T2M fields, computed by a 5000-iteration bootstrapping technique, without replacement. Bold dates at the top of panels d)–i) indicate the valid date for the forecasts.

Canada region helps to control the flow of cold air, with the difference between the low and high PV composite T2M anomalies showing statistically significant ($p < 0.1$) differences

across the Great Plains and the Arctic region (Figs. 5.5d – 5.5f). Overall, this analysis suggests that the Labrador Sea wave break is important for enabling the southward migration of cold air into the Great Plains, and is arguably the more important of the two wave breaks.

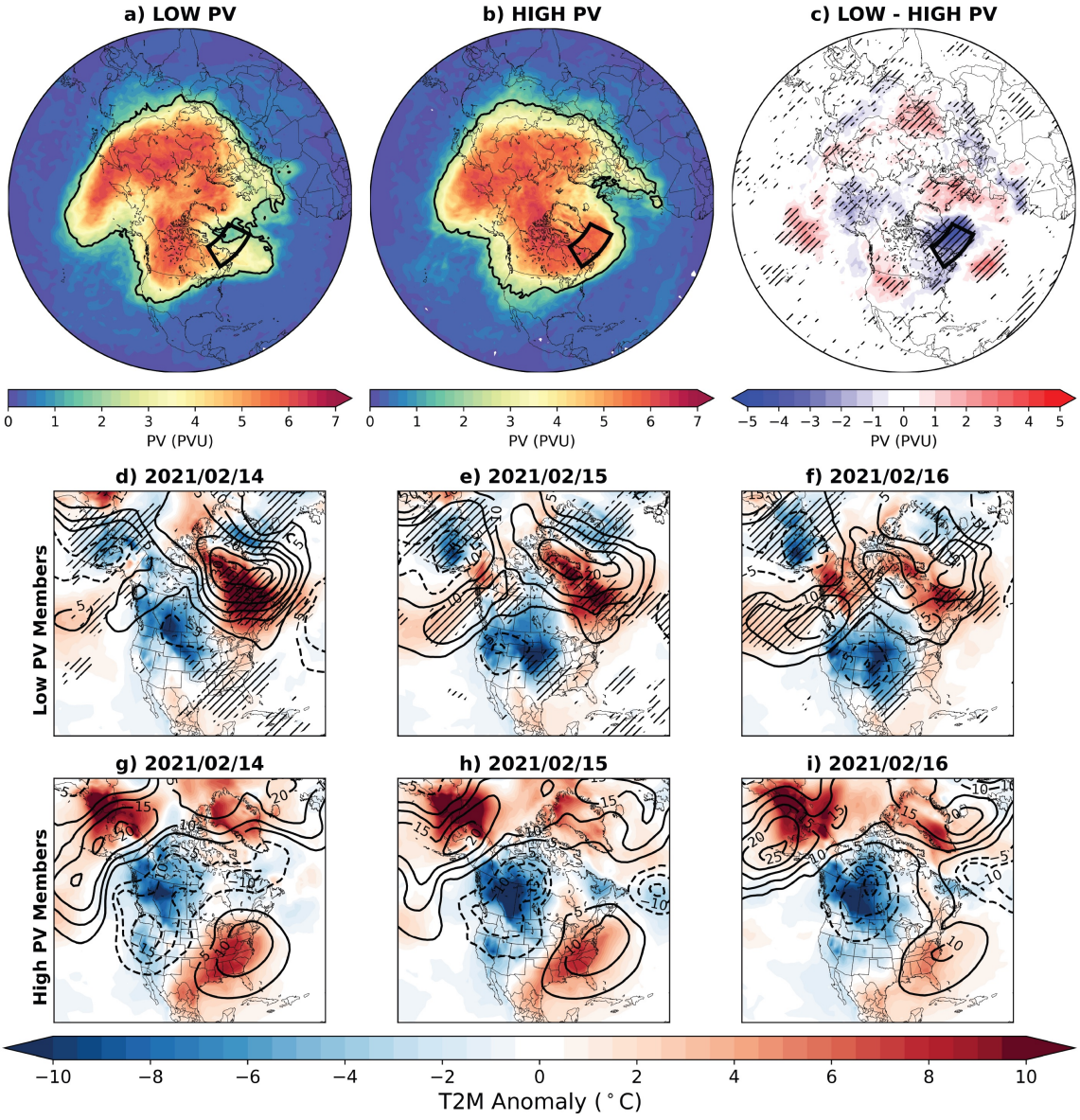


Figure 5.5: As in Fig. 5.4 but for the Labrador Sea wave break between 9-11 February 2021.

As an additional analysis, we also separated the ECMWF ensemble members into those featuring the lowest PV (<25th percentile) in both the East Siberian Sea *and* Labrador Sea wave break regions vs. the highest PV (>75th percentile) in both regions. The lowest

PV scenario features a tendency to draw cold anomalies south, with a more southward-extending Arctic high (Figure 5.6), demonstrating the importance of these two events in the forecast. Unfortunately, any further conclusions are difficult to make from this analysis due to the constraint on sample size in ensemble members.

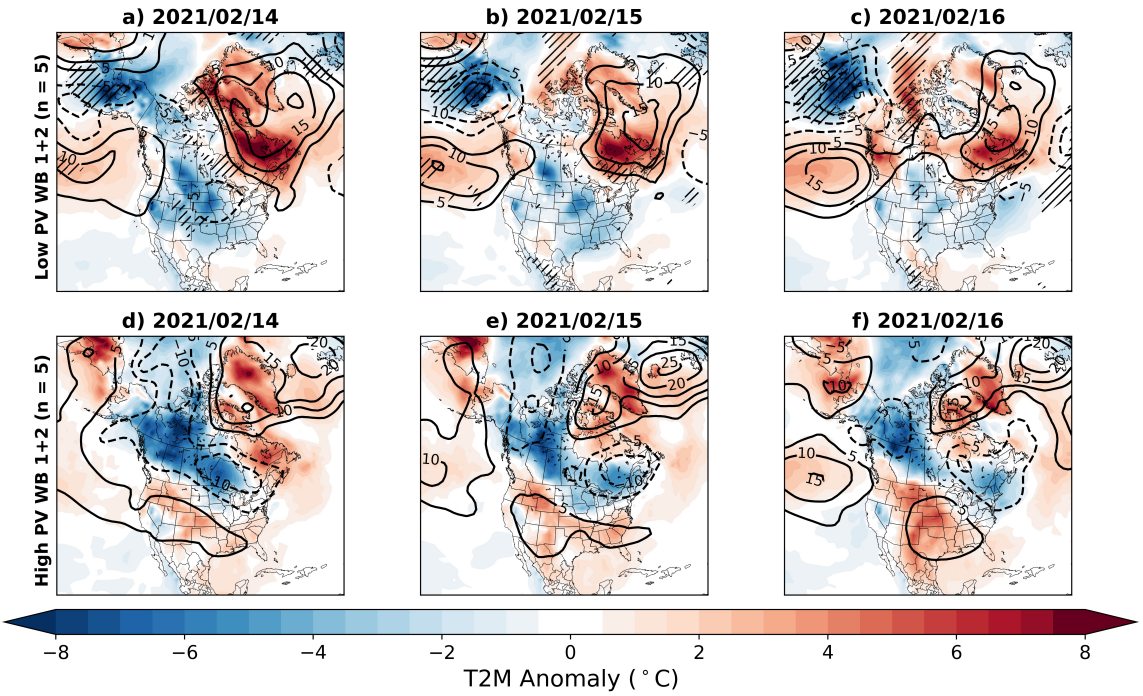


Figure 5.6: 28 January 2021 initialization ECMWF T2M anomaly (°C, shading) and 500 hPa geopotential height anomaly (dam, contours) composites between 14-16 February 2021 for ensemble members that a) – c) simulate both the East Siberian Sea and Labrador Sea wave breaks best (PV $<$ 25th percentile) and for ensemble members that d) – f) simulate both the East Siberian Sea and Labrador Sea wave breaks least effectively (PV $>$ 75th percentile). Solid contours represent positive anomalies and dashed contours represent negative anomalies. Hatching denotes statistical significance $(p < 0.1)$ in the composite difference (low - high PV) in the T2M field, computed by a 5000-iteration bootstrapping technique, without replacement. Bold dates at the top of panels a) – f) indicate the valid date for the forecasts.

5.4 Summary and Conclusions

The February 2021 CAO was a high-impact extreme weather event, the predictability of which is important to quantify and understand in operational S2S models. We find that two

significant Rossby wave breaks, one between 2-4 February 2021 in the East Siberian Sea and the second over the Labrador Sea between 9-11 February 2021 (Fig. 5.2), played a role in the evolution and forecast of this CAO. Ensemble members from both the ECMWF and NCEP models which correctly simulate these two wave breaks later simulate anomalous cold in the Great Plains in comparison to those ensemble members which do not, at a $\sim$3-week lead time (Figs. 5.4 and 5.5). These findings link with recent studies that suggested the February 2021 CAO was a result of tropospheric variability (Bolinger et al. 2022; Davis et al. 2022; Zhang et al. 2022). The analysis of the Rossby wave breaking events preceding the February 2021 event also demonstrates their links to persistent anomalous ridges over the East Siberian Sea and northern Canada/Greenland (Pelly and Hoskins 2003; Masato et al. 2011). The disparity between ensemble forecasts illustrates how S2S models can struggle to represent wave breaking occurrence and subsequent circulation/blocking features (e.g., Gray et al. 2014; Quinting and Vitart 2019). Thus, S2S forecasts of extreme weather events (e.g., heat waves, CAOs, heavy rainfall) may be improved by closely examining the entire ensemble suite, carefully considering the probability of occurrence of anomalous wave breaking events, and analyzing their role in enhancing extreme event likelihood in North America (e.g., Carrera et al. 2004; Guirguis et al. 2011; Huang et al. 2021; Lin et al. 2022). Other important forcing mechanisms (e.g., remote tropical forcing) can also contribute to changes in the large-scale circulation and thus the evolution of an extreme weather event. Furthermore, although representation of blocking from wave breaking may be improved by increased resolution (Gray et al. 2014; Schiemann et al. 2017; Quinting and Vitart 2019), targeted improvements in dynamical representation of physical processes, such as diabatic heating, may be key to unlocking S2S model performance (Gray et al. 2014; Quinting and Vitart 2019).

Future work should address how wave breaking simulations link to the development of large extratropical cyclones within S2S models, and whether dynamical/regional dependencies lead to ensemble disparity. Long-range skillful prediction of wave-breaking events

may also be tied to the models' ability to represent accurately the preceding hemispheric-scale circulation pattern. Pattern correlation analysis of the hemispheric-scale 500 hPa GPH anomalies between the ensemble members and reanalysis offers mixed results on this hypothesis, with conflicting results between the two wave breaks (not shown). Furthermore, the connections between the extreme US cold and tropopause polar vortex interactions (e.g., Cavallo and Hakim 2009, 2010, 2013) should be further studied for these events, since these features are known to facilitate mid-latitude CAOs (e.g., Papritz et al. 2019; Lillo et al. 2021; Biernat et al. 2021). Finally, applying our methodology to other CAOs and operational S2S models for other CAO cases would test the robustness of our findings.

Chapter 6

Application to Industry: Central United States Cold Air Outbreaks, Electricity Load, and Weather Regimes

The previous chapters found an array of important S2S precursors to extreme cold in the central US, alongside diagnosing their predictability in S2S forecasts and highlighting the importance of the stratosphere in the development of anomalous cold, both in observations and subseasonal forecasts. A highlight of the prior chapters is the importance of the North American winter weather regime both prior to and during extreme cold events in the central US. As stated previously, extreme cold events naturally place a strain on the electrical grid due to increased demand, and this was especially apparent during February 2021, as noted in the last chapter. Thus, we now provide a cross-disciplinary chapter to investigate the impact of North American winter weather regimes on central US electricity demand, tying our findings to extreme cold in the central US.

6.1 Introduction

Temperature is a key driver of variability for electricity demand, or consumption, across the world (Hor et al. 2005; Bessec and Fouquau 2008; Thornton et al. 2016; Cawthorne et al. 2021). Moreover, the exposure of electrical grids to warm season temperature variability is more apparent now due to anthropogenically-driven temperature increases and more frequent or prolonged heatwaves, leading to greater use of air conditioning (Perkins et al. 2012; Perkins-Kirkpatrick and Lewis 2020). However, wintertime extreme cold conditions can exhibit similar electrical demand surges associated with the increased use of heating systems (Doss-Gollin et al. 2021; Shaffer et al. 2022). Such strain of the electrical grid

has become notable during several CAOs that have occurred in recent years (Millin et al. 2022).

As we previously mentioned, CAOs often result in substantial socioeconomic and infrastructural impacts (e.g., Analatis et al. 2008; Smith and Sheridan 2019). One of the main impacts from these events is on the electrical grid through excessive demand and grid failures, such as in February 2021 (Doss-Gollin et al. 2021). The failures in natural gas production and electricity generation left millions of Texans without heat for an extended period, with power outages also reported throughout states in the Southwest Power Pool (SPP), which is a regional transmission organization that coordinates the electrical grid and power market in the central US (Doss-Gollin et al. 2021; Bolinger et al. 2022). Therefore, we need to understand how wintertime temperature affects central US electricity demand to better predict large-scale effects on the electrical grid, which this chapter will address.

Understanding the demand-temperature relationship could be key to those making decisions on energy trading, grid winterization, and preventing potential energy shortfalls on S2S lead times (Vitart and Robertson 2018). At these longer lead times, skillful predictions would aid societal impact mitigation and critical infrastructure protection. As such, this chapter uses North American winter weather regimes, that correlate with severe CAOs, to quantify their impact on extreme electricity demand in the SPP. The longer timescale nature of weather regimes, and their connection to extreme weather and S2S modes of variability such as the SPV, makes them appealing for extended-range predictions. Weather regimes have been shown to directly influence renewable energy production and demand in Europe and North America (van der Wiel et al. 2019; Garrido-Perez et al. 2020; Liu et al. 2023), and can even provide enhanced subseasonal ensemble forecast skill of European energy variables (Bloomfield et al. 2021), showcasing their usefulness to the energy sector (Grams et al. 2017; Millin et al. 2022).

Previous literature indicates a non-linear connection between temperature and electricity demand and consumption, often described by a parabolic relationship where cold and warmer ends of the temperature distribution drive increases in electricity demand, typically

in the winter and summer, respectively (e.g., Moral-Carcedo and Vicéns-Otero 2005; Mirasgedis et al. 2006; Bessec and Fouquau 2008; Deschênes and Greenstone 2011; Davis and Gertler 2015; Thornton et al. 2016; Auffhammer et al. 2017). However, limited research has focused on the impact of wintertime temperature on central US electricity demand and the associated cold-induced peak load variability. This chapter contributes to this gap by investigating how North American winter weather regimes correlate with electricity demand, highlighting the usefulness of regimes in predictive models of peak energy demand. Despite increased study of North American weather regimes, their influence on the electrical grid is relatively unexplored, but is an important aspect of their research-to-operation utilization potential. As such, advanced knowledge of the weather regime could be a primary driver for predicting electricity demand in the central US and allowing utility operators to anticipate extreme load demands in the winter.

6.2 Data and Methods

6.2.1 Electricity load data and North American weather regimes

To identify North American winter weather regimes, we follow the aforementioned method of Millin et al. (2022) in Section 2.2.3, which is an extended version of the regime identification used in Lee et al. (2019b), but for 1950–2023. As before, 500 hPa GPH anomalies are used to classify the regimes, obtained from ERA5 (Hersbach et al. 2020) data for November through March between 1950-2023 on a 0.5° horizontal resolution. Each day in the record is then assigned one of these five regimes based on minimum Euclidean distance from the cluster centroid.

We also collect information on hourly electricity demand for each balancing authority within SPP from Cicala (2022) for 1999–2010 and from SPP for 2011-2022. Cicala (2022) collects electricity demand data from Federal Energy Regulatory Commission (FERC) Form 714 and uses Least Absolute Shrinkage and Selection Operator (LASSO) to estimate

demand based on weather, population, and employment where the FERC data is missing or incomplete. Cicala (2022) cross-validates this LASSO estimation, finding that predicted load values are within 4% of their realized values. Our final sample includes the 13 balancing authorities (representing 91% of average annual SPP load) within SPP that could be consistently matched between the load data sources (Table 6.1). Specifically, we study peak load in 13 balancing authorities within SPP across parts of Oklahoma, Kansas, Missouri, Texas, Arkansas, New Mexico, Louisiana, and Nebraska. We keep all observations from November to March of each year during 1999-2022 and compute peak daily load as the maximum amount of electricity demanded in the SPP footprint on each day. We also calculate peak load anomalies for the SPP with respect to the 1999-2022 average, where we remove the seasonal cycle from each day.

To account for trends in the size of the population served by utilities operating in the SPP footprint, we normalize peak daily load by annual customer counts from Energy Information Administration (EIA) Form 861. Since utility names change throughout our sample period as firms merge or disband, we used administrative SPP reports to consistently document customers in the SPP footprint over time.

6.2.2 Polynomial model

To capture the non-linear connection between peak daily load anomaly and ERA5 T2M (both daily average and daily minimum), a linear regression is not suitable. As such, we fit a second degree polynomial (i.e., quadratic) function to the data to better describe the parabolic connection between the two variables. The coefficient of determination (R^2) for this model is computed by first calculating the sum of squares total (SST):

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{6.1}$$

Balancing Authority (1999-2011)	Balancing Authority (2012-2022)
Kansas City Power and Light	Kansas City Power and Light
Kansas City Power and Light	City of Independence, MO
Western Resources	Western Resources
Kansas City Public Utilities	City of Kansas City
Southwestern Electric Power Co.	Southwestern Public Service Co.
Oklahoma Gas and Electric	Oklahoma Gas and Electric
Southwestern Public Service Co.	Public Service Co. of Oklahoma
Public Service Co. of Oklahoma	Public Service Co. of Oklahoma
Western Plains Energy Kansas	Sunflower Electric Power Corporation
Sunflower Electric Power Corporation	Sunflower Electric Power Corporation
Western Farmers Electric Cooperative	Western Farmers Electric Cooperative
Empire District Electric Co.	Empire District Electric Co.
Nebraska Public Power District	Nebraska Public Power District
Omaha Public Power District	Omaha Public Power District
KCPL Greater Missouri Operations Co.	Evergy Missouri, West
Grand River Dam Authority	Grand River Dam Authority

Table 6.1: Corresponding balancing authority regions between two sources of electricity demand data. Correspondence is based on the authors' review of mergers and acquisitions documentation and administrative reports from SPP.

where y is the observed daily peak load anomaly and overbars denote the mean. Secondly, the sum of squared errors (SSE) between the predicted and observed load data is then calculated via the following equation:

$$SSE = \sum_{i=1}^{n} (y_i - f(T_i))^2 \tag{6.2}$$

where T is the daily average or daily minimum T2M, and f represents the polynomial function being applied. Finally, the coefficient of determination is then calculated by computing the difference between 1 and the division between the SSE and SST (Wilks 2011).

$$R^2 = 1 - \frac{SSE}{SST} \tag{6.3}$$

6.2.3 Probability density functions and risk ratios

To investigate the distributions of daily peak load anomalies by weather regime, we use Gaussian kernel density estimation to calculate the probability density functions, with 5-fold cross validation to select the optimal bandwidth. Additionally, we calculate the mean and standard deviation of the daily peak load anomalies in each regime, alongside the skewness and the Fisher (excess) kurtosis (i.e., normal kurtosis = 0). To explore the relationship between North American winter weather regimes and extreme daily peak load anomalies, we calculate risk ratios for exceeding three percentile thresholds: the 90th, 95th, and 97.5th percentiles. The risk ratio describes how much more likely a daily peak load extreme event is in each regime. To compute the risk ratios, we calculate the percentile thresholds in the full distribution of daily peak load anomalies and then calculate the probability of exceeding this threshold in each of the peak demand categories that are split by weather regime. The risk ratio, RR, is then calculated as in van der Wiel et al. (2019):

$$RR = \frac{P_{WR}}{P_{clim}} \tag{6.4}$$

where P_{WR} is the probability of an extreme peak load day in a given weather regime and P_{clim} is the probability of an extreme peak load day in the full distribution.

6.3 Results

6.3.1 North American winter weather regimes

To begin, we redefine the North American winter weather regimes using ERA5 but in the latest period of 1950–2023, using the k-means clustering analysis used in prior studies and prior chapters (Lee et al. 2019b; Millin et al. 2022). The five regimes are the same as

that in Chapter 2 (Fig. 2.4) but with slightly different occupation frequency statistics. The composite maps of these regimes can be seen in Figure A.13, where we have overlaid mean sea level pressure (MSLP) contours.

The five North American winter weather regimes are associated with distinct near surface temperature patterns (Fig. 6.1). The AkR regime is associated with the coldest weather pattern, with highly negative T2M anomalies extending from Canada into the central US, and across the SPP (Fig. 6.1a). Indeed, the AkR regime is linked to almost a third of the most extreme central US CAOs since 1950, due to the persistent nature of the anomalous Alaskan ridging (Chapter 2 and Millin et al. 2022) and its associated surface cold air advection from northwest North America, as shown in the MSLP field (Fig. A.13a). Cold signals through the central US are also present during WCR regime days, extending into the eastern US due to a north-to-northwesterly flow at the surface (Figs. A.13d and 6.1d). Furthermore, while the ArH demonstrates anomalous ridging which can produce blocking highs that promote CAO development (Lee et al. 2019b; Millin et al. 2022), negative T2M anomalies are of a lesser magnitude across the central US and SPP associated with a surface flow from a more north-to-northeasterly origin (Figs. A.13b and 6.1b). Finally, the PT and ArL winter weather regimes produce anomalous warmth across much of the US due to large-scale continental anomalous ridges and advection of milder maritime air onto the continent (Figs. 6.1c and 6.1e). Overall, North American winter weather regimes have clearly distinct near-surface temperature influences across the central US, which are likely to influence electricity load across the SPP.

6.3.2 Wintertime 2 m temperatures and peak electricity load

Before examining the relationship of electricity load with North American winter weather regimes, we explore how near-surface temperature in the central US relates to electricity load in this region. To quantify this relationship, we combine hourly electricity load

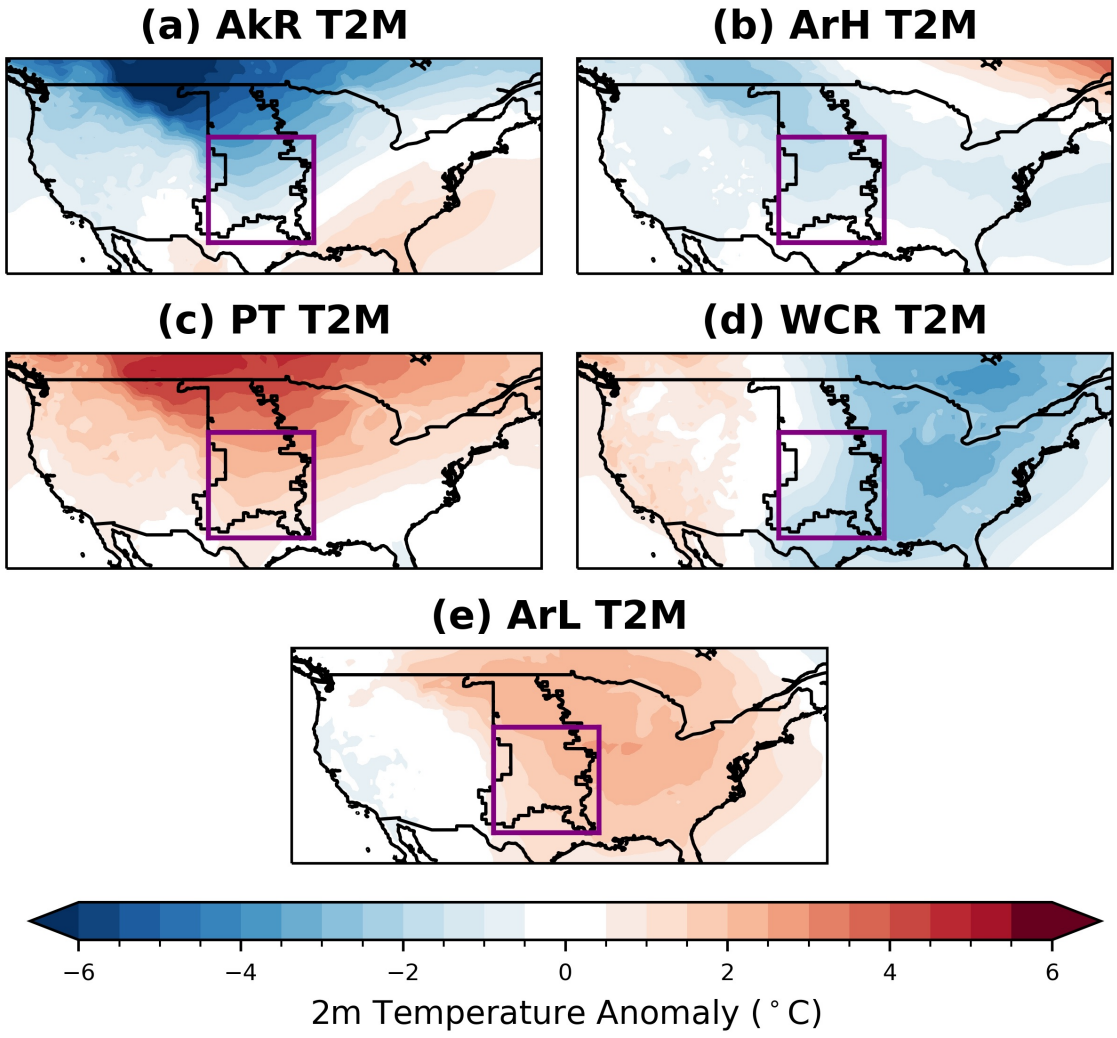


Figure 6.1: As in Fig. A.13 but for ERA5 linearly detrended T2M anomalies (shading, °C). The black lines show the region of the entire Southwest Power Pool (SPP) while the purple box (31-43°N, 256-268°E) represents the region of study for electricity load data in the SPP and the area in which T2M is averaged for Fig. 6.2. A sub-domain of the entire SPP region is used for analysis as it represents the 13 divisions within the SPP that could be consistently matched between load data sources (see methods). Anomalies are with respect to the 1979-2020 climatology also used to calculate the regimes.

data throughout the SPP across 13 balancing authorities between 1999–2022 for November through March (Table 6.1), scaled by yearly customer counts. The relationship between T2M and peak load anomalies exhibits a parabolic shape, a conclusion found in numerous prior studies across the globe (Moral-Carcedo and Vicéns-Otero 2005; Mirasgedis et al. 2006; Bessec and Fouquau 2008; Auffhammer et al. 2017; Thornton et al. 2016; Deschênes

and Greenstone 2011). Both the daily mean (Fig. 6.2a) and daily minimum (Fig. 6.2b) T2M show a similar relationship with SPP peak load anomalies, with coefficient of determination (R^2) values of 0.38 and 0.39, respectively. Thus, while we show that a quadratic polynomial model represents well the peak load anomaly as a function of T2M parameters, it could be argued that the extreme cold end of the distribution features the weakest non-linearity (Fig. 6.2). The minimum of this curve could also be useful for end users as it represents slightly below or relatively average peak load, but also a tipping point where small deviations in temperature can can lead to increases in electricity demand. For both daily mean and daily minimum T2M, the peak load anomalies are more positive between around 0 to -6°C compared to the minimum in the curve, with the most extreme positive peak load anomalies occurring when both metrics decrease below -6°C (Fig. 6.2). However, peak load anomalies of $\sim$0.5 MW/1000 customers are possible on the warmer side of the parabola, at $\sim$18°C for daily mean T2M and $\sim$12°C for daily minimum T2M. We also note that the inclusion of major holidays and weekends has a minimal impact on the relationship between load and daily mean T2M, and thus our results (Fig. A.14).

6.3.3 North American winter weather regimes and electricity load

To explore the regime-induced behaviour of SPP peak load anomalies, we first examine the mean daily peak load anomalies associated with each winter weather regime (Fig. 6.3a). The general impact of the winter weather regimes can be quantified by a shift in means, where the AkR, ArH, and WCR regime days feature statistically significant ($p < 0.05$) above average peak load, with AkR regime days having the greatest average peak load anomaly (around 0.13 MW/1000 customers; Fig. 6.3a). Subsequently, ArL and PT regime days result in significantly below average peak load, with mean peak load anomalies of around -0.10 and -0.15 MW/1000 customers, respectively (Fig. 6.3a).

We also investigate the probability density functions associated with categories of all days in each regime (Fig. 6.3b). The AkR and WCR regimes have broader right-hand tails

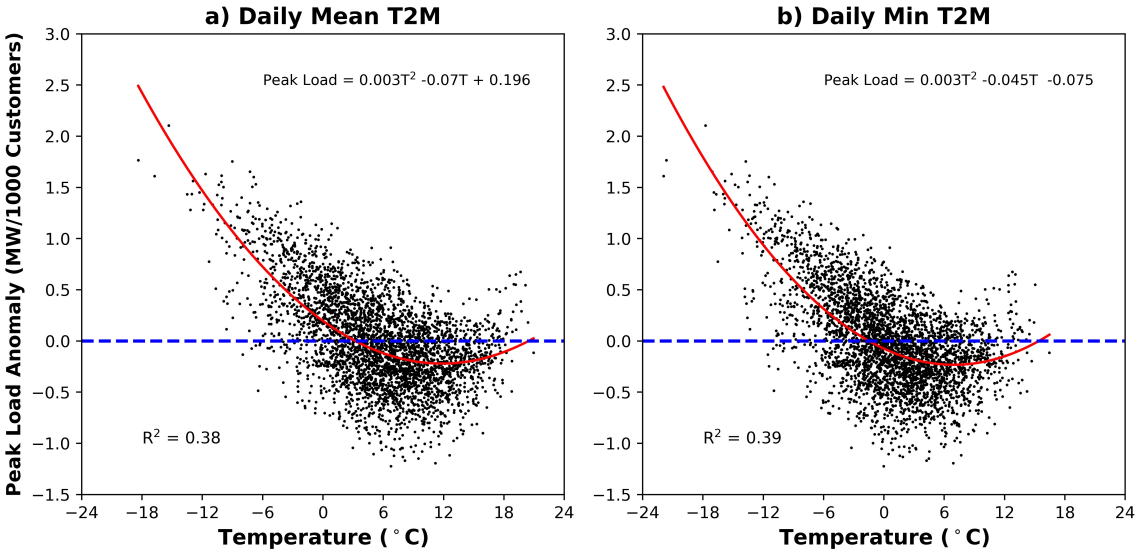


Figure 6.2: Scatter plots of peak load anomaly (MW/1000 customers) versus ERA5 (a) daily average T2M (°C) and (b) daily minimum T2M (°C), where both temperature variables have been averaged over the boxed region in Fig. 6.1. The red lines represent the second degree polynomial model for the relationship between the two parameters in each plot, with the equation of this polynomial shown on the plot, where T represents the temperature, as well as the the coefficient of determination (R^2) value for variance explained in peak load anomaly respective to the polynomial model. The blue dashed lines represent an average daily peak load between 1999-2022.

(Fig. 6.3b), which is not only due to their positive skewness ($\gamma = 0.36$ and $\gamma = 0.47$, respectively), but because their means are shifted toward higher peak load anomalies with slightly higher standard deviations than the other regimes ($\sigma = 0.48$ and $\sigma = 0.47$, respectively). The ArH regime also exhibits a mean shift toward higher peak load anomalies compared to the overall distribution with a more notable positive skew ($\gamma = 0.7$) and positive kurtosis ($\kappa = 0.94$). In fact, Fig. 6.3b also shows that the relatively higher skewness of the ArH substantially increases the skewness of that in the whole distribution ($\gamma = 0.55$). Since the mean and standard deviations for the ArH are much closer to that of the total distribution, the positive skewness and kurtosis means more weight is placed on higher peak load anomalies with a longer extension of the right-hand tail, which is likely associated with a number of extreme ArH-type CAOs in the region. Indeed, of the ten highest ArH peak load anomaly days, five are associated with Great Plains CAO events between 1950–2021 from Chapter 2 (Millin et al. 2022) and 3 belong to an extreme cold event during December 2022

(Yao et al. 2023). Therefore, excessive demand events may be unlikely in WCR and AkR regimes but are more typical of the circulation, whereas ArH-induced excessive demand events may require a more unique evolution not contained within the regime framework, such as *major* SSWs or reliance on other atmospheric blocks.

Conversely, the ArL and PT winter weather regime distributions are slightly positively skewed ($\gamma = 0.43$ and $\gamma = 0.45$, respectively) but have stronger positive kurtosis ($\kappa = 0.71$ and $\kappa = 1.51$, respectively) due to their more peaked distribution. However, these distributions occupy more of the negative peak load anomaly space due to their relatively lower means and standard deviations when compared with the other regimes. The ArL and PT distributions show that they are more likely to produce below average peak load anomalies across the SPP in the winter.

Using risk ratios, - i.e., the ratio of the probability of extreme peak load anomalies in a given regime to its sample-wide analog - we quantify the relative likelihood of extreme peak load anomalies in a given weather regime (van der Wiel et al. 2019). Risk ratios greater than 1 indicate an increased risk of extreme load compared to the total distribution (i.e., 1.2 equates to a 20% greater risk), whereas values less than 1 indicate a reduction in risk (i.e., 0.4 equates to a 60% reduced risk). Across all chosen extreme thresholds (90th, 95th, and 97.5th percentiles), the AkR, ArH, and WCR regime days have relative increases in likelihood of extreme peak load anomalies (Fig. 6.3c). These regimes are most associated with large negative temperature anomalies throughout the SPP (Fig. 6.1) and fit with our empirical relationship between temperatures (mean and minimum) and peak load anomalies in the region (Fig. 6.2). AkR regime days are approximately 1.75-2 times more likely to produce extreme peak load in the SPP. WCR regime days show a risk ratio of around 1.5 and have the largest risk ratio of all the regimes at the 97.5th percentile level (1.84; Fig. 6.3c). These relative increases in risk for WCR and AkR regime days are statistically significantly different from the baseline risk ratio of 1 at the 95% confidence level (Fig. 6.3c, gray bars), which also reflects the composite circulation pattern of all extreme days across various load thresholds, showing dominant patterns of anomalous ridging across Alaska and the West

Coast similar to their individual regime composites (Fig. 6.4). ArH regime days are on average associated with negative temperature anomalies in the SPP region, demonstrating a small, but non-significant, increase in likelihood of extremes with risk ratio values of around 1.2. This finding contrasts with the canonical relationship between SPV disruption and midlatitude extreme cold, potentially speaking to the reliance of the ArH on heightened GPH anomalies in other regions to produce extreme cold (Millin and Furtado 2022), whereas the AkR and WCR regimes seemingly have a much larger impact on US extreme cold development (Lee et al. 2019b) and thus anomalous peak load. Finally, across all thresholds, the ArL and PT weather regimes show a statistically significant reduction in risk of extreme peak load of around 50-80%. While ArL and PT extreme peak load events are unlikely to happen (only 17 and 9 events, respectively, at the 95th percentile), their circulation patterns are similar to that in Fig. 2.4, except central US anomalous troughing is stronger (Fig. 6.4).

The notable patterns in the risk ratio of extreme peak load relate back to the relationship between peak load anomalies and surface temperature. We find that the composite T2M anomaly for extreme peak loads in each regime that exceed their regime-relative 90th percentile show widely differing patterns (Fig. A.15). AkR and WCR regime extreme peak load days produce large, negative anomalies throughout the SPP, with the ArH also showing negative anomalies but of less magnitude. Subsequently, PT and ArL regime extreme peak load days show small negative anomalies throughout the SPP with even some positive T2M anomalies in the case of the northern SPP during PT regimes. Thus, North American winter weather regimes have distinct impacts on average and extreme peak electricity load throughout the SPP, with direct links to temperature. However, the upper bound of peak electricity load is significantly modulated by the regimes and their subsequent temperature anomalies.

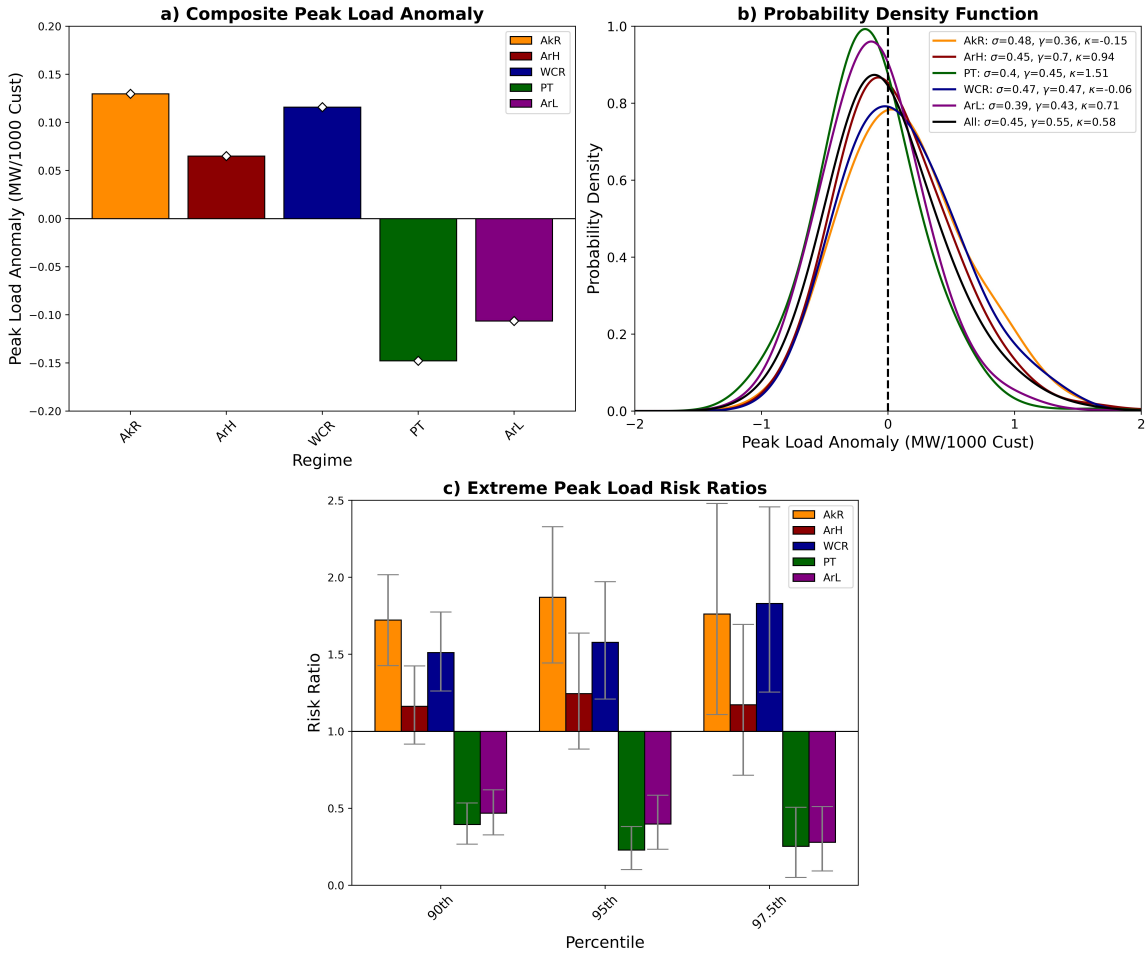


Figure 6.3: (a) The composite mean peak load anomaly (bars, MW/1000 customers) as a function of North American winter weather regime, where white diamonds indicate that the composite value is statistically significant at the 95th percentile via 5000-iteration bootstrapping with replacement. (b) The probability density function for peak load anomaly (lines, MW/1000 customers) as a function of North American winter weather regime via Gaussian kernel density estimation. σ, γ, and κ represent the probability density function standard deviation, skewness, and kurtosis, respectively. (c) The risk ratios of extreme peak load by North American weather regime (bars) for 90th percentile, 95th percentile, and 97.5th percentile extreme thresholds versus that of the entire distribution. Whiskers represent the 95th percentile confidence interval for each risk ratio from 5000-iteration bootstrapping with replacement. For all plots, AkR is orange, ArH is red, WCR is blue, PT is green, and ArL is purple.

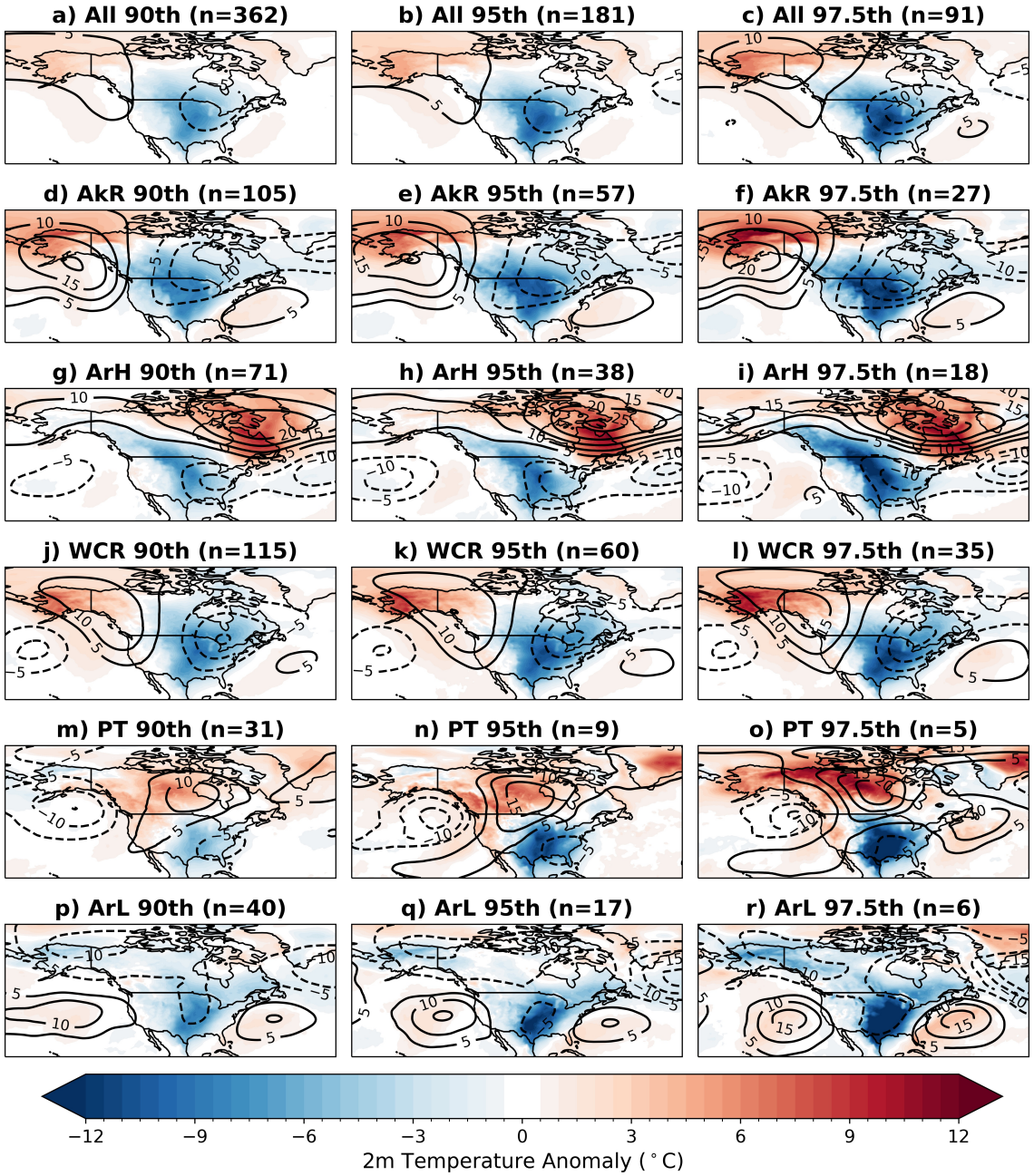


Figure 6.4: Composite maps of ERA5 T2M anomalies (shading, °C) and 500 hPa GPH anomalies (dam, contours), respective to the 1991-2020 climatology, for days exceeding the 90th, 95th, and 97.5th percentile thresholds of all peak load anomalies for a)–c) all days, d)–f) AkR days, g)–i) ArH days, j)–l) WCR days, m)–o) PT days, and p)–r) ArL days. The sample size is given for each composite.

6.4 Summary and Conclusions

Temperature is a key driver of electrical grid variability throughout the world (Hor et al. 2005; Bessec and Fouquau 2008; Thornton et al. 2016; Cawthorne et al. 2021). The February 2021 CAO demonstrated the vulnerability of power grids to extreme wintertime cold, where demand surges caused widespread outages across the South-Central US (Doss-Gollin et al. 2021). Our results demonstrate that central US wintertime mean and minimum temperature has a parabolic relationship with SPP peak load anomalies, with the coldest temperatures producing the most extreme peak load values (Fig. 6.2). Anomalous ridging in Alaska (the AkR regime) and along the West Coast of the US (the WCR regime) both yield the greatest wintertime positive peak load anomalies throughout the SPP (Fig. 6.3a), and the greatest risk for extreme peak load (Fig. 6.3c). Specifically, AkR and WCR regime days carry a 1.5–2 times greater likelihood of extreme peak load risk. Anomalous ridging near Greenland (the ArH regime) also causes positive peak load anomalies, but the enhanced risk of extreme peak load is not statistically significant. The relatively lower ArH risk ratios potentially owe to its dependence on other blocking features such as in the AkR (Millin and Furtado 2022), internal variability of the flow pattern, and that the connection between SPV disruption and ArH-induced cold conditions may not be as robust over North America. ArL and PT weather regimes are less likely to produce extreme cold, resulting in below average SPP peak load and reduced extreme peak load risk ratios. While wintertime weather regimes help determine the overall risk of extremely high load, they also may be useful for predicting periods of low electricity demand, which is useful for energy traders and showcases their potential for use alongside raw temperature forecasting. Overall, there is the need for ways of forecasting extremes, and regimes offer one avenue to that which is efficient, computationally inexpensive, but also could be used for utility purposes while allowing the interpretation of key dynamical temperature forcings.

The relationship between SPP peak load and North American winter weather regimes suggests that they could be a useful tool for predicting electrical grid demand on the challenging S2S forecast period of two weeks to two months (Vitart and Robertson 2018), especially given their longer timescale and more persistent nature, as well as their relation to key modes of variability associated with CAOs (Michelangeli et al. 1995; Lee et al. 2019b; Millin et al. 2022). An important caveat to our results is that this study only considers the SPP. Analyses for other electricity utility domains will likely yield different relationships with the North American winter weather regimes. However, if we can achieve improved prediction capability using regimes, decision makers could be provided with enhanced warning of conditions that produce wintertime electrical grid vulnerability on S2S timescales. Grid vulnerability to extreme cold has become more evident recently due to rising demand within the SPP and aging infrastructure, both adding unique stresses to the system (NERC 2023). Therefore, ways of forecasting extremes are needed, and regimes offers one avenue to that goal.

While we have not investigated how numerical weather prediction ensemble suites model peak load, future work could examine peak load as a function of temperature and weather regime in both numerical and machine learning-based forecasting models. Recent work has shown that the forecast skill for Atlantic-European weather regimes appears to be limited to two weeks (Büeler et al. 2021), but presently it's uncertain how well North American winter weather regimes can be forecast at different timescales. The unknown prediction skill of North American winter weather regimes is a caveat that future work should study to exploit their added value on the subseasonal timescale, especially regarding potential electricity load prediction. Furthermore, the AkR, ArH, and WCR regimes not only favor CAOs but also could be associated with anomalously windy conditions, which may have use for balancing increased demand from extreme cold via wind energy. Indeed, studies have shown that weather regimes modulate renewable energy production in both Europe and North America (van der Wiel et al. 2019; Garrido-Perez et al. 2020; Liu et al. 2023), but how this production could potentially offset shortfalls in CAO-related regimes

should be studied. Electricity shortfalls may also be the result of the prolonged nature of extreme cold events. Since extreme cold can often take several days to build, future work should consider how regime persistence affects electricity demand. Finally, some studies have shown that CAO measures, such as frequency, days, and duration, have decreased across North America and are projected to decrease, but the results are uncertain (Gao et al. 2015; Smith and Sheridan 2020, 2021). Therefore, future work should examine how changes in the climate system may influence peak load risk to critical energy infrastructure during winter cold extremes. It should also investigate how weather regimes have evolved, and may continue to evolve, including changes in their temperature responses and links to both cold events and demand extremes.

Chapter 7

Summary and Conclusions

Extreme cold events, namely CAOs, in the central US are high-impact events with serious societal and infrastructure consequences. Improving understanding of these events and their precursors is essential for enhancing predictability on the S2S timescale. This dissertation had three primary goals: (1) to develop a climatology of CAOs in the Great Plains and identify their key precursors, (2) to assess the predictability and the drivers of predictability for central US CAOs in S2S forecasts, and (3) to investigate the broader impacts of these events by linking them to electrical infrastructure vulnerability. To that end, this work built a comprehensive database of Great Plains CAOs from 1950–2021 and identified the SPV, MJO, North American winter weather regimes, and atmospheric blocking as key S2S precursors in both reanalysis and model experiments. It also provided a quantitative assessment of S2S forecast skill, showing that a prolonged and weakened SPV is a major preconditioning feature for confident and accurate week 3 cold spell forecasts, which are often accompanied by a persistent ArH regime. Finally, this dissertation established a clear link between North American weather regimes and changes in extreme electricity demand risk in the central US: the AkR, WCR, and ArH regimes increase risk, while the PT and ArL regimes reduce it. Overall, this work contributes significant new insights into the S2S predictability of CAOs in the central US and their broader societal implications.

Chapter 2 generated a database of CAOs in the Great Plains of the US to identify their key characteristics and S2S precursors. A total of 37 CAOs were identified between 1950–2021 using a threshold of five or more consecutive days with area-averaged, linearly detrended T2M anomalies below the 10th percentile. Composite analyses revealed two dominant onset weather regimes: the ArH and the AkR. AkR-type CAOs evolved rapidly on

synoptic timescales and were associated with stratospheric wave reflection and a strong SPV. In contrast, ArH-CAOs were preceded by weak SPV conditions several weeks in advance. Both types of CAOs showed anomalous upward WAF from Siberia, with downward WAF over Canada only occurring in the AkR-type CAOs. The rapid development and accompanying North Pacific wave train in AkR-type CAOs suggest a possible influence from tropical variability. These findings indicate that distinct dynamical pathways, operating on different timescales, can lead to Great Plains CAOs, each offering unique sources of S2S predictability. This chapter addressed the first goal of the dissertation by establishing a robust CAO climatology and identifying key precursors. It also highlighted the complex relationship between stratosphere–troposphere coupling and CAOs across North America. While the downward propagation of weak SPV conditions is a well-established link to mid-latitude CAOs (particularly in Europe), our results show that, in the central US, both the weak SPV and wave reflection modes are influential, each with distinct connections to the onset weather regime. These findings highlight the need for S2S prediction systems and climate models to accurately represent both modes of stratosphere–troposphere interaction, as well as their interplay with tropospheric regime dynamics. Improved representation of these processes could enhance predictive skill for central US cold spells and better constrain how their frequency and characteristics might evolve under future climate change.

Chapter 3 built upon earlier findings that both SPV and tropical variability may influence CAO development, by directly testing their impact on subseasonal forecast skill. To do this, we conducted targeted nudging experiments using CESM2 CAM6 S2S forecasts. In these experiments, either the stratosphere above 100 hPa (NUDGE-S) or the full atmospheric column in the tropics (NUDGE-T) was nudged toward MERRA-2 reanalysis values. A control experiment with no nudging was used for comparison (FREE). The experiments were initialized at several lead times before two CAO events: (1) the December 2017/January 2018 event, associated with wave reflection and strong Alaskan ridging; and (2) the January 2004 event, following a *major* SSW. For the 2017/18 case, the NUDGE-S experiment significantly improved forecasts of negative T2M anomalies over Canada

and the eastern US, along with better representation of North American winter weather regimes. The NUDGE-T experiment enhanced Alaskan ridging but led to more modest improvements in T2M forecasts. In contrast, for the 2004 event, both nudging approaches had limited impact on US T2M forecasts, despite some improvement in regime representation. These results suggest that while stratospheric and tropical variability both play a role in CAO predictability, their influence may vary by event and can interact constructively or destructively. As such, these important modes of variability should be monitored in an operational setting, but their combined influence is not always additive; forecast skill depends on how well models capture their interactions and downstream impacts. In this dissertation, we showed that weak SPV conditions and enhanced ArH frequency are key features of observed CAOs and confident correct S2S forecasts. In contrast, AkR-type CAOs tended to occur under stronger SPV conditions and were linked to wave reflection patterns potentially modulated by tropical convection. However, nudging experiments for the January 2004 event, which followed a major SSW with strong tropical forcing, had limited impact on forecast skill, highlighting that even favorable precursors may not enhance predictability unless their joint effects are constructively aligned and/or well represented in model forecasts. This chapter fulfilled the second goal of the dissertation by demonstrating how targeted representation of large-scale variability affects S2S forecast skill for CAOs.

Chapter 4 assessed the skill of six S2S models in forecasting central US cold spells, defined using the lower tercile of weekly mean T2M anomalies from 1993–2020. While all models performed well at short leads, skill declined beyond two weeks for CAO days, with the BOM model consistently underperforming, perhaps due to its coarse resolution and lower model lid. Despite reduced overall skill at week 3 for cold spells, models demonstrated some sharpness and reliability, though they tended to overforecast cold spells. Confident hit forecasts were linked to a persistent weak SPV signal propagating downward over several weeks and enhanced ArH regime frequency. In contrast, confident false alarms often simulated transitions into regimes inconsistent with observations, or even transitions that are hard to achieve in the confident strong SPV forecasts. Stratifying forecasts by SPV

strength at initialization revealed higher reliability in weak SPV cases, where confident hits showed robust ArH persistence and more prolonged stratospheric weakening. Strong SPV hit forecasts features a more localized strong SPV anomaly near initialization, resembling the AkR-type structure identified in Chapter 2, associated with wave reflection. Together, these findings establish key indicators of skillful Week 3 cold spell forecasts in terms of weather regimes and SPV strength. Future work can use these features as diagnostic benchmarks to evaluate how well models represent key precursors, design targeted forecast assessments, or develop hybrid prediction systems that leverage precursor-informed conditions to enhance forecast confidence for winter extremes. This chapter directly addressed dissertation goal (2) by quantifying S2S skill and demonstrating how precursor conditions, particularly the SPV strength and regime configuration, govern cold spell forecast success or failure.

Chapter 5 focused on the evolution and subseasonal predictability of the extreme February 2021 CAO, one of the most impactful cold events in recent decades. Reanalysis data showed that the event was facilitated by two distinct Rossby wave breaks: an anticyclonic wave break over the East Siberian Sea and a cyclonic wave break over the Labrador Sea. Using forecasts from the ECMWF and NCEP S2S models, we analyzed the evolution of weather regimes, extreme event probability, and the impact of these wave breaks on forecast skill at lead times of approximately 2–3 weeks. Ensemble members that successfully simulated the wave breaks produced more negative T2M anomalies over the Great Plains, consistent with improved representation of anomalous ridging. These results demonstrate that capturing persistent blocking anticyclones is critical for improving S2S forecasts of central US extreme cold. This chapter thus addressed the second goal of the dissertation by assessing the predictability of a high-impact CAO and identifying atmospheric blocking as a key precursor contributing to enhanced subseasonal forecast skill. Furthermore, this chapter reinforced key relationships identified in Chapter 2, showing that the February 2021 CAO followed a major SSW and its severity was closely tied to the configuration of atmospheric blocking and weather regimes, particularly the ArH.

Chapter 6 presents a cross-disciplinary analysis linking central US extreme cold, North American winter weather regimes, and electricity demand in the SPP region. We first showed a parabolic relationship between daily T2M and SPP peak load, with the coldest days driving the highest demand. Building on this, we demonstrated that winter weather regimes exert a strong influence on electricity load across 13 SPP balancing authorities. The AkR, WCR, and ArH regimes are associated with increased demand and elevated extreme demand risk, often coinciding with significant CAO days. In contrast, the ArL and PT regimes are linked to reduced risk due to warmer conditions. These results suggest that winter weather regimes not only serve as key S2S predictors of extreme cold but also provide actionable insights for anticipating electricity demand. Weather regimes could be integrated into S2S demand forecasting systems, where probabilistic forecasts of regime occurrence help improve load forecasting and resource planning on extended timescales. Additionally, regime-based diagnostics could support grid stress testing under worst-case demand spikes, inform strategies for managing electricity market price volatility, and aid long-term planning by evaluating how shifts in regime frequency under climate change may impact grid stress. Overall, this chapter primarily addresses goal (3) of the dissertation, while also contributing to goal (2) by illustrating the wide-ranging value of regime-based forecasts for both weather and energy applications.

Together, the chapters of this dissertation establish a cohesive framework for understanding, predicting, and contextualizing central US CAOs on the subseasonal timescale by characterizing events, identifying key precursors, evaluating forecasts, and analyzing associated weather regimes. By identifying distinct CAO types with differing dynamical origins, establishing the role of stratospheric and tropical variability in shaping predictability, and demonstrating the value of weather regimes as organizing structures for both CAO evolution and electrical grid impact, this work bridges fundamental climate dynamics with applied forecasting needs. This work shows that improved representation of the SPV, tropical teleconnections, and North American weather regimes sometimes enhance subseasonal forecast skill for anomalous-to-extreme cold in the central US, and that these same

regimes hold promise for anticipating electricity demand surges. Overall, the dissertation advances both scientific understanding and practical capability in the S2S prediction of extreme cold, emphasizing the interconnectedness of physical precursors, forecast performance, and energy-relevant outcomes.

Although this dissertation provides a strong foundation for understanding central US CAOs and their key S2S precursors, several open questions remain. A central challenge lies in how best to operationalize these findings to improve S2S forecasts of extreme cold. While one avenue is enhanced precursor monitoring, future work should explore integrating these signals, such as SPV anomalies and weather regime evolution, into statistical, machine-learning, and physics-based forecast models targeting both temperature extremes and electricity demand. Additionally, although this work demonstrates the relevance of North American winter weather regimes for cold event prediction, the degree to which improvements in cold forecasts depend on the forecastability of the regimes themselves remains an open area of research. For example, recent research has examined the predictability of European and North American weather regimes in some S2S prediction systems, finding that forecast skill declines beyond two weeks (Büeler et al. 2021; Molina et al. 2023). However, several studies have also demonstrated that circulation pattern-based approaches can extend skill at these longer lead times for European heatwaves and probabilistic energy-related metrics (Bloomfield et al. 2021; Rouges et al. 2024). A similar framework could be applied to improve the extended-range prediction of central US CAOs and associated energy impacts. However, a key theme that emerges across the chapters is the complex role of the ArH regime in shaping central US CAOs. Chapters 2 and 4 underscore the ArH regime's robust occurrence prior to CAOs, its persistence, and its association with weak SPV conditions, suggesting a favorable environment for enhanced subseasonal predictability. In contrast, results from Chapters 3 and 6 reveal a more nuanced picture: during the January 2004 event, nudging experiments show limited sensitivity to weak SPV conditions during this ArH-type CAO, and overall, the ArH regime is associated with a less cold mean near-surface temperature response compared to the AkR and WCR. This

apparent contradiction suggests that the influence of the ArH and weak SPV may depend on additional factors, such as the presence or transition of other regimes, or interactions with tropospheric wave activity. Understanding these regime interactions and their modulating effects should be a key focus of future work aimed at improving CAO prediction and interpretation.

Another consideration for future work is the spatial definition of CAOs and cold spells. While this study focused on events spanning the full meridional extent of the central US, not all CAOs exhibit such large-scale reach. Future studies could examine whether distinct dynamical pathways underlie North-Central US vs. South-Central US CAOs, or whether similar structures are present in cold extremes affecting other regions of the US. Model biases in the stratosphere, tropical teleconnections, regime transitions, and blocking may limit the skill of precursor-informed forecasts for extreme cold in the central US. S2S models often underestimate upward and downward stratosphere-troposphere coupling, the persistence of lower-stratospheric anomalies, and tropospheric stationary wave activity (Garfinkel et al. 2025; Vitart et al. 2025). They also struggle with MJO simulation, including weak amplitude, unrealistic phase speeds, and inaccurate teleconnection patterns in the North Pacific and Euro-Atlantic sectors (Wang et al. 2019; Wu and Jin 2021; Stan et al. 2022). Blocking is frequently underrepresented as well (Quinting and Vitart 2019). Improving S2S performance may require higher model lids, finer vertical and horizontal resolution, better representation of quasi-stationary waves, improved convective paramaterization, and more accurate representation of diabatic heating.

Together, these open questions provide fruitful ground for future research at the intersection of climate dynamics and societal resilience. By identifying key sources of predictability and connecting them to real-world impacts, this dissertation lays important groundwork for developing more actionable and reliable S2S forecasts of extreme cold.

Reference List

Afargan-Gerstman, H., and D. Domeisen, 2020: Pacific modulation of the North Atlantic storm track response to sudden stratospheric warming events. *Geophys. Res. Lett.*, **47**, https://doi.org/10.1029/2019GL085007.

Albers, J. R., M. Newman, A. Hoell, M. L. Breeden, Y. Wang, and J. Lou, 2022: The February 2021 cold air outbreak in the United States: A subseasonal forecast of opportunity. *Bull. Amer. Meteor. Soc.*, **103 (12)**, E2887–E2904, https://doi.org/10.1175/BAMS-D-21-0266.1.

Analatis, A., and Coauthors, 2008: Effects of cold weather on mortality: Results from 15 European cities within the PHEWE project. *Am. J. Epidemiol.*, **186 (12)**, 1397–1408, https://doi.org/10.1093/aje/kwn266.

Auffhammer, M., P. Baylis, and C. Hausman, 2017: Climate change is projected to have severe impacts on the frequency and intensity of peak electricity demand across the United States. *Proc. Natl. Acad. Sci. U.S.A.*, **114 (8)**, 1886–1891, https://doi.org/10.1073/pnas.1613193114.

Ayarzaguena, B., and Coauthors, 2020: Uncertainty in the response of sudden stratospheric warmings and stratosphere-troposphere coupling to quadrupled CO2 concentrations in CMIP6 models. *J. Geophys. Res. Atmos.*, **125 (6)**, https://doi.org/10.1029/2019JD032345.

Baldwin, M., and T. Dunkerton, 2001: Stratospheric harbingers of anomalous weather regimes. *Science*, **294 (5542)**, 581–584, https://doi.org/10.1126/science.1063315.

BBC News, 2021: Texas weather: Deadly winter storm sweeps Texas and US southern states. Accessed 22 July 2021, https://www.bbc.com/news/world-us-canada-56076686.

Bessec, M., and J. Fouquau, 2008: The non-linear link between electricity consumption and temperature in Europe: A threshold panel approach. *Energy Econ.*, **30 (5)**, 2705–2721, https://doi.org/10.1016/j.eneco.2008.02.003.

Biernat, K., L. Bosart, and D. Keyser, 2021: A climatological analysis of the linkages between tropopause polar vortices, cold pools, and cold air outbreaks over the central and eastern United States. *Mon. Wea. Rev*, **149 (1)**, 189–206, https://doi.org/10.1175/MWR-D-20-0191.1.

Bloomfield, H. C., D. J. Brayshaw, P. L. M. Gonzalez, and A. Charlton-Perez, 2021: Pattern-based conditioning enhances sub-seasonal prediction skill of European national energy variables. *Meteor. Appl.*, **28 (4)**, https://doi.org/10.1002/met.2018.

Bolinger, R., and Coauthors, 2022: An assessment of the extremes and impacts of the February 2021 South-Central U.S. Arctic outbreak, and how climate services can help. *Wea. Climate Extremes*, **36**, https://doi.org/10.1016/j.wace.2022.100461.

BoM, 2025: Tropical monitoring and outlooks. Accessed 26 March 2025, http://www.bom.gov.au/climate/mjo/#tabs=Monitoring.

Brady, R. X., and A. Spring, 2021: Climpred: Verification of weather and climate forecasts. *J. Open Source Software*, **6 (59)**, https://doi.org/doi:10/gh9646.

Butler, A., J. Sjoberg, D. Seidel, and K. Rosenlof, 2017: A sudden stratospheric warming compendium. *Earth Syst. Sci. Data*, **9 (1)**, 63–76, https://doi.org/10.5194/essd-9-63-2017.

Büeler, D., L. Ferranti, L. Magnusson, J. F. Quinting, and C. M. Grams, 2021: Year-round sub-seasonal forecast skill for Atlantic–European weather regimes. *Quart. J. Roy. Meteor. Soc.*, **147 (741)**, 4283–4309, https://doi.org/10.1002/qj.4178.

Carrera, M., R. Higgins, and V. Kousky, 2004: Downstream weather impacts associated with atmospheric blocking over the Northeast Pacific. *J. Climate*, **17 (24)**, 4823–4839, https://doi.org/10.1175/JCLI-3237.1.

Cavallo, S., and G. Hakim, 2009: Potential vorticity diagnosis of a tropopause polar cyclone. *Mon. Wea. Rev*, **137 (4)**, 1358–1371, https://doi.org/10.1175/2008MWR2670.1.

Cavallo, S., and G. Hakim, 2010: Composite structure of tropopause polar cyclones. *Mon. Wea. Rev*, **138 (10)**, 3840–3857, https://doi.org/10.1175/2010MWR3371.1.

Cavallo, S., and G. Hakim, 2013: Physical mechanisms of tropopause polar vortex intensity change. *J. Atmos. Sci.*, **70 (11)**, 3359–3373, https://doi.org/10.1175/JAS-D-13-088.1.

CAWCR, 2015: Forecast verification. Accessed 27 May 2025, https://www.cawcr.gov.au/projects/verification/.

Cawthorne, D., A. R. de Queiroz, H. Eshraghi, A. Sankarasubramanian, and J. F. DeCarolis, 2021: The role of temperature variability on seasonal electricity demand in the southern US. *Front. Sustain. Cities*, **3**, https://doi.org/10.3389/frsc.2021.644789.

Cellitti, M., J. Walsh, R. Rauber, and D. Portis, 2006: Extreme cold air outbreaks over the United States, the polar vortex, and the large-scale circulation. *J. Geophys. Res. Atmos.*, **111 (D2)**, https://doi.org/10.1029/2005JD006273.

Charlton, A. J., and L. M. Polvani, 2007: A new look at stratospheric sudden warmings. part i: Climatology and modeling benchmarks. *J. Climate*, **20 (3)**, 449–469, https://doi.org/10.1175/JCLI3996.1.

Charlton-Perez, A. J., L. Ferranti, and R. W. Lee, 2018: The influence of the stratospheric state on North Atlantic weather regimes. *Quart. J. Roy. Meteor. Soc.*, **144 (713)**, 1140–1151, https://doi.org/10.1002/qj.3280.

Cicala, S., 2022: Imperfect markets versus imperfect regulation in US electricity generation. *Am. Econ. Rev.*, **112 (2)**, 409–41, https://doi.org/10.1257/aer.20172034.

Cohen, J., L. Agel, M. Barlow, C. Garfinkel, and I. White, 2021: Linking Arctic variability and change with extreme winter weather in the United States. *Climate Dyn.*, **373 (6559)**, 1116–1121, https://doi.org/10.1126/science.abi9167.

Davis, L. W., and P. J. Gertler, 2015: Contribution of air conditioning adoption to future energy use under global warming. *Proc. Natl. Acad. Sci. U.S.A.*, **112 (19)**, 5962–5967, https://doi.org/10.1073/pnas.1423558112.

Davis, N. A., J. H. Richter, A. A. Glanville, J. Edwards, and E. LaJoie, 2022: Limited surface impacts of the January 2021 sudden stratospheric warming. *Nat. Commun.*, **13 (1136)**, https://doi.org/10.1038/s41467-022-28836-1.

Deschênes, O., and M. Greenstone, 2011: Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the US. *Am. Econ. J-Appl. Econ.*, **3 (4)**, 152–185, https://doi.org/10.1257/app.3.4.152.

Dias, J., S. N. Tulich, M. Gehne, and G. N. Kiladis, 2021: Tropical origins of weeks 2–4 forecast errors during the Northern Hemisphere cool season. *Mon. Wea. Rev.*, **149 (9)**, 2975–2991, https://doi.org/10.1175/MWR-D-21-0020.1.

Ding, X., G. Chen, P. Zhang, D. I. V. Domeisen, and C. Orbe, 2023: Extreme stratospheric wave activity as harbingers of cold events over North America. *Nat. Commun. Earth Env.*, **4 (1)**, https://doi.org/10.1038/s43247-023-00845-y.

Domeisen, D., L. Sun, and G. Chen, 2013: The role of synoptic eddies in the tropospheric response to stratospheric variability. *Geophys. Res. Lett.*, **40 (18)**, 4933–4937, https://doi.org/10.1002/grl.50943.

Domeisen, D. I., and Coauthors, 2019a: The role of the stratosphere in subseasonal to seasonal prediction: 1. Predictability of the stratosphere. *J. Geophys. Res. Atmos.*, **125 (2)**, https://doi.org/10.1029/2019JD030920.

Domeisen, D. I., and Coauthors, 2019b: The role of the stratosphere in subseasonal to seasonal prediction: 2. Predictability arising from stratosphere-troposphere coupling. *J. Geophys. Res. Atmos.*, **125 (2)**, https://doi.org/10.1029/2019JD030923.

Domeisen, D. I. V., and A. H. Butler, 2020: Stratospheric drivers of extreme events at the Earth's surface. *Nat. Commun. Earth Environ.*, **1 (49)**, https://doi.org/10.1038/s43247-020-00060-z.

Domeisen, D. I. V., A. H. Butler, K. Fröhlich, M. Bittner, W. A. Müller, and J. Baehr, 2015: Seasonal predictability over Europe arising from El Niño and stratospheric variability in the MPI-ESM seasonal prediction system. *J. Climate*, **28 (1)**, 256–271, https://doi.org/10.1175/JCLI-D-14-00207.1.

Doss-Gollin, J., D. J. Farnham, U. Lall, and V. Modi, 2021: How unprecedented was the February 2021 Texas cold snap? *Environ. Res. Lett.*, **16**, https://doi.org/10.1088/1748-9326/ac0278.

ECMWF, 2021: NCEP Model. Accessed 26 September 2022, https://confluence.ecmwf.int/display/S2S/NCEP+Model.

Erner, I., and A. Karpechko, 2024: Factors influencing subseasonal predictability of northern Eurasian cold spells. *Quart. J. Roy. Meteor. Soc.*, **150 (762)**, 2955–2975, https://doi.org/10.1002/qj.4744.

Erner, I., A. Karpechko, M. Kämäräinen, and H. Järvinen, 2022: Stratosphere–troposphere coupling enhances subseasonal predictability of Northern Eurasian cold spells. *Quart. J. Roy. Meteor. Soc.*, **148 (747)**, 2769–2783, https://doi.org/10.1002/qj.4335.

Furtado, J. C., J. Cohen, E. J. Becker, and D. C. Collins, 2021: Evaluating the relationship between sudden stratospheric warmings and tropospheric weather regimes in the NMME phase-2 models. *Climate Dyn.*, **56**, 2321–2338, https://doi.org/10.1007/s00382-020-05591-x.

Gao, Y., L. R. Leung, J. Liu, and G. Masato, 2015: Persistent cold air outbreaks over North America in a warming climate. *Environ. Res. Lett.*, **10**, https://doi.org/10.1088/1748-9326/10/4/044001.

Garfinkel, C. I., J. J. Benedict, and E. D. Maloney, 2014: Impact of the {MJO} on the boreal winter extratropical circulation. *Geophys. Res. Lett.*, **41 (16)**, 6055–6062, https://doi.org/10.1002/2014GL061094.

Garfinkel, C. I., S. B. Feldstein, D. W. Waugh, C. Yoo, and S. Lee, 2012: Observed connection between stratospheric sudden warmings and the Madden-Julian Oscillation. *Geophys. Res. Lett.*, **39 (18)**, https://doi.org/10.1029/2012GL053144.

Garfinkel, C. I., and Coauthors, 2025: A process-based evaluation of biases in extratropical stratosphere–troposphere coupling in subseasonal forecast systems. *Wea. Clim. Dyn.*, **6 (1)**, 171–195, https://doi.org/10.5194/wcd-6-171-2025.

Garrido-Perez, J. M., C. Ordóñez, D. Barriopedro, R. García-Herrera, and D. Paredes, 2020: Impact of weather regimes on wind power variability in western Europe. *Appl. Energy*, **264**, https://doi.org/10.1016/j.apenergy.2020.114731.

Gelaro, R., and Coauthors, 2017: The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). *J. Climate*, **30 (14)**, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1.

Grace, T. M., K. Pegion, and J. B. Basara, 2024: Atmospheric and surface dynamics during winter warm spells in the Southern Great Plains: Insights from the 2021 case study. *J. Geophys. Res. Atmos.*, **129 (24)**, https://doi.org/10.1029/2024JD040856.

Grams, C. M., R. Beerli, S. Pfenninger, I. Staffell, and H. Wernli, 2017: Balancing Europe's wind-power output through spatial deployment informed by weather regimes. *Nat. Clim. Chang.*, **7**, 557–562, https://doi.org/10.1038/nclimate3338.

Gray, S., C. Dunning, J. Methven, G. Masato, and J. Chagnon, 2014: Systematic model forecast error in Rossby wave structure. *Geophys. Res. Lett.*, **41 (8)**, 2979–2987, https://doi.org/10.1002/2014GL059282.

Green, M., and J. Furtado, 2019: Evaluating the joint influence of the Madden-Julian Oscillation and the stratospheric polar vortex on weather patterns in the Northern Hemisphere. *J. Geophys. Res. Atmos.*, **124 (22)**, https://doi.org/10.1029/2019JD030771.

Grimm, A. M., L. R. Hakoyama, and L. A. Scheibe, 2021: Active and break phases of the South American summer monsoon: MJO influence and subseasonal prediction. *Climate Dyn.*, **56**, 3603–3624, https://doi.org/10.1007/s00382-021-05658-3.

Grotjahn, R., and Coauthors, 2015: North American extreme temperature events and related large scale meteorological patterns: A review of statistical methods, dynamics, modeling, and trends. *Climate Dyn.*, **46**, 1151–1184, https://doi.org/10.1007/s00382-015-2638-6.

Guirguis, K., A. Gershunov, R. Schwartz, and S. Bennett, 2011: Recent warm and cold daily winter temperature extremes in the Northern Hemisphere. *Geophys. Res. Lett.*, **38 (17)**, https://doi.org/10.1029/2011GL048762.

Hamouda, M., C. Pasquero, and E. Tziperman, 2021: Decoupling of the Arctic Oscillation and North Atlantic Oscillation in a warmer climate. *Nat. Clim. Change*, **11**, https://doi.org/10.1038/s41558-020-00966-8.

Harnik, N., 2009: Observed stratospheric downward reflection and its relation to upward pulses of wave activity. *J. Geophys. Res. Atmos.*, **114 (D8)**, https://doi.org/10.1029/2008jd010493.

Henderson, S., and E. Maloney, 2018: The impact of the Madden–Julian Oscillation on high-latitude winter blocking during El Niño–Southern Oscillation events. *J. Climate*, **31 (13)**, 5293–5318, https://doi.org/10.1175/jcli-d-17-0721.1.

Henderson, S. A., E. D. Maloney, and E. A. Barnes, 2016: The influence of the Madden–Julian Oscillation on Northern Hemisphere winter blocking. *J. Climate*, **29 (12)**, 4597–4616, https://doi.org/10.1175/JCLI-D-15-0502.1.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146 (730)**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hitchcock, P., and I. Simpson, 2014: The downward influence of stratospheric sudden warmings. *J. Atmos. Sci.*, **71 (10)**, 3856–3876, https://doi.org/10.1175/jas-d-14-0012.1.

Hitchcock, P., and Coauthors, 2022: Stratospheric nudging and predictable surface impacts (SNAPSI): a protocol for investigating the role of stratospheric polar vortex disturbances in subseasonal to seasonal forecasts. *Geosci. Model Dev.*, **15 (13)**, 5073–5092, https://doi.org/10.5194/gmd-15-5073-2022.

Hor, C., S. J. Watson, and S. Majithia, 2005: Analyzing the impact of weather variables on monthly electricity demand. *IEEE Trans. Power Syst.*, **20 (4)**, 2078–2085, https://doi.org/10.1109/TPWRS.2005.857397.

Huang, C., and N. Nakamura, 2013: Dynamics of Rossby wave breaking. *AGU Fall Meeting Abstracts*, NG23A–1487.

Huang, J., P. Hitchcock, A. Maycock, C. McKenna, and W. Tian, 2021: Northern Hemisphere cold air outbreaks are more likely to be severe during weak polar vortex conditions. *Commun. Earth. Enviro.*, **2 (147)**, https://doi.org/10.1038/s43247-021-00215-6.

Huang, J., P. Hitchcock, W. Tian, and J. Sillin, 2022: Stratospheric influence on the development of the 2018 late winter European cold air outbreak. *J. Geophys. Res. Atmos.*, **127 (5)**, https://doi.org/10.1029/2021JD035877.

Johnson, N. C., D. C. Collins, S. B. Feldstein, M. L. L 'heureux, and E. E. Riddle, 2014: Skillful wintertime North American temperature forecasts out to 4 weeks based on the state of ENSO and the MJO. *J. Climate*, **29 (1)**, 23–38, https://doi.org/10.1175/WAF-D-13-00102.1.

Jung, T., M. J. Miller, and T. N. Palmer, 2010: Diagnosing the origin of extended-range forecast errors. *Mon. Wea. Rev.*, **138 (6)**, 2434–2446, https://doi.org/10.1175/2010MWR3255.1.

Karpechko, A. Y., P. Hitchcock, D. H. W. Peters, and A. Schneidereit, 2017: Predictability of downward propagation of major sudden stratospheric warmings. *Quart. J. Roy. Meteor. Soc.*, **143 (704)**, 1459–1470, https://doi.org/10.1002/qj.3017.

Karpechko, A. Y., F. Vitart, I. Statnaia, M. A. Balmaseda, A. J. Charlton-Perez, and I. Polichtchouk, 2023: The tropical influence on sub-seasonal predictability of wintertime stratosphere and stratosphere–troposphere coupling. *Quart. J. Roy. Meteor. Soc.*, **150 (760)**, 1357–1374, https://doi.org/10.1002/qj.4649.

Kautz, L., I. Polichtchouk, T. Birner, H. Garny, and J. Pinto, 2019: Enhanced extended-range predictability of the 2018 late-winter Eurasian cold spell due to the stratosphere. *Quart. J. Roy. Meteor. Soc.*, **146 (727)**, 1040–1055, https://doi.org/10.1002/qj.3724.

Kodera, K., H. Mukougawa, and S. Itoh, 2008: Tropospheric impact of reflected planetary waves from the stratosphere. *Geophys. Res. Lett.*, **35 (16)**, https://doi.org/10.1029/2008GL034575.

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 (1)**, 80–94, https://doi.org/10.1002/2015JD023359.

Kolstad, E., T. Breiteig, and 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 (649)**, 886–893, https://doi.org/10.1002/qj.620.

Konrad, C., and S. Colucci, 1989: An examination of extreme cold air outbreaks over eastern North America. *Mon. Wea. Rev.*, **117 (12)**, 2687–2700, https://doi.org/10.1175/1520-0493(1989)117⟨2687:AEOECA⟩2.0.CO;2.

Kretschmer, M., J. Cohen, V. Matthias, J. Runge, and D. Coumou, 2018a: The different stratospheric influence on cold-extremes in Eurasia and North America. *npj Clim. Atmos. Sci.*, **1 (44)**, https://doi.org/10.1038/s41612-018-0054-4.

Kretschmer, M., D. Coumou, L. Agel, M. Barlow, E. Tziperman, and J. Cohen, 2018b: More-persistent weak stratospheric polar vortex states linked to cold extremes. *Bull. Amer. Meteor. Soc.*, **99 (1)**, 49–60, https://doi.org/10.1175/BAMS-D-16-0259.1.

Lee, S., 2021: The January 2021 sudden stratospheric warming. *Weather*, **76 (21)**, 135–136, https://doi.org/10.1002/wea.3966.

Lee, S., A. Charlton-Perez, J. Furtado, and S. Woolnough, 2019a: Abrupt stratospheric vortex weakening associated with North Atlantic anticyclonic wave breaking. *J. Geophys. Res. Atmos.*, **124 (15)**, 8563–8575, https://doi.org/10.1029/2019JD030940.

Lee, S., J. Furtado, and A. Charlton-Perez, 2019b: Wintertime North American weather regimes and the Arctic stratospheric polar vortex. *Geophys. Res. Lett.*, **46 (24)**, 14 892–14 900, https://doi.org/10.1029/2019gl085592.

Lee, S. H., A. J. Charlton-Perez, S. J. Woolnough, and J. C. Furtado, 2022: How do stratospheric perturbations influence North American weather regime predictions? *J. Climate*, **35 (18)**, 5915–5932, https://doi.org/10.1175/JCLI-D-21-0413.1.

Lillo, S., S. Cavallo, D. Parsons, and C. Riedel, 2021: The role of a tropopause polar vortex in the generation of the January 2019 extreme Arctic outbreak. *J. Atmos. Sci.*, **78 (9)**, 2801–2821, https://doi.org/10.1175/JAS-D-20-0285.1.

Limpasuvan, V., D. Hartmann, D. Thompson, K. Jeev, and Y. Yung, 2005: Stratosphere-troposphere evolution during polar vortex intensification. *J. Geophys. Res.*, **110 (D24)**, https://doi.org/10.1029/2005jd006302.

Limpasuvan, V., D. Thompson, and D. Hartmann, 2004: The life cycle of the Northern Hemisphere sudden stratospheric warmings. *J. Climate*, **17 (13)**, https://doi.org/10.1175/1520-0442(2004)017⟨2584:TLCOTN⟩2.0.CO;2.

Lin, H., R. Mo, and F. Vitart, 2022: The 2021 western North American heatwave and its subseasonal predictions. *Geophys. Res. Lett.*, **49 (6)**, https://doi.org/10.1029/2021GL097036.

Liu, M., D. Hu, and Z. Guan, 2022: Modulation of a long-lasting extreme cold event in Siberia by a minor sudden stratospheric warming and the dynamical mechanism involved. *Climate Dyn.*, **60**, 797–811, https://doi.org/10.1007/s00382-022-06353-7.

Liu, Y., S. Feng, Y. Qian, H. Huang, and L. Berg, 2023: How do North American weather regimes drive wind energy at the sub-seasonal to seasonal timescales? *npj Clim. Atmos. Sci.*, **6 (100)**, https://doi.org/10.1038/s41612-023-00403-5.

Loikith, P., and A. Broccoli, 2012: Characteristics of observed atmospheric circulation patterns associated with temperature extremes over North America. *J. Climate*, **25 (20)**, 7266–7281, https://doi.org/10.1175/jcli-d-11-00709.1.

Madden, R. A., and P. R. Julian, 1971: Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific. *J. Atmos. Sci.*, **28 (5)**, 702–708, https://doi.org/10.1175/1520-0469(1971)028⟨0702:DOADOI⟩2.0.CO;2.

Madden, R. A., and P. R. Julian, 1972: Description of global-scale circulation cells in the Tropics with a 40–50 day period. *J. Atmos. Sci.*, **29 (6)**, 1109–1123, https://doi.org/10.1175/1520-0469(1972)029⟨1109:DOGSCC⟩2.0.CO;2.

Mariotti, A., and Coauthors, 2020: Windows of opportunity for skillful forecasts subseasonal to seasonal and beyond. *Bull. Amer. Meteor. Soc.*, **101 (5)**, E608–E625, https://doi.org/10.1175/BAMS-D-18-0326.1.

Masato, G., B. Hoskins, and T. Woollings, 2011: Wave-breaking characteristics of midlatitude blocking. *Quart. J. Roy. Meteor. Soc.*, **138 (666)**, 1285–1296, https://doi.org/10.1002/qj.990.

Matthias, V., and M. Kretschmer, 2020: The influence of stratospheric wave reflection on North American cold spells. *Mon. Wea. Rev.*, **148 (4)**, 1675–1690, https://doi.org/10.1175/mwr-d-19-0339.1.

Maycock, A., G. Masukwedza, P. Hitchock, and I. Simpson, 2020: A regime perspective on the North Atlantic eddy-driven jet response to sudden stratospheric warmings. *J. Climate*, **33 (9)**, 3901–3917, https://doi.org/10.1175/JCLI-D-19-0702.1.

McIntyre, M., and T. Palmer, 1983: Breaking planetary waves in the stratosphere. *Nature*, **305**, 593–600, https://doi.org/10.1038/305593a0.

McManus, G., 2009: Oklahoma Climate Winter 2008-2009. Tech. rep., Oklahoma Climatological Survey, 30 pp.

Messori, G., M. Kretschmer, S. H. Lee, and V. Wendt, 2022: Stratospheric downward wave reflection events modulate North American weather regimes and cold spells. *Wea. Clim. Dyn.*, **3 (4)**, 1215–1236, https://doi.org/10.5194/wcd-3-1215-2022.

Met Office, 2025: Reliability and sharpness diagrams. Accessed 22 May 2025, https://www.metoffice.gov.uk/research/climate/seasonal-to-decadal/gpc-outlooks/user-guide/interpret-reliability#:~:text=Reliability%20diagrams%20provide%20additional%20information,chance%20of%20observing%20the%20event.

Michel, C., and G. Rivière, 2011: The Link between Rossby wave breakings and weather regime transitions. *J. Atmos. Sci.*, **68 (8)**, 1730–1748, https://doi.org/10.1175/2011JAS3635.1.

Michelangeli, P., R. Vautard, and B. Legras, 1995: Weather regimes: Recurrence and quasi stationarity. *J. Atmos. Sci.*, **52 (8)**, 1237–1256, https://doi.org/10.1175/1520-0469(1995)052⟨1237:WRRAQS⟩2.0.CO;2.

Millin, O. T., and J. C. Furtado, 2022: The role of wave breaking in the development and subseasonal forecasts of the February 2021 Great Plains cold air outbreak. *Geophys. Res. Lett.*, **49 (21)**, https://doi.org/10.1029/2022GL100835.

Millin, O. T., J. C. Furtado, and J. B. Basara, 2022: Characteristics, evolution, and formation of cold air outbreaks in the Great Plains of the United States. *J. Climate*, **35 (14)**, 4585–4602, https://doi.org/10.1175/JCLI-D-21-0772.1.

Millin, O. T., J. C. Furtado, and C. Malloy, 2024: The impact of North American winter weather regimes on electricity load in the central United States. *npj Clim. Atmos. Sci.*, **7 (254)**, https://doi.org/10.1038/s41612-024-00803-1.

Mirasgedis, S., Y. Sarafidis, E. Georgopoulou, D. P. Lalas, M. Moschovits, F. Karagiannis, and D. Papakonstantinou, 2006: Models for mid-term electricity demand forecasting incorporating weather influences. *Energy*, **31 (2–3)**, 208–227, https://doi.org/10.1016/j.energy.2005.02.016.

Mitchell, D., L. Gray, J. Anstey, M. Baldwin, and A. Charlton-Perez, 2013: The influence of stratospheric vortex displacements and splits on surface climate. *J. Climate*, **26 (8)**, 2668–2682, https://doi.org/10.1175/JCLI-D-12-00030.1.

Molina, M. J., J. H. Richter, A. A. Glanville, K. Dagon, J. Berner, A. Hu, and G. A. Meehl, 2023: Subseasonal representation and predictability of North American weather regimes using cluster analysis. *Art. Intel. Earth Syst.*, **2 (2)**, https://doi.org/10.1175/AIES-D-22-0051.1.

Moore, R., O. Martius, and T. Spengler, 2010: The modulation of the subtropical and extratropical atmosphere in the Pacific basin in response to the Madden–Julian Oscillation. *Mon. Wea. Rev.*, **138 (7)**, 2761–2779, https://doi.org/10.1175/2010MWR3194.1.

Moral-Carcedo, J., and J. Vicéns-Otero, 2005: Modelling the non-linear response of Spanish electricty demand to temperature variations. *Energy Econ.*, **27 (3)**, 477–494, https://doi.org/10.1016/j.eneco.2005.01.003.

Nakagawa, K. I., and K. Yamazaki, 2006: What kind of stratospheric sudden warming propagates to the troposphere? *Geophys. Res. Lett.*, **33 (4)**, https://doi.org/10.1029/2005GL024784.

Nakamura, N., and C. Huang, 2018: Atmospheric blocking as a traffic jam in the jet stream. *Science*, **361 (6397)**, 42–47, https://doi.org/10.1126/science.aat0721.

NCEI, 2025: Billion-dollar weather and climate disasters: Summary stats. Accessed 17 May 2025, https://www.ncei.noaa.gov/access/billions/summary-stats.

NERC, 2023: 2023–2024 winter reliability assessment. Accessed 20 May 2024 https://www.nerc.com/pa/RAPA/ra/Reliability%20Assessments%20DL/NERC_WRA_2023.pdf.

NWS, 2021a: February 2021: Historic Winter Storm and Arctic Outbreak. Accessed 8 July 2021, https://www.weather.gov/fwd/Feb-2021-WinterEvent.

NWS, 2021b: February 7, 2021 Winter Storm. Accessed 22 July 2022, https://storymaps.arcgis.com/stories/96964b6457794d2196dff24d68d4688d.

Papritz, L., E. Rouges, F. Aemisegger, and H. Wernli, 2019: On the thermodynamic preconditioning of Arctic air masses and the role of tropopause polar vortices for cold air outbreaks from Fram Strait. *J. Geophys. Res. Atmos.*, **124 (21)**, 11 033–11 050, https://doi.org/10.1029/2019JD030570.

Pedragose, F., and Coauthors, 2011: Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, **12 (85)**, 2825–2830.

Pegion, K., and Coauthors, 2019: The subseasonal experiment (subx): A multimodel subseasonal prediction experiment. *Bull. Amer. Meteor. Soc.*, **100 (10)**, 2043–2060, https://doi.org/10.1175/BAMS-D-18-0270.1.

Pelly, J., and B. Hoskins, 2003: A new perspective on blocking. *J. Atmos. Sci.*, **60 (5)**, 743–755, https://doi.org/10.1175/1520-0469(2003)060⟨0743:ANPOB⟩2.0.CO;2.

Perkins, S. E., L. V. Alexander, and J. R. Nairn, 2012: Increasing frequency, intensity and duration of observed global heatwaves and warm spells. *Geophys. Res. Lett.*, **39 (20)**, https://doi.org/10.1029/2012GL053361.

Perkins-Kirkpatrick, S. E., and S. C. Lewis, 2020: Increasing trends in regional heatwaves. *Nat. Commun.*, **11 (3357)**, https://doi.org/10.1038/s41467-020-16970-7.

Perlwitz, J., and N. Harnik, 2003: Observational evidence of a stratospheric influence on the troposphere by planetary wave reflection. *J. Climate*, **16 (18)**, 3011–3026, https://doi.org/10.1175/1520-0442(2003)016⟨3011:oeoasi⟩2.0.co;2.

Perlwitz, J., and N. Harnik, 2004: Downward coupling between the stratosphere and troposphere: The relative roles of wave and zonal mean processes. *J. Climate*, **17 (24)**, 4902–4909, https://doi.org/10.1175/jcli-3247.1.

Plumb, R., 1985: On the three-dimensional propagation of stationary waves. *J. Atmos. Sci.*, **42 (3)**, 217–229, https://doi.org/10.1175/1520-0469(1985)042⟨0217:ottdpo⟩2.0.co;2.

Quinting, J., C. M. Grams, E. K.-M. Chang, S. Pfahl, and H. Wernli, 2023: Warm conveyor belt activity over the Pacific: Modulation by the Madden-Julian Oscillation and impact on tropical-extratropical teleconnections. *EGUsphere*, **5 (1)**, https://doi.org/10.5194/egusphere-2023-783.

Quinting, J., and F. Vitart, 2019: Representation of synoptic-scale Rossby wave packets and blocking in the S2S Prediction Project Database. *Geophys. Res. Lett.*, **46 (2)**, 1070–1078, https://doi.org/10.1029/2018GL081381.

Rao, J., C. I. Garfinkel, H. Chen, and I. P. White, 2019: The 2019 new year stratospheric sudden warming and its real-time predictions in multiple S2S models. *J. Geophys. Res. Atmos.*, **124 (21)**, 11 155–11 174, https://doi.org/10.1029/2019JD030826.

Renwick, J., and J. Wallace, 1996: Relationships between North Pacific wintertime blocking, El Niño, and the PNA pattern. *Mon. Wea. Rev.*, **124 (9)**, 2071–2076, https://doi.org/10.1175/1520-0493(1996)124⟨2071:rbnpwb⟩2.0.co;2.

Rex, D., 1950a: Blocking action in the middle troposphere and its effect upon regional climate: An aerological study of blocking. *Tellus*, **2 (3)**, 196–211, https://doi.org/10.1111/j.2153-3490.1950.tb00331.x.

Rex, D., 1950b: Blocking action in the middle troposphere and its effect upon regional climate: The climatology of blocking action. *Tellus*, **2 (3)**, 275–301, https://doi.org/10.1111/j.2153-3490.1950.tb00339.x.

Rex, D., 1951: The effect of Atlantic blocking action upon European climate. *Tellus*, **3 (2)**, 100–112, https://doi.org/10.1111/j.2153-3490.1951.tb00784.x.

Richter, J. H., and Coauthors, 2022: Subseasonal earth system prediction with CESM2. *Wea. Forecasting*, **37 (6)**, 797–815, https://doi.org/10.1175/WAF-D-21-0163.1.

Rodney, M., H. Lin, and J. Derome, 2013: Subseasonal prediction of wintertime North American surface air temperature during strong MJO events. *Mon. Wea. Rev.*, **141 (8)**, 2897–2909, https://doi.org/10.1175/MWR-D-12-00221.1.

Rouges, E., L. Ferranti, H. Kantz, and F. Pappenberger, 2024: Pattern-based forecasting enhances the prediction skill of European heatwaves into the sub-seasonal range. *Climate Dyn.*, **62**, 9269–9285, https://doi.org/10.1007/s00382-024-07390-0.

Schiemann, R., and Coauthors, 2017: The resolution sensitivity of Northern Hemisphere blocking in four 25-km atmospheric global circulation models. *J. Climate*, **30 (1)**, 337–358, https://doi.org/10.1175/JCLI-D-16-0100.1.

Schutte, M. K., A. Portal, S. H. Lee, and G. Messori, 2025: Dynamics of stratospheric wave reflection over the North Pacific. *Wea. Clim. Dyn.*, **6 (2)**, 521–548, https://doi.org/10.5194/wcd-6-521-2025.

Schwierz, C., M. Croci-Maspoli, and H. Davies, 2004: Perspicacious indicators of atmospheric blocking. *Geophys. Res. Lett.*, **31 (6)**, https://doi.org/10.1029/2003GL019341.

Seo, K., and H. Lee, 2017: Mechanisms for a PNA-like teleconnection pattern in response to the MJO. *J. Atmos. Sci.*, **74 (6)**, 1767–1781, https://doi.org/10.1175/JAS-D-16-0343.1.

Seo, K., H. Lee, and D. Frierson, 2016: Unraveling the teleconnection mechanisms that induce wintertime temperature anomalies over the Northern Hemisphere continents in response to the MJO. *J. Atmos. Sci.*, **73 (9)**, 3557–3571, https://doi.org/10.1175/JAS-D-16-0036.1.

Shaffer, B., D. Quintero, and J. Rhodes, 2022: Changing sensitivity to cold weather in Texas power demand. *iScience*, **25 (4)**, https://doi.org/https://doi.org/10.1016/j.isci.2022.104173.

Shaw, T., J. Perlwitz, and N. Harnik, 2010: Downward wave coupling between the stratosphere and troposphere: The importance of meridional wave guiding and comparison with zonal-mean coupling. *J. Climate*, **23 (23)**, 6365–6381, https://doi.org/10.1175/2010jcli3804.1.

Smith, E., and S. Sheridan, 2018a: The influence of atmospheric circulation patterns on cold air outbreaks in the eastern United States. *Int. J. Climatol.*, **39 (4)**, 2080–2095, https://doi.org/10.1002/joc.5935.

Smith, E., and S. Sheridan, 2020: Where do cold air outbreaks occur, and how have they changed over time? *Geophys. Res. Lett.*, **47 (13)**, https://doi.org/10.1029/2020gl086983.

Smith, E. T., and S. C. Sheridan, 2018b: The characteristics of extreme cold events and cold air outbreaks in the eastern United States. *Int. J. Climatol.*, **38 (S1)**, e807–e820, https://doi.org/10.1002/joc.5408.

Smith, E. T., and S. C. Sheridan, 2019: The influence of extreme cold events on mortality in the United States. *Sci. Total Environ.*, **647**, 342–351, https://doi.org/10.1016/j.scitotenv.2018.07.466.

Smith, E. T., and S. C. Sheridan, 2021: Projections of cold air outbreaks in CMIP6 Earth system models. *Clim. Change*, **169 (14)**, https://doi.org/10.1007/s10584-021-03259-x.

Stan, C., and D. Straus, 2007: Is blocking a circulation regime? *Mon. Wea. Rev.*, **135 (6)**, 2406–2413, https://doi.org/10.1175/mwr3410.1.

Stan, C., and Coauthors, 2022: Advances in the prediction of MJO teleconnections in the S2S forecast systems. *Bull. Amer. Meteor. Soc.*, **103 (6)**, E1426–E1447, https://doi.org/10.1175/BAMS-D-21-0130.1.

Thompson, D., and J. Wallace, 2000: Annular modes in the extratropical circulation. Part i: Month-to-month variability. *J. Climate*, **13 (5)**, 1000–1016, https://doi.org/10.1175/1520-0442(2000)013⟨1000:amitec⟩2.0.co;2.

Thornton, H. E., B. J. Hoskins, and A. A. Scaife, 2016: The role of temperature in the variability and extremes of electricity and gas demand in Great Britain. *Environ. Res. Lett.*, **11**, https://doi.org/10.1088/1748-9326/11/11/114015.

Tripathi, O. P., and Coauthors, 2015: The predictability of the extratropical stratosphere on monthly time-scales and its impact on the skill of tropospheric forecasts. *Quart. J. Roy. Meteor. Soc.*, **141 (689)**, 987–1003, https://doi.org/10.1002/qj.2432.

UCAR, 2024: Meteorology for physics-based nudging. Accessed 26 March 2025, https://wiki.ucar.edu/display/camchem/Meteorology+for+physics-based+nudging.

van der Wiel, K., H. C. Bloomfield, R. W. Lee, L. P. Stoop, R. Blackport, J. A. Screen, and F. M. Selten, 2019: The influence of weather regimes on European renewable energy production and demand. *Environ. Res. Lett.*, **14**, https://doi.org/10.1088/1748-9326/ab38d3.

Vitart, F., and M. A. Balmaseda, 2024: Sources of MJO teleconnection errors in the ECMWF extended-range forecasts. *Quart. J. Roy. Meteor. Soc.*, **150 (761)**, 2028–2044, https://doi.org/10.1002/qj.4688.

Vitart, F., and A. Robertson, 2018: The Sub-seasonal to Seasonal Prediction Project (S2S) and the prediction of extreme events. *npj Clim. Atmos. Sci.*, **1 (3)**, https://doi.org/10.1038/s41612-018-0013-0.

Vitart, F., and Coauthors, 2017: The subseasonal to seasonal (S2S) prediction project database. *Bull. Amer. Meteor. Soc.*, **98 (1)**, 163–173, https://doi.org/10.1175/BAMS-D-16-0017.1.

Vitart, F., and Coauthors, 2025: The WWRP/WCRP S2S project and its achievements. *Bull. Amer. Meteor. Soc.*, **106 (5)**, E791–E808, https://doi.org/10.1175/BAMS-D-24-0047.1.

Wallace, J., and D. Gutzler, 1981: Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.*, **109 (4)**, 784–812, https://doi.org/10.1175/1520-0493(1981)109⟨0784:TITGHF⟩2.0.CO;2.

Wang, L., and W. Chen, 2010: Downward arctic oscillation signal associated with moderate weak stratospheric polar vortex and the cold December 2009. *Geophys. Res. Lett.*, **37 (9)**, https://doi.org/10.1029/2010GL042659.

Wang, S., A. H. Sobel, M. K. Tippett, and F. Vitart, 2019: Prediction and predictability of tropical intraseasonal convection: seasonal dependence and the Maritime Continent prediction barrier. *Climate Dyn.*, **52 (2)**, 6015–6031, https://doi.org/10.1007/s00382-018-4492-9.

Weijenborg, C., H. de Vries, and R. Haarsma, 2012: On the direction of Rossby wave breaking in blocking. *Climate Dyn.*, **39**, 2823–2831, https://doi.org/10.1007/s00382-012-1332-1.

Westby, R., and R. Black, 2015: Development of anomalous temperature regimes over the southeastern United States: Synoptic behavior and role of low-frequency modes. *Wea. Forecasting*, **30 (3)**, 553–570, https://doi.org/10.1175/waf-d-14-00093.1.

Wheeler, D. D., V. L. Harvey, D. E. Atkinson, R. L. Collins, and M. J. Mills, 2011: A climatology of cold air outbreaks over North America: WACCM and ERA-40 comparison and analysis. *Geophys. Res. Lett.*, **47 (D12)**, https://doi.org/10.1029/2020GL086983.

White, C. J., and Coauthors, 2017: Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteor. Appl.*, **24 (3)**, 315–325, https://doi.org/10.1002/met.1654.

White, I., C. Garfinkel, E. Gerber, M. Jucker, V. Aquila, and L. Oman, 2019: The downward influence of sudden stratospheric warmings: Association with tropospheric precursors. *J. Climate*, **32 (1)**, 85–108, https://doi.org/10.1175/jcli-d-18-0053.1.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. Elsevier, Boston.

WPC, 2022: WPC Event Review/Winter Storm Archive (Prototype). Accessed 22 July 2022, https://www.wpc.ncep.noaa.gov/storm_summaries/event_reviews.php?YYYYMMDD=20210207&product=snow.

Wu, J., and F. Jin, 2021: Improving the MJO forecast of S2S operation models by correcting their biases in linear dynamics. *Geophys. Res. Lett.*, **48 (6)**, https://doi.org/10.1029/2020GL091930.

Xie, Z., R.Black, and Y. Deng, 2017: The structure and large-scale organization of extreme cold waves over the conterminous United States. *Clim. Dyn.*, **49**, 4075–4088, https://doi.org/10.1007/s00382-017-3564-6.

Yan, M., and Y. Guo, 2023: Influence of ENSO on the ECMWF subseasonal prediction of summer rainfall over the Yangtze River. *Climate Dyn.*, **61**, 3955–3969, https://doi.org/10.1007/s00382-023-06784-w.

Yao, Y., and Coauthors, 2023: Extreme cold events in North America and Eurasia in November-December 2022: A potential vorticity gradient perspective. *Adv. Atmos. Sci.*, **40**, 953–962, https://doi.org/10.1007/s00376-023-2384-3.

Zhang, M., X. Yang, and Y. Huang, 2022: Impacts of sudden stratospheric warming on extreme cold events in early 2021: An ensemble-based sensitivity analysis. *Geophys. Res. Lett.*, **49 (2)**, https://doi.org/10.1029/2021GL096840.

1 Appendix A

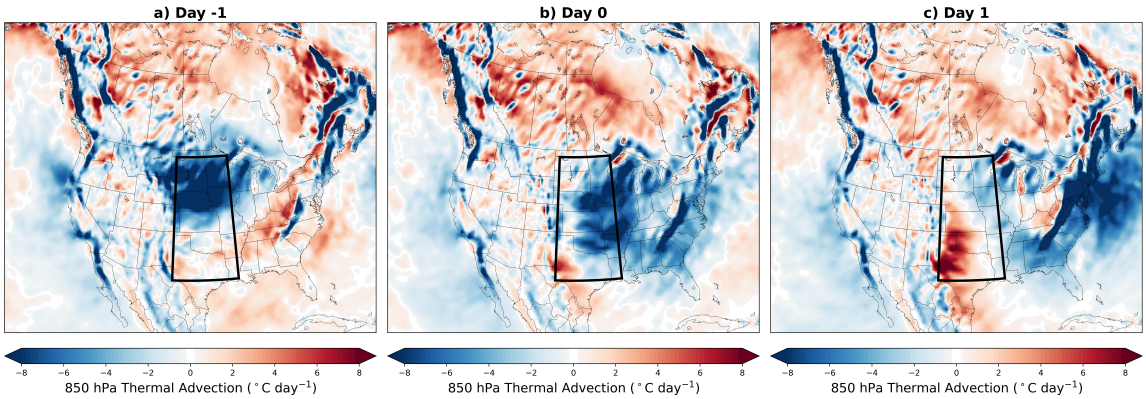


Figure A.1: Lagged composites of the 850 hPa horizontal thermal advection (shaded, °C day^{-1}) for all 37 CAOs for a) Day -1, b) Day 0, and c) Day 1. Positive lags represent days after CAO onset day (Day 0) and negative lags represent days prior to CAO onset. The thick, black box represents the region over which the CAOs were defined.

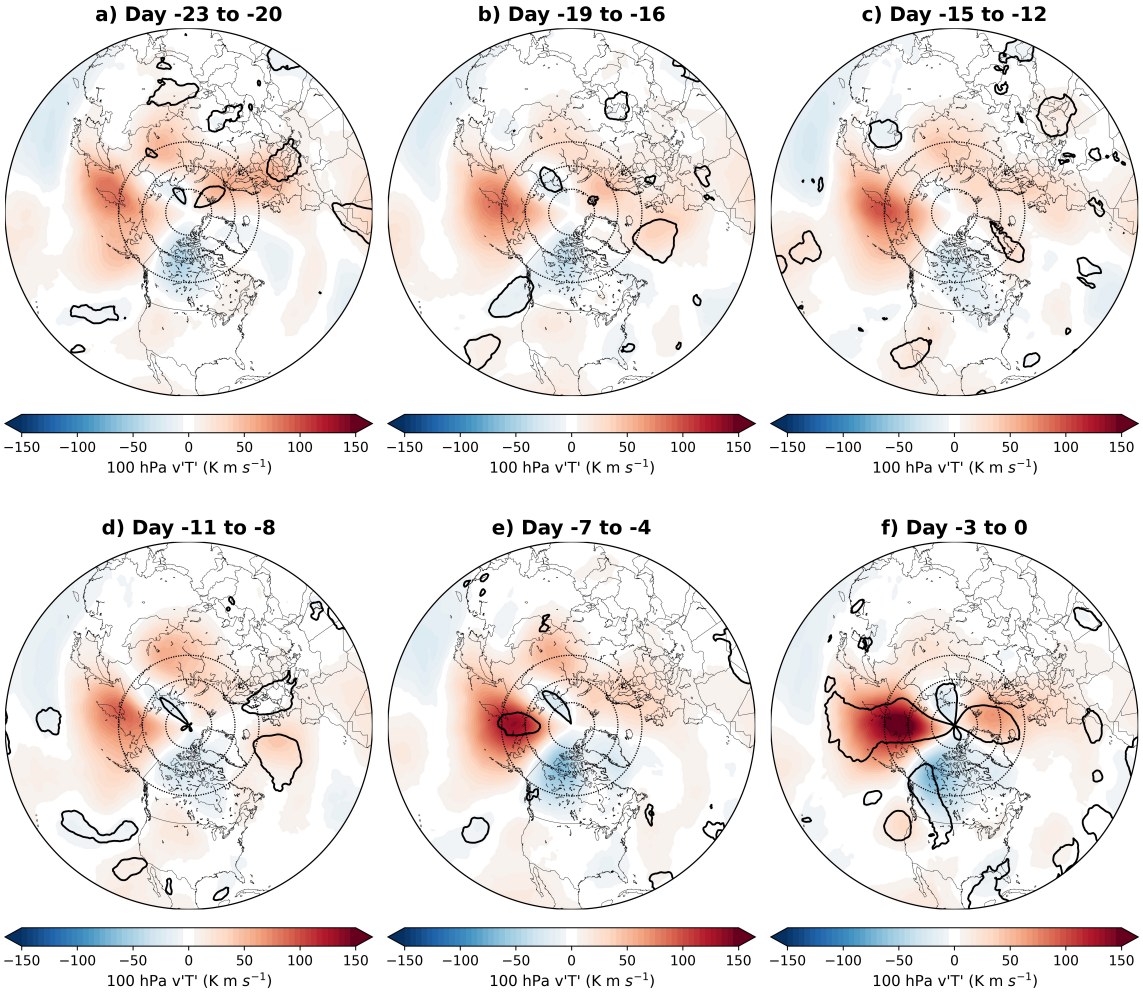


Figure A.2: Lagged composites of the 100 hPa *v'T'* (K m s^{-1}, shaded) across the Northern Hemisphere for ArH-CAOs, with respect to CAO onset (Day 0) at 4-day intervals from a) Days -23 to -20 through f) Days -3 to 0. Black contours represent significance at the 95th percentile via a bootstrapping approach of 5000 iterations.

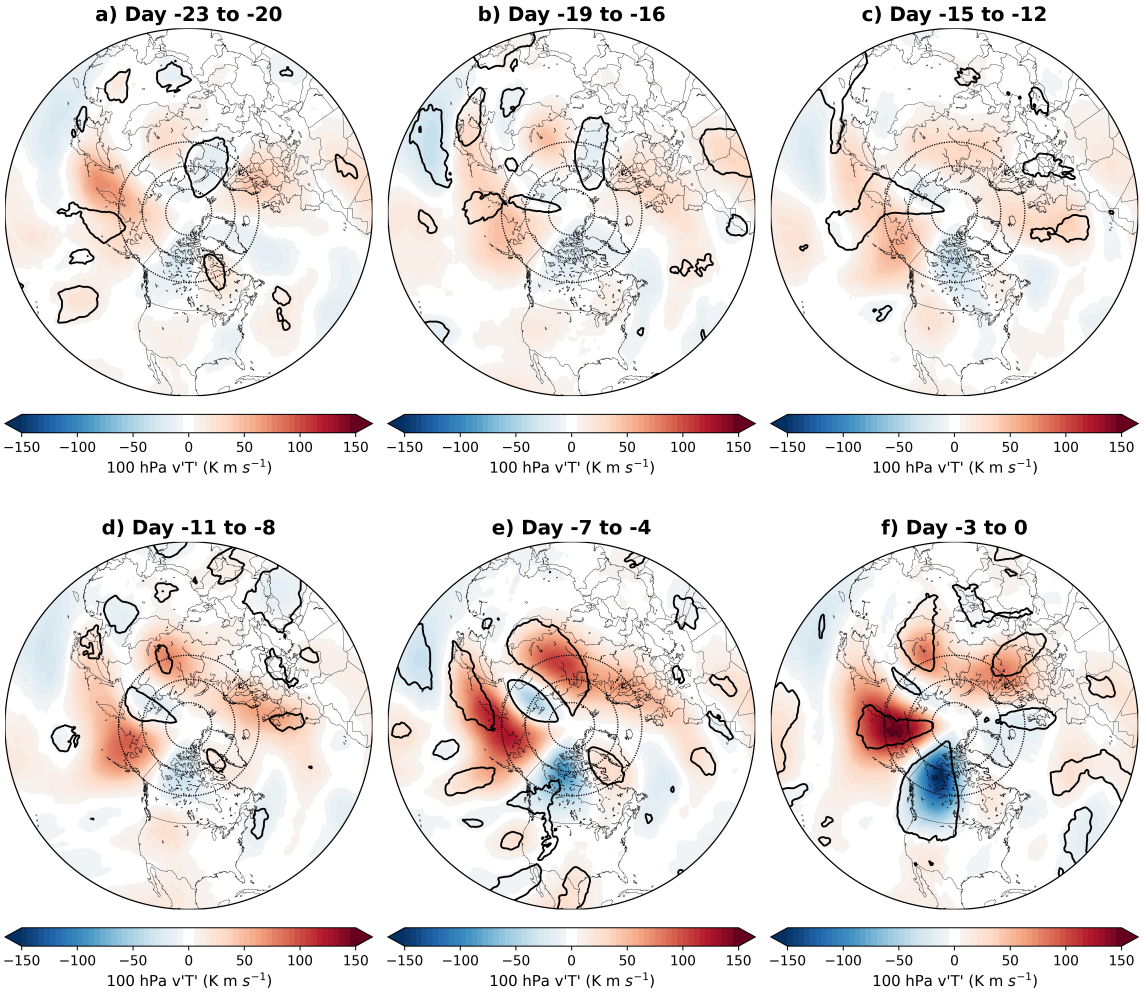


Figure A.3: Lagged composites of the 100 hPa *v'T'* (K m s^{-1}, shaded) across the Northern Hemisphere for AkR-CAOs, with respect to CAO onset (Day 0) at 4-day intervals from a) Days -23 to -20 through f) Days -3 to 0. Black contours represent significance at the 95th percentile via a bootstrapping approach of 5000 iterations.

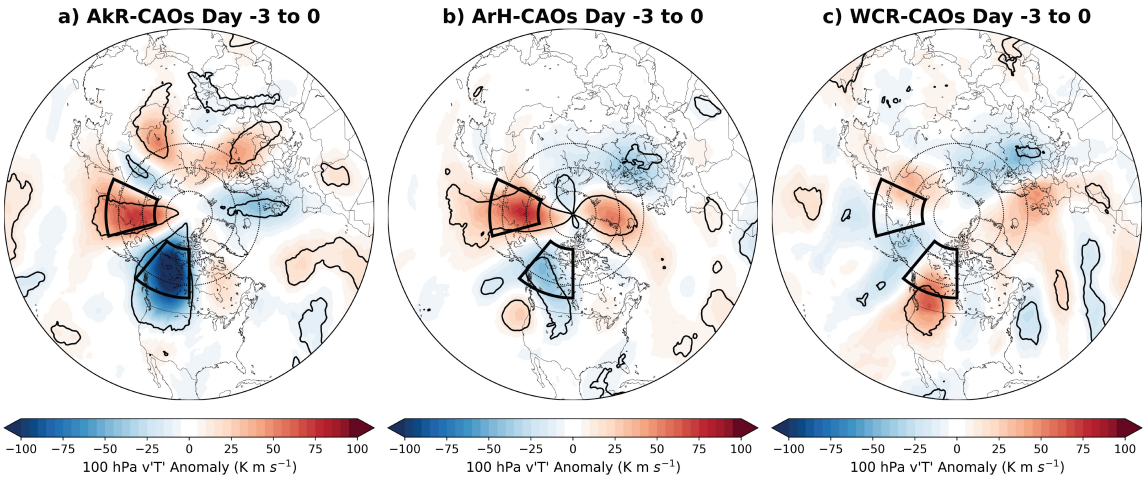


Figure A.4: Composites of the 100 hPa *v'T'* anomalies (K m s^{-1}, shaded) across the Northern Hemisphere for (a) AkR-CAOs, (b) ArH-CAOs, and (c) WCR-CAOs for Days -3 to 0. Black contours represent significance at the 95th percentile via a bootstrapping approach of 5000 iterations.

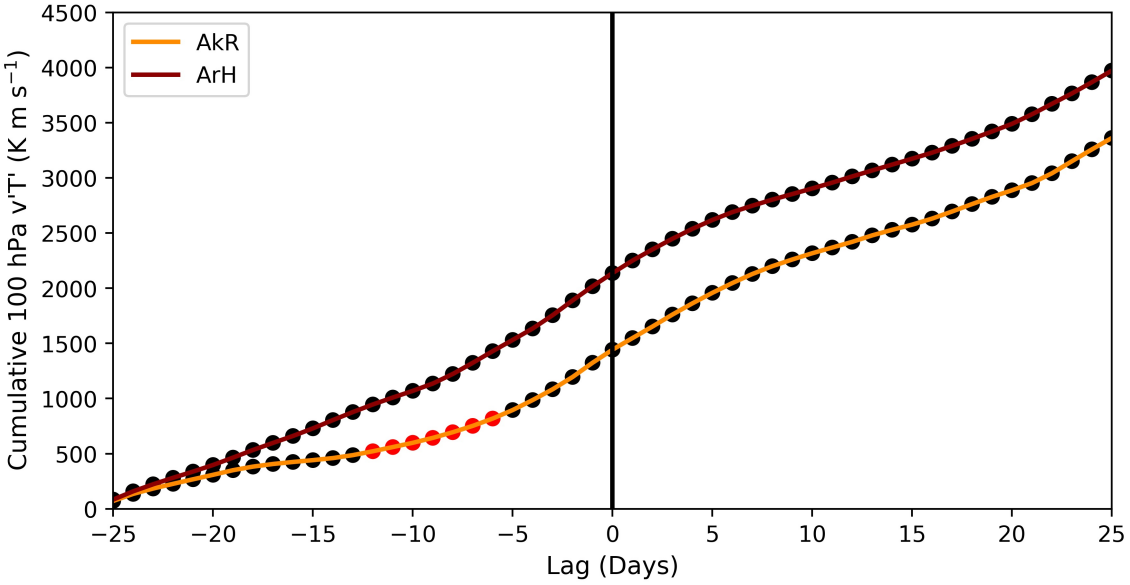


Figure A.5: Lagged composite of the cumulative total 100 hPa *v'T'* (K m s^{-1}) averaged over the Siberian region (55-75°N and 145-185°E) for AkR-CAOs (orange) and ArH-CAOs (dark red) with respect to CAO onset (Day 0). Statistical significance at the 95th percentile is shown via red dots via a 5000-iteration bootstrapping approach.

2 Appendix B

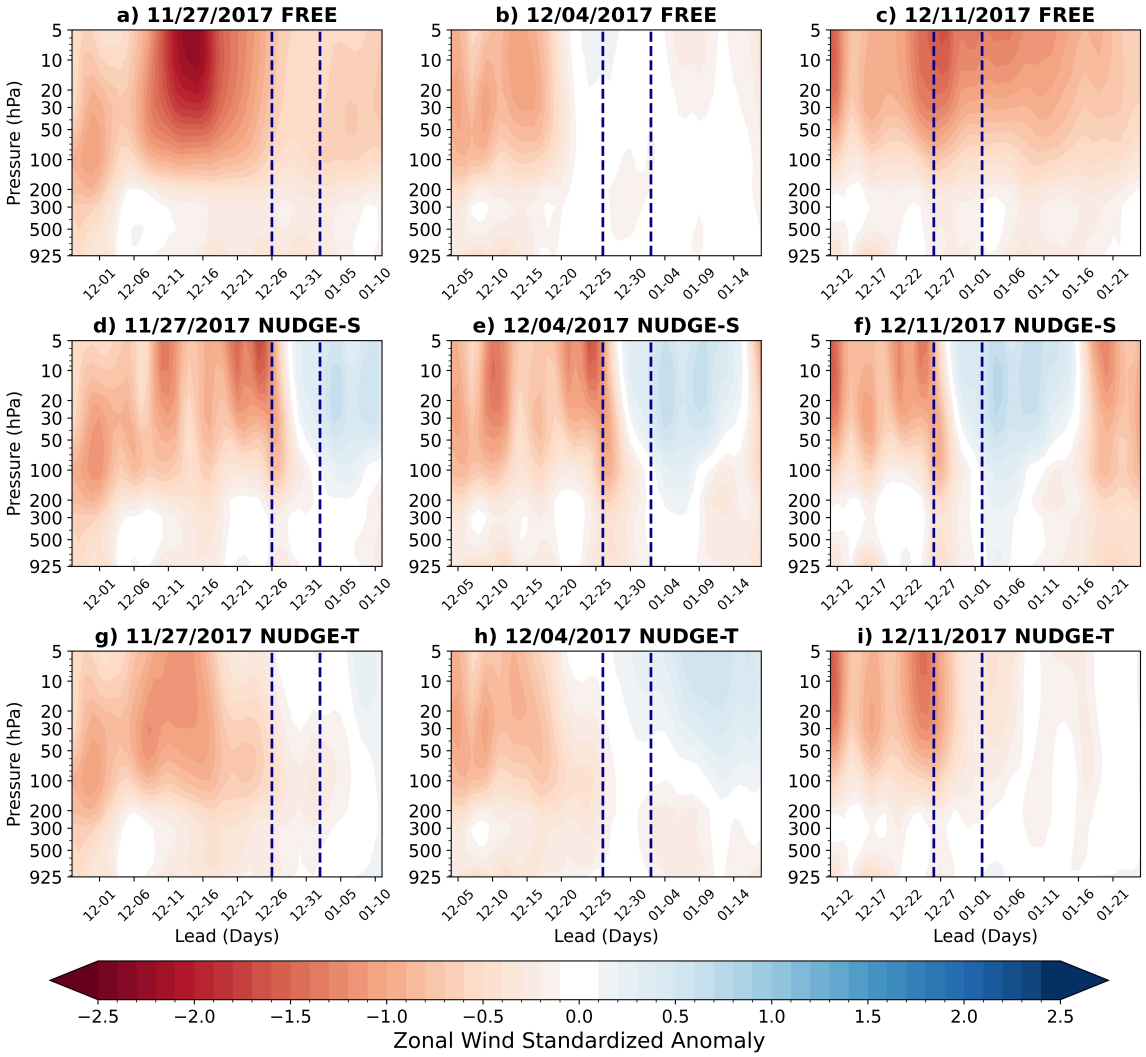


Figure A.6: The evolution of the ensemble mean zonal wind standardized anomaly (shading) averaged over the polar cap (60–90°N for the three subseasonal forecast initializations of 27 November 2017, 4 December 2017, and 11 December 2017 for a)–c) the FREE experiments, d)–f) the NUDGE-S experiments, and g)–i) the NUDGE-T experiments, respectively. The dashed lines show the start and end dates of the central US CAO.

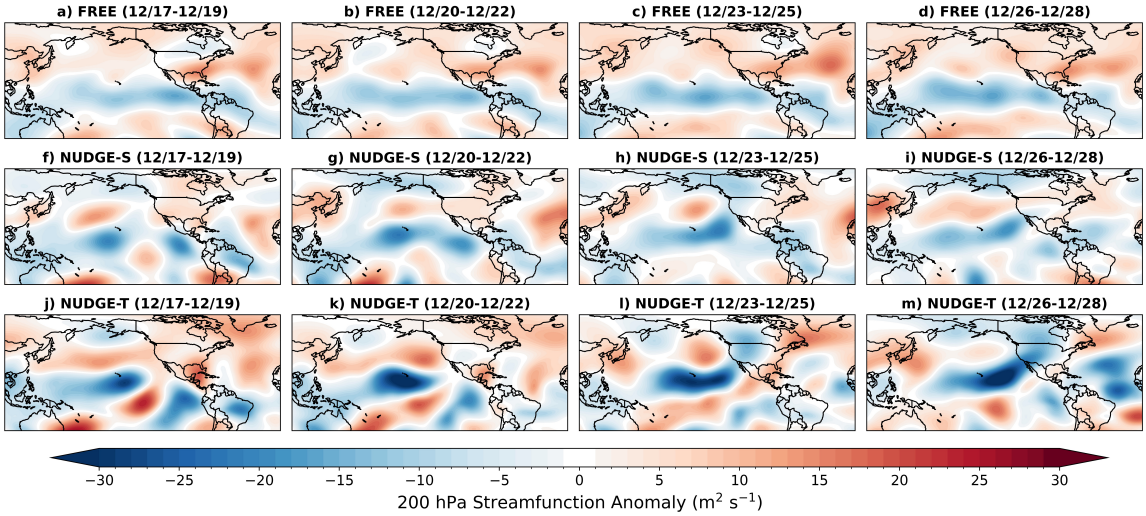


Figure A.7: The evolution of the ensemble mean 200 hPa streamfunction anomalies (shading, m^2 s^{-1}) in the 27 November 2017 initialization for 3-day averaging periods of 17–19 December 2017, 20–22 December 2017, 23-25 December 2017, and 26-28 December 2017 for a)–d) the FREE experiment, f)–i) the NUDGE-S experiments, and j)–m) the NUDGE-T experiments, respectively.

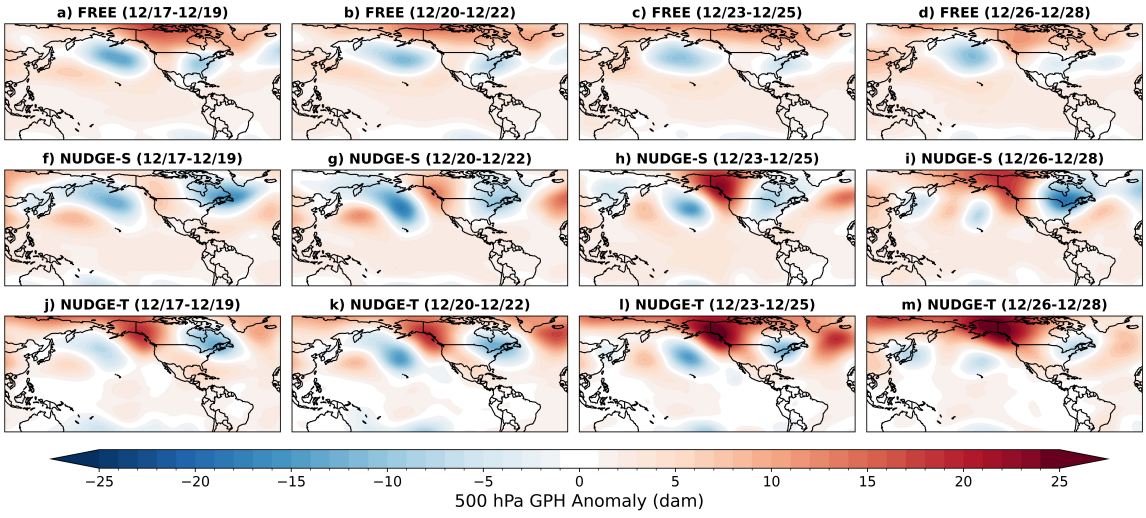


Figure A.8: The evolution of the ensemble mean 500 hPa GPH anomalies (shading, dam) in the 27 November 2017 initialization for 3-day averaging periods of 17–19 December 2017, 20–22 December 2017, 23-25 December 2017, and 26-28 December 2017 for a)–d) the FREE experiment, f)–i) the NUDGE-S experiments, and j)–m) the NUDGE-T experiments, respectively.

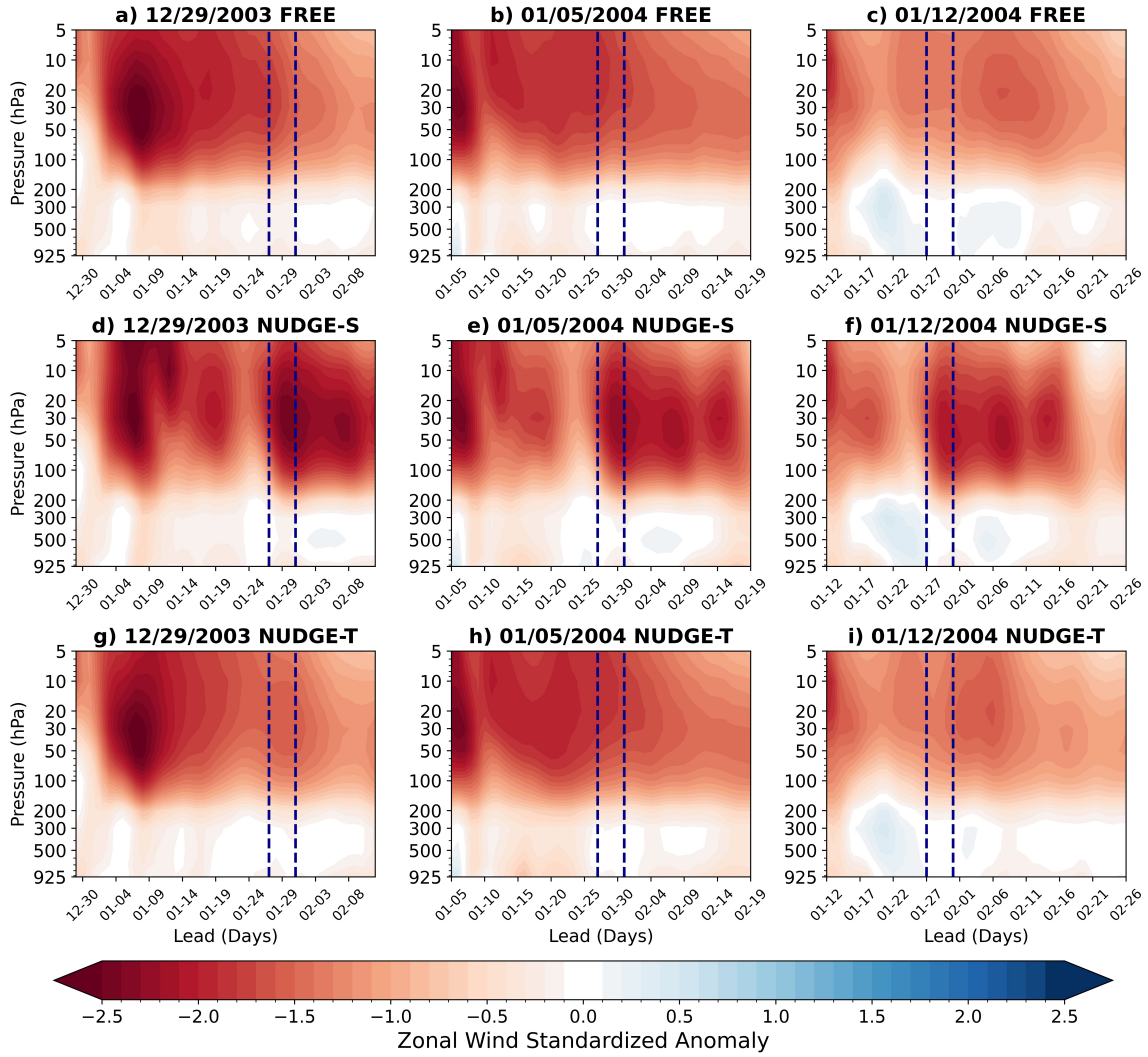


Figure A.9: The evolution of the zonal wind standardized anomaly (shading) averaged over the polar cap (60–90°N for the three subseasonal forecast initializations of 29 December 2003, 5 January 2004, and 12 January 2004 for a)–c) the FREE experiments, d)–f) the NUDGE-S experiments, and g)–i) the NUDGE-T experiments. The dashed lines show the start and end dates of the central US CAO.

3 Appendix C

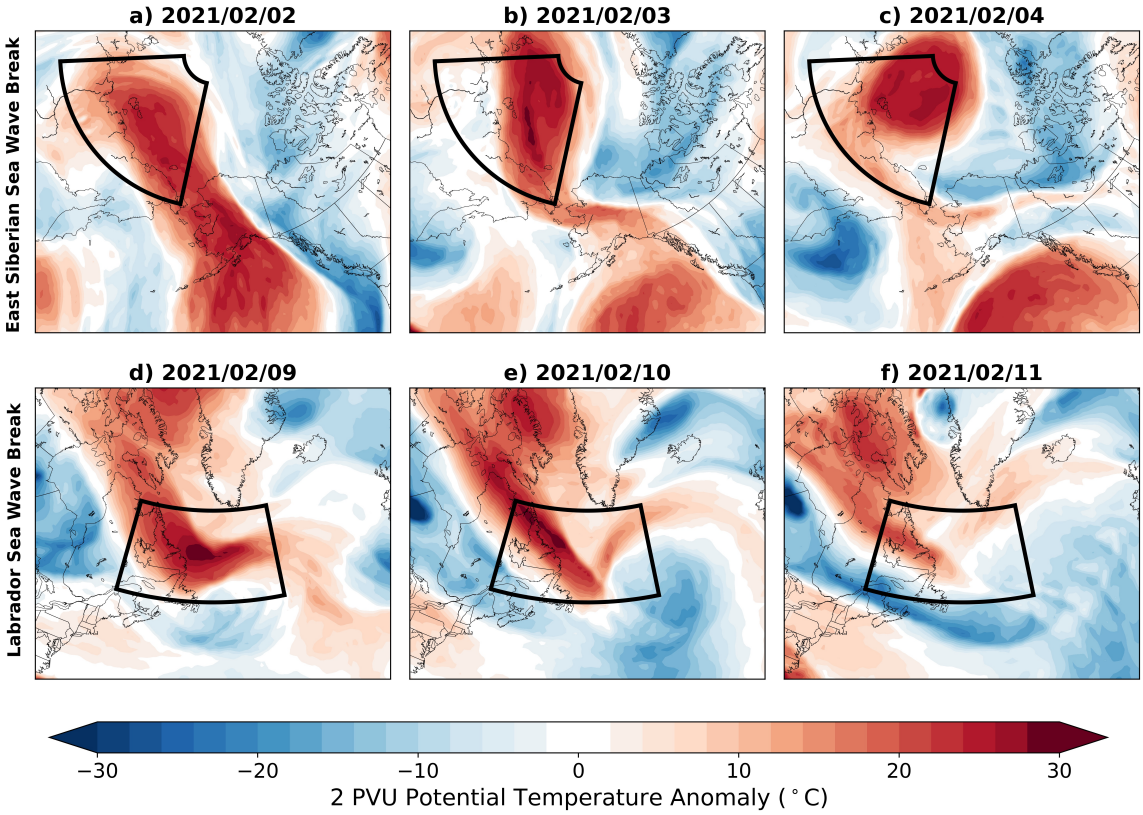


Figure A.10: ERA5 potential temperature anomalies on the 2 PVU surface (°C, shading) for two wave breaks during the evolution of the February 2021 CAO: a) – c) an East Siberian Sea anticyclonic wave break between 2-4 February 2021 and d) – f) a Labrador Sea cyclonic wave break between 9-11 February 2021. Black boxes in a) – c) and d) – f) represent the East Siberian Sea and Labrador Sea wave break regions, respectively.

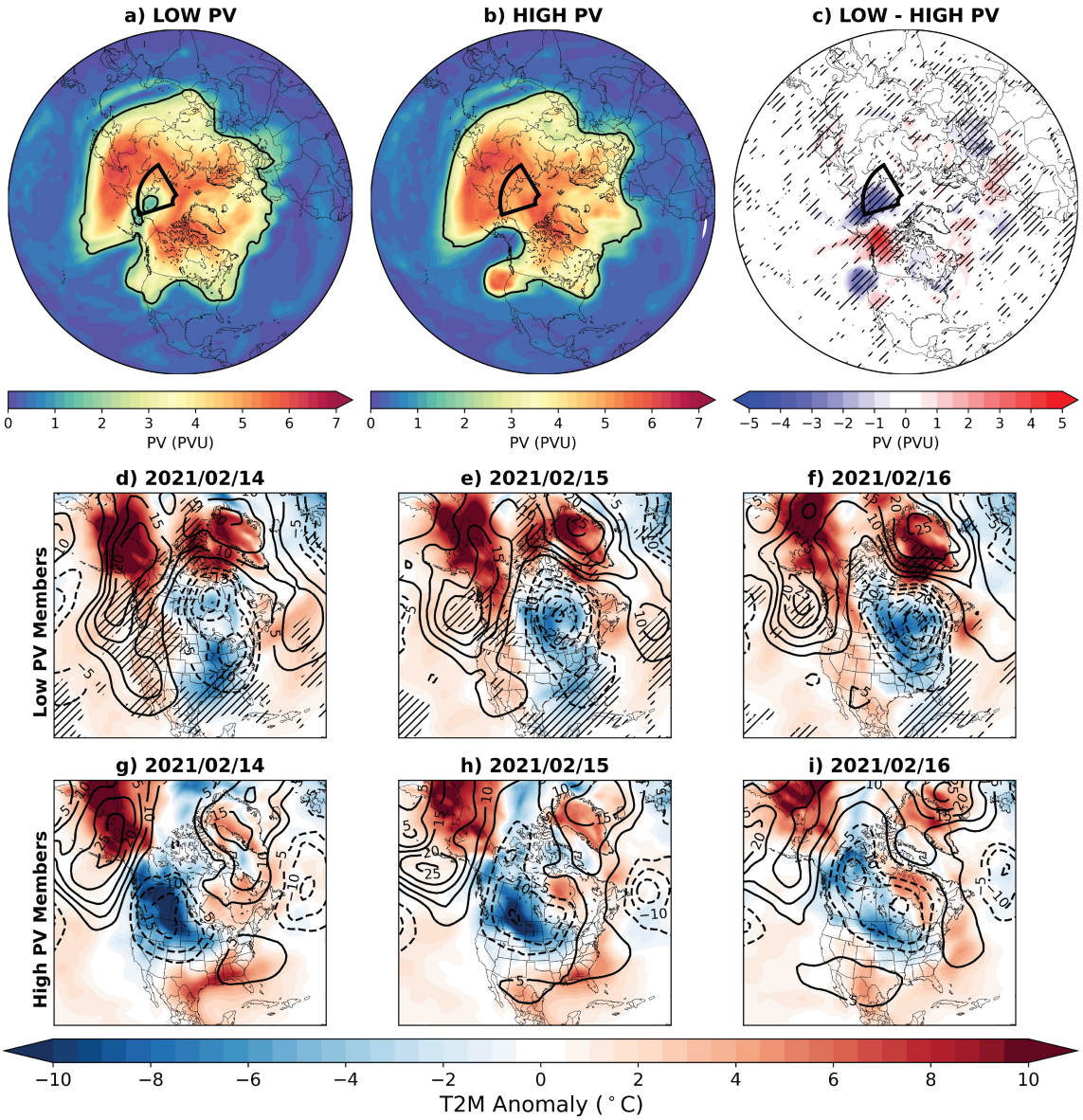


Figure A.11: As in Fig. 5.4 but for the 3 highest and lowest East Siberian Sea 320 K PV ensemble members in the NCEP model.

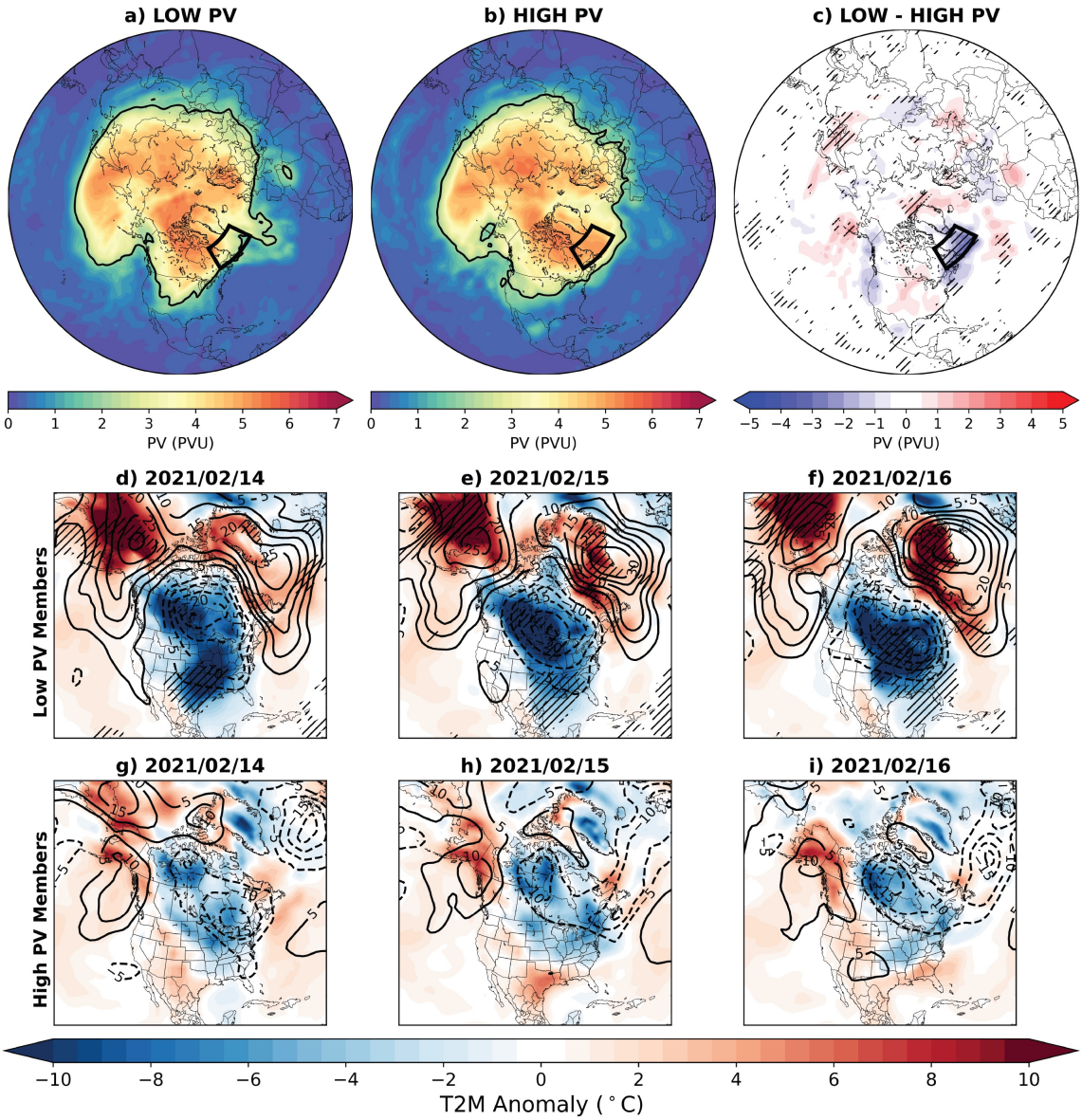


Figure A.12: As in Fig. 5.5 but for the 3 highest and lowest Labrador Sea 320 K PV ensemble members in the NCEP model.

4 Appendix D

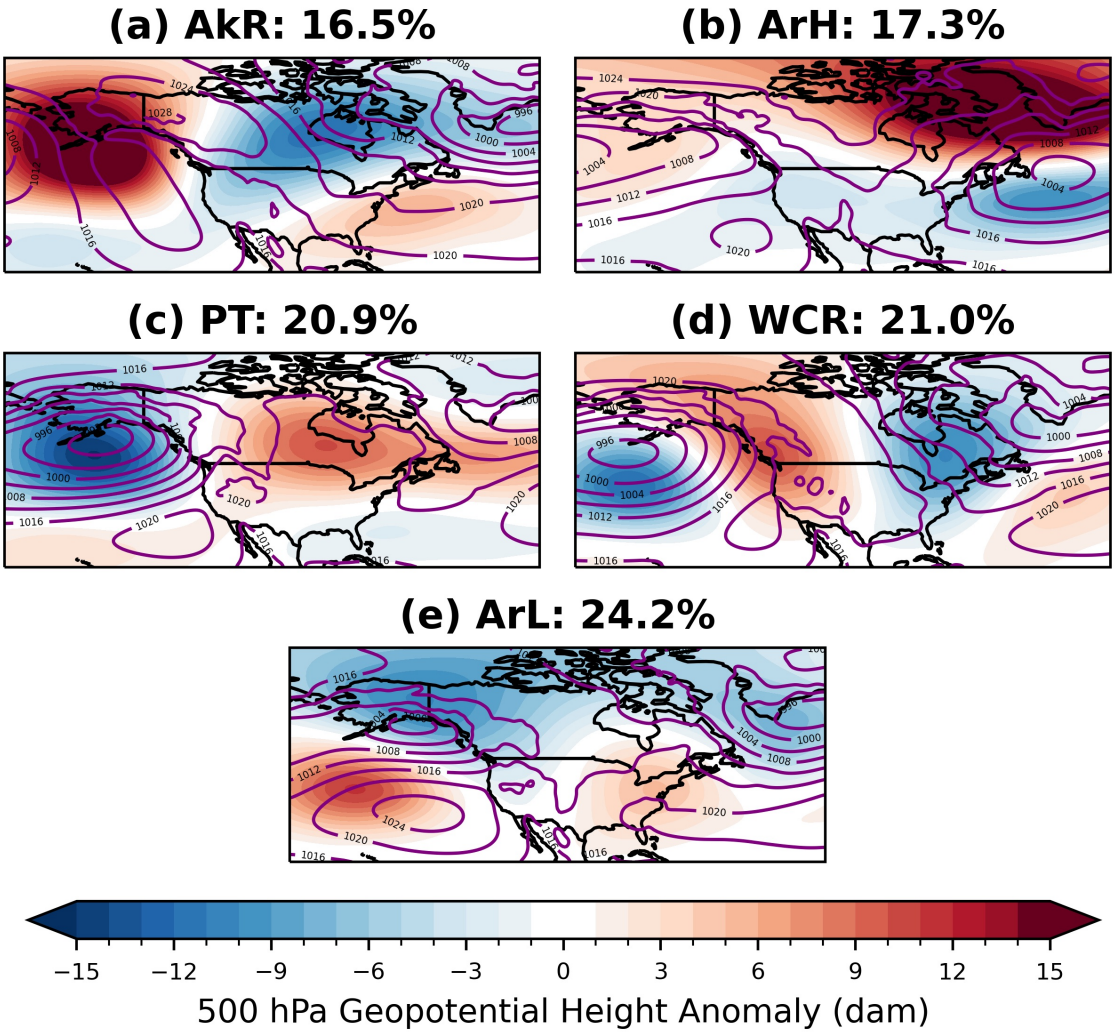


Figure A.13: Composite mean maps of ERA5 linearly detrended 500 hPa GPH anomalies (shading, dam) and full-field MSLP (contours, hPa) for a) the Alaskan Ridge (AkR), b) the Arctic High (ArH), c) the Pacific Trough (PT), d) the West Coast Ridge (WCR), and e) the Arctic Low (ArL), respectively. Percentages on panels represent the frequency of days occupied by a given winter weather regime in the 1950-2023 period. Anomalies are with respect to the 1979-2020 climatology.

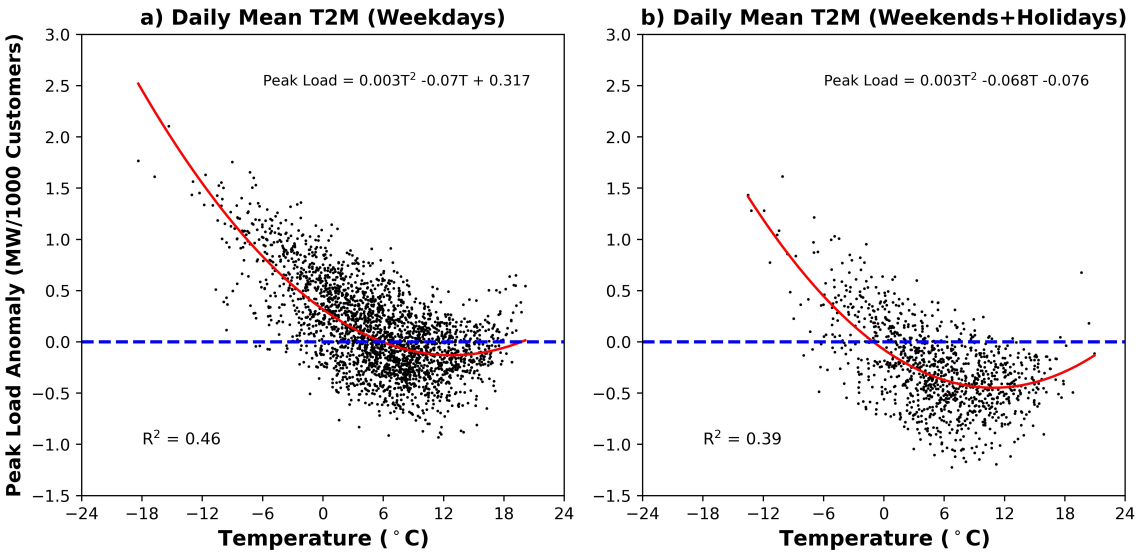


Figure A.14: Scatter plots of peak load anomaly (MW/1000 customers) versus daily mean ERA5 T2M (°C) for (a) non-holiday weekdays and (b) weekend days and non-weekend holidays. We have selected Christmas Day, Thanksgiving Day, and New Year's Day as the holidays to select. The daily average T2M has been averaged over the boxed region in Fig. 2. The red lines represent the second degree polynomial model for the relationship between the two parameters in each plot, with the equation of this polynomial shown on the plot, where T represents the temperature, as well as the coefficient of determination (R^2) value for peak load anomaly respective to the polynomial model. The blue dashed lines represent an average daily peak load between 1999-2022.

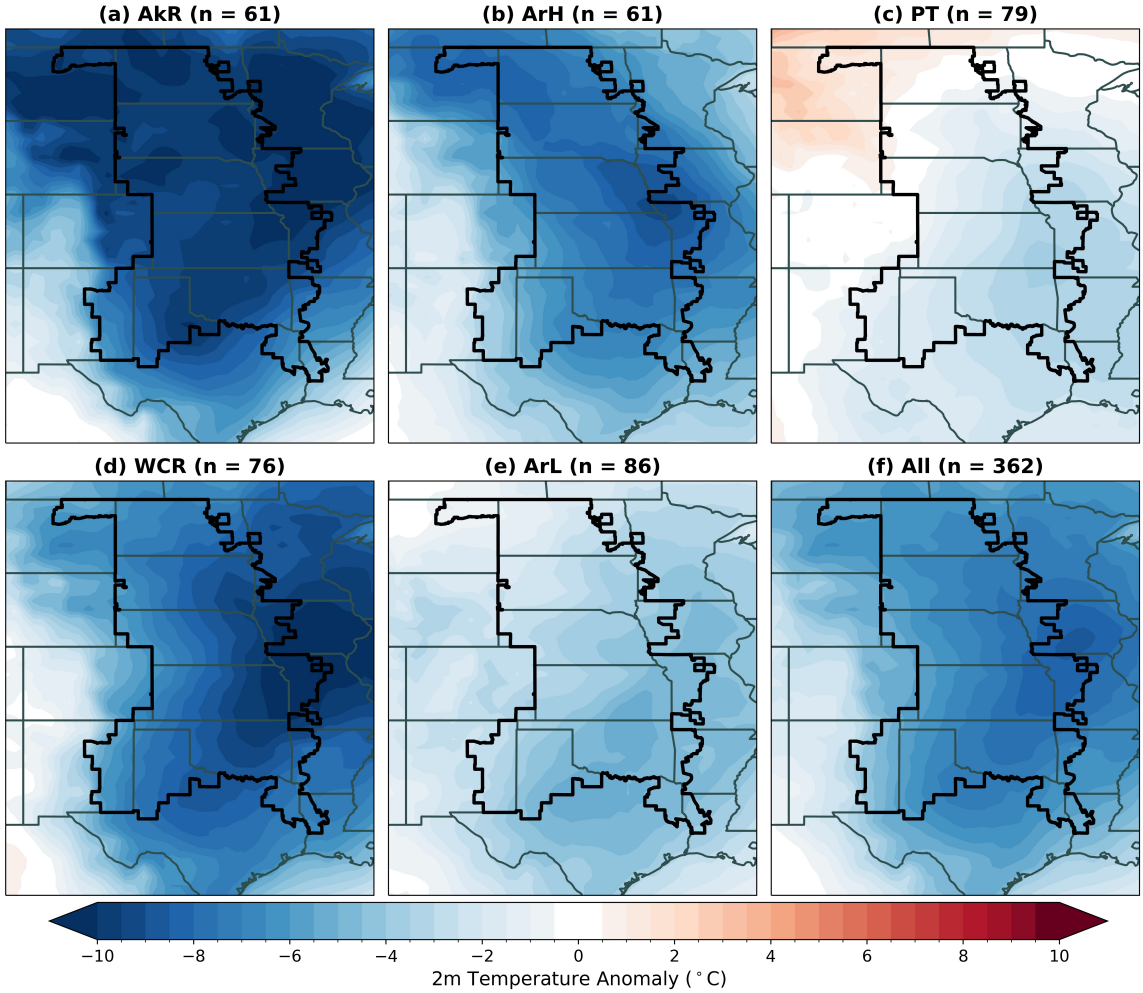


Figure A.15: Composite maps of ERA5 T2M anomalies (shading, °C), respective to the 1991-2020 climatology, for days exceeding the regime-specific 90th percentile of peak load anomalies in each weather regime of a) AkR, b) ArH, c), PT, d) WCR, e) ArL, and f) all days. The black contour represents the SPP region and the sample size is given for each composite.