

UNIVERSITY OF OKLAHOMA
GRADUATE COLLEGE

SKILL AND USABILITY: SUBSEASONAL TO SEASONAL PREDICTABILITY
OF 14-DAY EXTREME PRECIPITATION PERIODS IN THE CONUS

A DISSERTATION
SUBMITTED TO THE GRADUATE FACULTY
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy

By
MELANIE SCHROERS
Norman, Oklahoma
2024

SKILL AND USABILITY: SUBSEASONAL TO SEASONAL PREDICTABILITY
OF 14-DAY EXTREME PRECIPITATION PERIODS IN THE CONUS

A DISSERTATION APPROVED FOR THE
SCHOOL OF METEOROLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Elinor Martin, Chair

Dr. Jason Furtado

Dr. Renee McPherson

Dr. Harold Brooks

Acknowledgments

It is surreal that I am so near the end of my graduate school journey. Norman has truly become a home to me since moving here in 2018. I have had the pleasure of making lifelong friendships, found my place among the scientific world and created a network of like-minded professionals to share my love of science, all during my time in Norman. This was thanks to the wonderful and very large community I have built around me to support me through this process.

Firstly, I would like to thank my advisor, Dr. Elinor Martin, for seeing my potential during the summer of 2017 while in the NWC REU cohort. Although I knew I wanted to continue my research journey, I was very much lost in the new world of atmospheric science. The guidance and belief in my ability was paramount to the confidence needed to tackle a brand-new field coming from my undergraduate physics department. Thank you, Elinor, for all of the meetings spent talking through all my ideas and struggles. As well as all the time spent off topic, talking about academics or the latest funny thing happening in our lives. I truly believe I have grown not only as a scientist, but as an educator and collaborator under your advisement.

Next, I would like to thank my committee, Dr. Jason Furtado, Dr. Harold Brooks and Dr. Renee McPherson. To Dr. Jason Furtado, thank you for your support through the building of statistics skills and for guidance during my time as TA within your classroom. To Dr. Harold Brooks thank you for lessons learned on forecast verification in the classroom and beyond, as well as the many laughs and stories shared in the Weather Center. Finally thank you to Dr. Renee McPherson for suggesting I talk to Dr. Martin about an exciting graduate school opportunity in extreme precipitation while in the REU internship, as well as all the guidance given throughout the years in end-user work. Lastly thank you to the National Science Foundation, this work has been supported by NSF Grants #1663840 and #1944177.

My office mates in room 5700 have been a support system that I never envisioned having while in graduate school. Getting to enter the office and starting graduate school alongside Dr. Emily Tinney was such a comfort in times of high stress! Thank you to the former and current climate variability and change group – including Daniel

Tripp, Bryony Puxley, Margaret Hollis, Alethia Kielbasa and Dr. Ben Davis – for all the time spent listening when drawing out my many ideas on the office white boards. Lastly, I would like to thank other 5700 office mates Dr. Elisa Murillo-Greene, Andrea Gordon, Dr. Amanda Murphy and all others for always providing support and laughs throughout the years! Although the office (and maybe hallway) will be much quieter without my laugh, I know office 5700 will continue to be full of joy and hard work even after me.

Being a part of the PRES2iP team has been such a fun experience, and has shown me a positive example of collaborative science. Among the PRES2iP team members, I especially would like to thank Dr. Ty Dickinson for all the brain storming and support throughout our graduate school research. The support from both Dr. Paulina Ćwik and Olivia VanBuskirk was vital to the successful for the planning and coordination of the final end-user workshop. I cannot thank the PRES2iP Broader Impacts group, including Dr. Esther Mullens and Dr. Heather Lazrus, enough for believing in my management skills and for providing helpful insight into co-production of knowledge. Finally, I would like to thank the amazing end-users I have had the pleasure to work with online and in-person.

There are also many other people that have helped to guide my scientific identity to feel confident in my ability to achieve my dreams of completing my dissertation. Thank you to all the wonderful staff and faculty of the School of Meteorology for support and education that has been provided to me. Thank you to my REU advisors, Dr. Adrienne Wooten and Dr. Derek Rosendahl, for showing me the world of climate extremes in the summer of 2017. Thank you to my REU friends and mentors, Dr. Daphne LaDue and Briana Lynch, I made during REU that ushered me into the weather world and have continued to support me since 2017. Lastly, thank you to Dr. Robynne Lock, my undergraduate research advisor, and the rest of the faculty at Texas A&M University-Commerce. The foundation that was laid by the physics department at TAMUC helped increase my confidence to succeed as a student, a leader and finally as an interdisciplinary researcher.

To my friends in Norman – what a journey we have been on! Thank you for all the laughs, Friendsgivings, shark parties, game nights, book clubs, etc. My graduate school experience has been infinitely better because of the support system we have created together. I have grown into my favorite version of myself with you all, and

look forward to being your friend for the rest of my life. Next, I want to thank my roommates over the years, Joel McAuliffe and Dr. Sean Ernst. Thank you for splitting rent and putting up with Kits very cute, but very loud, guarding nature. To all the rest of my friends that have stayed with me through hard times and very good times, thank you for all your support and love!

Lastly, thank you to my family including my mom, Cary Schroers, and dad, Michael Schroers, for supporting my curious nature and always trying to help me answer my favorite question “Why?” To my siblings, nephews, and all of my aunts and uncles, thank you for supporting me on this long journey. The love and support from family has been a backbone for me through my many, many, many years of schooling and I am lucky and privileged to be able to have you all in my life! I have looked forward to every holiday that I could step away from graduate school to be with my family and look forward to finally being able to answer that tough question of “So, what next?”

Table of Contents

Acknowledgments	iv
List Of Tables	ix
List Of Figures	x
Abstract	xiii
1 Introduction	1
2 Verification of 14-day Extreme Precipitation Events in S2S Models	5
2.1 Background	5
2.2 Data	7
2.2.1 14-day Extreme Precipitation Event Database	7
2.2.2 S2S Hindcasts and Observation Data	7
2.3 Methods	10
2.3.1 Verification Methods	10
2.4 Verification Results	11
2.4.1 Precipitation Verification	11
2.4.2 Atmospheric Variable Verification	13
2.4.3 Extreme Events Bias	16
2.5 Case Study Analysis	23
2.5.1 Case Study Selection	23
2.5.2 Case Study Forecasts	27
2.6 Summary and Discussion	33
2.7 Conclusion	36
3 The Role of Extratropical Cyclones within 14-day Extreme Precipitation Events	38
3.1 Background	38
3.2 Data and Methods	41
3.2.1 14-day Extreme Precipitation Events	41
3.2.2 ETC tracking	41
3.2.3 Matching ETC Activity with Events and Associated Precipitation	42
3.3 Results	43
3.3.1 Comparison of Event ETC Activity within Event and Non-Event Periods	43
3.3.2 Precipitation Associated with ETC activity	47

3.4	Discussion	52
3.5	Conclusion	54
4	Listening to Stakeholders: S2S Precipitation Forecasts Usability	56
4.1	Background	56
4.2	Approach and Methods	58
4.3	Lessons Learned	62
4.3.1	Action versus Lead Time Activity	62
4.3.2	Text Product Activity	63
4.3.3	Precipitation by Day Activity	66
4.3.4	Precipitation by Amount Activity	66
4.3.5	Worst-Case Scenario Activity	68
4.3.6	Forecast Error Activity	69
4.3.7	Object-Oriented Activity	70
4.4	Summary	73
5	Discussion	77
	Reference List	84
	Appendix	91
1	Appendix A	91

List Of Tables

2.1	Group descriptions, including the number of events from 1996-2014. . .	9
2.2	S2S Project Hindcast Model Details	10
3.1	Number of events from 1979 to 2018 that had 0, 1, 2, 3, 4, or 5+ ETCs occurred near the event region. For all events and the six regional groups.	44
3.2	Number of events from 1979 to 2018 that had 0, 1, 2, 3, 4, or 5+ ETCs occurred near the event region per season for all events.	45
3.3	Summary of the ETC characteristics per group discussed within Section 3.3. Arrows indicate significant difference from non-event periods within the group domain.	54
4.1	Goals of each workshop activity and corresponding forecast product pro- vided.	61
A.1	Descriptions of 500hPa geopotential bias distributions	92
A.2	Descriptions of 850hPa specific humidity bias distributions	93

List Of Figures

2.1	Cluster grouping used for S2S model verification of CONUS 14-day extreme precipitation events, clustered in Dickinson et al. (2021)	8
2.2	Heat map of area-averaged ACC values for ensemble-averaged 14-day total precipitation anomalies, for whole model - CONUS domain (top) and event-specific within the six groups (bottom), for ECMWF (left), UKMO (middle), and CNRM (right).	12
2.3	Heat map of area-averaged ACC values for ensemble-averaged 14-day averaged 500 hPa Geopotential Height anomalies, for whole model - CONUS domain (top) and event-specific within the six groups (bottom), for ECMWF (left), UKMO (middle), and CNRM (right).	14
2.4	As in Figure 2.3, but for 850 hPa specific humidity anomalies.	15
2.5	As in Figure 2.3, but for 700 hPa specific humidity anomalies.	15
2.6	As in Figure 2.3, but for 500 hPa omega anomalies.	17
2.7	Distribution of ECMWF 500 hPa geopotential height bias for group 1 (left), 3 (middle), and 5 (right) for bootstrapped (n=2500) non-events Weeks 1/2 and Weeks 4/5, top panels (a), (b), and (c), and Events Weeks 1/2 and Weeks 4/5, bottom panels (d), (e), and (f). Note the differences in the range of specific humidity bias, along the x-axis. Zero bias is indicated as a dashed black line.	18
2.8	As in Figure 2.7 for group 2 (left), 4 (middle) and 6 (left).	19
2.9	ECMWF geospatial bias of 500 hPa geopotential height for group 5 for all lead times. Significant difference from the respective lead-time dependent bootstrapped (n=2500) non-events is shown as a fuchsia contour, at the 90% confidence interval.	20
2.10	As in Figure 2.7 but for 850 hPa specific humidity bias, for groups 1 (left), 3 (middle) and 5 (right)	21
2.11	As in Figure 2.7 but for 850 hPa specific humidity bias, for groups 2 (left), 4 (middle) and 6 (right)	22
2.12	As in Figure 2.9 but for 850 hPa specific humidity bias.	23
2.13	Individual event area-averaged ACC skill comparisons at all lead times for (a) 500 hPa geopotential height and precipitation and (b) 850 hPa specific humidity and precipitation. Case 1 skill event dots are circled in pink, while Case 2 skill dots are circled in black. Note the differences in range along each x-axis.	24

2.14	Precipitation 14-day totals for Case 1 (top) and Case 2 (bottom). For each case precipitation values are shown from PRISM observations (PRISM Climate Group, Oregon State University 2004), ERA5 regridded into S2S model resolution (used as observations for verification) , ECMWF forecasted totals at Week 1/2 and at Week 4/5 lead times. The extreme event polygon (EE polygon) for each case, red, can be seen over the observations and ERA5.	26
2.15	Event polygon of Case 1 (top left) and Hovmöller diagram of regridded ERA5 500 hPa geopotential height anomalies for the 14-day extreme event (bottom left). To the right, hovmöller diagrams of the ECMWF forecasted 500 hPa geopotential height for all four lead times are shown with their respective initialization date. The red bar seen on the top of the forecasts indicate the geospatial location of the event. Data was averaged between 30°N and 50 °N and plotted daily.	28
2.16	As in Figure 2.15, but for 850 hPa specific humidity anomalies.	29
2.17	As in Figure 2.15, but for Case 2's 850 hPa specific humidity anomalies. Data was averaged between 25°N and 45°N and plotted daily.	31
2.18	As in Figure 2.15, but for Case 2's 500 hPa v-wind (not anomalies). Data was averaged between 25°N and 45°N and plotted daily.	32
3.1	Track density over the CONUS of ETC tracks found using Tempest Extremes, as shown in (McCrary et al. 2024), for the time frame of interest, 1979 to 2018.	42
3.2	Three step process to use 6-hourly ETC tracks to find daily event precipitation (ERA5) associated with ETCs. Example is for one day within an event in group 4.	43
3.3	Distributions of all ETC track speeds, 6 hourly, when the track was passing within 10 degrees of the event polygons for event ETCs (blue, left) and for non-event ETCs passing within 8 degrees of the group domain (orange, right). Shown for all 6 groups. Significant difference between the distributions, using a t-test at a $p\text{-value} \leq 0.05$, is denoted as a * next to the group label.	46
3.4	As in Figure 3.3 for the ETC track (a) intensities and (b) maximum intensities.	47
3.5	Distributions of Bootstrapped random 14-day periods mean ETC frequencies are shown for all six groups. The shaded regions show the 95% confidence interval. Red dashed line is the observed event group mean with the value stated at the top right for each group. Note the differences in the ranges of the x-axis.	49
3.6	Distribution of all events ETC associated precipitation, % of total event precipitation associated with ETCs. Events are broken down by group, labeled in the top right. Note the first bar only contains percent values of 0%, while other are bins of 5% increments.	50

3.7	As in Figure 3.6 but broken down by the season of the beginning date of the event.	51
4.1	Summary of the five forecasts periods used for the experimental products that were given to the participants. As an example, model initialization dates are included for Scenario 1, that occurred from 6 – 19 December, 2015.	62
4.2	S2S experimental forecast products designed by the PRES2iP research team as a result of participant feedback in workshops 1 and 2. Scenario 2 (21 April to 4 May 2017) is displayed, showing forecasts for Missouri and surrounding states at a lead time of 7 days prior to the start of the extreme precipitation period. (a) Ensemble-mean, 14-day total precipitation and (b) ensemble-mean, 14-day precipitation anomalies are shown. Precipitation values are in inches for workshop participants. . .	63
4.3	As in Figure 2, but for the time series product that includes (a) daily precipitation, (b) accumulated precipitation, (c) ensemble standard deviation and (d) precipitation anomalies for 21 April to 4 May 2017. . .	65
4.4	As in Figure 2, but probability of 14-day total precipitation above 5 inches (127mm).	67
4.5	As in Figure 2, but for 90th percentile of 14-day total precipitation, labeled for participants as the Worst-Case.	69
4.6	Schematic examples of hypothetical Forecast (F) – Observation (O) combinations, from Davis et al. (2006, Fig 1)	71
4.7	Various pairings of forecasts (black) and truth (red). Colored dots represent the polygon centroid. Distances between centroids are about (a) 150 miles, (b) 220 miles, and (c) 350 miles.	72
4.8	As in Figure 7, but for polygons with similar centers but different spatial scales.	73
5.1	Observed tracks from ETC ERA5 track database used in Chapter 3 for one event are shown in the top panel, event polygon shown in red. Tracks from ECMWF, tracked using MSLP, for four different lead times are below. Only 5 members of the 11 members are shown. The event used in this example took place in group 4 from 10 to 23 April, 2001. .	81

Abstract

Long periods of extreme precipitation can cause significant impacts to many economic sectors, including transportation, water resource management, agriculture production, and infrastructure. Planning for these impacts would ideally start between the typical weather (out to 10-14 days) and seasonal (1-3 months) forecast periods. This gap in forecasting periods, referred to as the subseasonal to seasonal timescale (S2S), is known for relatively low accuracy in precipitation prediction when compared to other extremes, such as heatwaves and drought. Thus, very few forecast products exist for stakeholder use for decision-making regarding S2S precipitation. This dissertation (1) quantifies the current prediction capabilities of atmospheric conditions associated with extreme precipitation within S2S models, (2) demonstrates how extra-tropical cyclones play a role in driving extreme periods, and (3) engages with stakeholder communities to learn what is needed for usable S2S products.

Studies have shown that S2S precipitation forecasts have very low skill beyond 2 weeks. Yet studies quantifying the skill associated with atmospheric variables coincident with and preceding extreme precipitation periods have not been conducted. A database of 14-day extreme precipitation periods across the US was used to assess the current predictive skill of synoptic variables during the extreme periods for the first time. The highest skill is seen for geopotential height (representative of atmospheric pressure variations). Since the anomaly correlation coefficient drops below climatology by Week 2 for all variables, model bias during the extreme periods was examined. During the extreme periods, there is a dry bias that increases with lead time for specific humidity. The models also have an inability to realize the magnitude of 500hPa geopotential height dipole, common to all extreme periods. While, on average, skill is low past Week 2, there are cases where the precipitation period is driven by large-scale atmospheric conditions and has skill out to Week 4.

Case studies of these 14-day extreme precipitation periods revealed that they are often created by multiple rounds of storms, which are associated with extratropical cyclone (ETC) activity. Studies have shown that ETC activity accounts for 50% of

Northern Hemisphere precipitation; thus, understanding ETC activity during the 14-day extreme periods would help reveal future avenues of S2S prediction of long-duration precipitation extremes. Using a database of ETC tracks within ERA5 observational data, frequency, speed, intensity and event precipitation contribution of ETC activity during our extreme precipitation periods was analyzed and compared to non-event periods ETC activity. Characteristics of event ETC activity was highly dependent on the region. The Northern Plains events have slow moving, intense ETC that, on average, contributed more than 50% of the events precipitation. However, the West Coast, Great Lakes and Southeast have intense ETCs, that are more frequent during event periods than non-event periods. While the Northeast has many events with intense ETC activity, with higher precipitation contributions, the speed and frequency of ETCs was not significantly different than non-event periods. The Southern Plains did not have any differences between event and non-event ETC activity, with high variance of ETC precipitation contribution. Therefore, there is a strong connection to ETC activity within precipitation events in many regions that would need to be explored within S2S models in the future.

Since S2S forecasts have had low skill, forecast products have been underutilized. This creates an opportunity to work alongside stakeholders to develop forecast products that are useful and usable. The Prediction of Rainfall Extremes at Subseasonal to Seasonal Periods (PRES2iP) project team conducted three workshops over six years to engage with stakeholders to learn what is needed for decision-making for subseasonal precipitation. The PRES2iP team used the knowledge gained from the first two workshops to design a suite of extreme precipitation subseasonal forecast products for the third and final workshop in April 2023. Participants identified that impact-based text products alongside quantitative precipitation forecasts helped to improve their interpretation of products at longer lead times and greater ranges of uncertainty. Participants also noted that regional time-series of precipitation were useful when a measure of uncertainty was also included in the product. Uncertainty tolerance relating to the subseasonal forecast location and size of the 14-day extreme precipitation period was also found using an object-oriented verification scheme. Overall, there was no single product that worked best, but instead, a combination of spatial and temporal products with associated narratives, including impacts is needed. This work helps to close the gap associated with S2S prediction of impactful extreme precipitation, while

also engaging potential users throughout the research process to co-produce usable and trustworthy S2S precipitation forecast products.

Chapter 1

Introduction

Skillful prediction of extreme precipitation is a topic of significance for a variety of end-user sectors, including civil protection (Curriero et al. 2001), water resource management (Feldman et al. 2021), flood disaster prevention (Wilhelmi and Morss 2013), agricultural production (Klemm and McPherson 2017), transportation planning (Liu et al. 2022), and infrastructure security (McPhillips et al. 2021). When a precipitation period lasts multiple weeks compounding impacts are seen within these sectors over the region affected (Schroers and Martin 2022). Planning for these impacts would ideally begin at the subseasonal to seasonal timeframe (S2S; i.e. 2 weeks to 3 months), ahead of the typical weather timeframe (i.e. 0-10 days ahead). Application of operational and experimental S2S forecasts have shown that “horizon scanning” for signs of potential impacts caused by precipitation extremes is useful for many industries including agriculture, water management and emergency management (White et al. 2017; Merz et al. 2020; White et al. 2022; VanBuskirk et al. 2023a). Yet, this relies on S2S precipitation forecast products that are skillful, interpretable, and aid current decision-making processes. Thus there has been an increasing demand by decision makers for forecasts of precipitation extremes given that daily and multi-day extreme precipitation events are increasing, primarily due to the increase in the most extreme values (Kunkel et al. 1999; Groisman et al. 2012; Barlow et al. 2019).

The S2S timeframe is highly important to decision makers for planning and preparation for extreme weather hazards as regional disaster preparedness actions require

time for execution. For example, it can take several weeks to procure disaster-response or mitigation supplies (e.g., first aid kits, cleaning supplies, water-purification tablets, sandbags to reinforce riverbanks; White et al. 2017). Additionally, decision-makers within water management could utilize S2S forecasts of rainfall as input into stream-flow modeling to show potential flood risk (White et al. 2015). Forecasts targeting the S2S timeframe can extend time to pre-purchase and preposition essential materials in the at-risk region, reducing response time. Yet, weather and climate prediction products are often under-used by decision makers (Rayner et al. 2005; Morss et al. 2008; Weaver et al. 2013), most likely because forecasters and end-users perceive scientific information in different ways (Lemos and Morehouse 2005). For example, information that forecasters deem useful may not be usable by decision makers due to regulatory, political, or resource constraints (Morss et al. 2005; VanBuskirk et al. 2023b). Engaging users of the forecast products throughout the research development process has increased information adoption (Lemos and Morehouse 2005) and resulted in “actionable science,” i.e., science that is directly applied by decision makers and, by extension, benefits society (e.g. Vogel et al. 2016; Norström et al. 2020; Nyboer et al. 2021).

An ongoing challenge in the prediction of extreme precipitation on the S2S timeframe is the relatively low skill of precipitation forecasts and, subsequently, the ability to convey the associated risks and uncertainties to decision-makers. To increase research at this time frame the S2S prediction project created a database of S2S real-time and hindcast forecasts, out to a lead time of 60 days (Vitart et al. 2017). These models tend to have low skill in midlatitude precipitation past Week 1, especially when compared to the skill of other climate extremes within the United States, such as heat waves (de Andrade et al. 2019; Domeisen et al. 2022). Skill of long-duration precipitation extremes also showed negligible skill within the S2S project database (McAfee et al. 2023). Recent studies have shown improvements of S2S precipitation skill using

post-processing of S2S model output, including atmospheric variables and precipitation forecasts (Dickinson 2023; Lyu et al. 2023; Zhang et al. 2023). The low forecast skill seen for S2S precipitation forecasts may limit the trust and usability of potential forecast products at S2S lead times.

To address the gap in skillful and useful S2S extreme precipitation forecasts and encourage knowledge exchange among researchers and decision makers, the Prediction of Rainfall Extremes at Subseasonal to Seasonal Periods (PRES2iP) project, funded by the National Science Foundation (NSF), applied the stakeholder engagement approaches of “participate” and “empower,” as defined by Bamzai-Dodson et al. (2021). Specifically, the PRES2iP team collaborated with decision makers during workshops to steer PRES2iP research questions and co-develop S2S precipitation forecast products to be analyzed during the final workshop.

The long-duration extreme precipitation events analyzed within this dissertation come from a database of 14-day extreme precipitation periods, defined in Dickinson et al. (2021). Events were defined using a three criteria approach: accumulation, duration and size of area affected. These events are commonly associated with a broad area of large positive moisture anomalies and a trough-ridge anomaly dipole centered over the region to help drive ascent (Jennrich et al. 2020; Schroers and Martin 2022; Dickinson 2023). Further, looking at the daily evolution of the precipitation events shows multiple peaks in area-averaged rainfall (Pruitt et al. 2023), suggesting the 14-day events have multiple systems of precipitation that compound to an impactful two week period. There are multiple types of precipitation systems that can cause rainfall within a region including: extra-tropical cyclones (ETC) and their associated frontal systems, the North American Monsoon, mesoscale convective systems (MCS), small-scale unstable air mass thunderstorms, orographic precipitation and tropical cyclone associated precipitation (Kunkel et al. 2012). Out of these systems, ETCs, and their

related fronts, account for a majority of the climatological total precipitation over the midlatitudes (Hawcroft et al. 2012; Kunkel et al. 2012; Pfahl and Wernli 2012; Agel et al. 2015), with ETC activity accounting for up to 80% of daily extreme precipitation events in parts of the CONUS (Pfahl and Wernli 2012).

This goal of this study is to examine skill and usability of S2S prediction of 14-day extreme precipitation events. This is accomplished through three connected research projects, organized in three chapters. In Chapter 2, S2S predictability of long-duration extreme precipitation events and their associated atmospheric variables is analyzed. In Chapter 3, drivers of the events are examined through the role of ETCs. Lastly, in Chapter 4, an understanding of the potential usability of S2S precipitation forecasts, to help support the decision making process at S2S lead times, is gained through discussions with end-users.

Chapter 2

Verification of 14-day Extreme Precipitation Events in S2S Models

2.1 Background

Although model forecast of precipitation have low skill on average (de Andrade et al. 2019; Domeisen et al. 2022), there have been improvements using post-processing of S2S model output, that utilize atmospheric variables and precipitation. Bias-correcting S2S model precipitation fields alone have shown to either slightly improve or leave skill unchanged (Baker et al. 2019; Kolachian and Saghafian 2019; Li et al. 2019). Other post-processing techniques commonly involve the use of atmospheric variables, that have longer lasting skill than precipitation alone. Regime-based forecasting has shown skillful forecasts of 500 hPa geopotential height regimes associated with precipitation out to 10-15 days in advance (Robertson et al. 2020). Machine learning techniques have been used to improve skill by using inputs of S2S forecasted precipitation along with other S2S model outputs. For example, Lyu et al. (2023) used a U-net model approach to improve South China summer precipitation skill using ECWMF precipitation and atmospheric variable fields. Their analysis on predictor importance revealed that precipitation supported improved skill at Week 1, whereas atmospheric variables contributed to skill improvement past Week 1. Zhang et al. (2023) found that the addition of S2S forecasted surface temperatures and geopotential heights into their model

improved skill out to Week 4, when compared to their model using S2S forecasted precipitation alone. Further, Dickinson (2023) found that convolutional neural networks may be able to increase prediction skill of long-duration precipitation extremes when trained on S2S model output, including geopotential height, specific humidity, temperature, wind fields and vertical velocity.

It is known that synoptic forcing, geopotential height dipoles and anomalous moisture in a region, can be responsible for driving long-duration extreme precipitation events (Jennrich et al. 2020; Schroers and Martin 2022). One study found that on average 500 hPa forecasts within the S2S project models have skill out to day 13 in the Northern hemisphere, with higher skill seen in the winter (Son et al. 2020). Regime based verification has shown that S2S models can better predict persistent synoptic patterns but struggle to simulate regime transitions (Ferranti et al. 2015; Tuel and Martius 2023). Therefore there is potential for skillful forecasts of atmospheric variables within S2S models, also shown by the use of these variables in post-processing techniques in order to improve precipitation skill. Yet no studies have quantified the skill and biases of atmospheric variables associated with extreme precipitation with S2S models.

This chapter seeks to address the following research questions:

- What is the skill of atmospheric variables associated with 14-day extreme precipitation events within the S2S models and how does it compare to overall model skill?
- How do the bias distributions of these variables compare to overall model bias?
- Can case studies reveal preferred synoptic set ups for skillful S2S precipitation forecast and how it may differ from a forecast bust?

To address these questions, three S2S project models hindcasts are used to quantify the skill and bias of synoptic conditions associated with 14-day extreme precipitation events within the CONUS.

2.2 Data

2.2.1 14-day Extreme Precipitation Event Database

A database from 1915-2018 of 14-day extreme precipitation events across the CONUS created by Dickinson et al. (2021) was used for this study. Events were found using three thresholds: (1) 14-day precipitation totals have to be above the 99th percentile of the 14-day climatology, (2) 7 days of the 14-day period receive precipitation above their daily-mean and (3) the area of the event (identified by the event polygon) must be greater than or equal to 200,000 km². These criteria ensure that the event is extreme, large, and persistent. A total of 851 events were found throughout the CONUS. The resulting event polygons were then clustered, using k-means clustering, into 15 clusters. To increase the sample size of events, but keep regional similarity, events evaluated within the S2S hindcasts were placed into 6 regional groups, from the 15 clusters (Figure 2.1). Groupings of clusters were based on similar synoptic connections seen during extreme events, while also balancing a similar number of events covered during S2S model verification, Table 2.1.

2.2.2 S2S Hindcasts and Observation Data

Three model ensembles from the S2S project database were investigated (Vitart et al. 2017). Models used are from European Centre for Medium-Range Weather Forecast (ECMWF), the Met Office (UKMO) and Météo-France/Centre National de Recherche

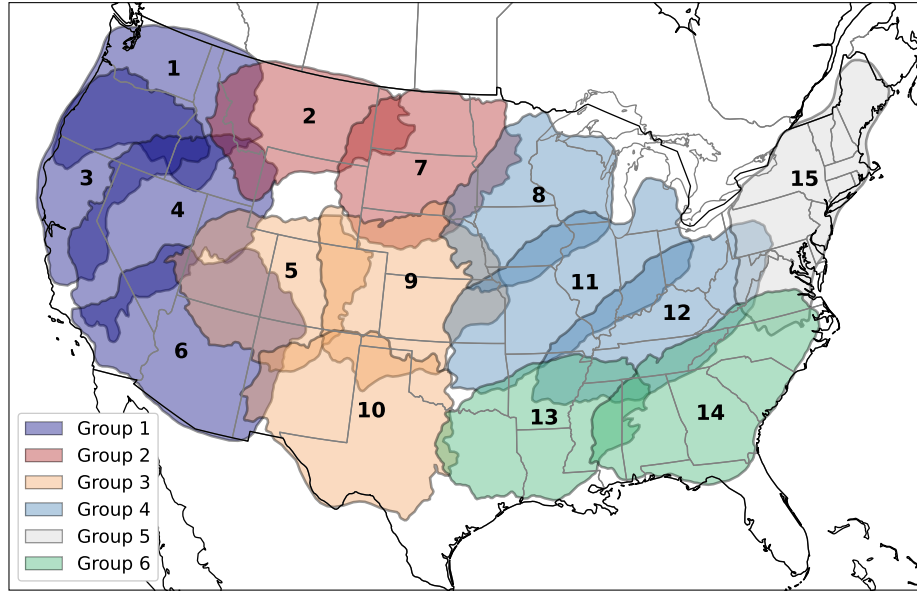


Figure 2.1: Cluster grouping used for S2S model verification of CONUS 14-day extreme precipitation events, clustered in Dickinson et al. (2021)

Météorologiques (CNRM). Models were chosen due to their higher skill seen in prediction of precipitation extremes relative to other S2S models (McAfee et al. 2023). The main characteristics of the three models can be found in Table 2.2. Verification was conducted on the hindcast common period of 1996-2014. This period covers 171 extreme events across all groups, Table 2.1. Along with precipitation, atmospheric variables chosen for verification include 500 hPa geopotential height, 700 hPa and 850 hPa specific humidity. Previous studies have shown the importance of mid-level geopotential height and moisture during these events (Jennrich et al. 2020; Dickinson et al. 2021; Schroers and Martin 2022) and these variables have been widely used as inputs within machine learning models for precipitation prediction (Dickinson 2023; Lyu et al. 2023; Zhang et al. 2023). Model anomalies were created by subtraction of the model climatology for all lead times. Model climatology was found by (1) calculating the

Group	Region Covered	Number of Events
1	West Coast	33
2	Northern Plains	25
3	Southern Plains	32
4	Midwest	39
5	Northeast	17
6	Southeast	25

Table 2.1: Group descriptions, including the number of events from 1996-2014.

mean of each day-of-year, during the model common period of 1996-2014, and (2) applying a 15-day triangular smoother similar to the methodology found in Pegion et al. (2019).

All hindcast variables were evaluated against the European Centre for Medium-Range Forecasts ERA5 reanalysis regridded to the common S2S model resolution of $1.5^\circ \times 1.5^\circ$. Anomalies of ERA5 were found using the same climatological period and triangular smoother as the model hindcast anomalies.

Model	Time Range	Reforecast Frequency	Ensemble Size	Resolution
European Centre for Medium-Range Weather Forecast (ECMWF)	1996-2014	Weekly	11	1.5° x 1.5°
Met Office (UKMO)	1996-2015	4/month	2	1.5° x 1.5°
Météo-France/Centre National de Recherche Météorologiques (CNRM)	1993-2014	Weekly	15	1.5° x 1.5°

Table 2.2: S2S Project Hindcast Model Details

2.3 Methods

2.3.1 Verification Methods

S2S models are typically not initialized on the same days or at the same frequency (Table 2.2) and therefore can make verification of specific periods across multiple models difficult. To verify precipitation total and atmospheric variables during the 14-day extreme precipitation events we employed lagged average ensembles (Pegion et al. 2019). This allows for the direct comparison of two-week ensemble-average model anomalies to two-week averaged observational anomalies during the extreme events. This technique causes small differences in lead-times associated with each verification period between the models used. As we are targeting two week events, skill values were found for Weeks 1/2, Weeks 2/3, Weeks 3/4, and Weeks 4/5. Thus, the lead-times associated with the beginning of each verification period vary between lead day 0-7 for Weeks 1/2, 8-14 for Weeks 2/3, 15-21 for Weeks 3/4, and 22-28 for Weeks 4/5. To address whether the skill within the extreme events differ from the model skill of all 14-day

periods (referred to as full model), 14-day averages were found within the models and skill was found using the climpred package (Brady and Spring 2021) for the same lead times as the events.

Anomaly Correlation Coefficient (ACC) was used to assess the model skill of prediction of the 14-day averaged ensemble-mean of each model. Although ACC is a measure of forecast accuracy and not a direct measure of forecast skill, it is used in this study in order to compare accuracy of the models across variables. ACC is also commonly used to measure skill within S2S models (Ferranti et al. 2015; Pegion et al. 2019; Zheng et al. 2019) When finding area-averaged ACC values a CONUS wide domain was used for geopotential height [10°N to 60°N and 135°W to 60°W]. A regional domain was used for remaining variables by creating a domain box using the maximum and minimums of latitude and longitude of the group being assessed.

Although ACC skill values can give us information on how well observations and forecasts are correlated, it does not tell us about differences in magnitude of the fields. Therefore, bias of the raw fields, model - observations, during the event periods were found and compared to the distribution of bootstrapped non-events (n=2500) from the same common period for multiple lead times. Non-events were found by randomly sampling 14 consecutive days from the common period, excluding all event days. Geospatial significant differences were found between the bootstrapped randomly sampled non-event bias and the event bias at the 90 % confidence level.

2.4 Verification Results

2.4.1 Precipitation Verification

To highlight the need to look into other atmospheric variables in tandem with precipitation, Figure 2.2, top row, shows the skill of ensemble-averaged 14-day total precipitation

anomalies for the three S2S project models used in this study. With the exception of Week 1/2 in ECMWF (ACC of 0.52) the model skill of all 14-day periods is negligible in all models, for all lead times. There is also a large drop in model skill seen between Week 1/2 and Week 2/3. During event periods (Figure 2.2, bottom panels), a similar pattern in skill across lead times is seen. The largest skill values, between 0.57 and 0.67, are seen in group 1 (West Coast) at Week 1/2. Although ECMWF has ACC values above 0.5, which is considered as skillful as climatology, for all groups at Week 1/2 on average there is no skill in 14-day precipitation anomalies past Week 1/2 for all three models. These skill values are similar to the precipitation skill when complex bias correction techniques are applied to model precipitation output McAfee et al. (2023).

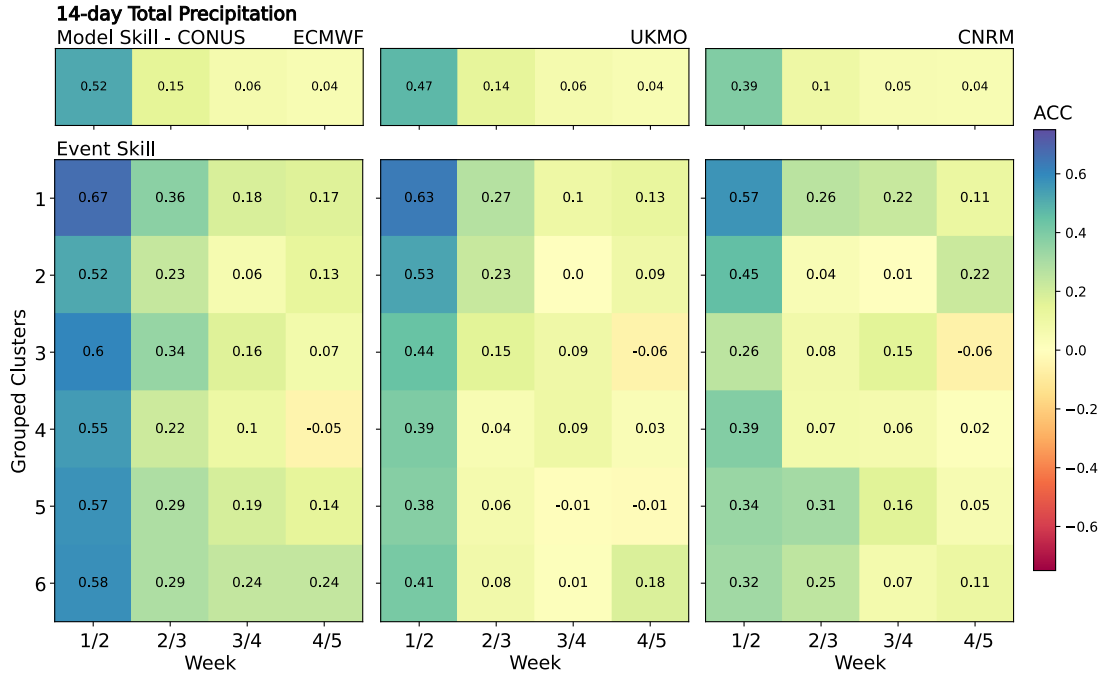


Figure 2.2: Heat map of area-averaged ACC values for ensemble-averaged 14-day total precipitation anomalies, for whole model - CONUS domain (top) and event-specific within the six groups (bottom), for ECMWF (left), UKMO (middle), and CNRM (right).

2.4.2 Atmospheric Variable Verification

Predictive skill of 500 hPa geopotential height anomalies for the full model and event periods are shown in Figure 2.3. The overall full model skill of 14-day periods, (i.e. not event specific), is comparable for all three models. The ACC skill values begin near 0.8 at Week 1/2 and decrease with lead time as expected, with ECMWF having the highest skill out of the three models. This pattern in model skill across lead times is similar for event specific skill for all three models. The event skill across the six groups is above 0.6 during events at Weeks 1/2 for all models, with small differences between groups. At Week 2/3, ECMWF is the only model that has ACC values greater than 0.5 for all six groups. All three models show skill, ACC values above 0.5, out to Weeks 2/3 for group 1 (West Coast). By Weeks 3/4 and Week 4/5 all models have ACC values below 0.5, with the exception of an ACC value of 0.52 in Week 3/4 for group 6 (Southeast) in ECMWF. The skill values for events is often lower than model skill at shorter lead times, but can have similar or higher magnitudes of ACC in some cases at lead times of Week 3/4 and Week 4/5. On average across the models, groups 1 and 6 have similar skill magnitudes to the overall model skill, due to the faster drop off in skill seen in Week 2/3 for the remaining groups.

The skill of the full model and extreme events ensemble-averaged specific humidity anomalies at 850 hPa are shown in Figure 2.4. The model skill of all 14-day periods is highest at Week 1/2, with ACC values above 0.5 for all three models shown. Skill drops below skillful values past Week 1/2, with a larger drop in skill seen in UKMO and CNRM. With overall lower ACC values for 850 hPa specific humidity than skill of 500 hPa geopotential height. Comparing model to event skill, we see differences between magnitude of skill and pattern of skill change across lead times for all models.

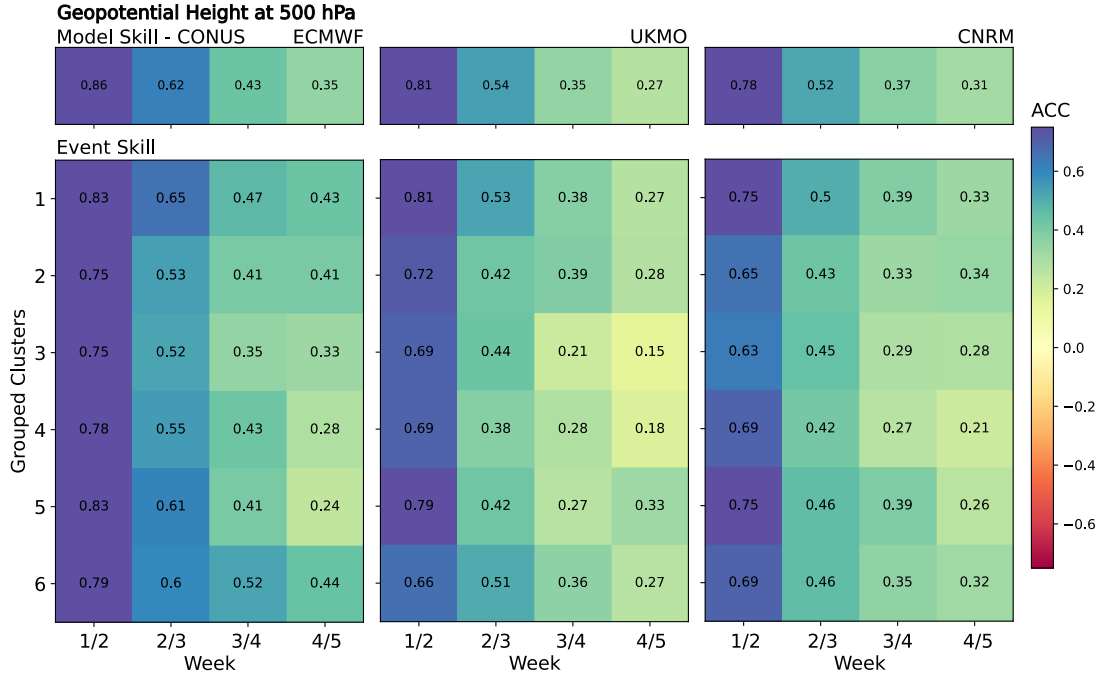


Figure 2.3: Heat map of area-averaged ACC values for ensemble-averaged 14-day averaged 500 hPa Geopotential Height anomalies, for whole model - CONUS domain (top) and event-specific within the six groups (bottom), for ECMWF (left), UKMO (middle), and CNRM (right).

With events having generally lower skill, with a larger drop in skill in between lead times, with an exception of group 1 (West Coast) that has higher skill values seen during events for all lead times in the three models. The event skill of 850 hPa specific humidity has large difference in skill values across models and groups. The highest skill during events is seen in Week 1/2 in ECMWF, where all ACC values are above 0.5. While UKMO and CNRM have low skill for most groups beginning at Week 1/2, that quickly drops to near zero ACC at Week 2/3 for all groups, with the exception of group 1 (West Coast) which has the highest skill across all models with ACC values of 0.76 for ECMWF, 0.49 for UKMO, and 0.71 for CNRM. Conversely, groups 5 and 6, the East Coast, tend to have much lower skill overall, with the only skill values above 0.5 seen in during Week 1/2 in ECMWF. A similar trend is seen for 700 hPa specific

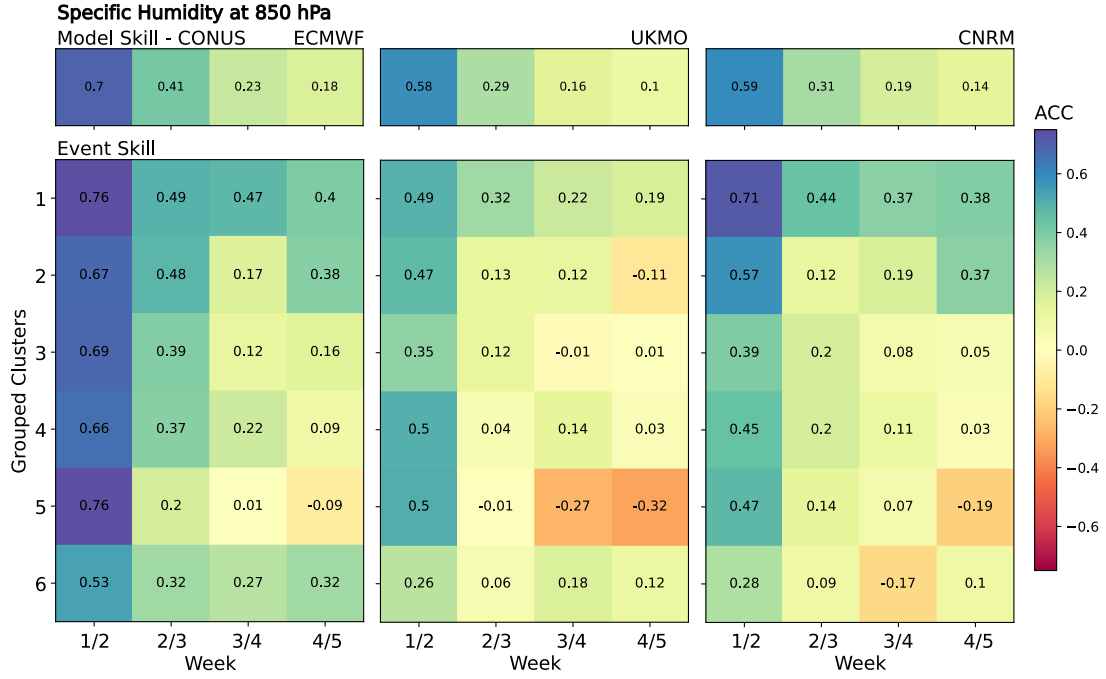


Figure 2.4: As in Figure 2.3, but for 850 hPa specific humidity anomalies.

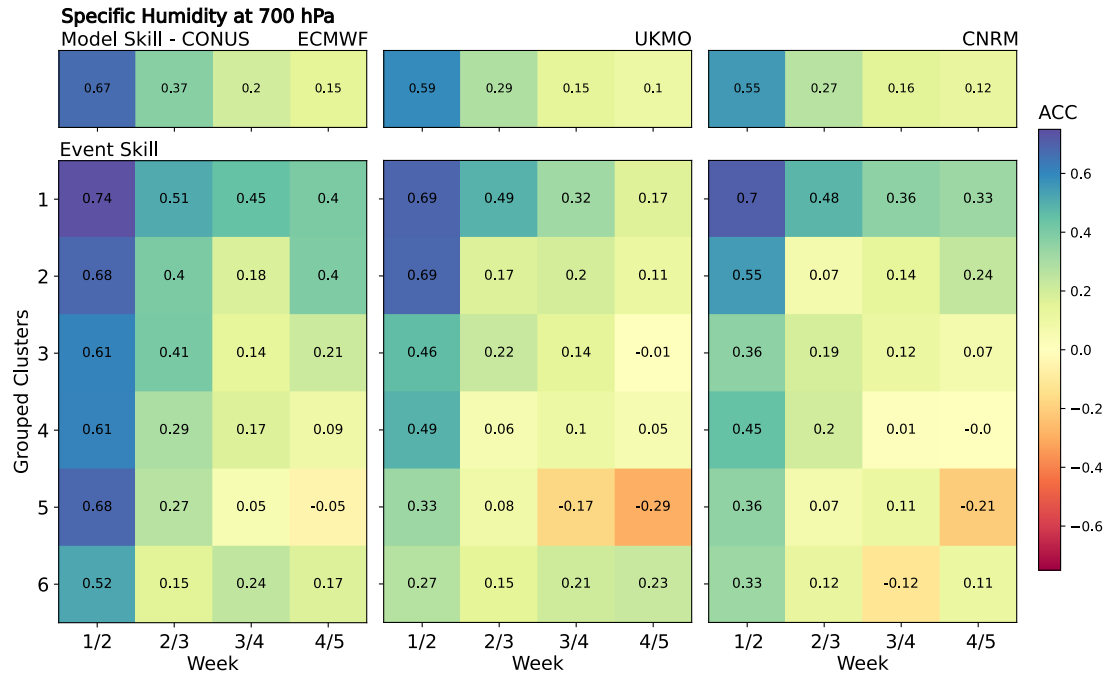


Figure 2.5: As in Figure 2.3, but for 700 hPa specific humidity anomalies.

humidity anomalies, Figure 2.5, with the only skillful predictions seen in Week 1/2. On average, skill of specific humidity during events at 850 hPa and 700 hPa drops with increasing lead time quicker than the overall model skill of 14-day periods, with negligible event skill starting at Week 1/2 in UKMO and CNRM.

Figure 2.6 shows the predictive skill of ensemble averaged Omega anomalies at 500 hPa for all 14-day periods and the 14-day extreme events. The maximum in skill is skill in ECMWF, with an ACC of 0.43 at Week 1/2, which is much smaller than skill seen in both geopotential height and specific humidity. The skill has similar values across all lead times for both the model skill and event skill. Across all models and groups, there are no skillful ACC values seen for any lead time, with the highest values occurring in Week 1/2. For Week 2/3 and beyond skill values are near zero across all groups. Although the low skill of omega might be due to averaging over a 14-day period, there are also other considerations. Omega is highly variable on spatial and temporal time scales, often associated with convection that may not be realized given the large grid spacing of the S2S models.

For the remaining sections results will be shown only for ECMWF, due to the higher skill values seen during event periods for all variables investigated. Bias analysis will focus on low level specific humidity (850 hPa) and 500 hPa geopotential height given these variables have the highest skill and have been previously shown as predictors of precipitation (Schroers and Martin 2022).

2.4.3 Extreme Events Bias

In order to compare event bias to overall model bias, bias distributions were analyzed. The distributions of ECMWF bias for 500 hPa geopotential height for non-event and event periods within groups 1, 3 and 5 can be seen in Figure 2.7, and groups 2, 4 and

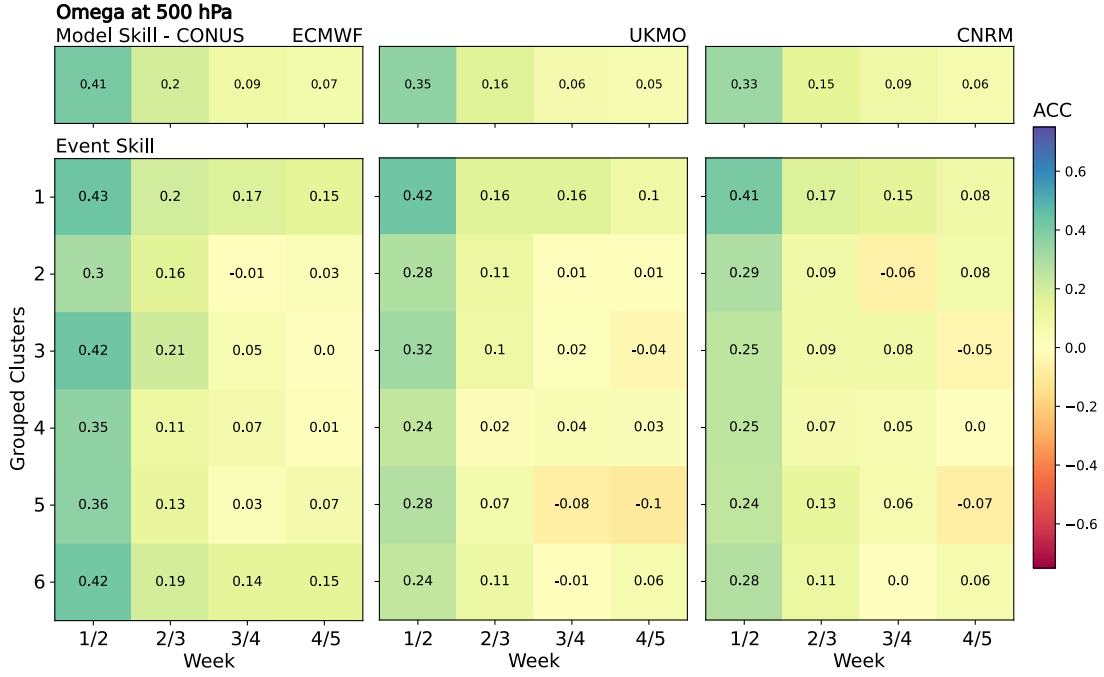


Figure 2.6: As in Figure 2.3, but for 500 hPa omega anomalies.

6 can be seen in Figure 2.8. Details of the moments of all distributions for Weeks 1/2 and Weeks 4/5 can be found in Appendix A, Table A.1. The distributions at Week 1/2 and Week 4/5 of non-event bias for all groups are similar, with a mean value near 8 gpm, signifying higher heights on average within the model. In addition, the non-event distributions have increasing variance with lead time, with increases seen in positive bias. The event bias distributions have similar mean and variance to the non-event distributions at Week 1/2. Yet at longer lead times, there is higher probability densities toward the positive and negative tails of the distributions, indicating that bias, both positive and negative, is increasing with lead time. This is also seen as increasing variance with lead time within the event distributions. Although all groups have similar patterns in increasing bias, group 5 (Northeast) has the largest range of bias, from -40 to 70 gpm, in both event and non-event distributions.

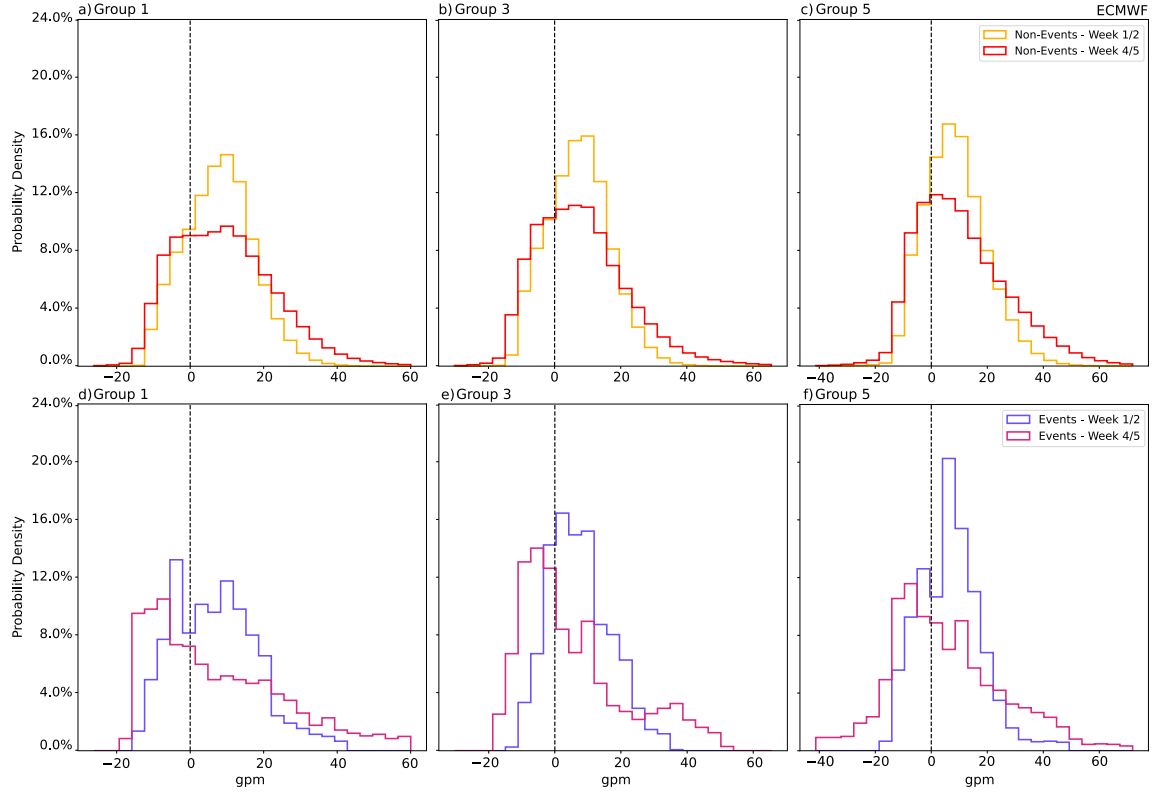


Figure 2.7: Distribution of ECMWF 500 hPa geopotential height bias for group 1 (left), 3 (middle), and 5 (right) for bootstrapped ($n=2500$) non-events Weeks 1/2 and Weeks 4/5, top panels (a), (b), and (c), and Events Weeks 1/2 and Weeks 4/5, bottom panels (d), (e), and (f). Note the differences in the range of specific humidity bias, along the x-axis. Zero bias is indicated as a dashed black line.

To understand where the larger biases from the distributions are seen, specifically within group 5 (Northeast), the spatial distributions of bias were examined (Figure 2.9). We know that these extreme events are often associated with an anomaly dipole centered over the region of the extreme event (Dickinson 2023; Schroers and Martin 2022; Jennrich et al. 2020). Specifically for the Northeast, the composites of the events days result in a anomaly dipole with a positive height anomaly off the coast and a negative height anomaly over the Eastern United States. Therefore the geospatial bias past Week 1/2 shows the model has an overall weaker anomaly dipole. At Week 1/2 the model is overestimating heights in the northeast (the region of extreme precipitation),

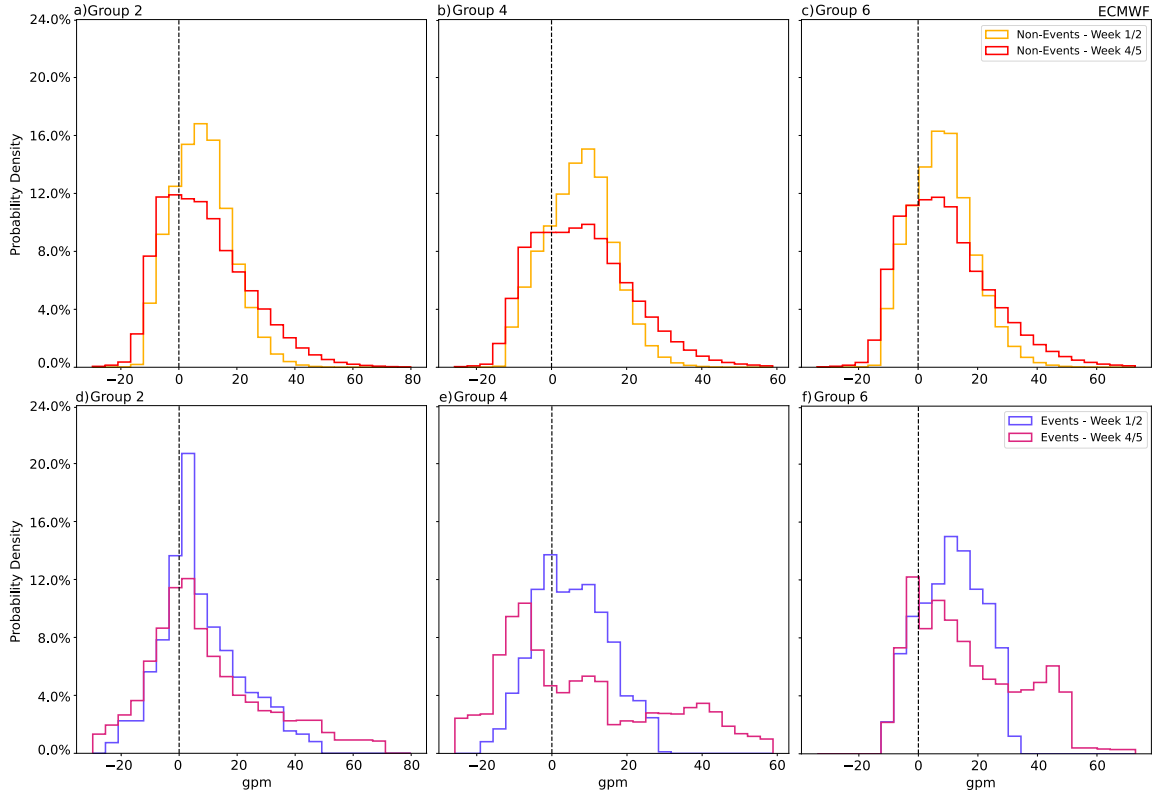


Figure 2.8: As in Figure 2.7 for group 2 (left), 4 (middle) and 6 (left).

with biases from 10 to 20 gpm, although these are not significantly different from the non-event biases. Past Week 1/2 the positive bias increases and a dipole bias develops with positive values over the Eastern United States, maximum bias of 40 gpm, and significant negative values to the east off the coast, maximum bias of -35 gpm. Although both observed positive and negative anomalies within the dipole play a role in event precipitation, the negative anomaly within Eastern United States is the dominate feature in the Northeast, often appearing ahead of the 14-day event (Jennrich et al. 2020). Therefore the positive bias in this region seen starting in Week 1/2 may misrepresent the formation of the 14-day event. This bias dipole oriented over the region of interest is seen for all groups, not shown. This suggests that the model is misrepresenting the extreme events dipole magnitude or orientation, as well

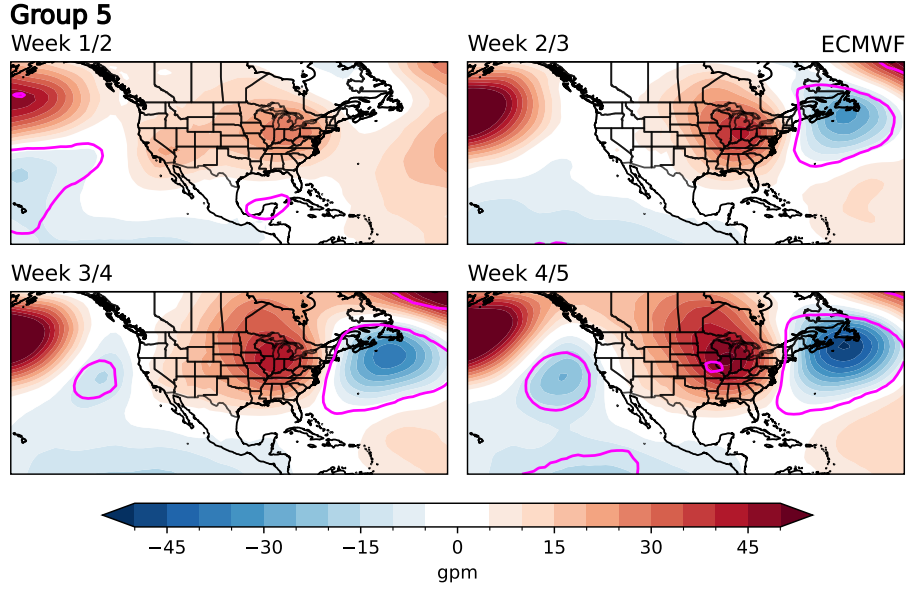


Figure 2.9: ECMWF geospatial bias of 500 hPa geopotential height for group 5 for all lead times. Significant difference from the respective lead-time dependent bootstrapped ($n=2500$) non-events is shown as a fuchsia contour, at the 90% confidence interval.

as being significantly different than the model systematic bias. In addition, there is a large positive height bias seen in the Northern Pacific, greater than 50 gpm for all lead times. Although it is seen during event days, it is not significantly different than the model bias. This suggest that the model is producing a weaker trough in this region that may have downstream affects.

The distributions of ECMWF 850 hPa specific humidity bias are shown for non-event and event periods for odd numbered groups in Figure 2.10 and even numbered groups in Figure 2.11. Details of the moments of all bias distributions can be found in Appendix A, Table A.2. The distributions of bias of non-event periods does not vary much from the first to the last lead times, seen as similar means and variances in Table A.2, with the largest shift seen in Group 2. Yet during our event periods, the distributions see a larger shift toward a dry bias with increasing variance as from

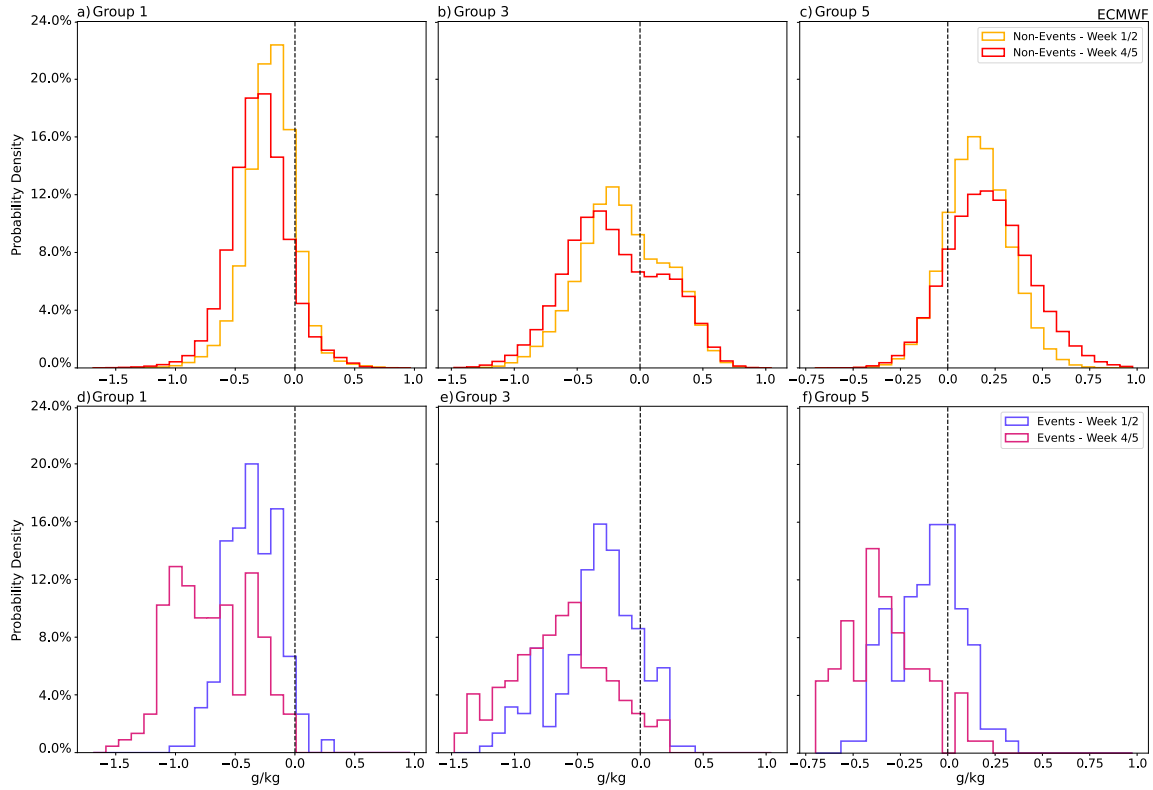


Figure 2.10: As in Figure 2.7 but for 850 hPa specific humidity bias, for groups 1 (left), 3 (middle) and 5 (right)

Week 1/2 to Week 4/5. For example, both non-events and events in group 3 have a bias distribution that has a negative mean, -0.13 and -0.34 respectively, at Week 1/2. Going to Week 4/5 the non-event mean becomes slightly more negative at -0.20, Figure 2.7b, but the event distribution mean is -0.77, Figure 2.7e. The trend to larger dry biases with increasing lead time during event periods is seen for the all groups. It is also important to note, that groups 4, 5 and 6 have differing signs of mean bias in non-events and events. During non-event periods the model trend is increasing wet bias with lead time, although the opposite is true during the event periods, although event periods bias is highly variable.

The spatial distributions of low level specific humidity shows the increasingly dry biases with lead time for all regions, shown only for group 5 in Figure 2.12. At Week

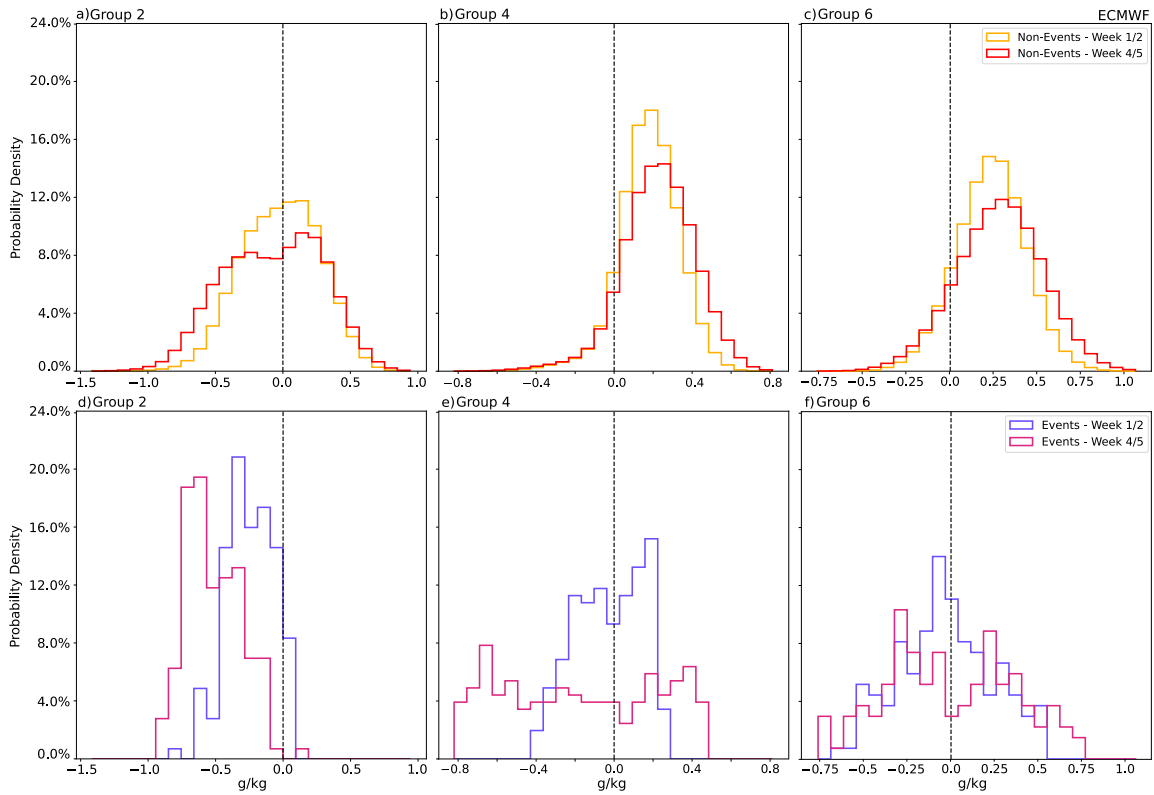


Figure 2.11: As in Figure 2.7 but for 850 hPa specific humidity bias, for groups 2 (left), 4 (middle) and 6 (right)

1/2 the spatial bias is negative throughout most of the region and is significantly different from non-events, which follows given that the non-event bias distribution had a positive mean, Figure 2.7c. The bias continues to become more negative throughout the region past Week 1/2, with a larger area of significant difference. The largest magnitudes of bias, reaching -0.9 g/kg in Week 3/4 and 4/5, are seen along the coast, from New Jersey to Massachusetts. Although there are difference in areas of significant differences between groups, the trend of the model to under represents the low level moisture during the extreme events is seen in all group, with larger biases seen at longer lead times.

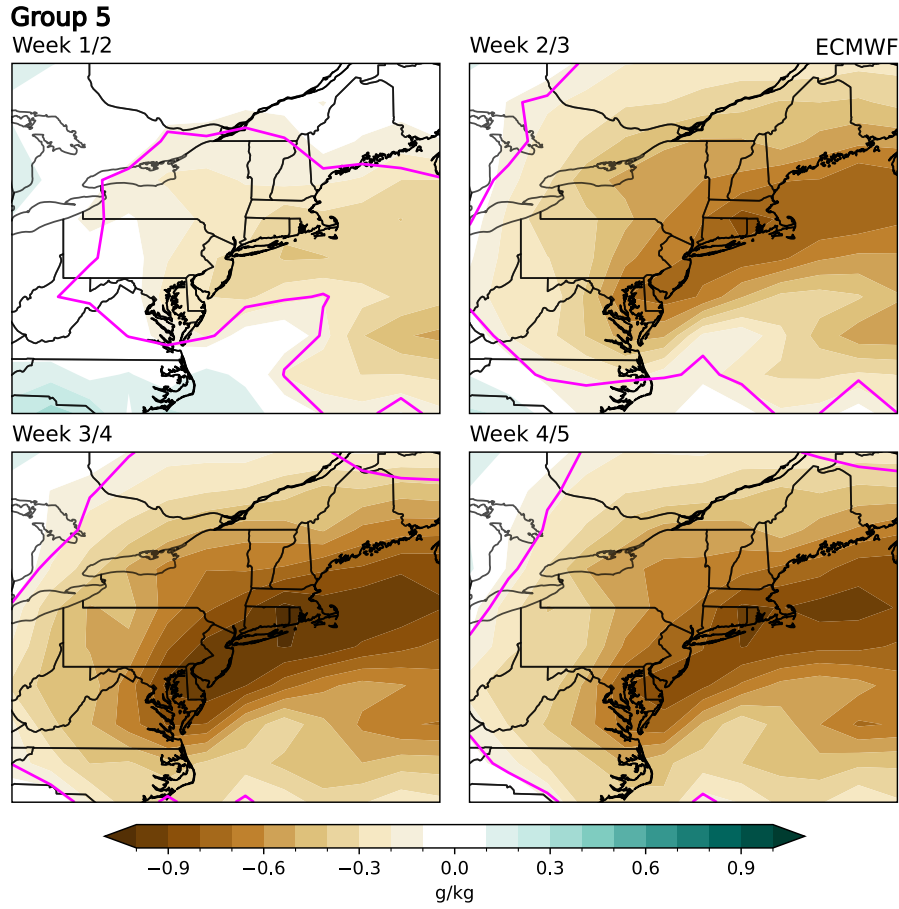


Figure 2.12: As in Figure 2.9 but for 850 hPa specific humidity bias.

2.5 Case Study Analysis

2.5.1 Case Study Selection

Although quantifying the skill and bias of the extreme event periods on average is important for understanding regional differences and model capabilities, in practice, use of S2S forecasts by decision makers would be on an event-by-event basis. Therefore, skill was also assessed for individual extreme events. Methods for finding ACC skill values for each event periods are the same as used the Atmospheric Variable Verification, but each day was considered rather than a 14-day average. Skill of precipitation

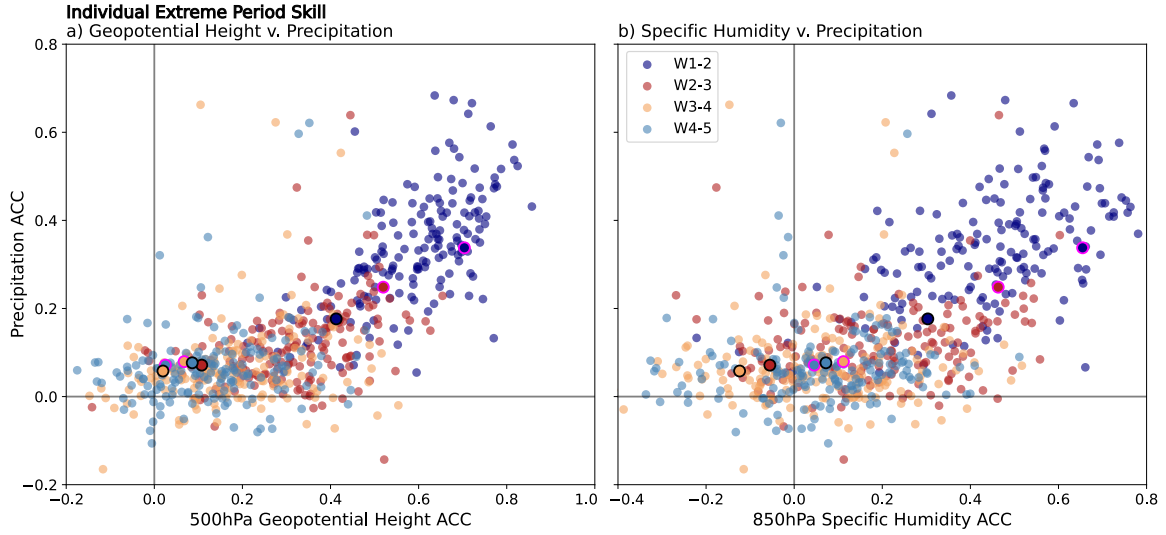


Figure 2.13: Individual event area-averaged ACC skill comparisons at all lead times for (a) 500 hPa geopotential height and precipitation and (b) 850 hPa specific humidity and precipitation. Case 1 skill event dots are circled in pink, while Case 2 skill dots are circled in black. Note the differences in range along each x-axis.

was found by isolating the extreme period's precipitation anomalies using their event polygons with a 3° padding and verifying it against ERA5 precipitation. We know decision makers can tolerate spatial shifts in precipitation on S2S lead times, therefore a 3° padding was used to capture event precipitation that was spatially shifted from observations (Schroers et al. 2024). Significance testing of Spearman's correlation coefficient between the variables was done at the p-value <0.05 level.

The atmospheric variables used in this study are commonly used in post-processing techniques to improve S2S precipitation skill, therefore we want to determine if S2S model skill of atmospheric variables relates to skill of precipitation. Figure 2.13 shows skill comparisons of all extreme event periods for the four different lead time. Very similar to the skill seen in the 14-day averages, Figures 2.3 and 2.4, individual event skill for 500 hPa geopotential height is the highest at Week 1/2 and has a significant positive correlation with precipitation skill at Week 1/2, Spearman's correlation of 0.65 (Fig 2.13a, blue dots). Past the first lead time, both precipitation and geopotential height

have much lower skill for most events. Although correlation between the variables decreases with lead time it remains significant for all lead times, with a Spearman's correlation of 0.45 at Week 2/3 and 0.25 at Week 3/4. For the comparison between low level specific humidity and precipitation skill, Figure 2.13b, there is more spread in skill values with a smaller significant correlation of 0.54 between the variables seen for Week 1/2. At Week 2/3 the correlation is weak (0.25) but significant, but drop to no significant correlation for remaining lead times. Past Week 1/2, there are many events that have skill near zero skill for both variables. This is similar to the drop off in skill seen in 14-day average skill seen in Figure 2.4.

Although skill, on average, for atmospheric variables associated with the 14-day extreme events is low past Week 1/2, there are events that have skillful synoptic forecasts out to Week 3/4 with better precipitation forecasts. Using the individual event skills, we chose two event periods to further analyze the S2S synoptic variable forecasts from the model. Cases were chosen to compare an event with higher skill (Case 1) to an event with lower skill (Case 2) of precipitation, 500 hPa geopotential height and 850 hPa specific humidity, highlighted in Figure 2.13.

Case 1, the forecast with higher skill (Figure 2.14a), occurred along the West Coast from 18 December to 31 December 2005. Precipitation maximums above 280 mm were seen along the coast and in lower elevation areas in Oregon and Northern California. Another maximum was also seen around Boise, Idaho with precipitation values reaching 180mm, seen in PRISM precipitation and ERA 5 precipitation regridded to S2S model resolution, Figure 2.14a. This event had multiple landfalling atmospheric rivers that lead to very saturated soil, with a very heavy one-day precipitation event on 31 December that caused devastating flooding. During this event, ENSO was in a

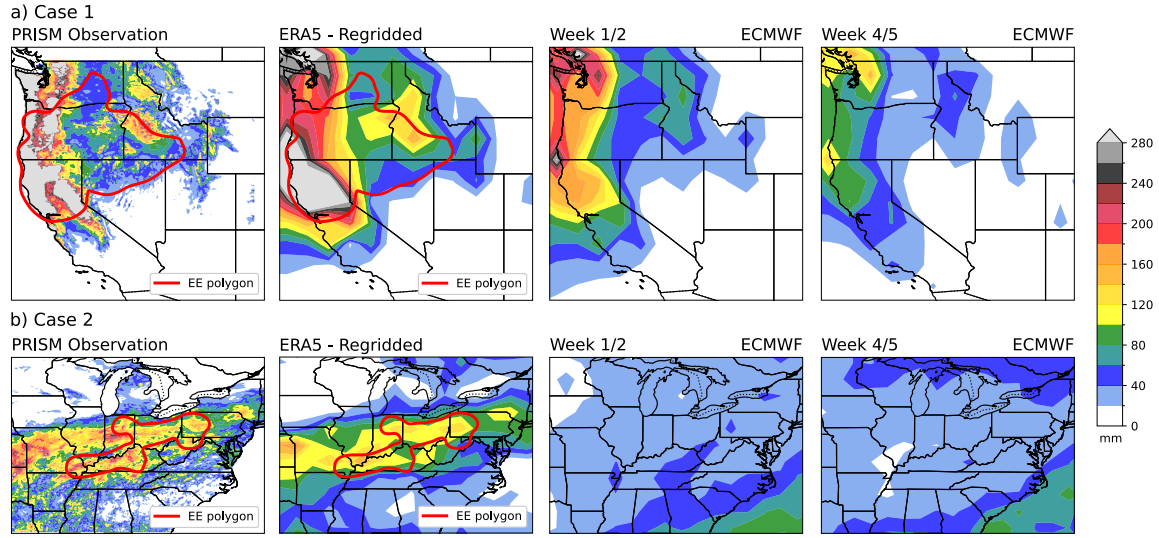


Figure 2.14: Precipitation 14-day totals for Case 1 (top) and Case 2 (bottom). For each case precipitation values are shown from PRISM observations (PRISM Climate Group, Oregon State University 2004), ERA5 regridDED into S2S model resolution (used as observations for verification), ECMWF forecasted totals at Week 1/2 and at Week 4/5 lead times. The extreme event polygon (EE polygon), red, can be seen over the observations and ERA5.

neutral phase developing into La Niña. Although Bao et al. (2006) found that moisture transport is enhanced during the neutral ENSO phase, other studies have shown more winter atmospheric river activity along the CONUS West Coast during El Niño (Mundhenk et al. 2016; Payne and Magnusdottir 2014). Impacts from the precipitation were seen throughout Northern California, Oregon and even into Nevada (Parrett and Hunrichs 2006). ECMWF precipitation forecast at Week 1/2 had maximums in precipitation along the coast, reaching only 180mm compared to the observed 280mm, and a much smaller secondary maximum in Idaho of 60mm, compared to an observed 180mm. Although the ECMWF precipitation forecast did not accurately represent the magnitude of the precipitation, with an ACC skill of 0.38 at Week 1/2, the location and shape of the observed precipitation is similar to the forecast, which still may be useful to decision makers at the S2S time frame (Schroers et al. 2024).

Case 2, the less skillful precipitation forecast with an ACC value of 0.18 at Week 1/2, occurred from 22 August to 4 September 2003 across the Midwest extending from Southern Missouri to Western Pennsylvania (Figure 2.14b). Maximums in precipitation, reaching 180mm, occurred in Northern Kentucky and Southwestern Missouri. There was a secondary swath of precipitation that stretched from northern Indiana to Western Pennsylvania that had precipitation values of 120mm. Both observational datasets, PRISM and regrided ERA5, have a similar pattern of precipitation totals for the 14-day periods. This event was a result of tropical moisture transport poleward ahead of Tropical Storm Grace in the Gulf of Mexico, that made landfall on the coast of Texas on 31 August. Ahead of landfall, a predecessor rainfall event occurred around the western portion of the event polygon. Predecessor rainfall events are historically hard to forecast due to the mesoscale nature of the convection (Galarneau et al. 2010). At the end of the 14-day event, the combination of a frontal boundary and the remnants of the tropical storm led to rainfall across the region. The forecasted totals of precipitation was unrepresentative of the observational totals beginning at Week 1/2 and continued to dry out with increasing lead time.

2.5.2 Case Study Forecasts

To further examine the forecast of the synoptic conditions during each case study, Figure 2.15 shows the ERA5 and forecasted 500 hPa geopotential height anomalies for the 14-days of Case 1. ERA5 reanalysis shows that this event had a stationary anomaly dipole centered over the event polygon that was present throughout most of the 14 days. The positive height anomaly is associated with a ridge that is present along the West Coast throughout the event. In both the Week 1/2 and Week 2/3 forecasts the anomaly dipole is seen throughout the entire forecast period, with an

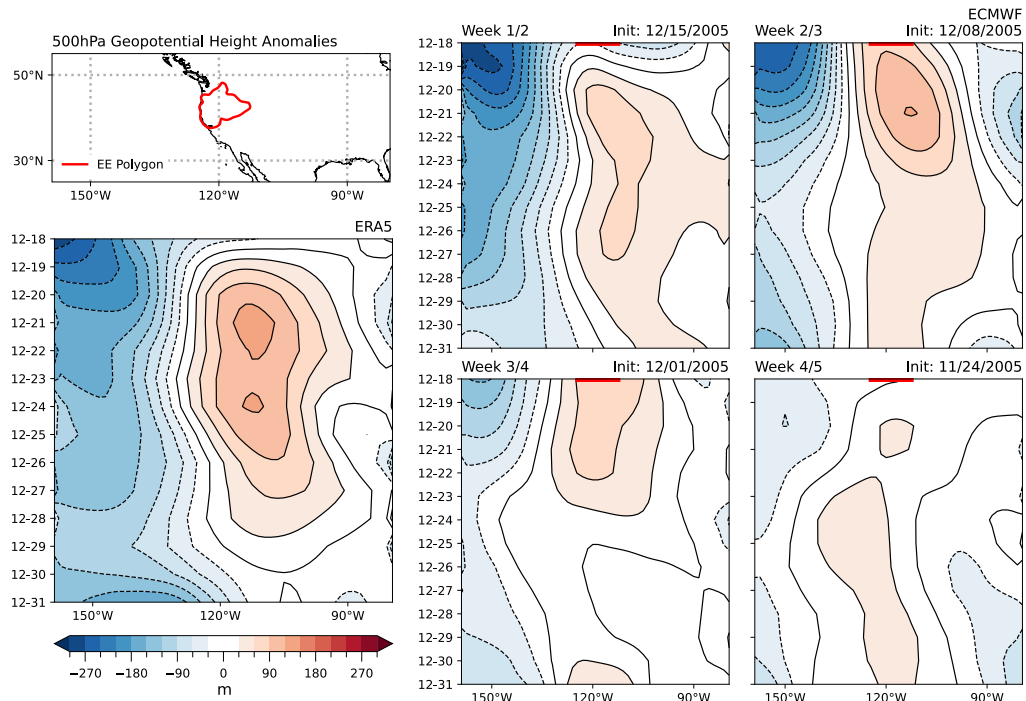


Figure 2.15: Event polygon of Case 1 (top left) and Hovmöller diagram of regridded ERA5 500 hPa geopotential height anomalies for the 14-day extreme event (bottom left). To the right, hovmöller diagrams of the ECMWF forecasted 500 hPa geopotential height for all four lead times are shown with their respective initialization date. The red bar seen on the top of the forecasts indicate the geospatial location of the event. Data was averaged between 30°N and 50 °N and plotted daily.

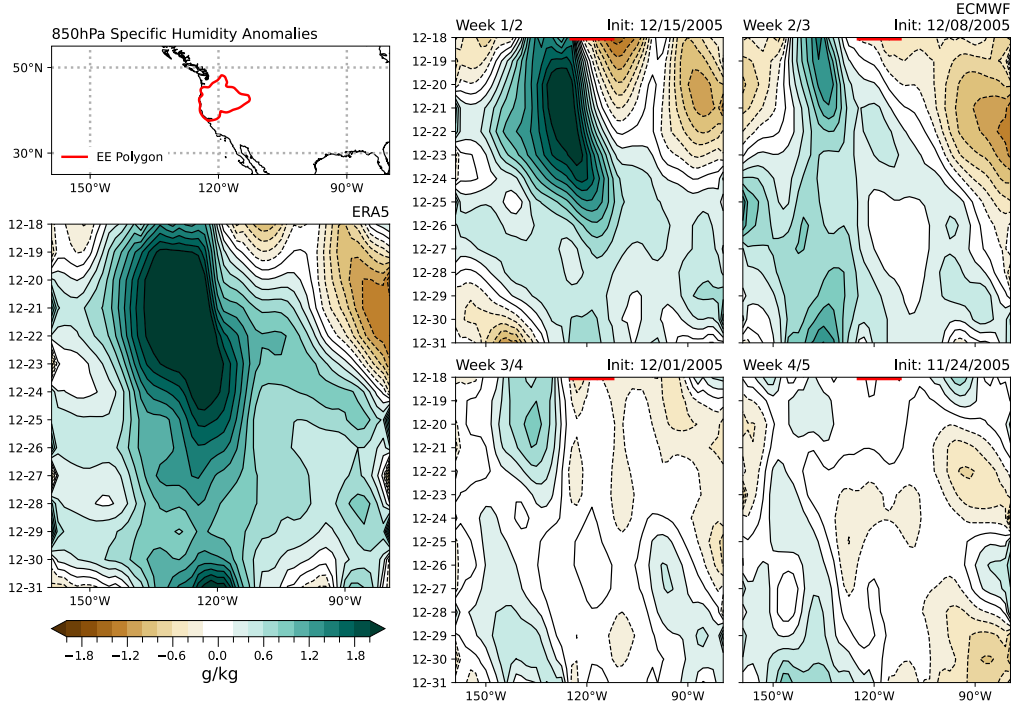


Figure 2.16: As in Figure 2.15, but for 850 hPa specific humidity anomalies.

ACC value of 0.70 and 0.52 respectively. The forecasted magnitude of the positive height anomaly along the West Coast is smaller than what is seen within ERA5 for all lead times, with negative height bias of -20m in most of the region. By Week 3/4 the dipole is only seen for the first five days of the forecast period, before showing near zero anomalies. By Week 4/5 there is near zero anomalies for all 14 days.

The persistence of the observed geopotential anomalies, led to large positive moisture anomalies, exceeding 1.9 g/kg, within the region throughout the event duration (Figure 2.16). Within the ECMWF forecast for Week 1/2 and Week 2/3 the positive moisture anomaly is present near 110°W, albeit with a smaller magnitude of 0.7g/kg in Week 2/3, similar to the dry bias seen in group 4 in Figure S3. By Week 3/4 the forecast is a near zero moisture anomaly after the first five days of the forecast period. Both the forecast height and low level moisture are representative of observations until Week 3, with an ACC of 0.66 and 0.46 at Week 1/2 and Week 2/3 respectively. Similar

values of ACC were seen for group 1 in Figure 2.4. Although the synoptic patterns are similar to observations, there is a negative bias present in both geopotential heights and specific humidity that grow with lead time, resembling the changes in bias distributions. Therefore, it follows that the precipitation forecast would closely resemble observations given the skillful forecast of the persistent geopotential height pattern with associated moisture transport into the region.

In contrast, Case 2 had no skill starting at Week 1/2 for geopotential height (ACC of 0.41) or specific humidity (ACC of 0.30), Figure 2.13. This was due to a misrepresentation of tropical moisture transport within the region because the forecast did not show the observed tropical storm within the Gulf of Mexico. This is an example of the impacts on precipitation amount when the ECMWF model underestimates tropical cyclone activity within the Atlantic Basin (García-Franco et al. 2023). To illustrate the incorrect tropical moisture transport observations and forecasts of 850 hPa specific humidity and 500 hPa v-wind components are shown, Figures 2.17 and 2.18 respectively. Observations of 850 hPa specific humidity, Figure 2.17, show positive moisture anomalies within the region from 24 August to the end of the event on 4 September. This anomalous moisture is not present within any of the four lead time ECMWF forecasts, with Week 1/2 actually showing negative moisture anomalies. The lack of moisture anomalies within the model is explained by forecasts of the v-wind component, Figure 2.18. Within observations, there is southerly wind at 100°W that moved into the region beginning on 26 Aug. This channel of southerly wind moved moisture poleward from Tropical Storm Grace until the end of the 14-day period. Within all four ECMWF forecasts, the southerly winds at 100°W never shifted into the region. The lack of tropical moisture transport lead to forecasts that had dry specific humidity anomalies with no resulting precipitation within the forecast.

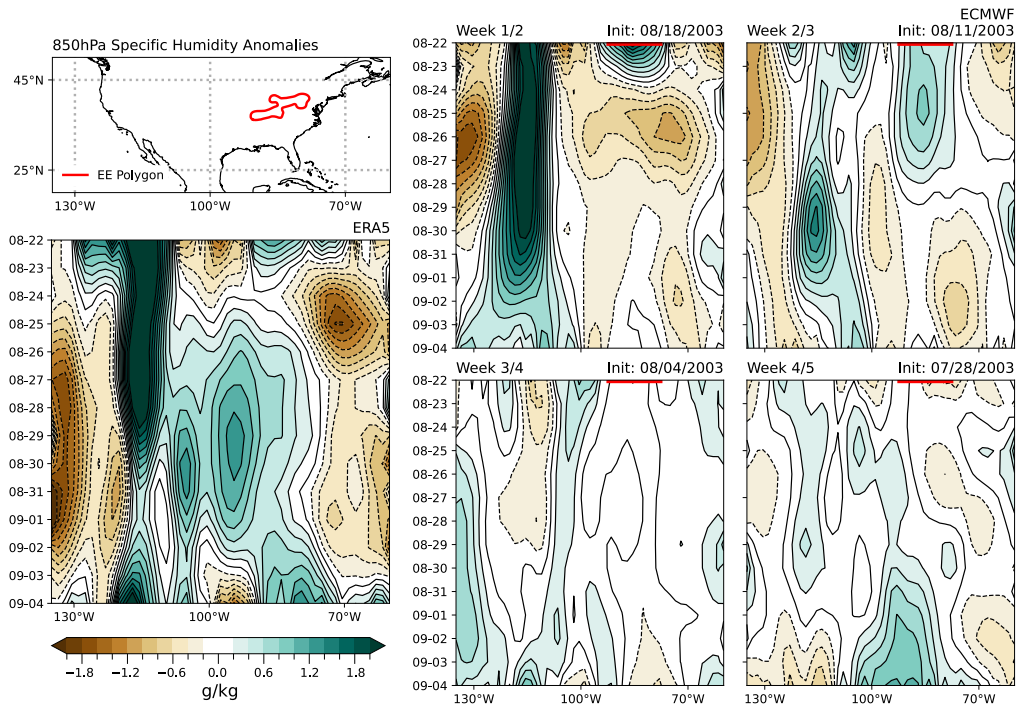


Figure 2.17: As in Figure 2.15, but for Case 2's 850 hPa specific humidity anomalies. Data was averaged between 25°N and 45°N and plotted daily.

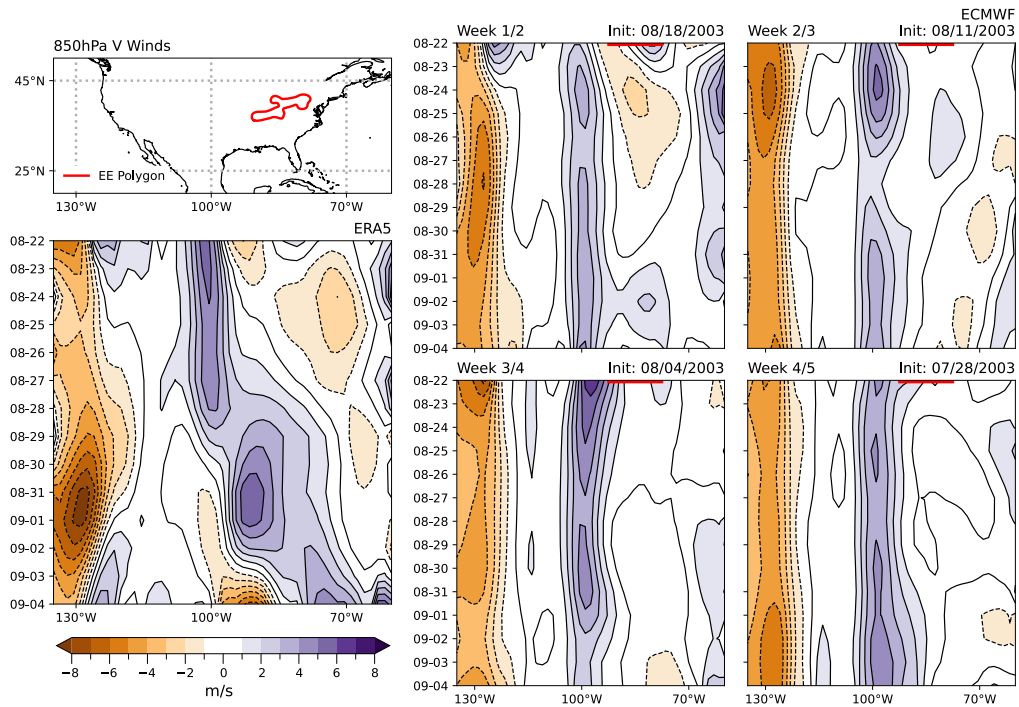


Figure 2.18: As in Figure 2.15, but for Case 2's 500 hPa v-wind (not anomalies). Data was averaged between 25°N and 45°N and plotted daily.

2.6 Summary and Discussion

In this study we quantified the average skill of 14-day averaged anomalies of 500 hPa geopotential height, 700 hPa and 850 hPa specific humidity within three S2S models (ECMWF, UKMO and CNRM) during 14-day extreme precipitation events across six regions in the CONUS. On average, 500 hPa geopotential height has skill out to Week 2/3, with ECMWF having the highest skill. Groups 1 and 6, West Coast and Southeast respectively, retain skill longer than other regions. We also saw no large differences between overall model skill and event specific skill of 500 hPa geopotential heights in all three models. Therefore the models, on average, are skillful at the prediction of 14-day averaged 500 hPa geopotential height in both event and non-event periods.

Conversely, skill for regional low-level specific humidity is lower on average than 500 hPa geopotential height but still larger than that for precipitation itself. As opposed to 500 hPa geopotential height, there are differences in overall model and event specific skill for specific humidity. For example, Group 1, West Coast, has event specific skill that is more skillful than the overall model skill for most lead times, suggesting that during 14-day precipitation extremes the models are better at predicting the moisture than when there is not an extreme. The larger skill values seen within group 1, West Coast, for both geopotential heights and specific humidity is consistent with larger skill values seen in this region for the precipitation related to 14-day extreme events, seen in Figure 2.2. Lastly, 14-day averaged omega at 500 hPa has no skill for both overall model and extended extreme precipitation events in any groups within any of the three models analyzed. The low skill might be associated with averaging over 14-days, since omega can be highly variable on daily timescales, but also shows the models inability to resolve some drivers of precipitation.

To understand the model skill within extreme events, the model bias for 500 hPa geopotential height and 850 hPa specific humidity was analyzed for the extreme event periods and compared to non-event bias in the ECMWF model. Bias of 500 hPa geopotential height in groups 1, 3 and 5 grows in both positive and negative directions as lead time increases during events and non-event, although the probability densities on the tails of the event bias distributions are larger.

Spatially, the larger magnitude biases are seen as a dipole in all regions past Week 1/2. Areas of significant difference either encapsulate both parts of the bias dipole (group 1 through 4), the downstream ridge (group 5) or the upstream trough (group 6). The bias dipole is collocated with an anomaly dipole which is a common feature for 14-day extreme precipitation events (Schroers and Martin 2022). While the bias of low level specific humidity at Week 1/2 is significantly different than non-event bias for a large portion of the region for all groups. For the Eastern United States (groups 4, 5 and 6), the average model bias is positive (excess moisture), while bias seen during events is negative (dry bias). This is particularly important when using bias correction techniques, since the sign of the systemic model bias might differ during extreme periods. For example, if forecasters are aware of the positive moisture bias, they may assume that any forecast of excessive moisture is being "overdone" by the model and be overcautious of predicting extreme precipitation. The models inability to realize the magnitude and/or location of the geopotential height anomaly dipole past Week 1, along with an increasingly dry bias at longer lead times seen within ECMWF, strongly impacts the forecasted precipitation totals for long duration events.

Examining individual event skill showed similar trends of skill to the averages for both 500 hPa geopotential height and 850 hPa specific humidity. To illustrate an example of a strong S2S forecast and a weaker one, two case studies were analyzed. Case 1 captured the shape of precipitation but missed the extreme magnitudes. The

atmospheric variables during the 14-day event had a persistent pattern in observations that were skillfully forecast in the models, supporting studies that showed S2S models may skillfully predict persistent patterns (Ferranti et al. 2015; Tuel and Martius 2023). While Case 2 misrepresented tropical moisture transport into the region, leading to no precipitation forecasted at S2S lead times in the region of the observed event. This forecast bust was likely related to the ECMWF model missing Tropical Storm Grace beginning at Week 1/2, which is consistent with studies that showed a misrepresentation of tropical cyclone activity in the Atlantic/Gulf of Mexico basin (García-Franco et al. 2023). Although the forecasts of precipitation in both cases had differing skill values, both showed a dry bias within specific humidity that increased with lead time shown in the bias distributions.

Although this study provides insights into three S2S project models capability to predict 14-day precipitation extremes and their associated atmospheric conditions, there are limitations to consider. The 14-day extreme events are rare and the short common period of S2S hindcasts, 1996-2014, limits the amount of extreme events that can be analyzed by region and thus limits the ability to have a more thorough understanding of forecasts of opportunity. Future studies could look into high and low skill cases to identify possible sources of predictability that lead to higher skill cases. In addition, we know decision makers have a higher level of spatial uncertainty tolerance at the S2S lead times, but the use of grid based skill metrics (e.g. anomaly correlation coefficient) penalizes spatial shifts in forecasted precipitation (Schroers et al. 2024). Therefore, a similar study could be conducted for long-duration precipitation extremes using an object-oriented framework to account S2S uncertainty tolerance.

2.7 Conclusion

This study aimed to quantify the skill and bias of three S2S models (ECMWF, UKMO and CNRM) during 14-day extreme precipitation events by answering three research question.

Firstly, what is the skill of atmospheric variables associated with 14-day extreme precipitation events within the S2S models and how does it compare to overall model skill? On average skill is below 0.5 past Week 1/2 of 14-day averaged atmospheric variables during 14-day precipitation events, but still higher than the skill of the precipitation itself. The skill is highest along the West Coast, which also is the region with the greater skill in precipitation.

Secondly, how do the bias distributions of these variables compare to overall model bias? It was shown that biases exist in the atmospheric variables associated with 14-day extreme precipitation events. Further, the atmospheric variables biases during events, geopotential height and specific humidity, are larger than model systemic biases. Therefore, when utilizing these variables within post-processing techniques for precipitation, it is important to address these biases, especially within low level moisture.

Thirdly, are there preferred synoptic set ups for skillful S2S precipitation forecast and how does it differ from a forecast bust? Two case studies of events within the ECMWF model were analyzed, with differing precipitation skill. Case 2 did not have any precipitation forecasted, due to a misrepresentation of tropical moisture advection into an area of mesoscale convection. Case 1 captured the shape and size of the observed precipitation field, but missed the magnitude of the observed rainfall. Atmospheric variables were well forecast most likely due to the persistent large-scale set-up present for the 14-day period. This shows that models may be able to create usable S2S

forecasts especially given decision makers typically use S2S forecasts as a planning tool and no deterministic decisions are made at these lead times, as shown in Chapter 4.

Forecasts of long-duration extreme precipitation events on the S2S timescale would benefit decision makers to help plan for the lasting impacts seen within the CONUS. Yet skill of S2S model precipitation is low. Post-processing techniques have been utilized in order to improve precipitation forecasts, often using more skillful S2S model atmospheric variables with biases. In this study we have highlighted the differing skill and biases of atmospheric variables associated with precipitation events to guide future model developments for more skillful precipitation predictions. While there is still much to learn about predicting long-duration extreme precipitation events at the S2S lead time, this study is necessary step to improving future predictions.

Chapter 3

The Role of Extratropical Cyclones within 14-day Extreme Precipitation Events

3.1 Background

Persistent precipitation, lasting two weeks, can lead to lasting impacts over the affected region (Schroers and Martin 2022), with many end-user communities desiring longer range, trustworthy precipitation forecasts, discussed further in Chapter 4. Understanding the mechanism by which extended extreme precipitation events occur is imperative to improving model forecast quality. It has been shown that the 14-day precipitation events have multiple peaks in area-averaged precipitation that compound to create the long-duration event (Pruitt et al. 2023). Multiple types of precipitation systems exist that may be cause rainfall within a region including: extratropical cyclones and their associated frontal systems, tropical cyclones (Schroers and Martin 2022), the North American Monsoon, mesoscale convective systems (MCS), small-scale unstable air mass thunderstorms, and orographic precipitation (Kunkel et al. 2012). Extratropical cyclones (ETCs) and the related fronts account for a majority of the climatological total precipitation over the midlatitudes, up to 80% of daily extreme precipitation events in parts of the CONUS (Hawcroft et al. 2012; Kunkel et al. 2012; Pfahl and Wernli 2012; Agel et al. 2015). Given ETCs importance for daily extremes, ETCs may play a role in extended extremes as well.

Precipitation, wind extremes, and temperature changes are often seen during passage of an ETCs over a region. Precipitation related to ETCs is typically seen to the east of the lowest pressure, along the frontal zones (Hawcroft et al. 2016). Multiple studies have shown the precipitation amount within an ETC is highly dependent on cyclone intensity and available moisture (Field and Wood 2007; Rudeva and Gulev 2011; Pfahl and Sprenger 2016). The largest contribution of ETC related precipitation to total precipitation within the CONUS is seen during the winter (Ulbrich et al. 2009; Hawcroft et al. 2012). This is especially true within the Northeastern United States, where ETC related precipitation can make up to 85% of total wintertime precipitation (Hawcroft et al. 2012; Agel et al. 2015). Although the southeast has less ETCs on average, ETCs in the region are highly associated with daily precipitation extremes (Pfahl and Wernli 2012). In the Central and Western CONUS ETCs are not a significant contributor to winter precipitation, but may play a larger role in warmer months (Hawcroft et al. 2012; Pfahl and Wernli 2012). Atmospheric rivers are also highly correlated with ETC activity, with 82% of atmospheric rivers being associated with an ETC in the Western United States (Zhang et al. 2019). Therefore all regions in the United States can have precipitation related to ETC passage.

There are multiple ways in which a region may experience extended precipitation periods due to ETC activity. When multiple ETCs pass within the same region within a short period of time, known as cyclone clustering, precipitation from each cyclone may compound to create an extended precipitation event (Karwat et al. 2024). On the other hand, slow moving cyclones over a region of moisture advection may create long lasting precipitation from one cyclone. Therefore understanding ETC intensity in combination with speed and frequency will help to reveal how ETC precipitation may impact a region, given both processes may be contributing to precipitation within the CONUS.

Although ETCs play a large role in precipitation variability within midlatitudes, prediction of the location and intensity of a single ETC track is only skillful out to day seven (Froude et al. 2007). Thus Zheng et al. (2019) quantified the wintertime S2S prediction skill of extratropical cyclone activity, defined as the mean square of the 24 hour difference of MSLP. It was found that single models from the S2S project, (Vitart et al. 2017), have areas of skillful forecasts over the CONUS at Week 2. When multiple models were combined into a multi-model ensemble, skillful ACC values were seen over CONUS at Week 3-4 (Zheng et al. 2019). Other studies have focused on climate projections of ETCs (Colle et al. 2015; Hawcroft et al. 2016, 2018). Under a warming climate, the total number of ETCs has shown to steadily decrease, while the most intensely precipitating ETCs increase into the mid to late 21st century (Hawcroft et al. 2018). Colle et al. (2015) reviewed many studies to show climate models tend to have biases related to ETCs due to grid spacing and misrepresentation of dynamics related to genesis and evolution of ETCs. Although the precipitation associated with ETCs within climate models are comparable with observations but with an under representation of weaker ETC precipitation (Hawcroft et al. 2016). Therefore, the potential increase in intensely precipitating ETCs highlights the need to understand the connection of ETCs seen within impactful long-duration extreme precipitation events in the CONUS.

Investigating ETC activity during 14-day events is a needed step to understand what types of precipitation systems are compounding to cause these long duration events. This chapter seeks to address the following research questions:

- What is the ETC activity seen within events across the regional groups?
- Is the ETC activity seen during events significantly different than non-event periods?

- What is the contribution of ETC related precipitation to 14-day extreme precipitation events?

3.2 Data and Methods

3.2.1 14-day Extreme Precipitation Events

Fourteen day extreme precipitation events within the CONUS, as defined in (Dickinson et al. 2021) and shown in Chapter 2. For an event to be counted the 14-day total precipitation must be over the 99th percentile of 14-day precipitation climatology, at least half of the event days must be above the daily-mean and the area of the event must be greater than or equal to 200,000km². Once events were found, they were clustered into 15 clusters, using k-means clustering. The time period of interest for this study is 1979-2018, which accounts for 371 total events. In order to increase the sample size of the regional domains, the 15 clusters were placed into 6 regional groups, as in Figure 2.1. Each event has an associate polygon that encompassed all extreme points.

3.2.2 ETC tracking

The ETCs within ERA5 were tracked using TempestExtremes, an automated feature tracking software (Ullrich and Zarzycki 2017; Ullrich et al. 2021). Tracks of ETCs were founded at a 6-hour time step from 1979 to 2018, as shown in (McCrary et al. 2024). Candidate ETCs are identified with the following criteria: (1) local maximum in the variance of 24 hour difference in MSLP, (2) surrounded by a closed isobar of 4.2 hPa² less than the maximum value and (3) the closed isobar and local maximum must be within a 6° radial cap. Tracks from candidate ETCs are stitched together by scanning

ahead 6 hours for ETCs within 6° . Any track that did not persist for 60 hours was removed from the dataset. ETC track density of the database used for this study is shown in Figure 3.1.

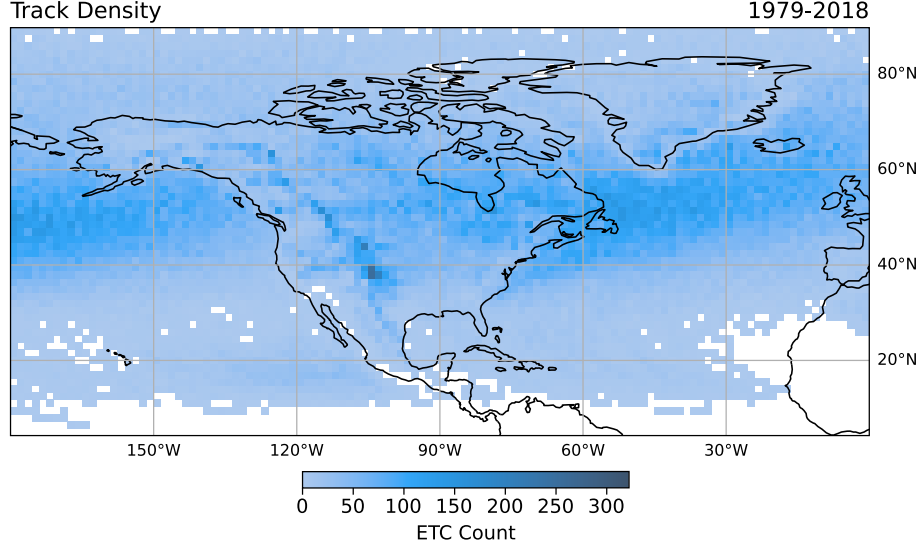


Figure 3.1: Track density over the CONUS of ETC tracks found using Tempest Ex-tremes, as shown in (McCrory et al. 2024), for the time frame of interest, 1979 to 2018.

3.2.3 Matching ETC Activity with Events and Associated Precipitation

ETCs were associated with a specific event if they (1) occurred during the 14-day event and (2) passed within 10 degrees of the event polygon. The event associated ETC tracks were compared to a distribution of non-event ETC activity from 1979-2018 to test if event ETC activity is significantly different in speed, intensity, and maximum intensity than non-event periods. The distribution of non-event ETCs includes all ETCs (1) not found within event days and (2) passed within 8 degrees of the group domain, which is a comparable size to the event polygon with a 10 degree padding that was used to find event ETCs.

Once ETCs were matched with an event, precipitation masks were created by using a 1000 km radius around the center of the track every 6 hours, Figure 3.2a. Associated event precipitation was found by applying the 6 hourly masks, Figure 3.2b, and the event polygon, Figure 3.2c, to daily precipitation data. ERA5 daily precipitation, 0.5° resolution, was used to find the percentage of the events total ERA5 precipitation that was related to ETCs. Although the simple 1000 km radius around the ETC center for associated precipitation might miss trailing fronts, this distance was shown to be a good approximation by Hawcroft et al. (2012).

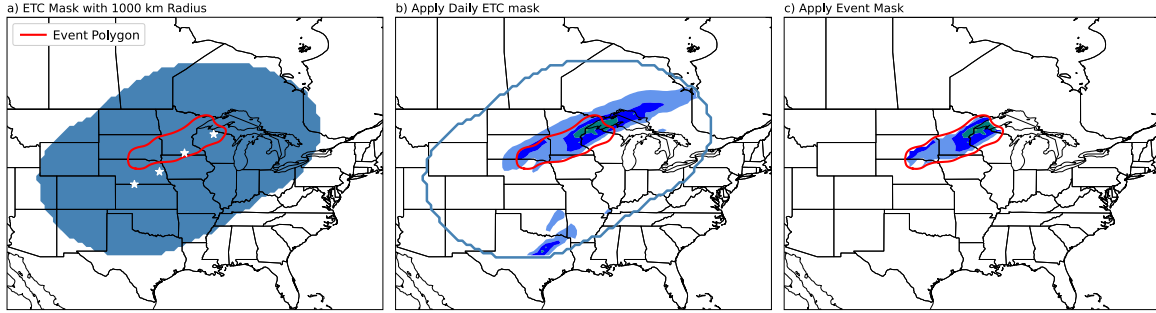


Figure 3.2: Three step process to use 6-hourly ETC tracks to find daily event precipitation (ERA5) associated with ETCs. Example is for one day within an event in group 4.

3.3 Results

3.3.1 Comparison of Event ETC Activity within Event and Non-Event Periods

The number of ETCs during all events were counted from 1979-2018. The counts of events with differing number of associated ETCs are detailed for all events, as well as each group in Table 3.1. For all events across the CONUS, the average number of ETCs seen near the event within the 14-day period is 3.0, or about 1 ETC every 5

days. Looking at specific groups we see that group 4 (Great Lakes Region) and group 5 (Northeast) have the highest average number of ETCs during events with 3.9 and 3.4 respectively, or an ETC every 4 days. Although group 5 has fewer events within the time period than other regions, due to the grouping of clusters, 15 of the 28 total events have more than 4 ETCs associated with the events. Group 4 has 28 events that have more than 5 ETCs seen during the events, about one ETC every 2.5 days. Conversely, group 3 (Southern Plains) has the lowest average ETC activity during events, with a mean of 2.2 ETCs (an ETC every 7 days) seen during events. This is due to having 9 events with no ETC activity and 18 events with only 1 associated ETC track.

ETCs per Event	All Events	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
0	26	6	3	9	2	3	3
1	64	14	9	18	3	3	16
2	66	17	14	8	12	2	13
3	73	20	9	12	21	5	6
4	73	16	5	8	18	7	19
5+	69	14	9	6	28	8	3

Table 3.1: Number of events from 1979 to 2018 that had 0, 1, 2, 3, 4, or 5+ ETCs occurred near the event region. For all events and the six regional groups.

ETC activity was also counted by season for the entire CONUS event dataset, detailed in Table 3.2. Matching with ETC climatology, we see the lowest amount of ETC activity during the events within JJA, with an average of 1.5 ETCs during events. Most CONUS events during JJA have either have 1 or 0 ETC passing near the event. While DJF, MAM and SON have double the average number of ETCs during event periods. Within DJF the distribution of ETCs during events is skewed toward higher

ETCs per Event	DJF	MAM	JJA	SON
0	1	3	18	4
1	10	9	27	18
2	16	18	9	23
3	28	23	7	15
4	21	25	4	23
5+	19	22	4	24

Table 3.2: Number of events from 1979 to 2018 that had 0, 1, 2, 3, 4, or 5+ ETCs occurred near the event region per season for all events.

numbers of ETCs, which makes sense given the climatology of ETCs peaks during this time. Although MAM and SON have many events with high numbers of ETCs, there are still many events with only 1 or 2 ETCs. This may be due to the transition into or out of the warm season.

Quantifying ETC activity during events helps to reveal regions where ETC activity may be the dominate driver of event precipitation, but it does not tell us how the ETCs during events compare with non-event ETC activity. Understanding the differences from climatology may help to reveal if ETC activity during events is unique and therefore a source of extended precipitation. Figure 3.3 shows the distributions of both event and non-event ETC speeds. Across all six groups, the mean speed is near 13 m/s for both event and non-event periods, although the distributions range from -3 m/s (westward motion) to 35 m/s. The only group where ETCs associated with events and non-event have significantly different speeds is group 2, Northern Plains. Within this region event ETCs are significantly slower than non-event ETCs, a means of 12.1 m/s and 13.4 m/s respectively.

On the other hand, the ETC intensities, defined as the MSLP along the 6 hourly tracks, associated with events have many groups that are significantly different than the groups non-event ETC activity, shown in Figure 3.4a. All groups, with the exception of group 3 (southern plains), have statistically significant lower intensities. In groups 5 (Northeast) and 6 (Southeast) this significant difference is driven by more lower pressure intensities seen in the tails of the distributions, most likely from tropical cyclone activity that may be included within the tracks. Yet when we focus on only the ETCs maximum intensities (Figure 3.4b) there are no significant differences for any group. This indicates that, for most groups, the ETC intensities are greater, but their maximum intensities and speeds are similar to non-event ETCs.

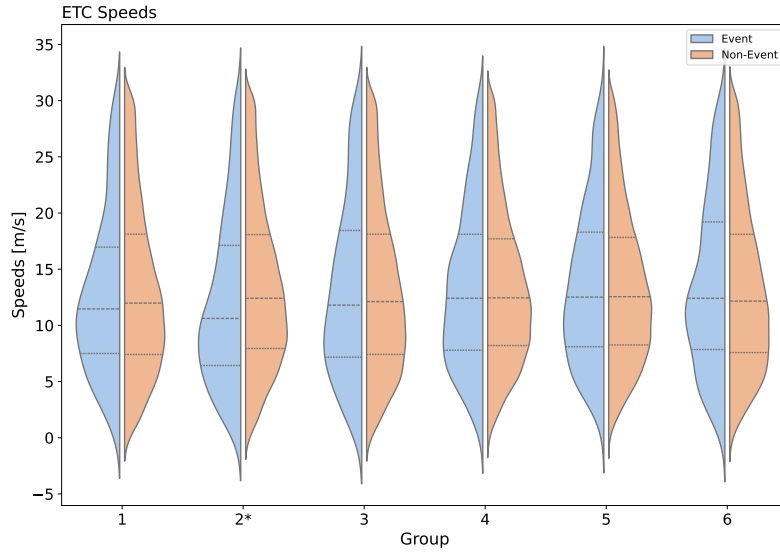


Figure 3.3: Distributions of all ETC track speeds, 6 hourly, when the track was passing within 10 degrees of the event polygons for event ETCs (blue, left) and for non-event ETCs passing within 8 degrees of the group domain (orange, right). Shown for all 6 groups. Significant difference between the distributions, using a t-test at a $p\text{-value} \leq 0.05$, is denoted as a * next to the group label.

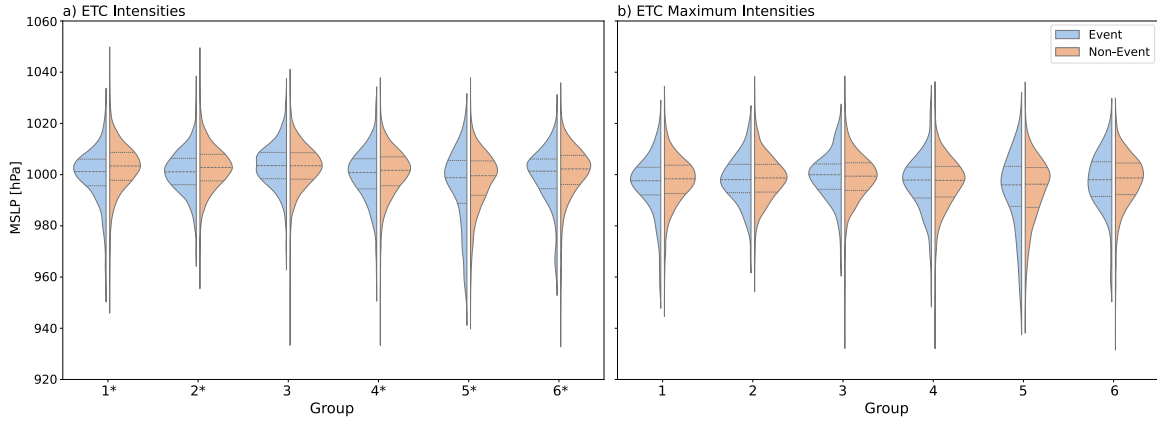


Figure 3.4: As in Figure 3.3 for the ETC track (a) intensities and (b) maximum intensities.

Lastly, a comparison of mean ETC frequency within events and random 14-day periods was conducted using bootstrapping with 5000 iterations, shown in Figure 3.5. Groups 1, 4 and 6 have significantly higher mean ETC frequency during event periods. This indicates that, on average, we see significantly more ETCs passing near the group during the events 14-day periods, than random 14-day periods. However, groups 2, 3, and 5 have observed means within the bootstrapped distributions. Group 2 and 5 have observed event means on the lower side of the distribution which, although not significant, indicates on average they have lower ETC counts within events. More ETCs seen within the events 14-day period shows spatial clustering of ETCs, which may compound precipitation risk (Karwat et al. 2024).

3.3.2 Precipitation Associated with ETC activity

For all events from 1979 to 2018, the percent of total precipitation associated with ETCs was found, broken down by group in Figure 3.6 and season in Figure 3.7. Out of the 371 events, there are 47 events (13% of events) that have no ETC associated precipitation, while another 41 (11% of events) have less than 5% of the event's total precipitation associated with ETCs. When counting ETCs per event, only 26 events

had 0 ETCs, shown in Table 3.1. Therefore there were 21 events that had an ETC associated with the event, but due to proximity or location of the ETC, there was no contribution of precipitation to the event. The remaining events have a peak in the distribution at 25% to 30%. While there are only 24 events (6% of all events) that have greater than 75% of their total precipitation associated with an ETC.

When the distribution is broken down by group, there are emerging trends, Figure 3.6. Groups 1, 3 and 6 have the most events with 0% ETC associated precipitation. This suggests that these regions have other precipitation mechanisms that contribute to the event total precipitation, or the simple 1000 km radius is missing frontal zones that might extend south into the regions. Groups 4 and 2 have events seen within most bins in the distribution, with no clear peak in ETC associated precipitation. Lastly, Group 5 (Northeast) has the fewest events for all groups due to the groupings of the clusters but the majority of the events (15 of 18) lie above 40% of total precipitation from ETCs.

The seasonal distributions of the ETC associated precipitation percent, Figure 3.7, matches trends seen in previous work regarding ETC precipitation climatology. Out of the 88 events that had less than 5% ETC associated precipitation, 42 of those events occurred in the summer months of JJA. The SON season also has more events within the lower percent values, with a maximum in the distribution at 5% to 15%. MAM and DJF are more evenly distributed across the binned percent values, with a maximum near 35%.

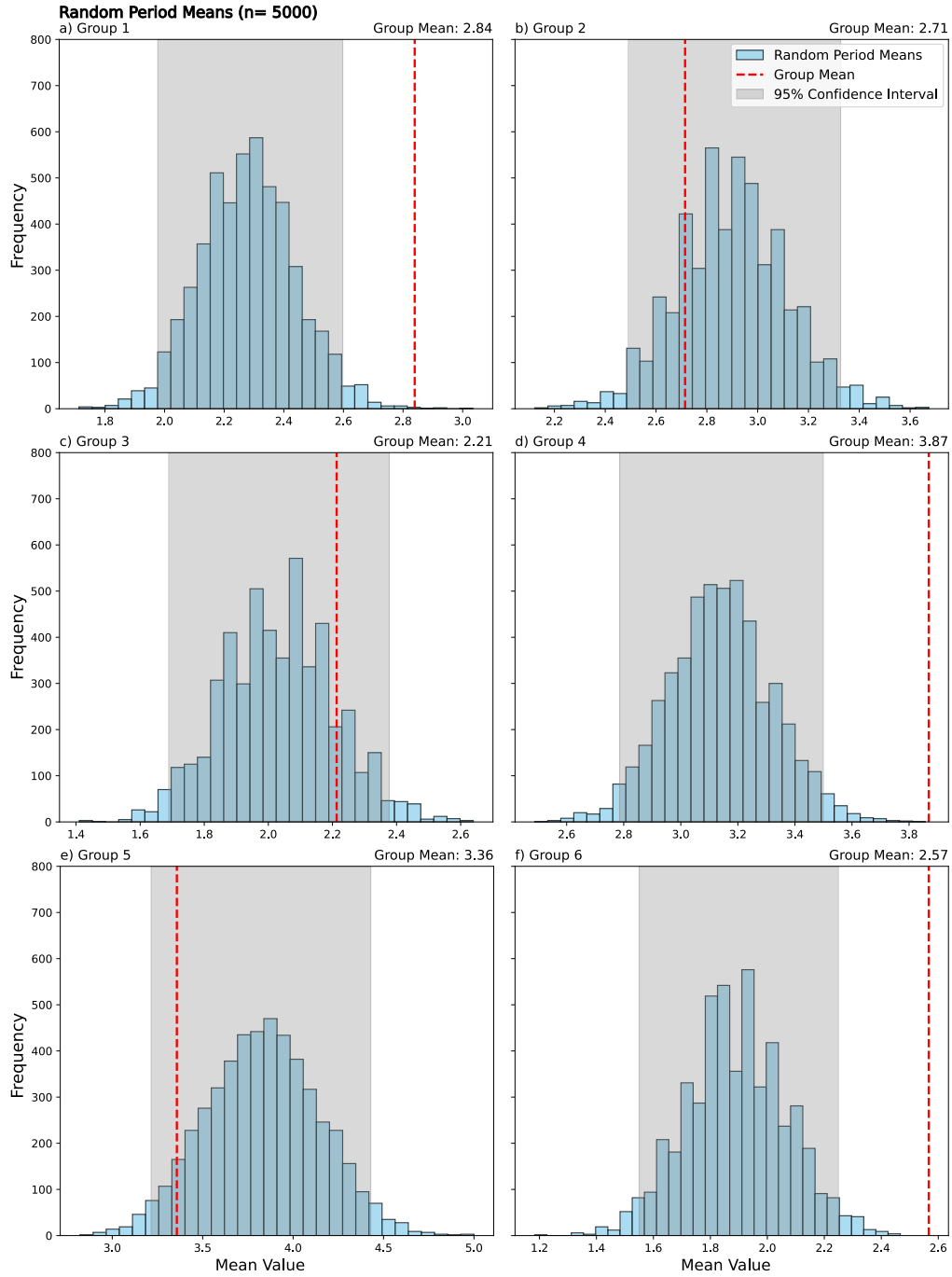


Figure 3.5: Distributions of Bootstrapped random 14-day periods mean ETC frequencies are shown for all six groups. The shaded regions show the 95% confidence interval. Red dashed line is the observed event group mean with the value stated at the top right for each group. Note the differences in the ranges of the x-axis.

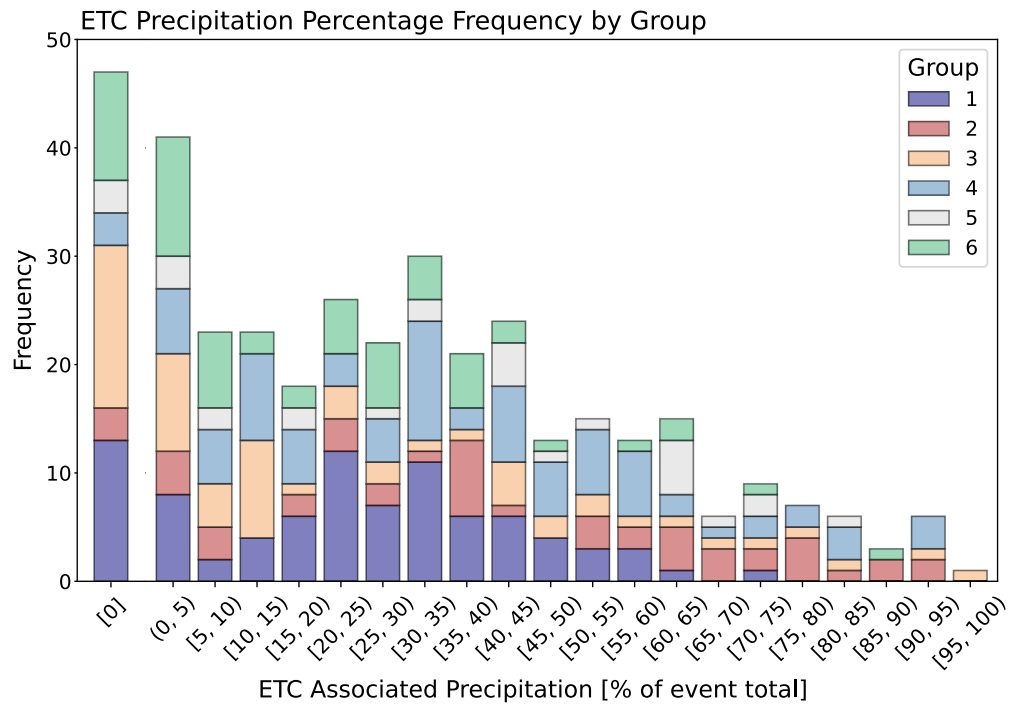


Figure 3.6: Distribution of all events ETC associated precipitation, % of total event precipitation associated with ETCs. Events are broken down by group, labeled in the top right. Note the first bar only contains percent values of 0%, while other are bins of 5% increments.

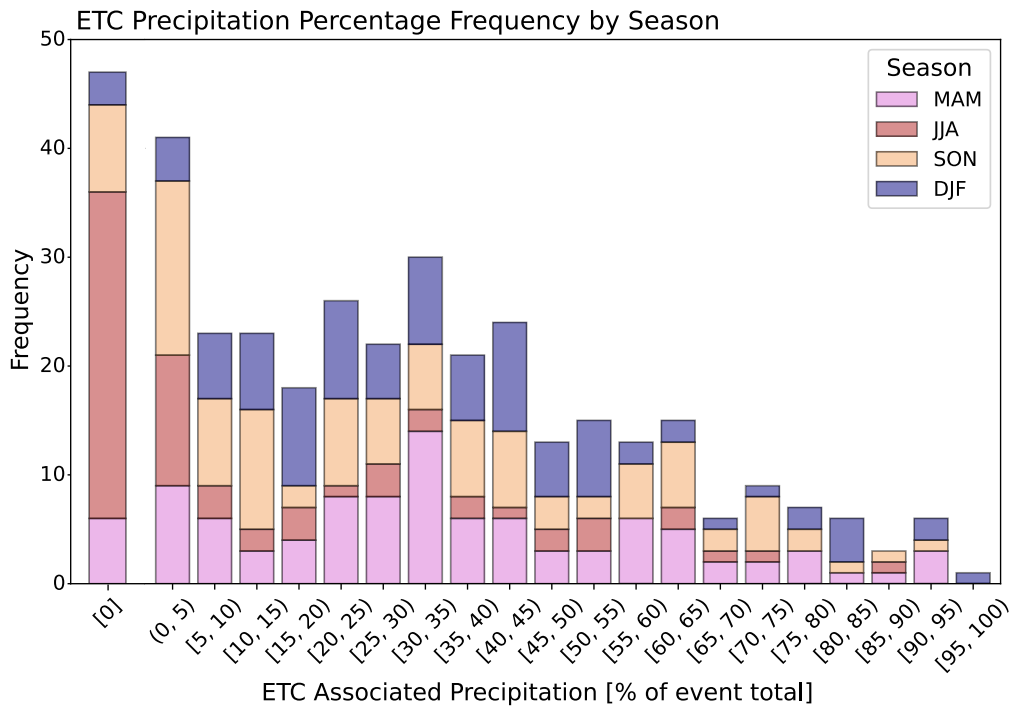


Figure 3.7: As in Figure 3.6 but broken down by the season of the beginning date of the event.

3.4 Discussion

This study aimed to quantify the role of ETCs within 14-day extreme precipitation events. For each event from 1979-2018, ETCs within 10 degrees of the event polygon were counted. The largest number of ETCs per event were seen in group 4 (Great Lakes Region) and group 5 (Northeast). Both of these regions are along the CONUS storm track region that often see higher frequencies of ETCs. Group 3 (Southern Plains) has the the most events with a low frequencies of ETCs. Similar to previous studies, events within JJA had less ETC activity with more ETCs seen in events within DJF, MAM and SON (Ulbrich et al. 2009; Hawcroft et al. 2012).

Next, the characteristics of ETCs within the events, including speed, intensity, maximum intensity and frequency, were compared to non-event periods. Group 3 (Southern Plains) had the lowest amount of ETC activity compared to all other groups, with similar ETC frequencies, speeds and intensities to non-event periods. Therefore ETC activity, although present in some events, is not be the main driver of precipitation during most events in the Southern Plains, indicating other mechanisms may be playing a larger role in event precipitation.

Aside from Group 3, all other groups had higher intensities seen for event ETCs when compared to non-event ETCs, although no significant difference was seen for the maximum intensities. This could have direct implication on the event precipitation given that ETC precipitation is dependent on intensity and moisture availability (Field and Wood 2007; Rudeva and Gulev 2011; Pfahl and Sprenger 2016). Group 1 (West Coast), group 4 (Great Lakes Region) and group 6 (Southeast) had significantly more ETCs on average, with significantly higher intensities, during the 14-day event periods. This suggests that the number of intense ETCs seen during the 14-day events are

compounding to drive event precipitation in these regions, although group 6 has the lowest average event precipitation associated with ETCs.

Group 2 (Northern Plains) was the only group that had significantly slower ETCs during event periods, although the 14-day frequency of ETCs was not significantly less than random 14-day periods. Therefore during events in the Northern Plains we see slow moving, intense ETCs that lead to extended extreme precipitation periods. Lastly, group 5 (Northeast) did have more ETC activity than other groups, but the only significant difference from non-event periods was seen in ETC intensity.

The ETC associated event precipitation was found as a percent of total event precipitation. Two peaks are seen in the percent distribution. The first occurring from 0% to 5%, accounting for 24% of the events from 1979-2018. The next peak, smaller than the first, is seen at 30% - 35%. This pattern is also seen for events within groups 1, 3 and 6. While groups 4 and 5 have more spread in their percent distributions. Although group 2 (Northern Plains) only had a mean of 2.71 ETCs per event, events within this region had the highest percent of their precipitation associated with ETCs. Looking at the distribution in terms of seasonality of events matches shows similar results to ETC counts seen in Table 3.2. We see less ETC activity in JJA, with overall less precipitation associated with ETCs. While the remaining events within other seasons have their peak percent at 30% - 35% ETC precipitation.

The use of a radial cap to find event ETC activity and its associated precipitation does have limitations. This method may miss long trailing frontal zones, or it may over count precipitation unrelated to the ETC. In future studies using objective frontal tracking may better quantify ETC related precipitation. Yet, results match well with general ETC climatologies found in previous studies (Hawcroft et al. 2012; Pfahl and Wernli 2012).

3.5 Conclusion

Understanding the drivers behind extended precipitation periods it imperative for future S2S model forecast improvement and also may provide additional avenues for proxy forecasting at the S2S timeframe. Therefore the the characteristics of 14-day event ETCs were quantified and compared to non-event periods.

ETC activity was highly dependent on the region, as summarized in Table 3.3. The Northern Plains had slow-moving, intense ETCs within events. In contrast, the West Coast, Great Lakes and Southeast have significantly more ETCs within event periods that are also more intense. The Northeast had more intense ETCs with a higher average of ETC activity, but the high frequency of ETCs is not unique to event periods. Lastly, the South Plains had no differences seen between event and non-event periods.

Group	14-day Mean ETC	14-day Frequency	Speed	Intensity	Mean % ETC Precip
1-West Coast	2.84	↑	-	↑	24.8%
2-North Plains	2.71	-	↓	↑	43.9%
3-South Plains	2.21	-	-	-	22.1%
4-Great Lakes	3.87	↑	-	↑	36.4%
5-Northeast	3.36	-	-	↑	36.7%
6-Southeast	2.57	↑	-	↑	20.2%

Table 3.3: Summary of the ETC characteristics per group discussed within Section 3.3. Arrows indicate significant difference from non-event periods within the group domain.

Lastly, the contribution of ETC related precipitation to the total precipitation seen in 14-day extreme precipitation events was found. Percentage of ETC related precipitation was dependent on both the season of the event and the region. The Northern

Plains, Great Lakes and Northeast has the highest average percentage of ETC related precipitation. While the remaining regions have lower average ETC precipitation percentages, it is important to note there are still events in these regions that have larger percentages. This indicates that there is more variability in the type of precipitation drivers in these regions.

These results highlight there are regions where 14-day extreme precipitation events are often associated with greater ETC activity and thus higher ETC related precipitation. Although ETC activity during events needs to be analyzed within S2S models, a previous study has shown skill of general ETC activity out to week 3-4 (Zheng et al. 2019). Therefore ETC activity may be a future avenue for S2S prediction of long-duration extreme precipitation.

Chapter 4

Listening to Stakeholders: S2S Precipitation

Forecasts Usability

This chapter has been published in the following peer-reviewed journal article:

Schroers, M. A., T. A. Dickinson, P. Ówik, R. A. McPherson, and E. R. Martin, 2024: Listening to Stakeholders III: Potential Users Evaluate Product Content and Design for Subseasonal Extreme Precipitation Forecasts. *Bull. Amer. Meteor. Soc.*, **105**, E1422–E1437, <https://doi.org/10.1175/BAMS-D-23-0233.1>.

Mentions of supplementary material within this chapter, specifically the activity packet, can be found at the journal article cite.

4.1 Background

The PRES2iP researchers used best practices for co-development of new science and products (models, tools, graphics, etc.) with practitioners. We hosted workshops and applied proven methods, such as icebreakers, focus groups, testbed, and role-playing activities, to engage participants and produce knowledge with our stakeholders to be incorporated in future research.

Workshop 1, the Research Priorities Workshop, took place in person on 12-13 July 2018 in Norman, Oklahoma, with 21 practitioners participating. The workshop was designed to prioritize the types of S2S extreme precipitation events to study throughout the PRES2iP project. Discussions included: (1) how participants defined “extreme

precipitation,” (2) how these events impacted their jurisdiction, (3) how participants accounted for uncertainties within their decision processes, and (4) how they interpreted precipitation forecast products (VanBuskirk et al. 2021). The PRES2iP team learned the importance of place-based experiences of participants regarding extreme precipitation events. For example, the decision-making process and forecast uncertainty tolerance among stakeholders may change for differing regions, with various timelines of actions based on historical local or regional impacts (Schroers and Martin 2022). Further, statistically extreme rainfall that caused little damage, few injuries, and no fatalities was not considered “extreme” by participants. In addition, established relationships with forecasters or scientists, and personal experience or familiarity with S2S forecast products, were identified as the two most important aspects for building their confidence in making decisions on the S2S timeframe.

Workshop 2, the Product Definition Workshop, occurred virtually from 4-6 October 2021, with 19 participants. The workshop aimed to (1) identify the products participants use at longer lead times, (2) gather feedback on prototype fact sheets designed to communicate the climatology of extreme precipitation periods and regional impacts, and (3) understand how the participants perceived uncertainty at the S2S timeframe (VanBuskirk et al. 2023a). The PRES2iP team heard how participants had limited knowledge of S2S products, regardless of the participant’s job and jurisdiction. Participants stated that probabilistic products would be the best fit for those produced for S2S lead times, although the desired probabilities to trigger participant actions tended to be high—above 50% in most cases. Lastly, participants stressed the need for a suite of products that showed location, timing, intensity, and antecedent conditions for S2S extreme precipitation. The results from both workshops informed our research goals and guided our next steps with focus on meeting stakeholders needs.

4.2 Approach and Methods

The PRES2iP team hosted our third stakeholder workshop in person on 13-14 April 2023, in Norman, Oklahoma. This workshop built on knowledge gathered during the prior two workshops. Logistical planning for this workshop occurred from November 2022 to April 2023. All 31 participants who attended one or both previous PRES2iP workshops were invited to this workshop; four attended alongside five new participants. Backgrounds of the nine participants were: emergency managers (2), water resource managers (6), and a decision-maker who worked with water utilities (1). Low response rate from our previous participants, due to spring scheduling conflicts and the inability to travel during their storm season, created the need for us to reach out to new participants. In order to increase the sample size of stakeholders, we hosted a three-hour virtual mini-workshop in February 2023 with three participants from prior workshops. These three participants evaluated the planned activity sheet and corresponding product website, providing necessary feedback to ensure our materials were clear. As a result, we added more antecedent conditions to each case study, and replaced titles of “1 in 10” to our 90th percentile products to “Worst-Case” throughout all materials. Although the small sample size of stakeholders available for this workshop may limit the breadth of knowledge we may gain, it will provide a good foundation for future research.

The full workshop consisted of three sessions. Session 1 introduced the project’s goals and results from previous workshops, focusing on how the information learned from participants in prior workshops had been incorporated into the research itself. The participants also networked with fellow participants and PRES2iP team members through an icebreaker activity. Lastly, to inspire thinking on the S2S timeframe,

participants were placed in small groups and tasked to rank various meteorological phenomena, (e.g. tornadoes, hurricanes, monsoons) based on their respective time scale, from shortest to longest.

The remaining two sessions examined experimental S2S forecast products designed by the PRES2iP team using two impactful 14-day extreme precipitation case-studies from differing seasons and locations sourced from the PRES2iP 14-day extreme precipitation database (Dickinson et al. 2021). The first scenario (S1) session covered an extreme precipitation period across the Pacific Northwest that occurred from 6-19 December 2015. During this two-week period, Oregon experienced an estimated \$21.6 million in property damages due to flooding - both coastal and inland, debris flow, and high winds. Furthermore, an additional \$14.9 million in property damages were reported in Washington (NCEI 2023). S1 was characterized by wet antecedent conditions combined with an anomalously wet forecast 7 and 30 days ahead of the event. The second scenario (S2) occurred from 21 April to 4 May 2017, predominantly across Missouri, with heavy rains also affecting parts of surrounding states. Property damages in Missouri alone were estimated to surpass \$137 million as extensive destruction was observed in the form of damage to roads, residential homes, businesses, and crop fields across the entire state (NCEI 2023). S2 exhibited relatively normal antecedent conditions with the longer-range forecasts depicting a dry weather forecast.

During the scenario-based sessions, participants were placed in groups of 3-4 people to complete a six-part guided activity packet facilitated by a member of the PRES2iP team (packet found in supplementary material). Another team member took notes of participant statements and reactions. The session packet contained relevant background information on each case study, instructions about how to access the available products, and questions for each activity (described in Table 4.1). The PRES2iP team

developed an accompanying website housing the forecast products, antecedent conditions, and outlooks for both scenarios; all products included were identified as being used by participants in prior workshops. To highlight possible impacts, the following antecedent conditions were provided: soil moisture conditions, the U.S. Drought Monitor (Svoboda et al. 2002), and precipitation totals from the prior month. We also included snowpack information for S1. We provided the following commonly used outlooks: seasonal, monthly, day 8-14, and day 6-10 precipitation from the Climate Prediction Center (CPC); CPC hazard outlooks; and the Day 3 and Day 1 excessive rainfall outlooks from the Weather Prediction Center (WPC). The website simulated how participants would gradually access products over time in an operational setting. Although our focus was receiving feedback on our experimental forecast products, our PRES2iP team encouraged participants to use any of the existing operational products provided.

PRES2iP's experimental S2S products were derived from real-time forecasts using the National Center for Environmental Prediction (NCEP; Saha et al. 2014) model from the S2S model database (Vitart et al. 2017). The NCEP model, version 2 of the Climate Forecast System (CFSv2), was initialized daily and had 16 ensemble members. Precipitation anomalies were created using a daily precipitation climatology using a hindcast period of 1991-2010 following the methodology of ?. On each case-study website, we linked our experimental forecast products at lead times of 30, 21, 14, 7, and 3 days ahead of the start of the extreme precipitation period, summarized with respect to Scenario 1 in Fig. 4.1. Table 4.1 and Figs. 4.2-4.5 detail the forecast products in each packet. Participants were not told which model we used until the conclusion of the workshop to prevent any a priori biases that could affect their responses.

Table 4.1: Goals of each workshop activity and corresponding forecast product provided.

Activity Name	Activity Goal	Corresponding Experimental Forecast Product or Activity Focus
#1 Action vs. Lead Time	Connect stakeholder actions related to extreme precipitation to forecast lead time using forecasted precipitation.	Ensemble mean 2-week total precipitation and precipitation anomaly maps.
#2 Benefits and Limitations of Text Products	Gather feedback on three different types of text products associated with S2S forecast products.	Descriptions of the 2-week period following various styles.
#3 Precipitation Forecasts by Day	Interpret and discuss pros/cons of forecasted daily precipitation values at a given location for various lead times.	Time series of best case, mean, and worst-case daily precipitation and accumulated precipitation.
#4 Precipitation Forecasts by Amount	Compare and contrast probabilistic precipitation forecasts by amount to ensemble mean forecasts.	Probability precipitation >5 in, 10 in (Scenario 1); 3 in, 5 in (Scenario 2)
#5 Worst-Case Scenario	Examine “Worst-Case Scenario” forecast product at various lead times and compare to ensemble mean forecasts.	90th percentile of 2-week total precipitation
#6.1 Forecast Error - Model Skill	Examine estimated forecast error within an S2S model to discern model uncertainty tolerance.	Table of mock POD and FAR (see Table 2)
#6.2 Forecast Error - Object Oriented	Rank forecast and observation pairs to show spatial uncertainty tolerance on the S2S timeframe.	Fig. 1 from Davis et al. (2006) and self-drawn hypothetical forecast/observation pairs.

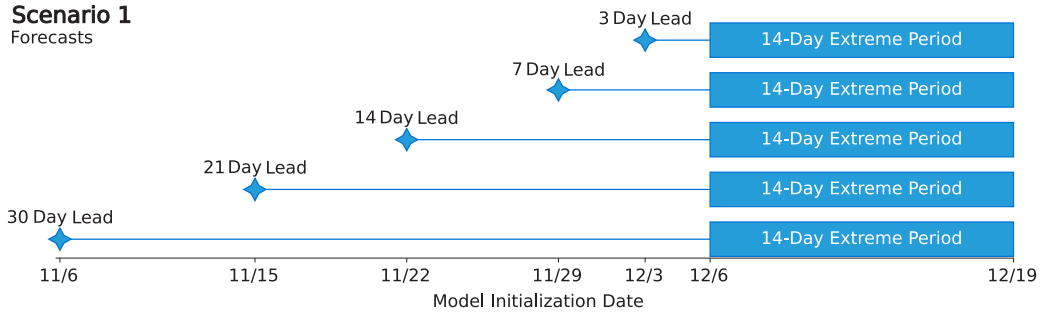


Figure 4.1: Summary of the five forecasts periods used for the experimental products that were given to the participants. As an example, model initialization dates are included for Scenario 1, that occurred from 6 – 19 December, 2015.

4.3 Lessons Learned

In the following sections, the design, implementation, and feedback of the six activities that the groups completed for both scenarios is discussed.

4.3.1 Action versus Lead Time Activity

Based on discussions from our previous workshops, we understand that users make decisions at specific lead times ahead of a potential event. Therefore, this activity tasked the participants to match possible actions to specific lead times (see supplementary material). For each lead time, we provided the ensemble-mean total precipitation and anomalies for the 14-day period of interest (Fig. 4.2). We designed this product to emulate the WPC’s quantitative precipitation forecast (QPF) to ensure easy interpretation, as QPFs are one of the most used forecast products with regard to precipitation-related decision making (VanBuskirk et al. 2023b).

For each lead time, participants reviewed the list of actions - adding others that were relevant to their work, and analyzed the forecast products to note which of these possible actions they might take. Reported actions were distinctly different between the two scenarios. In S1 (wet antecedent conditions), participants indicated that they

would begin actions at lead day 30 through 14, primarily preparing dam-failure emergency action plans, inspecting dams or other flood-mitigation equipment, and engaging in discussions on probability of extreme precipitation with colleagues. On lead days 7 and 3, more actions would be taken, including preparing the region for hazard mitigation, briefing public officials, and increasing the dialogue with colleagues and staff. In S2 (normal antecedent conditions), three of seven participants did not list any actions until lead day 7; the actions listed for the shorter lead times closely resembled those in S1.

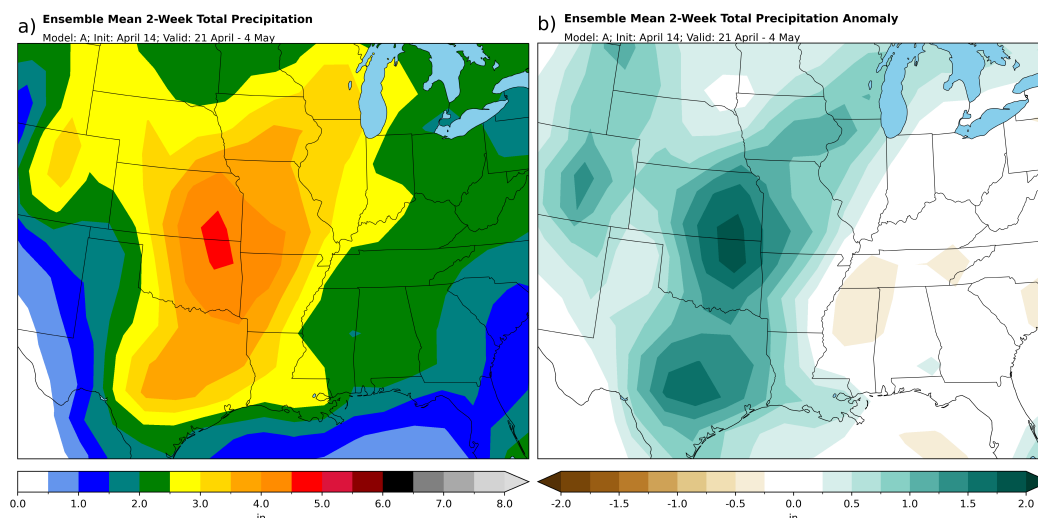


Figure 4.2: S2S experimental forecast products designed by the PRES2iP research team as a result of participant feedback in workshops 1 and 2. Scenario 2 (21 April to 4 May 2017) is displayed, showing forecasts for Missouri and surrounding states at a lead time of 7 days prior to the start of the extreme precipitation period. (a) Ensemble-mean, 14-day total precipitation and (b) ensemble-mean, 14-day precipitation anomalies are shown. Precipitation values are in inches for workshop participants.

4.3.2 Text Product Activity

In previous workshops, participants emphasized the need for supporting narratives associated with S2S forecast products. Specifically, during the first workshop many

participants stated that having a narrative would help with interpretation of unfamiliar products (VanBuskirk et al. 2021). Furthermore, participants reaffirmed in workshop 2 that having text-based information regarding possible losses could facilitate proactive planning for an extreme period (VanBuskirk et al. 2023a). Therefore, in this activity, we drafted three different text forecast products for each scenario at a 3-7-day lead time, found in supplementary material. Each text product was designed intentionally to address prior workshop participants' requests in different formats. The first product was modeled after a National Weather Service forecast discussion, encompassing a synopsis and a long-term forecast discussion. The second product was the shortest, had the least jargon, and focused on amounts, timing, and potential impacts of forecasted precipitation. The final and lengthiest product included the content of the second product as well as technical information on the synoptic scale set-up observed in S2S models and existing teleconnection patterns that could support extreme precipitation.

After reading each text product, participants were tasked to detail its advantages and disadvantages. For both scenarios, participants favored the second text product because its information regarding expected impacts and the forecasted range of precipitation at different locations was the most useful. While many participants appreciated the first text product, some felt it was too technical or jargon-heavy, making it challenging to extract useful information. The third text product was considered excessively detailed and lacking in actionable information. Overall, participants found S2S forecast-related text products to be most useful when they included "calls to action" with anticipated impacts, forecasted precipitation ranges over specific dates, and location-specific differences when necessary. This result is consistent with studies on the weather timeframe that show impact-based severe weather warnings increase the

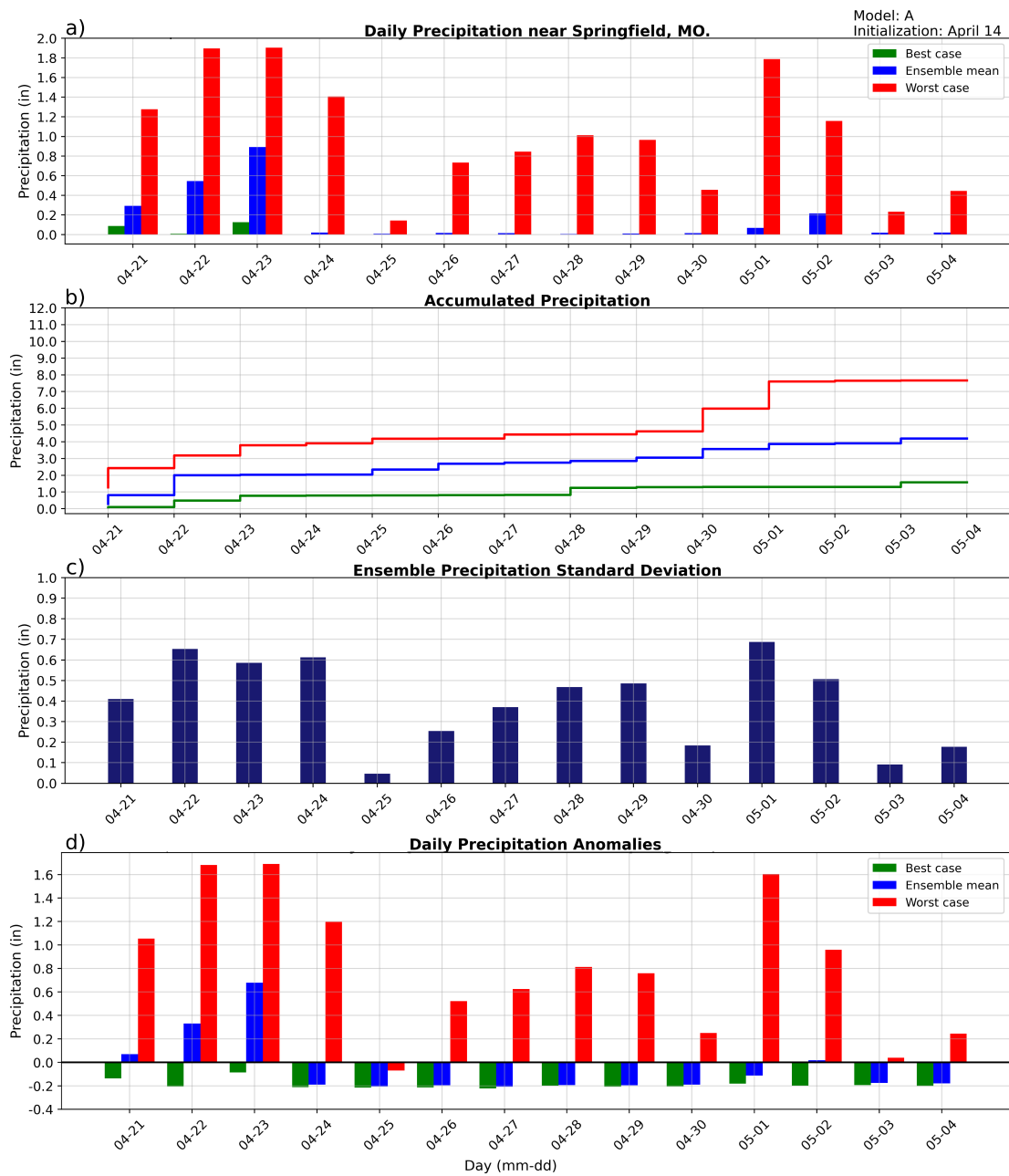


Figure 4.3: As in Figure 2, but for the time series product that includes (a) daily precipitation, (b) accumulated precipitation, (c) ensemble standard deviation and (d) precipitation anomalies for 21 April to 4 May 2017.

public's understanding and intention to take protective action (Ripberger et al. 2015; Casteel 2016; Potter et al. 2018).

4.3.3 Precipitation by Day Activity

An important lesson from workshop 2 was that a suite of forecast products must include both the spatial distribution of precipitation and information on timing and intensity (VanBuskirk et al. 2021). In response, this activity gave participants daily precipitation forecasts for Seattle, WA, in S1 and Springfield, MO, for S2. The PRES2iP team designed a product with four graphs (Fig. 4.3) that showed precipitation values, accumulated precipitation values, ensemble standard deviation, and precipitation anomalies for each of the 14 days. Daily values were presented for the best case (10th percentile of the ensemble) and the worst-case scenario (90th percentile of the ensemble). We included daily values of the ensemble’s standard deviation to show model uncertainty, as prior workshop participants indicated that uncertainty information might help to increase trust in the forecast product (VanBuskirk et al. 2023a).

Overall, all participants provided positive feedback on this activity’s product, stating that it was useful to have both map and graph products (and not either alone) to increase their confidence in the forecast. Many participants emphasized that the ensemble standard deviation information helped to clarify the forecast, though some preferred the information displayed as whiskers on the daily precipitation product rather than a bar graph. In contrast, the daily anomalies were deemed less useful, as they required additional interpretation; the other three panels were considered sufficient.

4.3.4 Precipitation by Amount Activity

During workshop 2, all participants preferred probabilistic forecasts over deterministic products because the former conveyed forecast uncertainty (VanBuskirk et al. 2023a). Accordingly, the PRES2iP team created two map products that displayed the probability that 2-week precipitation totals would exceed two different thresholds (e.g. Fig.

4.4). Prior workshop participants stated the importance of place-based thresholds tailored to their specific jurisdiction (VanBuskirk et al. 2021); however, limited resources prohibited us from creating a user-selected threshold. Instead, we chose our lower threshold to align with the region’s climatology, set at 5 inches for S1 and 3 inches for S2. Higher thresholds of 10 and 5 inches, respectively, represented a more “extreme” value.

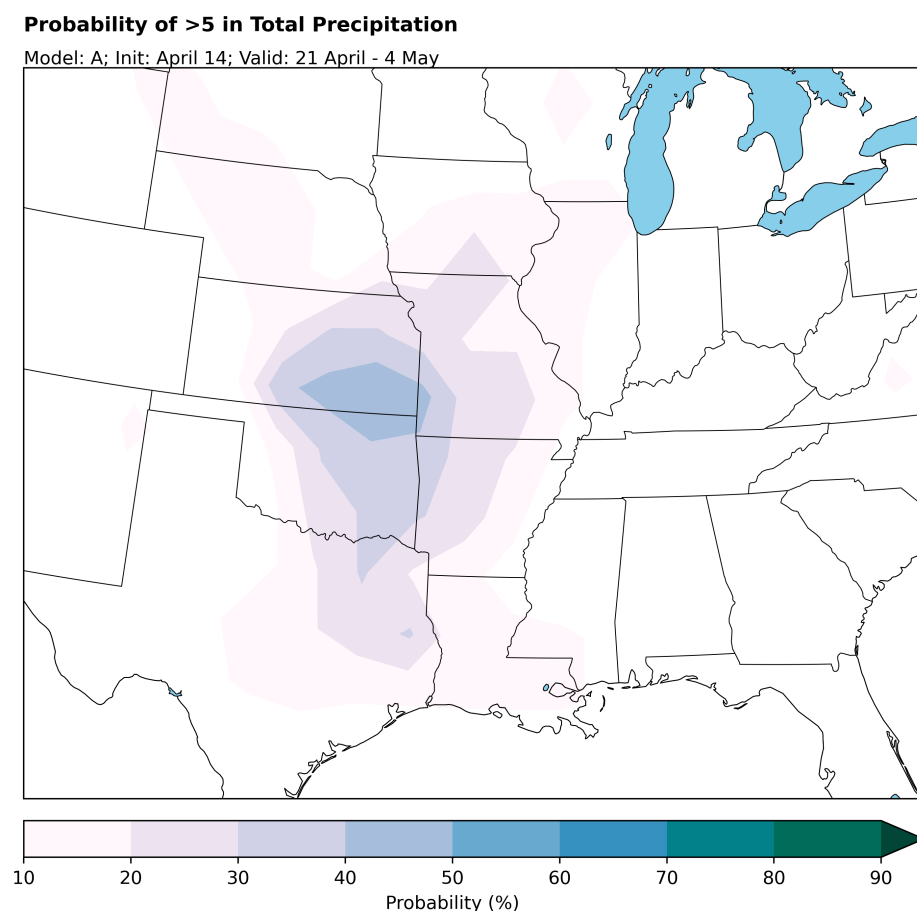


Figure 4.4: As in Figure 2, but probability of 14-day total precipitation above 5 inches (127mm).

The participants were tasked to compare this product to the other available map products and discuss potential applications. Participants thought the products effectively depicted where the highest probability of large precipitation amounts would

occur but noted that the location information was not useful without timing information. For example, one participant stated, “That could mean less than 1 inch per day,” while another added, “Or it could be 5 inches in one day.” All participants agreed that a suite of products that included QPF, text description, and daily time series products would be more useful than this probabilistic product.

4.3.5 Worst-Case Scenario Activity

During the “Precipitation by Day” activity, we provided daily precipitation values for the worst-case scenario (90th percentile) at a single point, which did not inform participants about the spatial distribution of the precipitation. During workshop 1, participants noted that extreme-event planning might involve the use of a worst-case scenario map product (VanBuskirk et al. 2021). Therefore, we asked participants to discuss whether the worst-case map product (Fig. 4.5) added useful information to the “Precipitation by Day” products. All participants agreed that the worst-case map product was useful at longer lead times to show the possible extent of the extreme period, but the product was not useful alone because it did not highlight when the precipitation would fall within the 14-day period. One participant stated, “[The bar chart of daily precipitation] is good in short range, but further in time it is better to have a map. With a point forecast for Seattle 30 days out, you might miss what is happening in the area surrounding unless it’s a map.” Ahead of day two of the workshop, participants also had discussed changing the language of “Worst-Case Scenario” to “Maximum in Precipitation over 2 weeks” because it would allow decision makers to decide what is the worst-case for their region, given current conditions.

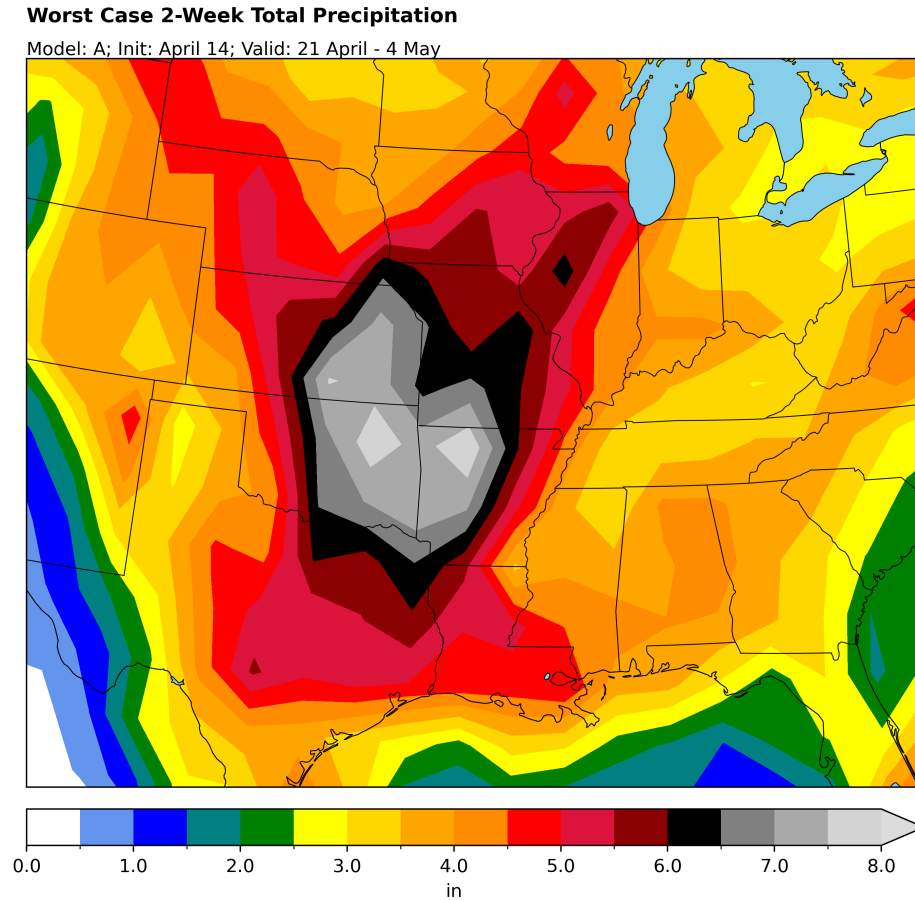


Figure 4.5: As in Figure 2, but for 90th percentile of 14-day total precipitation, labeled for participants as the Worst-Case.

4.3.6 Forecast Error Activity

Through discussion with stakeholders in workshop 2, we understand that stakeholders may be reluctant to use forecast products if past forecasts were perceived as “poor.” Conversely, trust can be built over time with models that produce reliable forecast products (VanBuskirk et al. 2023a). During workshop 3, many participants saw S2S model output for the first time, prompting our interest in forecast-error tolerance of stakeholders on the S2S timescale using “traditional” skill metrics. Thus, for all lead times, we provided estimated values of Probability of Detection (POD) and False Alarm Rate (FAR) in a typical S2S model regarding our extreme precipitation periods (Table

2). Participants expressed caution when dealing with forecasts beyond 14 days and did not consider these metrics. They became more interested in lead times of 14 days (and shorter) and PODs of 50% and larger. Intriguingly, participants also preferred FARs of 50% or less, indicating that end-users desired confidence levels and skill metrics to be “more yes than no” before considering products to be actionable. These results coincide with those in workshop 2, where participants wanted probabilities of occurrence above 50% to prompt action (VanBuskirk et al. 2023a).

Lead Time	POD across 2,000 square-mile area	FAR across 2,000 square-mile area
30 days	30%	85%
21 days	40%	80%
14 days	50%	75%
7 days	70%	55%
3 days	80%	30%

Participants preferred high PODs rather than low FARs with one participant noting, “Everybody forgets the false alarm you prepared for, everybody remembers the [miss] you didn’t prepare for[...].” Yet, the metrics themselves were seen as not useful to understanding forecast error. Some participants stated that it would be more interesting to have a model-to-model comparison tool.

4.3.7 Object-Oriented Activity

One of the pivotal insights of workshop 2 was that end-users inherently anticipated forecasts changes over time and thus have a certain level of tolerance to spatial displacements in expected impacts (VanBuskirk et al. 2021). This result suggested that an object-oriented verification scheme—one that treats the forecast as a contiguous

region instead of a series of grid-boxes—might be more useful in evaluating forecast quality and usability. Therefore, our team designed an activity to understand the types and magnitudes of error tolerance in this context.

The activity used theoretical pairs of forecast and observation contiguous regions (Fig. 4.6), as in Davis et al. (2006). Each participant ranked the perceived forecast quality from best to worst in the context of subseasonal extreme precipitation periods. Three of our seven participants identified panel ‘a’ as the best. However, the remaining four participants ranked panel ‘e’ as the best. Conversely, most participants ranked panels ‘b’ and ‘d’ as the worst.

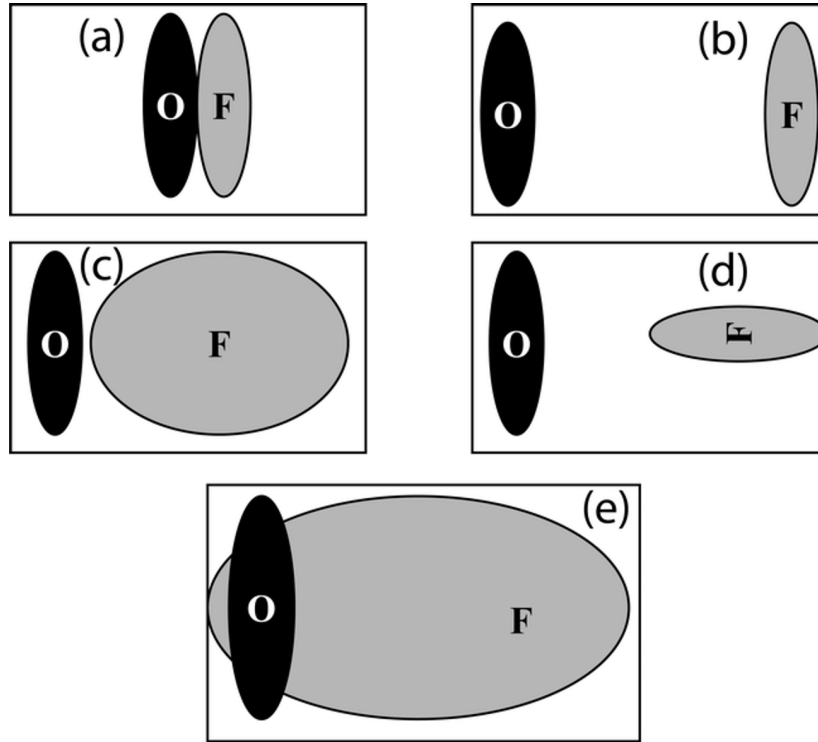


Figure 4.6: Schematic examples of hypothetical Forecast (F) – Observation (O) combinations, from Davis et al. (2006, Fig 1)

Participants then were given more realistic forecasts with a lead time of 3 days (Figs. 4.7-4.8, black) and a polygon of the actual extreme precipitation region (Figs. 4.7-4.8,

red), with distances between centroids given to participants in miles (see Dickinson et al. 2021). The forecast objects were adapted from an experimental model from the PRES2iP team. Three maps (Fig. 4.7) focused on the spatial proximity of forecast-observation pairs and utilized polygons with similar sizes. In Fig. 4.7a, the polygon centroids were approximately 150 miles apart; all participants agreed the forecast was a “hit.” When polygon centroids were 220 miles apart (Fig. 4.7b), participants were more hesitant to call the forecast a “hit,” with one participant definitively identifying it as a “miss.” One participant described Fig. 4.7a,b as “really good” forecasts while others made comments such as “good on a state level” and “the forecast generally covers the same area...it is ok.” The northernmost shift in the forecast (Fig. 4.7c), with centroids about 350 miles apart, prompted unanimous agreement that the forecast was too distant from the observed location. Participants also noted that, under otherwise equal conditions, the direction of spatial errors was unimportant; however, changes in antecedent conditions, for example, could affect their tolerance to translations in the forecast.

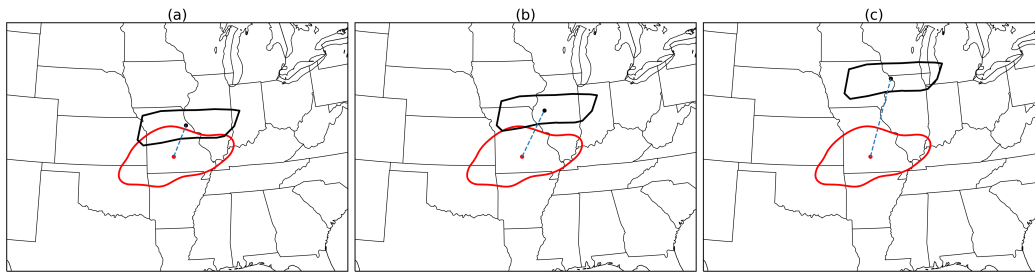


Figure 4.7: Various pairings of forecasts (black) and truth (red). Colored dots represent the polygon centroid. Distances between centroids are about (a) 150 miles, (b) 220 miles, and (c) 350 miles.

Finally, participants compared the relative sizes of the forecast and observed regions while maintaining a small distance between centroids (100 mi, Fig. 4.8). Six of seven participants classified Fig. 4.8a as a “hit” and generally preferred the larger forecast

region. However, one practitioner expressed a preference for forecasts with higher precision (e.g., Fig. 4.8b) if they were associated with higher likelihood. The consensus was that a forecast akin to Fig. 4.8a at long lead times would be useful for planning before using a forecast akin to Fig. 4.8b at shorter lead times.

Traditional metrics, such as critical success index or anomaly correlation coefficient, would yield scores near 0 for a scenario such as Fig. 4.7b despite nearly every participant considering the forecast to be a “hit.” Hence, this activity reaffirmed the utility of object-oriented verification schemes in the context of S2S extremes. However, participants assessed size and proximity in isolation, so future conversations are needed to diagnose how end-users would balance these two factors.

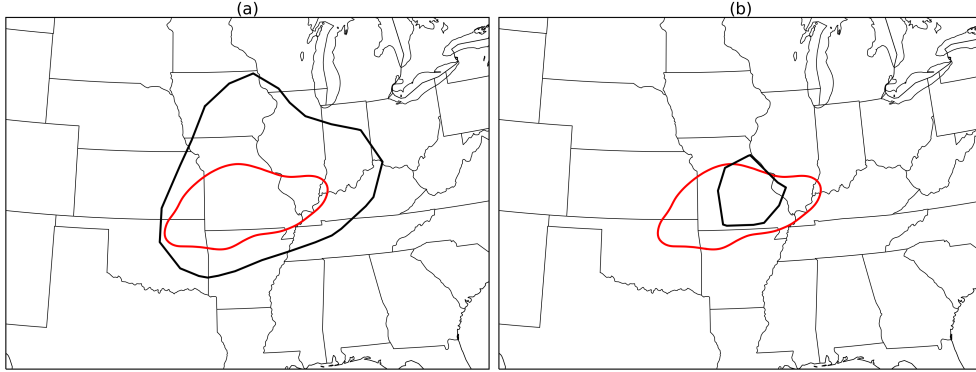


Figure 4.8: As in Figure 7, but for polygons with similar centers but different spatial scales.

4.4 Summary

During this workshop, the PRES2iP team presented a suite of experimental S2S 14-day precipitation forecast products developed using knowledge gained through collaboration with decision makers at two previous workshops. The products covered two different 14-day extreme period scenarios: one in the Pacific Northwest with wet antecedent

conditions and an anomalously wet forecast (S1) and another in the eastern Plains with normal antecedent conditions and a drier forecast (S2). Both scenarios resulted in an impactful observed event, although the participant perceptions of the severity of the 14-day period differed between scenarios, owing to the differences in forecast skill. All potential decision actions associated with the experimental S2S products focused on planning at lead times of 14 days or longer, with more actions for risk mitigation taken at lead times less than 14 days. The actions listed at S2S lead times for S1 were consistent with previous studies (e.g. White et al. 2017). For S2, participants stated they would take less action, or no action for three participants, until lead day 7. Thus, the inaccurately dry forecast for S2 at S2S lead times, 30, 21 and 14 days ahead of the extreme period, led to less planning occurring. The perceived usability of the products did not change based on the scenario, yet the outcomes changed with differences in forecast skill. This emphasizes the need to engage with decision-makers about product content and design, while also studying the improvement of precipitation skill at S2S lead times.

The extended nature of the 14-day extreme precipitation periods that were targeted during this workshop led to a lot of discussion regarding the importance of timing information. During every activity regarding a product the lack, or the inclusion, of timing information was brought up by participants. This highlights the need for a suite of products that cover the spatial extent, as well as timing information. Thus, the participants valued a combination of ensemble-mean map products (e.g., 14-day QPF), text narrative with “calls to action,” and daily time series products. The combination of different forms of products helped to increase their trust in the forecast and, thus, its usability. One participant stated, “All graphs had value for a singular perspective – but confidence is low. Looking at them in combination increased trust and confidence in what was going to happen.” Interestingly, these types of products are already heavily

used on the weather timeframe, and therefore there is less interpretation needed by the stakeholder that are already comfortable with weather products. On the other hand, all products showing anomalies were perceived as less usable due to the additional interpretation needed to gauge the severity of the anomaly. This result may be due to the stakeholders within the workshop not knowing other regions' precipitation trends, whereas they may be more comfortable working with anomalies within their own region.

Additionally, the PRES2iP team explored participants' tolerance of uncertainties on the S2S timeframe, finding multiple examples that depicted higher tolerance for differences in location and spatial coverage than seen during the weather timeframe. Despite some increased nuance in forecast evaluation, the implementation of object-oriented schemes to verify S2S forecasts presents a clear opportunity for increased education, awareness, and actionability on these timeframes.

Although the PRES2iP forecast products were experimental and not meant to be transitioned into operations, this work provides the foundation for future enhancement and development of products that are useful for and usable by decision makers. Due to low availability of our stakeholders, this workshop was only conducted with nine stakeholders and thus limits the ability to generalize the findings to all stakeholder communities. In the future, similar workshops should be conducted, avoiding peak storm season, with larger sample sizes to further validate our findings. Furthermore, future studies should strive to get input from more perspectives, such as agriculture, to ensure the results shown here are more widely applicable or determine additional needs from other decision maker communities. The PRES2iP team and resulting research have been immeasurably influenced by stakeholder input and engagement through our three workshops. By engaging with stakeholders throughout the entire project, we have conducted more meaningful research. Documentation of our three workshops - their design and implementation, and their results - is intended to encourage a broad range of

researchers to engage with decision makers, managers, underserved populations, or any parties deemed appropriate, throughout the entire research and development process. Our activity examples can be modified for similar uses in research projects or serve as examples in classroom coursework to educate the next generation of researchers on how to listen to product users.

Chapter 5

Discussion

Bridging the prediction gap in the S2S timeframe, between typical weather forecasting and seasonal outlooks, would benefit decision makers in many sectors to increase resilience and mitigation efforts for long lasting impacts. In this study long-duration precipitation events on the S2S timeframe were analyzed from multiple perspectives: model predictability, atmospheric drivers and end-user S2S forecast usability.

There has been an increasing need for improvement of S2S precipitation forecasts for future mitigation of risks associated with extended precipitation extremes (VanBuskirk et al. 2023a). Although improvements have been seen when utilizing post-processing of model output of atmospheric variables, along with precipitation (Lyu et al. 2023; Zhang et al. 2023; Dickinson 2023), no study has quantified the skill of atmospheric variables used that are related to extreme precipitation. Thus this study quantified the ACC and bias of multiple atmospheric variables. The following are the novel results for Chapter 2:

- Model skill of atmospheric variables during extreme precipitation events is low, on average, past Week 1/2, which is similar to overall model skill, although 500 hPa geopotential height has regions of skillful predictions out to Week 3/4.
- Although overall model skill of these variables is similar during events, larger model biases were seen during event periods in 500 hPa geopotential and 850 hPa specific humidity.

Within the midlatitudes, the majority of climatological precipitation is driven by Extratropical cyclones (Kunkel et al. 2012). Composite analysis has been conducted to highlight common atmospheric patterns related to the extended precipitation events (Jennrich et al. 2020; Schroers and Martin 2022; Dickinson 2023), yet no study has quantified the direct impact of ETCs on precipitation totals seen within these events. Thus, ETC activity and precipitation was analyzed and directly common to non-event ETC activity. The following are the novel results for Chapter 3:

- ETCs are present in many events but the characteristics of the event ETCs (frequency, intensity, speed) differ depending on region.
- The percentage of precipitation related to ETC activity is also highly dependent on the region, but also on the season of the event.

Lemos and Morehouse (2005) showed that engagement with end-users throughout the research process can lead to more product adoption. Although previous studies have shown how stakeholders may, or already have, adopted S2S products for differing hazards (White et al. 2017, 2022), none have engaged stakeholders about extended extreme precipitation. Thus we aimed to understand what is needed for S2S decision-making and how uncertainty tolerance may exist on the S2S timescale with regard to extended extreme precipitation. The following are the novel results for Chapter 4:

- End-users desire a range of products that cover spatial, temporal and possible impacts with regard to long-duration precipitation events on S2S lead times.
- There is more uncertainty tolerance for decision-making at S2S lead times, since S2S forecasts would primarily be used as a planning tool.

- Although end-user usability of forecast products were not related to forecast skill, actions taken were greatly decreased with low skill. Thus reducing the utility of the S2S forecast for planning purposes.

Forecasts on the S2S timeframe are being underutilized by stakeholders, as shown in Chapter 4. Therefore, quantifying the skill of S2S models is important to show models strengths and weakness. Yet typical grid based metrics, as used in Chapter 2, can penalize spatial shifts or lower magnitudes in precipitation that end-users have stated are tolerable at S2S lead times, when forecasts are primarily used for planning purposes. For example, case study 1, analyzed in Chapter 2, only had an ACC skill value of 0.38 for precipitation at Week 1/2, which would be seen as unskillful statistically. Yet the actual forecasted precipitation had an accurate shape and location, which given the end-user uncertainty tolerance would be a useful forecast. In the future, using skill metrics that do not penalize spatial shifts, such as object oriented skill, would be a more robust way to analyze skill at the S2S timeframe.

Stakeholders during Workshop 3 favored quantitative precipitation forecasts (QPFs), but S2S model precipitation has very low skill, on average, even with bias correction (McAfee et al. 2023). Therefore in Chapter 2, skill of S2S model's atmospheric variables was assessed due to the common use of these variables within post-processing techniques to improve S2S forecasts, including regime-based forecasting and machine learning models. When targeting forecasts of extended precipitation extremes this post-processing output typically produces probabilistic information that does not retain timing information. This creates an additional challenge for S2S forecasting. Through conversations with stakeholders, it was found that having a suite of products that contain timing, location and potential impacts is best in order to interpret the forecast and have greater trust in the model. Thus forecasting S2S extended extreme precipitation events should have multiple ways to illustrate a regions risk. Given the results shown

in Chapter 3, ETC activity may be another possible avenue to highlight areas with increased risk, while also helping to provide end-users a suite of products.

For ETC activity to be an option for potential S2S forecasting for extended precipitation risk, skill within S2S models must be analyzed. Due to regional differences in ETC activity seen during events, it would be useful to look at all characteristics shown in Chapter 3, including frequency, speed and intensity. This would require a database of tracks from S2S model output. Tracking ETCs within S2S models can be challenging due to temporal, only daily lead times, and spatial resolution, large grid spacing. Thus others have focused on skill of "storminess" (Zheng et al. 2019), but this does not retain information of ETC characteristics. To illustrate an example of potential ETC tracking using S2S model output, a case study of an event that took place in group 4 from 10 to 23 April, 2001 is shown in Figure 5.1. These tracks were also found using Tempest Extremes, but tracked using MSLP instead of variance of MSLP (used for the ERA5 tracks). This event had 91.4% ETC related precipitation, with 5 different ETCs near the event polygon over the 14-day period seen in ERA5 tracks. Although ETC tracks are present at every lead time, no ECMWF member shown captured multiple ETCs near the event polygon, during the 14-days of the observed event. At Week 1/2 there is the least amount of ETC activity near the event, while Week 4/5 had the most. Differences with lead time are very different than the skill found in Chapter 2, thus more work would be useful to have a more thorough understanding of ETC activity within S2S models.

Extreme event research inherently comes with sample size issues. Although the database of events has a long record, S2S models have a short hindcast period that limits the ability to have a large sample size. As S2S models improve, with a longer

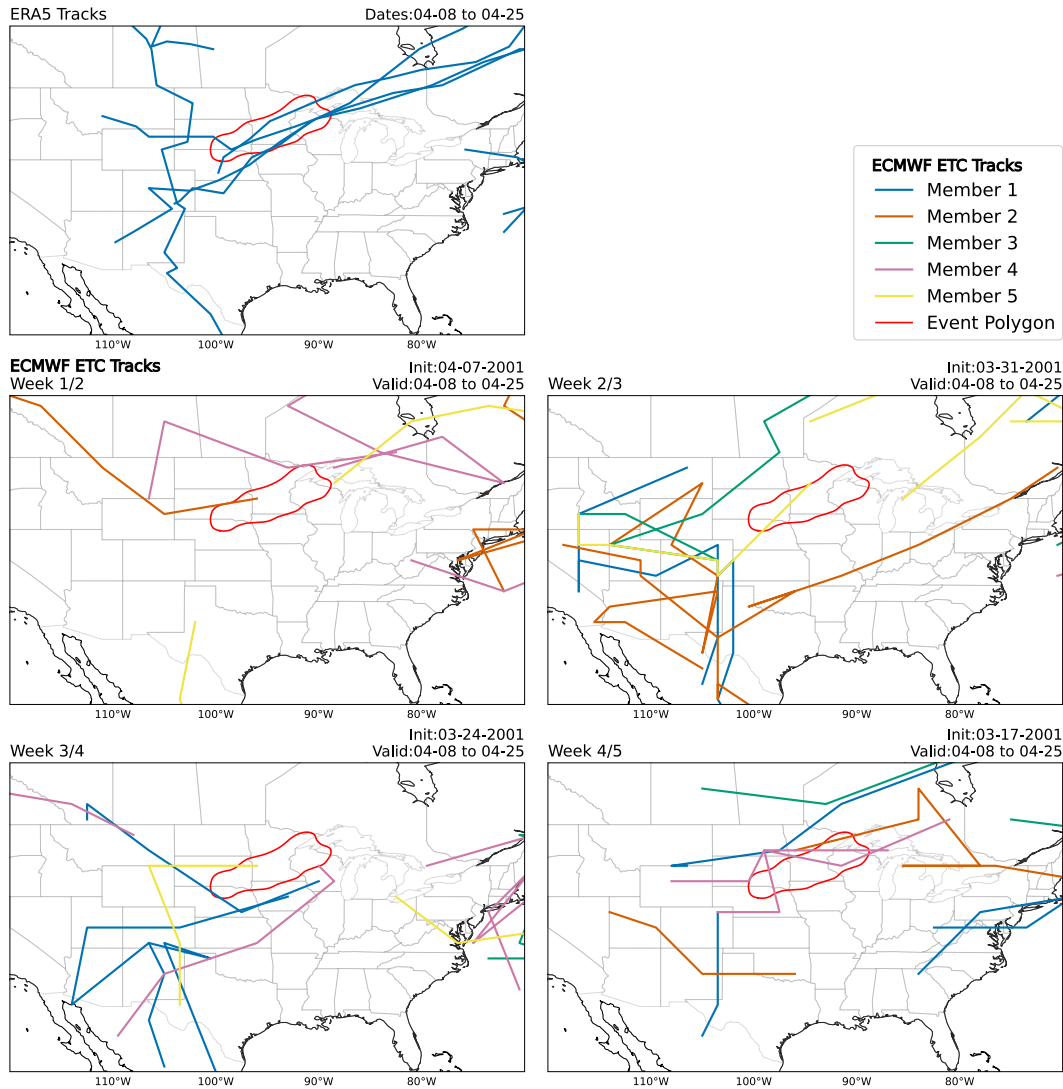


Figure 5.1: Observed tracks from ETC ERA5 track database used in Chapter 3 for one event are shown in the top panel, event polygon shown in red. Tracks from ECMWF, tracked using MSLP, for four different lead times are below. Only 5 members of the 11 members are shown. The event used in this example took place in group 4 from 10 to 23 April, 2001.

hindcast period, more studies that investigate extended extreme precipitation periods large-scale teleconnections, ETC activity, or object oriented skill of event periods, would be a great step toward useful and usable S2S forecast products. During S2S forecast product development, it is important to continue having conversations

with stakeholders so that developed products will be adopted into the decision-making progress (Lemos and Morehouse 2005). Therefore, discussions with a larger group of diverse stakeholders would benefit regarding forecasting of ETC activity would be a necessary step.

The goals of this study was to examine skill and usability S2S predictions of 14-day extreme precipitation events. This was done through analysis of model skill, investigation of potential large-scale drivers and co-production of knowledge with stakeholder communities that would benefit from S2S predictions. Although the work here helps to create a foundation for future S2S forecasts of extended precipitation extremes, there is still much to be done in order to increase S2S model trust and use within the stakeholder communities.

Reference List

- Agel, L., M. Barlow, J.-H. Qian, F. Colby, E. Douglas, and T. Eichler, 2015: Climatology of Daily Precipitation and Extreme Precipitation Events in the Northeast United States. *Journal of Hydrometeorology*, **16** (6), 2537 – 2557, <https://doi.org/10.1175/JHM-D-14-0147.1>.
- Baker, S. A., A. W. Wood, and B. Rajagopalan, 2019: Developing subseasonal to seasonal climate forecast products for hydrology and water management. *Journal of the American Water Resources Association*, **55** (4), 1024–1037, <https://doi.org/https://doi.org/10.1111/1752-1688.12746>.
- Bamzai-Dodson, A., A. E. Cravens, A. A. Wade, and R. A. McPherson, 2021: Engaging with stakeholders to produce actionable science: A framework and guidance. *Weather, Climate, and Society*, **13** (4), 1027–1041, <https://doi.org/10.1175/WCAS-D-21-0046.1>.
- Bao, J.-W., S. A. Michelson, P. J. Neiman, F. M. Ralph, and J. M. Wilczak, 2006: Interpretation of enhanced integrated water vapor bands associated with extratropical cyclones: Their formation and connection to tropical moisture. *Monthly Weather Review*, **134** (4), 1063 – 1080, <https://doi.org/10.1175/MWR3123.1>.
- Barlow, M., and Coauthors, 2019: North American Extreme Precipitation Events and Related Large-scale Meteorological Patterns: a Review of Statistical Methods, Dynamics, Modeling, and Trends. *Climate Dynamics*, **53** (11), 6835–6875, URL <https://doi.org/10.1007/s00382-019-04958-z>.
- Brady, R. X., and A. Spring, 2021: climpred: Verification of weather and climate forecasts. *Journal of Open Source Software*, **6** (59), 2781, <https://doi.org/10.21105/joss.02781>.
- Casteel, M. A., 2016: Communicating increased risk: An empirical investigation of the National Weather Service’s impact-based warnings. *Weather, Climate, and Society*, **8** (3), 219–232, <https://doi.org/10.1175/WCAS-D-15-0044.1>.
- Colle, B. A., J. F. Booth, and E. K. M. Chang, 2015: A Review of Historical and Future Changes of Extratropical Cyclones and Associated Impacts Along the US East Coast. *Curr Clim Change Rep*, **1** (3), 125–143, <https://doi.org/10.1007/s40641-015-0013-7>.
- Curriero, F. C., J. A. Patz, J. B. Rose, and S. Lele, 2001: The association between extreme precipitation and waterborne disease outbreaks in the United States, 1948–1994. *American journal of public health*, **91** (8), 1194–1199, <https://doi.org/10.2105/ajph.91.8.1194>.

- Davis, C., B. Brown, and R. Bullock, 2006: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Monthly Weather Review*, **134** (7), 1772–1784, <https://doi.org/10.1175/MWR3145.1>.
- de Andrade, F. M., C. A. S. Coelho, and I. F. A. Cavalcanti, 2019: Global precipitation hindcast quality assessment of the Subseasonal to Seasonal (S2S) prediction project models. *Clim Dyn*, **52** (9), 5451–5475, <https://doi.org/10.1007/s00382-018-4457-z>.
- Dickinson, T., 2023: Developing a framework for seamless prediction of subseasonal to seasonal extreme precipitation events in the united states. Ph.D. thesis, The University of Oklahoma.
- Dickinson, T. A., M. B. Richman, and J. C. Furtado, 2021: Subseasonal-to-Seasonal Extreme Precipitation Events in the Contiguous United States: Generation of a Database and Climatology. *Journal of Climate*, **34** (18), 7571–7586, <https://doi.org/10.1175/JCLI-D-20-0580.1>.
- Domeisen, D. I. V., and Coauthors, 2022: Advances in the Subseasonal Prediction of Extreme Events: Relevant Case Studies across the Globe. *Bulletin of the American Meteorological Society*, **103** (6), 1473–1501, <https://doi.org/10.1175/BAMS-D-20-0221.1>.
- Feldman, D. R., J. M. Tadić, W. Arnold, and A. Schwarz, 2021: Establishing a range of extreme precipitation estimates in California for planning in the face of climate change. *Journal of Water Resources Planning and Management*, **147** (9), 04021056, [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001410](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001410).
- Ferranti, L., S. Corti, and M. Janousek, 2015: Flow-dependent verification of the ecmwf ensemble over the euro-atlantic sector. *Quarterly Journal of the Royal Meteorological Society*, **141** (688), 916–924, <https://doi.org/https://doi.org/10.1002/qj.2411>.
- Field, P. R., and R. Wood, 2007: Precipitation and cloud structure in midlatitude cyclones. *Journal of Climate*, **20** (2), 233 – 254, <https://doi.org/10.1175/JCLI3998.1>, URL <https://journals.ametsoc.org/view/journals/clim/20/2/jcli3998.1.xml>.
- Froude, L. S. R., L. Bengtsson, and K. I. Hodges, 2007: The prediction of extratropical storm tracks by the ecmwf and ncep ensemble prediction systems. *Monthly Weather Review*, **135** (7), 2545 – 2567, <https://doi.org/10.1175/MWR3422.1>.
- Galarneau, T. J., L. F. Bosart, and R. S. Schumacher, 2010: Predecessor rain events ahead of tropical cyclones. *Monthly Weather Review*, **138** (8), 3272 – 3297, <https://doi.org/10.1175/2010MWR3243.1>.
- García-Franco, J. L., C.-Y. Lee, S. J. Camargo, M. K. Tippett, D. Kim, A. Molod, and Y.-K. Lim, 2023: Climatology of tropical cyclone precipitation in the s2s models. *Weather and Forecasting*, **38** (9), 1759 – 1776, <https://doi.org/10.1175/WAF-D-23-0029.1>.

- Groisman, P. Y., R. W. Knight, and T. R. Karl, 2012: Changes in Intense Precipitation over the Central United States. *Journal of Hydrometeorology*, **13** (1), 47–66.
- Hawcroft, M., E. Walsh, K. Hodges, and G. Zappa, 2018: Significantly increased extreme precipitation expected in Europe and North America from extratropical cyclones. *Environ. Res. Lett.*, **13** (12), 124006, <https://doi.org/10.1088/1748-9326/aaed59>.
- Hawcroft, M. K., L. C. Shaffrey, K. I. Hodges, and H. F. Dacre, 2012: How much Northern Hemisphere precipitation is associated with extratropical cyclones? *Geophysical Research Letters*, **39** (24), <https://doi.org/10.1029/2012GL053866>.
- Hawcroft, M. K., L. C. Shaffrey, K. I. Hodges, and H. F. Dacre, 2016: Can climate models represent the precipitation associated with extratropical cyclones? *Clim Dyn*, **47** (3), 679–695, <https://doi.org/10.1007/s00382-015-2863-z>.
- Jennrich, G. C., J. C. Furtado, J. B. Basara, and E. R. Martin, 2020: Synoptic characteristics of 14-day extreme precipitation events across the united states. *Journal of Climate*, **33** (15), 6423 – 6440, <https://doi.org/https://doi.org/10.1175/JCLI-D-19-0563.1>.
- Karwat, A., C. L. E. Franzke, J. G. Pinto, S.-S. Lee, and R. Blender, 2024: Northern hemisphere extratropical cyclone clustering in era5 reanalysis and the cesm2 large ensemble. *Journal of Climate*, **37** (4), 1347 – 1365, <https://doi.org/10.1175/JCLI-D-23-0160.1>.
- Klemm, T., and R. A. McPherson, 2017: The Development of Seasonal Climate Forecasting for Agricultural Producers. *Agricultural and Forest Meteorology*, **232**, 384–399, <https://doi.org/10.1016/j.agrformet.2016.09.005>.
- Kolachian, R., and B. Saghaian, 2019: Deterministic and probabilistic evaluation of raw and post processed sub-seasonal to seasonal precipitation forecasts in different precipitation regimes. *Theor Appl Climatol*, **137** (1), 1479–1493, <https://doi.org/10.1007/s00704-018-2680-5>.
- Kunkel, K. E., K. Andsager, and D. R. Easterling, 1999: Long-Term Trends in Extreme Precipitation Events over the Conterminous United States and Canada. *Journal of Climate*, **12** (8), 2515–2527.
- Kunkel, K. E., D. R. Easterling, D. A. R. Kristovich, B. Gleason, L. Stoecker, and R. Smith, 2012: Meteorological Causes of the Secular Variations in Observed Extreme Precipitation Events for the Conterminous United States. *Journal of Hydrometeorology*, <https://doi.org/10.1175/JHM-D-11-0108.1>.
- Lemos, M. C., and B. J. Morehouse, 2005: The co-production of science and policy in integrated climate assessments. *Global Environmental Change*, **15** (1), 57–68, <https://doi.org/10.1016/j.gloenvcha.2004.09.004>.

- Li, W., J. Chen, L. Li, H. Chen, B. Liu, C.-Y. Xu, and X. Li, 2019: Evaluation and Bias Correction of S2S Precipitation for Hydrological Extremes. *Journal of Hydrometeorology*, **20** (9), 1887–1906, <https://doi.org/10.1175/JHM-D-19-0042.1>.
- Liu, H., H. Jiang, C. Zhao, and X. Bian, 2022: Long-term responses of high-speed railway subjected to extreme precipitation events. *Transportation Geotechnics*, **37**, 100852, <https://doi.org/10.1016/j.trgeo.2022.100852>.
- Lyu, Y., S. Zhu, X. Zhi, Y. Ji, Y. Fan, and F. Dong, 2023: Improving subseasonal-to-seasonal prediction of summer extreme precipitation over southern china based on a deep learning method. *Geophysical Research Letters*, **50** (24), e2023GL106245, <https://doi.org/https://doi.org/10.1029/2023GL106245>.
- McAfee, D. M., E. R. Martin, and J. C. Furtado, 2023: Bias correction and evaluation of s2s model performance for forecasting us extreme precipitation events. *103rd AMS Annual Meeting*, URL <https://ams.confex.com/ams/103ANNUAL/meetingapp.cgi/Paper/412553>.
- McCrary, R. R., M. Bukovsky, L. Kessenich, C. M. Zarzycki, and S. McGinnis, 2024: Future changes in u.s. extratropical cyclones and their associated mean and extreme precipitation, snowfall, and surface winds. *104th AMS Annual Meeting*, URL <https://ams.confex.com/ams/104ANNUAL/meetingapp.cgi/Paper/439189>.
- McPhillips, L. E., M. Matsler, B. R. Rosenzweig, and Y. Kim, 2021: What is the role of green stormwater infrastructure in managing extreme precipitation events? *Sustainable and Resilient Infrastructure*, **6** (3-4), 133–142, <https://doi.org/10.1080/23789689.2020.1754625>.
- Merz, B., and Coauthors, 2020: Impact Forecasting to Support Emergency Management of Natural Hazards. *Reviews of Geophysics*, **58** (4), <https://doi.org/10.1029/2020RG000704>.
- Morss, R. E., J. K. Lazo, B. G. Brown, H. E. Brooks, P. T. Ganderton, and B. N. Mills, 2008: Societal and economic research and applications for weather forecasts: Priorities for the North American THORPEX program. *Bulletin of the American Meteorological Society*, **89** (3), 335–346, <https://doi.org/10.1175/BAMS-89-3-335>.
- Morss, R. E., O. V. Wilhelmi, M. W. Downton, and E. Gruntfest, 2005: Flood Risk, Uncertainty, and Scientific Information for Decision Making: Lessons from an Interdisciplinary Project. *Bulletin of the American Meteorological Society*, **86** (11), 1593–1602, <https://doi.org/10.1175/BAMS-86-11-1593>.
- Mundhenk, B. D., E. A. Barnes, and E. D. Maloney, 2016: All-season climatology and variability of atmospheric river frequencies over the north pacific. *Journal of Climate*, **29** (13), 4885 – 4903, <https://doi.org/10.1175/JCLI-D-15-0655.1>, URL <https://journals.ametsoc.org/view/journals/clim/29/13/jcli-d-15-0655.1.xml>.

- NCEI, 2023: Storm events database. National Oceanic and Atmospheric Administration. National Oceanic and Atmospheric Administration, URL <https://www.ncdc.noaa.gov/stormevents/>.
- Norström, A. V., and Coauthors, 2020: Principles for Knowledge Co-Production in Sustainability Research. *Nature Sustainability*, **3** (3), 182–190, <https://doi.org/10.1038/s41893-019-0448-2>.
- Nyboer, E. A., and Coauthors, 2021: Supporting Actionable Science for Environmental Policy: Advice for Funding Agencies From Decision Makers. *Frontiers in Conservation Science*, **2**.
- Parrett, C., and R. Hunrichs, 2006: Storms and flooding in california in december 2005 and january 2006-a preliminary assessment. *U.S. Geological Survey Open-File Report 2006-1182*.
- Payne, A. E., and G. Magnusdottir, 2014: Dynamics of landfalling atmospheric rivers over the north pacific in 30 years of merra reanalysis. *Journal of Climate*, **27** (18), 7133 – 7150, <https://doi.org/10.1175/JCLI-D-14-00034.1>, URL <https://journals.ametsoc.org/view/journals/clim/27/18/jcli-d-14-00034.1.xml>.
- Pegion, K., and Coauthors, 2019: The subseasonal experiment (subx): A multimodel subseasonal prediction experiment. *Bulletin of the American Meteorological Society*, **100** (10), 2043 – 2060, <https://doi.org/https://doi.org/10.1175/BAMS-D-18-0270.1>.
- Pfahl, S., and M. Sprenger, 2016: On the relationship between extratropical cyclone precipitation and intensity. *Geophysical Research Letters*, **43** (4), 1752–1758, <https://doi.org/10.1002/2016GL068018>.
- Pfahl, S., and H. Wernli, 2012: Quantifying the Relevance of Cyclones for Precipitation Extremes. *Journal of Climate*, **25** (19), 6770 – 6780, <https://doi.org/10.1175/JCLI-D-11-00705.1>.
- Potter, S. H., P. V. Kreft, P. Milojev, C. Noble, B. Montz, A. Dhellemmes, R. J. Woods, and S. Gauden-Ing, 2018: The influence of impact-based severe weather warnings on risk perceptions and intended protective actions. *International Journal of Disaster Risk Reduction*, **30**, 34–43, <https://doi.org/10.1016/j.ijdr.2018.03.031>.
- PRISM Climate Group, Oregon State University, 2004: URL <http://prism.oregonstate.edu>.
- Pruitt, E. E., T. Dickinson, and M. A. Schroers, 2023: Synoptic analysis of two 14-day extreme precipitation events in the northwestern united states. *103rd AMS Annual Meeting*, URL <https://ams.confex.com/ams/103ANNUAL/meetingapp.cgi/Paper/422586>.

- Rayner, S., D. Lach, and H. Ingram, 2005: Weather Forecasts are for Wimps: Why Water Resource Managers Do Not Use Climate Forecasts. *Climatic Change*, **69** (2), 197–227, <https://doi.org/10.1007/s10584-005-3148-z>.
- Ripberger, J. T., C. L. Silva, H. C. Jenkins-Smith, and M. James, 2015: The influence of consequence-based messages on public responses to tornado warnings. *Bulletin of the American Meteorological Society*, **96** (4), 577–590, <https://doi.org/10.1175/BAMS-D-13-00213.1>.
- Robertson, A. W., N. Vigaud, J. Yuan, and M. K. Tippett, 2020: Toward identifying subseasonal forecasts of opportunity using north american weather regimes. *Monthly Weather Review*, **148** (5), 1861 – 1875, <https://doi.org/10.1175/MWR-D-19-0285.1>.
- Rudeva, I., and S. K. Gulev, 2011: Composite analysis of north atlantic extratropical cyclones in ncep–ncar reanalysis data. *Monthly Weather Review*, **139** (5), 1419 – 1446, <https://doi.org/10.1175/2010MWR3294.1>, URL <https://journals.ametsoc.org/view/journals/mwre/139/5/2010mwr3294.1.xml>.
- Saha, S., and Coauthors, 2014: The NCEP Climate Forecast System Version 2. *Journal of Climate*, **27** (6), 2185–2208, <https://doi.org/10.1175/JCLI-D-12-00823.1>.
- Schroers, M., and E. Martin, 2022: Synoptic connections and impacts of 14-day extreme precipitation events in the United States. *Journal of Applied Meteorology and Climatology*, **61** (7), 877–890, <https://doi.org/10.1175/JAMC-D-21-0174.1>.
- Schroers, M. A., T. A. Dickinson, P. Ćwik, R. A. McPherson, and E. R. Martin, 2024: Listening to Stakeholders III: Potential Users Evaluate Product Content and Design for Subseasonal Extreme Precipitation Forecasts. *Bulletin of the American Meteorological Society*, <https://doi.org/10.1175/BAMS-D-23-0233.1>.
- Son, S.-W., H. Kim, K. Song, S.-W. Kim, P. Martineau, Y.-K. Hyun, and Y. Kim, 2020: Extratropical prediction skill of the subseasonal-to-seasonal (S2S) prediction models. *Journal of Geophysical Research: Atmospheres*, **125** (4), e2019JD031 273, <https://doi.org/https://doi.org/10.1029/2019JD031273>.
- Svoboda, M., and Coauthors, 2002: The Drought Monitor. *Bulletin of the American Meteorological Society*, **83** (8), 1181–1190, <https://doi.org/10.1175/1520-0477-83.8.1181>.
- Tuel, A., and O. Martius, 2023: Weather persistence on sub-seasonal to seasonal timescales: a methodological review. *Earth System Dynamics*, **14** (5), 955–987, <https://doi.org/10.5194/esd-14-955-2023>, URL <https://esd.copernicus.org/articles/14/955/2023/>.
- Ulbrich, U., G. C. Leckebusch, and J. G. Pinto, 2009: Extra-tropical cyclones in the present and future climate: a review. *Theor Appl Climatol*, **96** (1), 117–131, <https://doi.org/10.1007/s00704-008-0083-8>.

- Ullrich, P. A., and C. M. Zarzycki, 2017: TempestExtremes: a framework for scale-insensitive pointwise feature tracking on unstructured grids. *Geoscientific Model Development*, **10** (3), 1069–1090, <https://doi.org/10.5194/gmd-10-1069-2017>.
- Ullrich, P. A., C. M. Zarzycki, E. E. McClenney, M. C. Pinheiro, A. M. Stansfield, and K. A. Reed, 2021: TempestExtremes v2.1: a community framework for feature detection, tracking, and analysis in large datasets. *Geoscientific Model Development*, **14** (8), 5023–5048, <https://doi.org/10.5194/gmd-14-5023-2021>.
- VanBuskirk, O., P. Ćwik, R. A. McPherson, H. Lazrus, E. Martin, C. Kuster, and E. Mullens, 2021: Listening to Stakeholders: Initiating Research on Subseasonal-to-Seasonal Heavy Precipitation Events in the Contiguous United States by First Understanding What Stakeholders Need. *Bulletin of the American Meteorological Society*, **102** (10), E1972–E1986, <https://doi.org/10.1175/BAMS-D-20-0313.1>.
- VanBuskirk, O. G., T. A. Dickinson, M. A. Schroers, R. A. McPherson, and E. R. Martin, 2023a: Listening to Stakeholders II: Adapting Research Products on Subseasonal to Seasonal Heavy Precipitation Events by Exploring Options with Users. *Bulletin of the American Meteorological Society*, **104** (4), E781–E791, <https://doi.org/10.1175/BAMS-D-22-0229.1>.
- VanBuskirk, O. G., R. A. McPherson, and L. E. Mullenbach, 2023b: What floodplain managers want: Using weather and climate information for decision-making. *Weather, Climate, and Society*, **15** (3), 759–772, <https://doi.org/10.1175/WCAS-D-22-0080.1>.
- Vitart, F., and Coauthors, 2017: The Subseasonal to Seasonal (S2S) Prediction Project Database. *Bulletin of the American Meteorological Society*, **98** (1), 163–173, <https://doi.org/https://doi.org/10.1175/BAMS-D-16-0017.1>.
- Vogel, J., E. McNie, and D. Behar, 2016: Co-Producing actionable Science for Water Utilities. *Climate Services*, **2-3**, 30–40, <https://doi.org/10.1016/j.cliser.2016.06.003>.
- Weaver, C. P., R. J. Lempert, C. Brown, J. A. Hall, D. Revell, and D. Sarewitz, 2013: Improving the Contribution of Climate Model Information to Decision Making: The Value and Demands of Robust Decision Frameworks. *WIREs Climate Change*, **4** (1), 39–60, <https://doi.org/10.1002/wcc.202>.
- White, C. J., S. W. Franks, and D. McEvoy, 2015: Using Subseasonal-to-Seasonal (S2S) Extreme Rainfall Forecasts for Extended-Range Flood Prediction in Australia. *Proc. IAHS*, **370**, 229–234, <https://doi.org/piahs-370-229-2015>.
- White, C. J., and Coauthors, 2017: Potential Applications of Subseasonal-to-Seasonal (S2S) Predictions. *Meteorological Applications*, **24** (3), 315–325, <https://doi.org/10.1002/met.1654>.

- White, C. J., and Coauthors, 2022: Advances in the Application and Utility of Subseasonal-to-Seasonal Predictions. *Bulletin of the American Meteorological Society*, **103** (6), 1448–1472, <https://doi.org/10.1175/BAMS-D-20-0224.1>.
- Wilhelmi, O. V., and R. E. Morss, 2013: Integrated Analysis of Societal Vulnerability in an Extreme Precipitation Event: A Fort Collins Case Study. *Environmental Science & Policy*, **26**, 49–62, <https://doi.org/10.1016/j.envsci.2012.07.005>.
- Zhang, L., T. Yang, S. Gao, Y. Hong, Q. Zhang, X. Wen, and C. Cheng, 2023: Improving Subseasonal-to-Seasonal forecasts in predicting the occurrence of extreme precipitation events over the contiguous U.S. using machine learning models. *Atmospheric Research*, **281**, 106 502, <https://doi.org/10.1016/j.atmosres.2022.106502>.
- Zhang, Z., F. M. Ralph, and M. Zheng, 2019: The relationship between extratropical cyclone strength and atmospheric river intensity and position. *Geophysical Research Letters*, **46** (3), 1814–1823, <https://doi.org/https://doi.org/10.1029/2018GL079071>.
- Zheng, C., E. K.-M. Chang, H. Kim, M. Zhang, and W. Wang, 2019: Subseasonal to Seasonal Prediction of Wintertime Northern Hemisphere Extratropical Cyclone Activity by S2S and NMME Models. *Journal of Geophysical Research: Atmospheres*, **124** (22), 12 057–12 077, <https://doi.org/10.1029/2019JD031252>.

1 Appendix A

Bias Distribution Biases

Table A.1: Descriptions of 500hPa geopotential bias distributions

Group	Period	Lead[wks]	Mean	Variance	Skewness	Kurtosis
1	Non-Event	1	7.90	87.8	0.233	-0.853
		4	9.41	184.3	0.639	0.494
	Event	1	6.94	130.4	0.472	-0.164
		4	9.59	442.6	1.042	0.527
2	Non-Event	1	7.91	101.2	0.391	0.026
		4	9.09	227.3	0.800	0.677
	Event	1	7.41	190.7	0.565	0.099
		4	6.37	544.1	0.223	0.382
3	Non-Event	1	7.33	89.1	0.251	-0.054
		4	8.19	195.7	0.845	1.231
	Event	1	7.46	86.3	0.462	-0.231
		4	5.59	260.2	0.962	0.043
4	Non-Event	1	7.36	80.03	0.167	-0.169
		4	8.36	174.0	0.672	0.553
	Event	1	5.30	94.71	0.163	-0.551
		4	6.42	489.9	0.542	-0.648
5	Non-Event	1	8.66	121.3	0.491	0.284
		4	11.04	274.2	0.788	0.867
	Event	1	6.71	123.6	0.603	0.757
		4	4.89	447.8	0.411	0.384
6	Non-Event	1	8.22	108.4	0.448	0.207
		4	9.51	236.9	0.902	1.150
	Event	1	11.09	108.8	-0.060	-0.891
		4	17.18	336.5	0.604	-0.487

Table A.2: Descriptions of 850hPa specific humidity bias distributions

Group	Period	Lead[wks]	Mean	Variance	Skewness	Kurtosis
1	Non-Event	1	-0.202	0.044	-0.217	1.63
		4	-0.299	0.060	-0.046	1.28
	Event	1	-0.356	0.046	0.012	-0.05
		4	-0.704	0.109	0.055	-0.90
2	Non-Event	1	-0.018	0.084	-0.156	-0.33
		4	-0.073	0.133	-0.156	-0.60
	Event	1	-0.245	0.032	-0.246	-0.41
		4	-0.510	0.042	0.382	-0.52
3	Non-Event	1	-0.128	0.115	-0.012	-0.38
		4	-0.200	0.148	0.079	-0.53
	Event	1	-0.339	0.106	-0.464	-0.23
		4	-0.774	0.248	-0.326	-0.56
4	Non-Event	1	0.179	0.024	-0.499	1.39
		4	0.225	0.036	-0.267	1.12
	Event	1	-0.026	0.029	-0.267	-1.03
		4	-0.236	0.184	0.067	-1.32
5	Non-Event	1	0.159	0.028	0.085	0.14
		4	0.219	0.047	0.214	0.086
	Event	1	-0.099	0.031	-0.099	-0.54
		4	-0.429	0.073	0.0075	-0.60
6	Non-Event	1	0.232	0.040	-0.088	0.12
		4	0.280	0.063	-0.123	0.070
	Event	1	-0.030	0.074	0.057	-0.56
		4	-0.073	0.180	-0.285	-0.48