The Most Important Meteorological Processes To The Generation Of Extreme Precipitation Over The CONUS

By

LEIF M SWENSON

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Atmospheric Science

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Richard Grotjahn, Chair

Matthew Igel

Erwan Monier

Committee in Charge

2022

Contents

Abstract

Extreme precipitation (PEx) events have major societal impacts. These events are rare and can have small spatial scale, making statistical analysis difficult. To mitigate these difficulties a methodology was developed to objectively define "coherent" regions wherein data points have matching annual cycles of precipitation. Regions are found by training self-organizing maps (SOMs) on the annual cycle of precipitation for each grid point across the contiguous United States (CONUS). Multiple criteria are applied to identify useful numbers of regions for our future application. This methodology is applicable across data sets and is tested on both reanalysis and gridded observational data. This method of regionalization is then used, in conjunction with two automatic methods of determining the meteorological cause of PEx events, to determine the relationship between mean precipitation seasonality and the different types of PEx events. The first automatic method uses simple metrics, derived from the literature, which are ultimately unable to clearly distinguish between different types of PEx events. The second uses the Quasi-Geostrophic (QG) omega equation to identify fundamental weather patterns associated with different types of PEx events. These weather patterns are identified in a novel way using a SOM trained on a pressure-time series of vertical velocity from each of the advective forcing terms in the QG omega equation for each PEx event. Using the unsupervised learning of the SOM allows for the identification of the most descriptive set of 9 patterns in vertical velocity associated with precipitation extremes in the current climate.

Acknowledgments

I'd like to thank Dr. Richard Grotjahn for his continued support and guidance throughout my time in Davis, both were invaluable to me. To the family and friends that have shared this journey with me, your support kept me going. I'd also like the thank each of my committee members for their time and effort: Paul Ullrich, Erwan Monier, Richard Grotjahn, Matthew Igel, and John Chiang.

Chapter 1 Introduction

1.1 Processes and Extreme Precipitation

Extreme precipitation events are one of the leading causes of weather-related economic losses. This risk has a strongly seasonal component and both the risk as well as the seasonality is known to vary geographically. The frequency of various meteorological phenomena (e.g. midlatitude cyclones, convection, fronts, ...) associated with precipitation extremes also varies geographically and seasonally (Kunkel et al., 2012). Understanding which are the most frequent drivers of extreme precipitation in the current climate is a primary goal of this work.

The work presented in subsequent chapters will allow us to test the, often assumed, central hypothesis of the dissertation: *The seasonality of mean precipitation is predictive of the distribution of extreme precipitation generating mechanisms.* This hypothesis is seen in papers like Kunkel et al. (2012) but never tested. This relationship is thought to exist because extreme events make a large contribution to the mean precipitation during the time of year they occur, and that certain types of extreme events are more likely to occur at certain times of year. In this work we develop the tools to assess both aspects of the hypothesis individually and test their relationship.

In chapter 2 we create a methodology to define regions with similar annual cycles of precipitation. Extreme precipitation is rare by definition, which makes small sample sizes a considerable hurdle in the research process. To address this hurdle we create a process to define coherent regions to aggregate events within, increasing the sample sizes. This process is based on the annual cycle of daily precipitation and helps us understand the geographic variation in precipitation seasonality. These regions are based only on mean daily precipitation and so are not biased by foreknowledge of any kind of extreme precipitation. This will allow us to assess how the seasonality of mean precipitation changes in space, which is the first tool needed to test the central hypothesis of this work.

We will consider many different approaches to identify the meteorological process or processes that give rise to individual precipitation extremes. The work of chapter 3 focuses on simple metrics associated with each process of interest (convection, fronts, and cyclones) that can be calculated efficiently at grid cell experiencing an extreme. These metrics are physically related to the mechanism generating extreme precipitation and are in common use (Dowdy and Catto, 2017; Parfitt et al., 2017). We hypothesize that *These commonly used and physically based metrics can distinguish the primary process causing an extreme event.* Our investigations will show that over a continuous area experiencing extreme precipitation all three of the metrics can be found with noteworthy strength, and that at a single grid cell the primary processes can change up to 5 times in a 24 hour rainy period. Because the commonly used metrics are interrelated and variable in time they are unable to distinguish unambiguously the primary process causing an extreme event. This makes them unsuitable to test our central hypothesis.

It has been found that extreme precipitation can be scaled as the vertical advection of the saturation specific humidity holding saturation equivalent potential temperature constant (O'Gorman and Schneider, 2009). The most basic statement of this scaling consists of only two parts: 1) vertical velocity, and 2) atmospheric temperature. Changes in atmospheric temperature and moisture are well captured by climate models so our work focuses on the dynamical part of extreme precipitation, the vertical velocity. This consideration leads us to adopt the QG omega equation as our primary tool to identify meteorological processes associated with extreme precipitation events. We separately consider the two primary forcings on QG omega to identify the set of 9 most descriptive weather patterns associated with extreme precipitation in the CONUS. This approach naturally focuses on patterns with strong QG forcings. One notable result is the clear identification of a Large-Scale Meteorological Pattern (LSMP) corresponding to a strong pacific northwest atmospheric river from only a local time series of vertical velocity. Another is the identification of the seasonal and geographic variation of which weather pattern is most frequently associated with extreme precipitation. These weather patterns are interpreted and discussed further in chapter 4. This approach makes up the final tool we need to test our central hypothesis. This will be done by assessing the inter-regional differences in the mixtures of weather patterns as well as

the intra-regional similarities, both of which are needed to find whether *The seasonality of mean precipitation is predictive of the distribution of extreme precipitation generating mechanisms.*

This dissertation is presented in 3 main chapters. Chapter 2 deals with the creation and testing of a new methodology to define regions suitable for aggregating extreme precipitation events within (Swenson and Grotjahn, 2019). These regions are the first step in a process assess the predictability of the distribution of meteorological mechanisms behind extreme precipitation from mean precipitation seasonality. Chapter 3 discusses an approach to process identification that we found to be unsuitable for separating events by cause due to the interrelation between the three metrics we tested (Swenson, 2023, In Preparation). Chapter 4 presents an algorithm for the identification of the key meteorological process driving an individual precipitation extreme based on the omega equation. This, combined with the results of chapter 2 allows us to test the central hypothesis that *The seasonality of mean precipitation is predictive of the distribution of extreme precipitation generating mechanisms.* The Omega equation based identification uses commonly available variables in all gridded datasets, can be applied across datasets, and is readily interpretable (Swenson and Grotjahn, 2023, In Review). The final chapter provides a summary of the conclusions.

Chapter 2 Aggregation of Events Within Regions Of Similar Precipitation Seasonality

2.1 Introduction

One of the fundamental problems in researching extreme events is finding a large enough sample size to make the statistics robust. A common way to build sample size is to aggregate events within some geographic area (e.g.,Kunkel et al. (1993); Karl and Knight (1998); Grotjahn and Faure (2008); Kunkel et al. (2012)). Aggregation can be effective for precipitation extremes because precipitation varies across smaller scales than other atmospheric variables (e.g., temperature and pressure) (Hewitson and Crane (2005)).

Other climate regionalizations have been made before, most notably in Karl and Knight (1998), Kottek et al. (2006), and Bukovsky (2011). Previous regionalizations are not suitable to aggregating extreme precipitation events for one or more reasons: the regions are too large, the regions are too discontinuous over mountainous regions, the regions are partly defined from elevation, the regions are defined from combinations of meteorological variables, or the regions are partially based on local vegetation. In addition, these regionalizations are created or modified by subjective factors like consensus among researchers or intuition. Our method uses precipitation solely and the shape and number of the regions result from predefined criteria.

This paper presents an objective way to select geographic areas for grouping extreme precipitation events by training self-organizing maps (SOMs) on the normalized annual cycle of precipitation. Therefore, each region contains points having similar seasonal cycle. The seasonal cycle is normalized so that the regions are not influenced by the size of the total annual precipitation.

Removing the total allows us to find larger scale patterns and not have SOM-based regions that merely show topographic elevation or proximity to the ocean. Seasonality is emphasized because extreme precipitation events and the meteorological drivers behind them can be mainly seasonal (Kunkel et al. (2012)). In some areas wintertime precipitation is almost exclusively caused by frontal systems whereas other areas receive most of their summertime precipitation from convective systems. Training a SOM on the normalized annual cycle of precipitation may contain similar extreme precipitation events occurring at different places in the region (Kunkel et al., 2012). The paper is organized as follows. Section 2 details the different datasets used. Section 3 describes the methodology. Section 4 tests the approaches. Sections 5 and 6 apply the criteria and compare reanalyses. Section 7 compares the maps obtained by our method to other regionalizations in the literature,and concluding remarks are provided in section 8.

2.2 Data

Climate Forecast System Reanalysis (CFSR) precipitation data (Saha and Coauthors (2010a)) are emphasized for training the self organizing map. CFSR uses the Climate Forecast System model (CFS) data to generate a continuous best estimate of the state of the ocean–atmosphere system (Saha and Coauthors (2010b)). CFSR is chosen since it has all the fields we need to create algorithms (in later work) that diagnose the meteorological drivers of precipitation. The CFSR temporal resolution of four times a day will allow our diagnostics to capture the individual process(es) driving each extreme event. CFSR incorporates hourly input data. The CFS model has T382, or approximately 35 km horizontal resolution, using a sigma-pressure hybrid vertical coordinate with 64 levels and a top pressure of; 0.266 hPa. The CFSR has 0.58 resolution in both latitude and longitude and temporal resolution of 6 h. Our time period is 1 January 1979 – 31 December 2010.

As a cross check we also apply our methodology to the Climate Prediction Center's (CPC) unified precipitation data (Chen et al. (2008a)). The CPC data are based on a rain gauge network spanning the conterminous United States (CONUS) and have been interpolated to a latitude–longitude grid using an optimal interpolation objective analysis technique as in Xie et al.

(2007) and Chen et al. (2008b). These gridded precipitation data have a resolution of 0.258 in both latitude and longitude and are recorded daily as an accumulation from1200 UTC of the day before to 1200 UTC of the current day. We use data from 1 January 1950 through 31 December 2018.

2.3 Methodology

The goal is to train a SOM to divide the CONUS into regions with similar precipitation characteristics. These regions are called "coherent regions" in this report. Because extreme precipitation is likely seasonal, we use the annual cycle of precipitation to create our regions. SOMs are a type of artificial neural network first introduced by Kohonen (1982). SOMs utilize a competitive and unsupervised learning algorithm to produce a lower dimensional representation of the input data; in this case almost 6000 average annual cycles, one for each grid point, are grouped into a much more manageable and representative; 15 average annual cycles, one for each region. Each region's annual cycle is the average of the average annual cycle of each grid point within that region. Another feature is that the 15 representative annual cycles are ordered by similarity, that is, the annual cycles of region $K = 1$ and of the highest K value region (e.g., $K = 15$) are the most dissimilar. This feature makes it simple to see the full range of patterns extracted from the input data. It should be noted that the two most dissimilar regions are often adjacent, as seen in Figs. 2.1a–f. This indicates that two decidedly different regimes abut one another. Where this occurs with a jagged or messy border suggests that, perhaps due to interannual variations, the boundary is uncertain. Precipitation has a highly skewed distribution, which makes any measure of the annual cycle very noisy. It is therefore useful to take the cube root of precipitation before creating the measure of the annual cycle as that operation is observed to transform precipitation data to an approximately normal distribution (Stidd (1953)). Processes are described in the next section that were discarded (sections 4a–c) in favor of using the cube root, which will be discussed further in section 4d. All leap days are then removed, and a long-term daily mean (LTDM) is created by averaging all the 1 January data from every year, the 2 January data from every year, and so on. This is done for each grid point individually so the end result is a time series of the cube root of precipitation with 365 values for each grid point. We care most about the timing of precipitation and want the SOM

to be able to easily compare climatologically wet and dry areas. Because of these concerns, each time series is adjusted so that the range of the data is from 0 to 1. The procedure is to subtract the minimum value of each time series from every value in the time series and divide the result by the new maximum of that time series. This normalized annual cycle of precipitation will henceforth be referred to as the LTDM-n. The LTDM-n allows the methodology to compare the occurrence of the wettest days at different locations instead of simply creating regions based on annual rainfall amount.

Without some form of LTDM it is very common to get long strings of zeros in a daily precipitation accumulation time series. These strings create a false similarity between distant, climatologically dry, areas. Taking an LTDM reduces the chances of zero values occurring and separates climatologically dry areas better. An added benefit is that a representative seasonal cycle will be created for each region. Future applications may desire to focus on a specific season (e.g., winter in California) and having seasonality built into the SOMs is advantageous.

To test that the regions created by the SOM are statistically distinguishable, a test based upon the "false discovery rate" (FDR) is used (Johnson, 2013). This test checks if the LTDM-n at each grid point in a region is significantly different than the LTDM-n in every other region at the 5% level. This method provides a useful upper bound on the number of regions we can find across the CONUS. This upper bound is 32 regions for the CFSR data and 63 regions for the CPC data. To find an "optimal" number of regions (K_B) rather than merely rely on the upper bound presented by Johnson (2013), five metrics of four criteria are considered: connectedness, robustness, compactness, and the number of extreme events in each region during the record.

We prefer that each SOM-based region be contiguous, a property we label *connectedness*. Connectedness has two elements. First, we want to minimize the number of separate areas that comprise each SOM region. Second, we prefer each SOM region to be mainly a single larger area and that any other areas be individually and collectively small. To measure this connectedness attribute, two metrics were designed. The first counts the number of isolated areas that belong to each region, where an isolated area is a continuous group of grid points from one single region entirely surrounded by grid points from other regions. This metric for the mean number of isolated ar-

eas composing the regions will be called the isolated area count (IAC). This average number of separate areas per region varies from a minimum of 1 to a maximum of about 10 in CFSR data and about 30 in CPC data. The second measure of connectedness recognizes that a region broken into several small disconnected areas and one large connected area is preferable to a region broken into approximately equally sized disconnected areas. To this end the second metric computes, for each region, the ratio of grid points not belonging to the largest isolated area to the region's total number of grid points. This metric of the fraction of areas in the region that are not part of the largest area will be called the minor areas fraction (MAF). The CONUS average of the MAF is used. We specify that no region may have connectedness metrics above 6 and 0.25, respectively, for the two metrics discussed above.

Another criterion is *robustness*, which refers to how much each SOM region boundary changes when the number of SOM regions allowed is incremented; the less change to existing region boundaries by the addition of another region the better. To measure this quantity, for each grid point we count the other grid points that are in the same region as the particular grid point for a given value of K. If, for $K+1$ regions, any of the counted grid points are no longer in the same region as the particular grid point, we discard them from our count and divide the reduced count by the original count (for K regions). Even if the region were to grow in size, we are only considering grid points that were part of the original count. Hence, 1 is the largest this metric can be. It is expected that regions will shrink when K is increased but it is possible for a region to grow in size. If a particular region grows in size from a map with K regions to a map with $K+1$ regions then robustness can also be one for grid points in that region. However, one or more adjacent regions will have shrunk, thereby lowering the robustness scores of their grid points. The map average of these ratios (i.e., the average of the ratios at all the grid points) is between 0 and 1. A map average near 1 is ideal,with lower values indicating that a map with K regions is not as robust. We specify that a map average ratio be 0.65 for the map to be considered adequately robust.

The last criterion, *compactness*, is intended to foster our eventual goal of compositing events within each region. Long and thin regions are a hindrance to compositing because when a region extends too far in one direction, lining up the origins of particular events becomes problematic. In

each region we calculate the nondimensional ratio (or compactness ratio) of the square root of the total area encompassing the largest connected group of grid points divided by the perimeter of that area to evaluate how compact each region is. We specify that no region may have a compactness ratio below 0.075. For reference a perfectly circular region would have a compactness ratio of; 0.28. The set of $K = 15$ regions has a median compactness ratio of 0.14.

Thresholds introduced are not general. After looking at maps for K values of 2–63 we subjectively decided on these thresholds of 6 and 0.25 for connectedness, 0.65 for robustness, and 0.075 for compactness. The thresholds are a way of establishing minimal qualifications for each criterion.

2.4 Approaches considered to constrain the annual cycle

This study uses the normalized LTDM of the cube root of precipitation, but it is useful to see other approaches considered and rejected. Different methods can find an annual cycle. These include different types of harmonic analysis where more or fewer harmonics are retained depending on the time scale of interest. One could also use a measure of central tendency for each day of the year across all years in the data. Four measures of the annual cycle of precipitation are discussed.

One notes that very robust, connected, and compact regions can be made simply by training a SOM on the full, unprocessed, daily precipitation record at each grid point. For our purposes there are two issues with this method of regionalization. First, using the full record does not emphasize seasonality, which is undesirable for the reasons discussed above. Second, this method yields a precipitation time series for each region that is 32 years long and much less readily interpretable than a measure of the annual cycle. A figure of the regions produced by training a SOM on the full, unprocessed, daily precipitation records is shown in figure Supplement1.

2.4.1 Harmonic LTDM to capture the annual cycle

Because we are most interested in what time of year precipitation tends to occur in a specific area, the logical first step in creating SOM-based precipitation regions is to train the SOM on the annual cycle of precipitation. Prior studies of the annual cycle of temperature (Grotjahn, 2011; Grotjahn and Zhang, 2017) found that a limited set of harmonics captures a smoothly varying LTDM.

Harmonics work because the day-to-day variability remains quite large even when averaging 60 years of data. Given their success in creating a smoothly varying annual cycle by retaining only the first six harmonics of the LTDM of temperature, wind, and geopotential height, we tested a similar filtering on precipitation. This filtering creates a smoothly varying annual cycle of precipitation, here called the harmonic long-term daily mean (HLTDM). Choosing the number of harmonics to retain in order to adequately represent the annual cycle of precipitation is unclear. In Wang and LinHo 12 harmonics are used to capture the onset of the Asian monsoon, but in an earlier paper (Wang and Wang) only 4 harmonics are used. The older paper was not concerned with the precise timing of the onset, but rather with how the intensity changes from year to year.The newer paper sought to identify the time of monsoon onset, which required a more detailed representation of the annual cycle. To choose the number of harmonics here, we observed the average difference between the LTDM and the HLTDM at each grid point versus how many harmonics we kept. This analysis led us to use six harmonics because after the sixth harmonic the additional reduction in difference between the LTDM and the HLTDM, in our map average, becomes small relative to the reduction in difference gained when adding each of the first six harmonics.

Maps created using six harmonics (Figs. 2.1a,d) were unsatisfactory in that they consistently failed our tests of connectedness by exceeding the thresholds mentioned above. The most prominent cause of disconnectedness is a persistent link between Wyoming and Illinois, which belong to disconnected parts of the same region. This is illustrated in Fig. 2.1d for region 8 where southern Illinois and eastern Missouri are part of the same region as most of Wyoming. The similarity between the HLTDMs of the geographically separate areas of Wyoming and Illinois, among other disconnected regions, indicates that gross seasonality is not enough to define coherent precipitation regions. Sub seasonal variations should play a stronger role in identifying our regions if we desire connectedness, especially on smaller scales. With that in mind we experimented with keeping more harmonics when constructing our HLTDM and we did find that the more harmonics that were kept, the better the map scored in each of our criteria. However, part of that better score is coming from large variations on very short time scales still present after averaging the data over the full period of record, whereas we prefer matching the broader seasonal cycle.

2.4.2 Long-term daily median of precipitation

To retain day-to-day variation, we considered constructing a time series from a central tendency of the precipitation distribution for each day of the year for each grid point. Because precipitation is known to have a very skewed distribution (Ison et al.) we tested the median as our central tendency. Unfortunately,there are multiple locations in Arizona and New Mexico where the median precipitation is 0 for every day of the year. This does not reflect the seasonality in precipitation that exists in those areas and makes the long-term daily median a poor choice to represent the annual cycle of precipitation.

2.4.3 LTDM of precipitation with and without a 3-day running mean

Since a long-term daily median is a poor choice to measure the annual cycle we trained SOMs on the LTDM of precipitation at each grid point (Figs. 2.1b,e).This LTDM is much noisier and less intuitive than a HLTDM or long-term daily median but does avoid the issues discussed in the previous two subsections. This trade off also creates regions that have better connectedness than regions made using a HLTDM or long-term daily median but there are still areas of significant disconnectedness. Even averaging 32 years of data, there remains significant variation from one day to the next in the LTDM. Large daily variations are problematic so we attempt to soften them with a simple nonrecursive smoother so we used a simple nonrecursive smoother in time on the raw precipitation data before creating the LTDM. The smoothed value for day X equals one-quarter of day $X-1$ plus one-half of day X plus one-quarter of day $X+1$. A LTDM is then created from the smoothed data at each grid point and used to train the SOM. The smoothing altered the shapes of the SOM regions, but the connectedness metric was not sufficiently improved. Disconnected regions moved, but their number was not appreciably reduced. The smoothing had insufficient benefit to any of the four criteria overall.

2.4.4 Working with the cube root of daily precipitation data

It is well known that precipitation data fit a gamma distribution (Ison et al.) better than a normal distribution. Therefore, transforming the gamma distributed precipitation data to a more normal distribution by taking the cube root of precipitation (Stidd (1953)) was tested. This operation has a much larger effect on large precipitation values than on small precipitation values, effectively reducing the impact of extreme data. This reduction is valuable because we want the SOMs to be based on the seasonal cycle, which facilitates coherence in space and has links to extreme precipitation mechanisms mentioned earlier. Additionally, we found that even the LTDMs of neighboring grid points had large differences in their peak values. Taking the cube root of the raw precipitation data deemphasizes the spatial variation of these peak values. This method still retains the seasonal cycle of precipitation and incorporates a reduced form of subseasonal variation leading to maps that consistently score better in all four criteria than the maps based on any of the methods discussed above. For low K values the map using the cube root of precipitation performs marginally better than the map without taking a cube root.The superiority of this method over using the LTDM without taking the cube root grows as K increases and is easily seen for maps with higher values of K (Fig. 2.1).

2.5 Results

2.5.1 Criteria values as a function of K

When making composites we increase the number of events sampled by expanding the area we aggregate over. Because we are analyzing extreme events here we use the raw (i.e., no normalization or taking of the cube root) precipitation data. We define an extreme event to be a 6-h period that exceeds the 95^{th} percentile for precipitation accumulation at a particular grid point, after discarding time steps with zero precipitation. For each region we count the number of time steps when any grid point reports an extreme as a "regional event." Creating a threshold for the number of these regional events that each region must have would necessarily base the threshold on the driest (fewest precipitation periods) region found by the SOM. Instead, we want each region to aggregate more

extreme events than could be found from any single grid point within it. To this end we calculate a regional extremes ratio (RER). The RER is the number of time periods when a point somewhere in the region exceeds the 95^{th} percentile divided by the number of time periods when there is rain somewhere in the region. Since RER is a ratio based on exceeding the 95^{th} percentile, then the RER for an individual grid point would be essentially 0.05 or 5%. We would like our SOM regions to capture at least 4 times as many extreme events as an individual grid point would,so we apply a RER threshold of 0.2 or 20% criterion to our analysis. The RER allows comparison of extremely dry and extremely wet areas of CONUS more than would a fixed number of events. This ratio also allows intercomparison of different datasets having differing periods of data, grid intervals, and/or time intervals. Figure 2.2 shows the relationship in CFSR data between K and RER in the region with the lowest RER along with the 20% threshold. From Fig. 2.2 it is clear that using more regions means that each of those regions aggregate fewer events and asymptotically approach the value (RER = 5%) for a single grid point. This simple RER threshold indicates that values of $K > 15$ are eliminated from further consideration in the CFSR data.

However, Figs. 2.3a–c show that as the number of regions is increased the compactness ratio and IAC improve dramatically and MAF sees modest improvement in the worst region's ratio. But increasing the number of regions decreases the number of events in each region by decreasing the average areal extent of each region over which we can aggregate events. Therefore, a balance is sought between two competing goals. The first goal is to create regions that are large enough to have meaningfully more events to aggregate compared to considering a single grid point. The second goal is to create regions that attain high scores in compactness and low (better) scores in IAC and MAF. Creating a map with fewer regions helps the first goal while creating a map with more regions helps the second goal.

Of the K values that pass the event threshold described (2–15) the values that do not meet our robustness criteria (values < 0.65 in Fig. 2.3d) are discarded. The remaining values ($K = 3$, 5–13, and 15) are ranked by their median score in compactness ratio, and their worst region's scores in MAF and IAC. The worst region's score was judged to be more important than the median for the purpose of selecting the optimal value of K due to the lack of variation in the median scores of

MAF and IAC. In the case of IAC this was particularly true for $K > 10$. The "optimal" K value with the best average rank is 15 in this paper. One notes that $K = 12$ also does very well in this comparison and is the most compact of all the maps considered.

An odd characteristic is a persistent link between the Florida (FL) peninsula and the New Mexico/Texas (NM/TX) border area, especially when K is small ($K < 10$). At first glance this seems very strange because peninsular Florida has a warm, humid, tropical climate whereas the New Mexico and Texas border area is arid (Kottek et al. (2006)). This linkage also shows up in the CPC dataset, which is based on observations, ruling out model error as causing this unexpected similarity. Our procedure to normalize the data at each grid point causes the grouping of New Mexico with Florida. Both areas share a late summer relative peak in precipitation and, while Florida is much wetter, the normalization of precipitation magnitudes causes Florida's humid maritime seasonal cycle to closely match New Mexico's arid continental cycle. For compositing events, having similarly large spatially separated areas grouped together is very undesirable. We design a process to identify and label as separate regions large, spatially separate areas that the SOM analysis assigns to one region. This process is discussed further in section 6.

2.5.2 Variations within the CONUS of the annual cycle

Figure 2.4 illustrates how the normalized annual precipitation cycle varies across the CONUS; the spatial variation is similar for smaller and larger numbers of regions. The $K = 12$ regions version is emphasized because it has the most compact regions of all maps that were considered ($K = 2$–15). All along the west coast the wet season starts in November and ends around May or June. Regions 10, 11, and 12 are then primarily differentiated by the relative strength of their dry seasons compared to the median value of their annual cycle.This strength decreases as we move from north to south. In the southwest, region 1 encompasses most of Arizona and New Mexico; it has both a weak wintertime wet season (especially January–March) and much stronger late summer to early fall wet season (July–September). Hence, this area gets some wintertime precipitation,probably from Pacific storm tracks, but is more influenced by the North American monsoon (NAM; Adams and Comrie) in the late summer. The large late summer peak in precipitation makes the annual cycle

here similar to the annual cycle in the Florida peninsula as discussed above. The northern Great Plains are dominated by wintertime precipitation with a fairly strong dry season from July into November (region 9). Moving south through the Great Plains we see a gradual flattening of the annual cycle down into south Texas (regions 6, 5, and 4). In the southeast and inland from Florida, region 3 has a weak dry period (October–December) leading into a winter and spring that are near the annual mean precipitation rate with a slightly wetter summer. The northeast (region 7) is fairly wet until August and then becomes wet again in November. Eight of the twelve regions in this map are to some degree wintertime dominated when it comes to precipitation.

2.6 Comparison between CFSR and CPC

When our criteria with respect to K are examined to find the optimal value of K for SOMs trained on the CPC data, broadly similar results are found, with a few key differences (Figs. 2.5 and 2.6). Comparing the regions with the fewest unique events, for a given value of K, in the CPC and CFSR-based maps (not shown) shows that the CPC-based maps have far fewer unique events. If the region with the lowest RER, for a given value of K, in the CPC-based maps is compared to the region with the lowest RER, for the same value of K, in the CFSR-based maps, they will have a very similar value of RER (comparing Figs. 2.5 and 2.2). So, even though the different regions have different numbers of unique events due to the temporal and spatial sampling and length of the data, the RER is comparable, as intended. The median compactness ratio of the CPC-based maps generally increases with K apart from a dip into lower values for $K = 8$ and 9 (Fig. 2.6a). Unlike CFSR-based maps, the worst region's compactness ratio generally decreases with increasing K for $K > 5$. As the number of regions increases both the median and worst region's IAC decreases steadily (Fig. 2.6b). The worst region in the CPC-based maps exhibits this trend much more strongly than does the worst region in the CFSR-based maps. The median region's MAF is very comparable to that of the CFSR-based maps for $K > 5$, both having small values that vary little. The worst region's MAF in the CPC-based maps is fairly constant between 0.4 and 0.5 except for $K = 2$, 3, and 5 where it falls below 0.3 (Fig. 2.6c). This is very different from the CFSR-based maps, which have low values (~ 0.1) from $K = 12$ through $K = 19$ and generally higher values

(0.3–0.4) elsewhere (Fig. 2.3c).The relationship between robustness and K of the CPC-based maps is similar to the CFSR-based maps but with overall higher robustness and different local peaks (Fig. 2.6d).

To find the optimal value of K for the CPC-based maps we used the same 20% threshold of regional events divided by periods with nonzero precipitation. Applying this threshold requires $K < 17$, excluding $K = 14$. This threshold is similar to the CFSR value despite the large difference in the upper bounds provided by the false discovery rate for the CFSR-based maps compared to the CPC-based maps. Again, this is because RER is a far more limiting factor than FDR.

One does not expect the metrics presented here to find the same optimal value of K for both datasets, but one expects the values of K to be comparable. The median compactness is lowest for $K = 8$ and 9. The worst compactness ratio is highest for $K = 2$–4 and 6. The worst compactness ratios show a decreasing trend with increasing K, which is opposite from the CFSR data. Additionally, the compactness ratios found for CPC data area bout half as large as those found for CFSR data. The median region's IAC is very noisy for $K = 2$–10 with $K = 10$ being the largest value. For $K > 10$ the median region's IAC is much smaller. The CPC IAC is generally larger than for the CFSR data. As with IAC, the median region's MAF is small for $K > 5$. The worst region's MAF is lowest (i.e., best) for $K = 3$ and 5 while the other values are higher, between 0.4 and 0.5. The CFSR MAF values are similar except for the worst regions from $K = 12$–19, where they are around half as much. The robustness of CPC-based maps is highest (i.e., best) for $K = 8$, 11, and 15. Compared to CFSR-based maps the CPC-based map shave smaller differences between the most and least robust maps. By these considerations CPC-based maps with $K = 15$ regions perform best.

In CPC data, $K = 15$ is not large enough to have Florida (FL) separate from the New Mexico–Texas (NM/TX) border region through the SOM analysis; instead, $K \geq 41$ is required for this to happen. However, maps with ≥ 41 regions have far fewer events per region than ideal. Again, the link between FL and the NM/TX border region is caused by those separate areas experiencing their respective wet seasons at nearly the same time of year. Because of the distance between FL and NM, the relatively similar sizes of these two subregions, the region they both belong to should

be separated into two distinct regions for analysis and making of composites. This approach is recommended for all regions with a minor area that is greater than a quarter the size of the major area; for reference, the FL part of region 4 is 0.41 the size of the NM/TX part in Fig. 2.7a. The FL part of region 5 is 0.53 the size of the NM/TX part in Fig. 2.7c. In the analysis shown in this paper, the only regions that would meet this requirement encompass FL and NM/TX and the separation results in a map with 12 and 16 regions and is shown in Figs. 2.7b and 2.7d. One notes that this threshold does identify other regions for separation at high values of K($\geq$ 20). For some of these maps Michigan (MI) is its own region and the two parts of MI are disconnected from one another.

2.7 Comparison between SOM and other regionalizations

Karl and Knight (1998) used nine regions to analyze precipitation trends over the CONUS based on combining entire states. State boundaries do not necessarily correspond to meteorological "boundaries" or climatological zones. Nonetheless, their choice of regions has been popular so it is important to see how their regions compare with our SOM-based regions. Their map (Fig. 2.8a) is constrained by state boundaries, and therefore scores very well in our connectedness metrics.The Karl and Knight map also does a fairly good job of grouping areas by their annual cycle of precipitation in many parts of the CONUS, although this was not their stated goal. To compare the Karl and Knight map quantitatively to the SOM method, we created a SOM with nine regions (see Fig. 2.8c). Both Figs. 2.8c and 2.8e were created with eight regions, and the ninth region comes from separating Florida from the New Mexico–Texas border region as described in the previous section. We want a quantitative measure of how much each grid point is like the rest of the grid points in its region in terms of its annual cycle. For each of Figs. 2.8a and 2.8c we calculate the root-mean-square difference (RMSD) between the LTDM-n at each grid point and the mean of each other grid point's LTDM-n within the region the original grid point belonged to. The RMSD is found by taking the squared difference in LTDM-n at each day between one grid point and the average value of the LTDM-n within the region on that day, then taking the average of all days. Small values of this RMSD indicate that a specific grid point is very much like the rest of its region,

and large values indicate a grid point that is quite different. The result for each map is plotted in Figs. 2.8b and 2.8d. We do not believe $K = 9$ creates regions that are small enough to confidently aggregate events within. This is because the $K = 9$ map has a smaller compactness ratio compared to a map with $K = 15$ (Fig. 2.3a). Nonetheless, Karl and Knight use nine regions so we make our comparison using nine SOM-based regions. Even using a small ($K = 9$) number of regions, the SOM regions have grid points with a more consistent annual cycle than do the Karl and Knight regions. While we see a number of areas with particularly poor consistency in Karl and Knight's regions, namely Mississippi, western Montana, and the Four Corners states, we also see that their choice performed very well in the Pacific Northwest.

The North American Regional Climate Change Assessment Program (NARCCAP; Mearns et al. (2012)) uses regions outlined in Bukovsky (2011) to capture North American regional climatology; we refer to these regions as the Bukovsky regions. For a more direct comparison to our regions, we have plotted only those 17 Bukovsky regions that exist over the CONUS (Fig. 2.9b). These Bukovsky regions closely follow those used by the National Ecological Observatory Network(NEON) put forward in Kampe et al. (2010). These regions are based on a statistical analysis of nine ecoclimate state variables, including temperature, precipitation, and solar insolation. One should not expect them to match exactly our regions since ours are based on normalized precipitation only. To compare our regions to the Bukovsky regions, we again match the number of regions, now $K = 17$ in Fig. 2.9. For each grid point we find the fraction of grid points that were in the same region as the target grid point under the Bukovsky regions compared to how many are still in the same region as the target grid point under the SOM regions, similar to our measure of robustness. This fraction varies between 0 and 1, with values of 1 indicating more agreement between the two sets of regions; the result of this calculation is shown in Fig. 2.9c. The map average result is less than 0.5, which means the average grid point is in a region with more than half of the grid points being different when we compare the Bukovsky regions to our SOM generated regions from CFSR data. While the maps, overall, are not very similar, one recognizes that there are areas of fairly good agreement. This is notable because of the differing ways the regions were produced.The areas of agreement are shown by bluer hues in Fig. 2.9c. The places where agreement is good are

similar to those found in comparison with Karl and Knight (1998), like the Pacific Northwest, New England, Southern California, parts of the Great Plains, southern Michigan, and Florida. Overall, the method used by Bukovsky produces a map that is quite different to the one produced by the SOM method. That is expected because the two maps were designed with different purposes: Bukovsky's purpose being to create regions sensitive to changes in temperature and precipitation to aid in North American climate change assessment, while ours is to group areas purely by the annual cycle of precipitation. It is nevertheless encouraging to find some commonalities between these regionalizations. The same comparison is shown between the Bukovsky regions and SOM regions based on CPC data (Figs. 2.9d,e). Figure 2.9d was created from a SOM with 16 regions with the 17^{th} region created by manually separating Florida from the NM/TX border region. The map average is again below 0.5 and is quite similar to Fig. 2.9c. The CPC and CFSR differences from the Bukovsky regions are generally similar; a couple of exceptions are for Florida and the very northernmost part of California.

2.8 Conclusions

This paper introduced a new and objective way to select contiguous geographic areas that experience precipitation at similar times of the year. Our purpose in devising this methodology is to identify regions for compositing information about extreme events. We have devised our method to focus on the timing of precipitation during the year because the timing is a key factor in determining which meteorological processes are primarily causing the precipitation. Our method creates "coherent" regions based only on the normalized annual cycle of precipitation, or LTDM-n. These regions are compact enough to aggregate extreme precipitation events within each region with confidence that the events are similar. The directness of interpretation of the LTDM-ns that represent each region makes comparisons to other regionalizations or datasets straightforward and easily quantifiable. Comparisons to other regionalizations are also aided by flexibility in the number of regions.

In an effort to optimize the choice for K (the number of SOM regions), six criteria were developed to measure four aspects of the precipitation regions. We first tested the statistical distinguishability,

at the 5% level, of each pair of regions using the false discovery rate to reveal the upper bound of the number of regions that could be created. This produced upper limits of 32 regions for CFSR data and 63 regions for CPC data. We next applied a criterion called the regional extremes ratio,whose 20% threshold means all the SOM regions must have at least 4 times as many time periods with at least one grid point reporting precipitation exceeding the 95^{th} percentile than would occur at a single grid point. In short, this RER criterion requires every region to provide notably more events for aggregation than if a single grid point was used. Also, RER facilitates comparison across datasets with different resolution and length. We created two criteria to measure connectedness because we want to track two distinct aspects of connectedness.The first aspect, measured by IAC, is how many groups of grid points are disconnected from the largest group in a region. The second aspect, measured by MAF, is how large are the disconnected groups relative to each other and to the largest group. Both criteria's map averages (i.e., the average over all K regions in a particular map) have a strong inverse trend with K. As more regions are added, the average region becomes more connected. The trend in the least connected region as K increases is much flatter. This indicates that there are only a few regions per map that have substantial values of IAC and MAF. These regions usually occur in mountainous or desert regions of the Southwest. Robustness refers to how much a map is altered by changing the value of K. The relationship between robustness and K does not exhibit a clear trend and is better described as a saw tooth function. For one to several consecutive values of K, robustness is relatively low and then for one to several K values it is relatively high. Compactness is measured by the ratio of the square root of the area of a region divided by the perimeter of that region. The map average of compactness tends to increase (improve) as the value of K increases.

Our methodology was applied to CFSR and CPC precipitation data. The criteria provide a guide for choosing an optimal value for K, but in practice there were additional problems. A persistent problem is that fairly high values of K are needed in our tested data to decouple an area including the Florida peninsula from an area near the NM/TX border that are placed in the same region by the SOM algorithm. In practice, one might either opt for a larger K or if that is not feasible (i.e., if the number of events within the regions becomes too small) then manually intervene

and treat the two areas as separate regions in later applications, as shown in Fig. 2.7.

For these CFSR data, $K = 15$ is optimal and the map has several notable features and quite compact regions (Fig. 2.1f). Both the East and Gulf Coasts are broken into three different regions (although they do share region 2) while the West Coast is divided into only two regions. For these CPC data, $K = 15$ is optimal and after separating the Florida and NM/TX border areas yields a map with 16 regions (Fig. 2.7d). In comparing the CPC map with 16 regions to the CFSR map with 15 regions the most obvious difference is in the Pacific Northwest. Regions 12 and 13 in the CFSR are largely combined into region 15 in the CPC data. Florida and the Gulf Coast are in separate regions for the CPC but the same region for the CFSR. Another notable difference is that throughout the Great Plains the lines between regions tend to run from north to south or from east to west in the CFSR map while in the CPC map they are diagonally oriented from southwest to northeast.

Maps created from CFSR and from CPC data were compared to regions described in Karl and Knight (1998) (Fig. 2.8). All three maps with nine regions performed least well in our RMSD test in similar areas of the country. Substantial differences between the local annual cycle and the regional annual cycle were found along the Rocky Mountains, in the Southwest deserts and along a line drawn from east Texas to Michigan. These differences have lower magnitudes in the two SOMs than in the Karl and Knight regions. All three maps did relatively well in the Pacific Northwest and parts of the High Plains. The two SOMs did much better in Florida, while the Karl and Knight regions did better in New York State.

Maps created from CFSR and CPC data were compared to the Bukovsky regions (Fig. 2.9). Both SOMs agree with the Bukovsky regions in the north and south parts of the West Coast as well as east Texas and New England. The SOMs and the Bukovsky regions disagree more strongly over much of the Great Basin, western mountains, and central plains. The CFSR map agrees with Bukovsky in Florida but not in other parts of the Southeast; this relationship is reversed between the CPC and Bukovsky regions. Overall, the agreements and disagreements between the CPC and Bukovsky regions are both stronger in magnitude than those between the CFSR regions and the Bukovsky regions.

This method can be applied elsewhere in the world and the size of the regions is tunable by changing the number of regions (K) specified in the SOM. This flexibility allows the technique be used to investigate problems over different spatial scales. Here our interest is in large-scale meteorology but in principle one could find a dense observational network and examine how precipitation varies within an individual watershed.

These regions are the first step in a process to automate identification of the primary meteorological mechanism or mechanisms creating an extreme precipitation event. This SOM-based approach can be used to assess how well climate models capture the spatial changes in the annual cycle of precipitation.

Acknowledgements U.S. Unified Precipitation data are provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado,USA,from their Website at https://www.esrl.noaa.gov/psd/. CFSR data are provided by the DOC/NOAA/NWS/NCEP/EMC on their Web site at https://rda.ucar.edu/datasets/ds093.0/. We also acknowledge funding from U.S. DOE Office of Science Award DE-SC0016605, and USDA NIFA Hatch project Ac-cession 1010971.

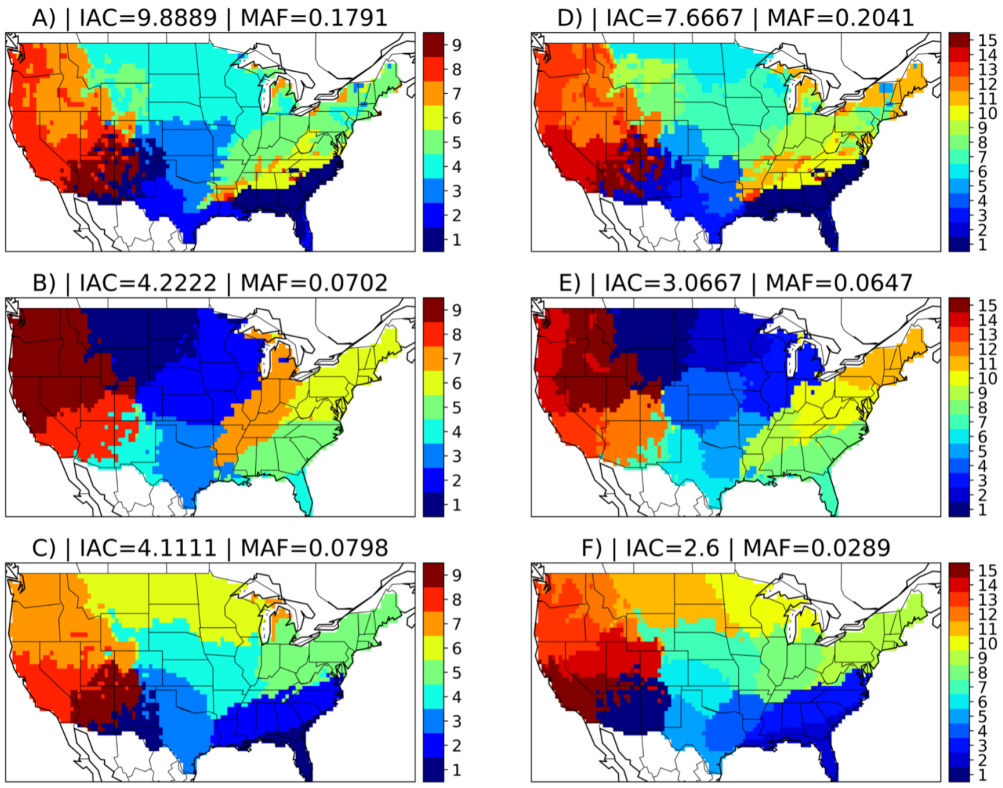

Figure 2.1: SOM regions created from A) The first 6 harmonics of the long-term daily mean (LTDM) of precipitation at each grid point, $K = 9$. B) The LTDM of precipitation, adjusted to vary from 0 to 1, at each grid point where $K = 9$. C) The LTDM of the cube root of precipitation, adjusted to vary from 0 to 1, at each grid point where $K = 9$. D) Same as in A) but $K = 15$. E) Same as in B) but $K = 15$. F) Same as in C) but $K = 15$. IAC is the mean value of the isolated area count and MAF is the mean value of the minor areas fraction. Smaller values are preferable for both IAC and MAF.

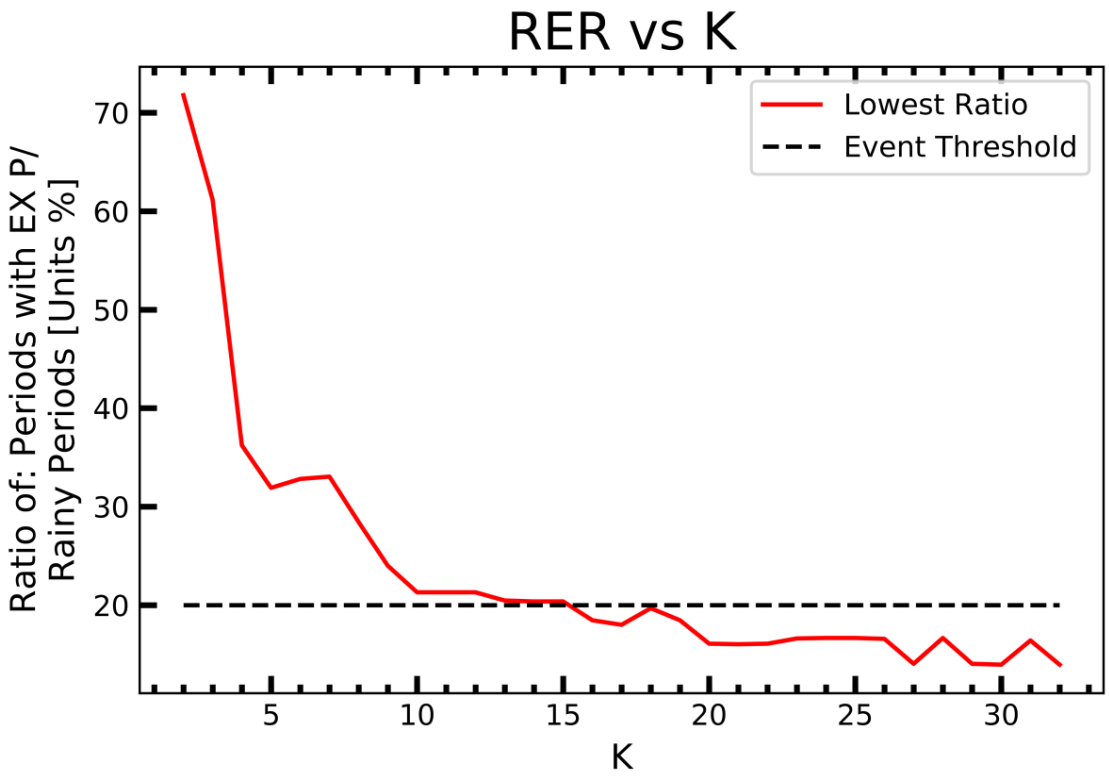

Figure 2.2: The ratio of periods with extreme precipitation ($> 95\%$ value) to periods with nonzero precipitation (RER) in the CFSR data contained within the region with the lowest RER for each value of K is shown in red. Our threshold of 20% (meaning four times as many periods with an extreme somewhere in the region as occur at a single grid point) is shown by the dashed line.

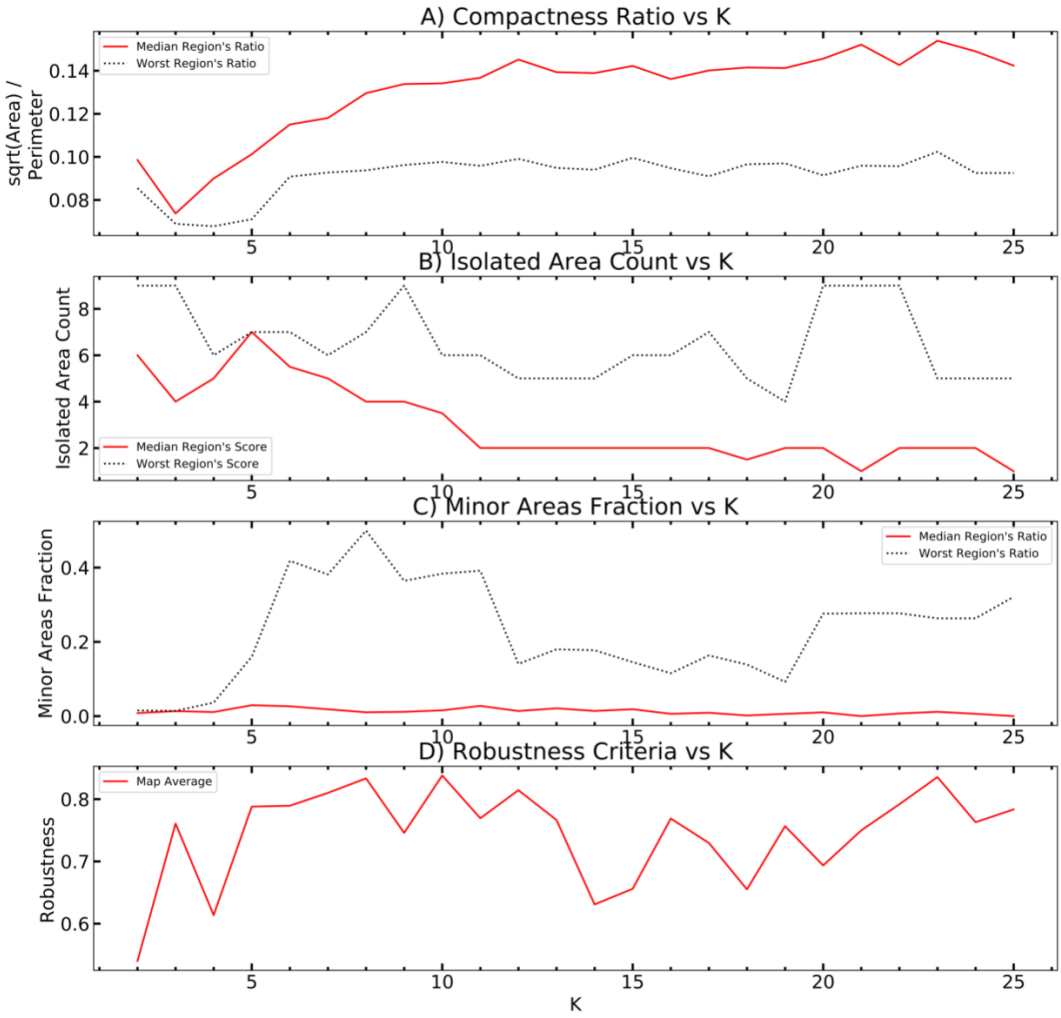

Figure 2.3: Performance criteria for CFSR based SOM regions created from the LTDM of the cube root of precipitation, adjusted to vary from 0 to 1, at each grid point. A) Number of regions vs compactness ratio. The dashed curve is the least compact region and the red curve is the median region's compactness ratio. B) Number of regions versus isolated area count. The dashed curve is the worst performing region and the red curve is the median isolated area count of the regions . C) Number of regions versus the minor areas fraction. The dashed curve is the worst preforming region and the red curve is the median minor areas fraction of the regions. D) Number of regions versus robustness. The red curve shows the map average for robustness for the shown values of K. Preferred metric values are higher in A) and D), and lower in B) and C).

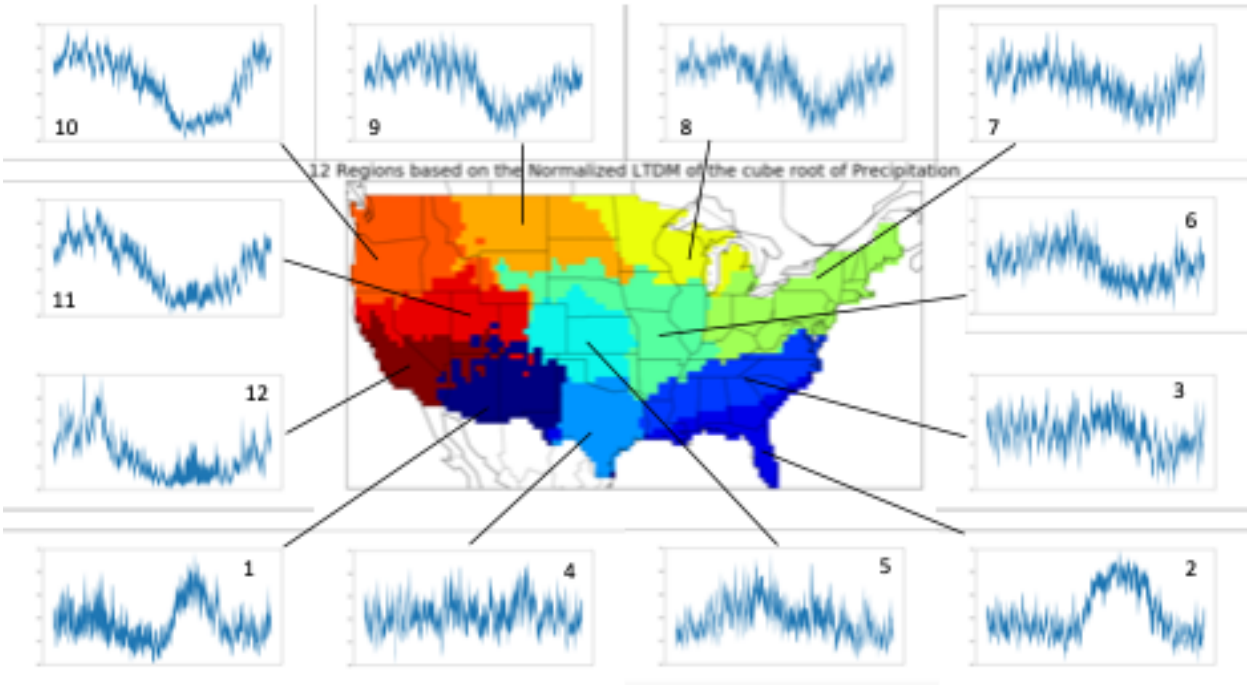

Figure 2.4: SOM regions with $K = 12$ created from the LTDM of the cube root of precipitation, adjusted to vary from 0 to 1, at each gridpoint. The surrounding plots show the representative LTDM of the cube root of precipitation, adjusted to vary from 0 to 1, at each gridpoint. The middle 50% of grid points in each region are contained in the shaded area of each subplot. The subplots each begin at 1 January on the left and end on 31 December on the right. The value $K = 12$ is chosen because it separates the NM/TX border region from Florida in these CFSR data and has the highest compactness ratio of all maps considered ($K = 2$–15).

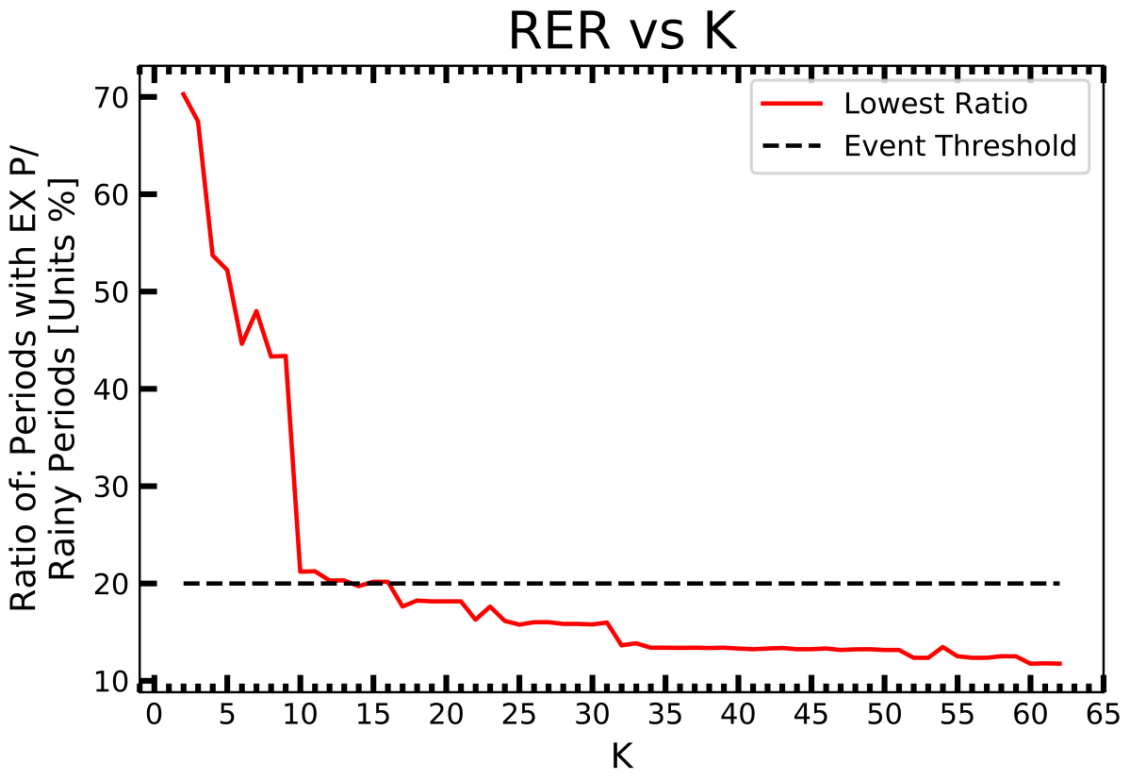

Figure 2.5: Same as Figure 2.2 except for CPC data.

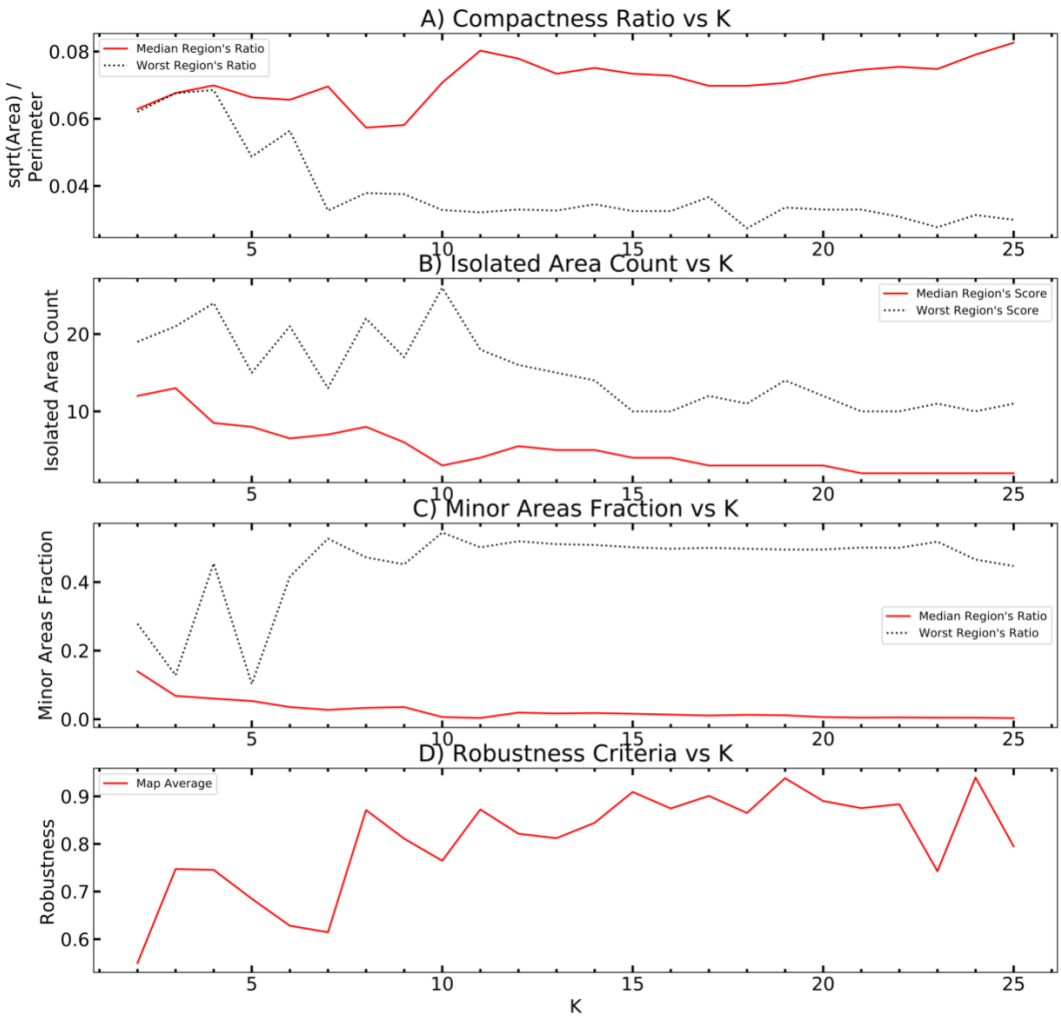

Figure 2.6: Similar to Figure 2.3 except for CPC based SOM regions.

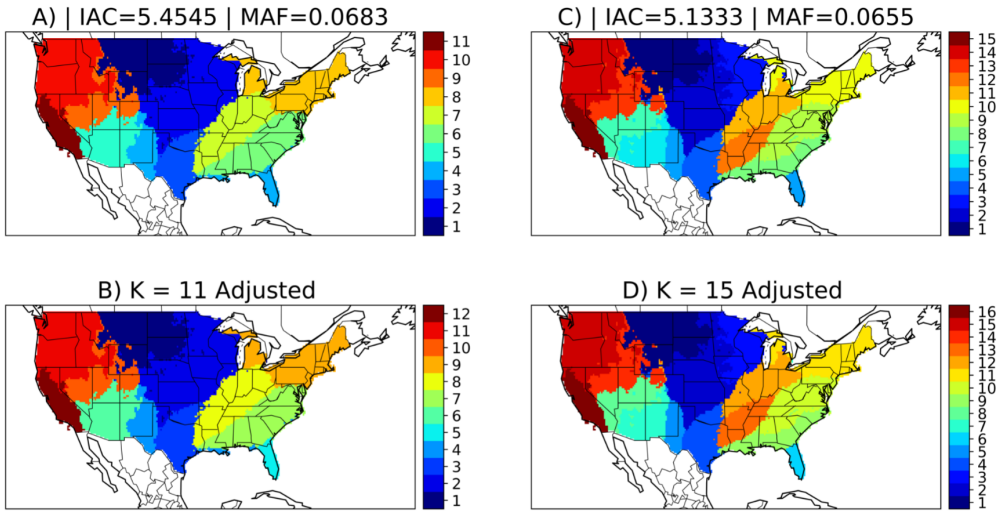

Figure 2.7: SOMs regions from the CPC data. A) $K = 11$, B) $K = 11$ again but with region 4 separated into two distinct regions to create 12 regions, C) $K = 15$, and D) $K = 15$ again but with region 5 separated into two distinct regions to create 16 regions. IAC is the mean value of the isolated area count and MAF is the mean value of the minor areas fraction. These CPC data are created from the LTDM of the cube root of precipitation, adjusted to vary from 0 to 1, at each grid point, same as was done for CFSR data.

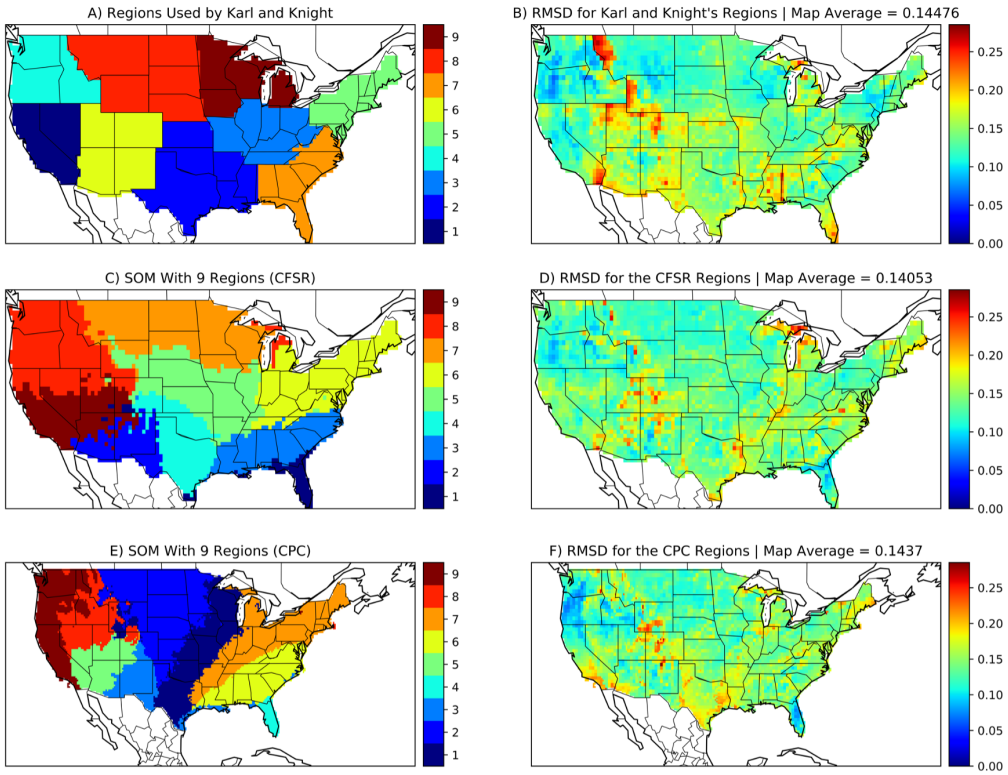

Figure 2.8: A) Our representation of the nine regions used by Karl and Knight in their 1998 paper. B) The root mean squared difference (RMSD), described in section 5b, of the regions in A). C) SOM regions created from the LTDM of the cube root of precipitation (CFSR Data), adjusted to vary from 0 to 1, $K = 9$ to match the number of regions in A). D) as in B) but for the regions shown in C). E) As in C) but for CPC data. F) as in D) but for the regions shown in E). Note regions in E) are calculated from CPC data but the RMSD is calculated with CFSR data to be comparable to B) and D). For both C) and E) FL was manually separated from the NM/TX border region to form the 9^{th} region. This does not substantially affect the results shown in D) and F). Smaller RMSD is desirable.

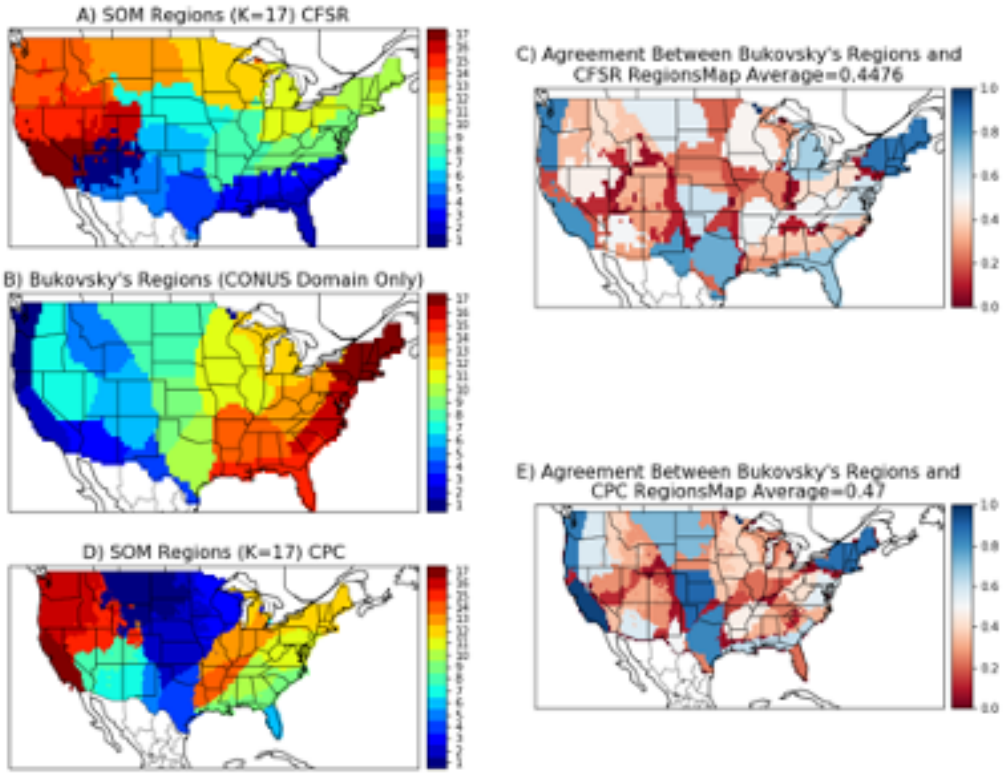

Figure 2.9: A) SOM regions created from the LTDM of the cube root of precipitation (CFSR data), adjusted to vary from 0 to 1, $K = 17$ to match the number of regions in B). B) Bukovsky regions plotted over the CONUS domain. C) The agreement between the maps in A) and B) using the robustness method described in section 5b except both maps being compared have 17 regions. Hence, larger (darker blue) values mean greater agreement and smaller (darker red) values mean less agreement. D) As in A) but for CPC data. E) As in C) but comparing maps B) and D).

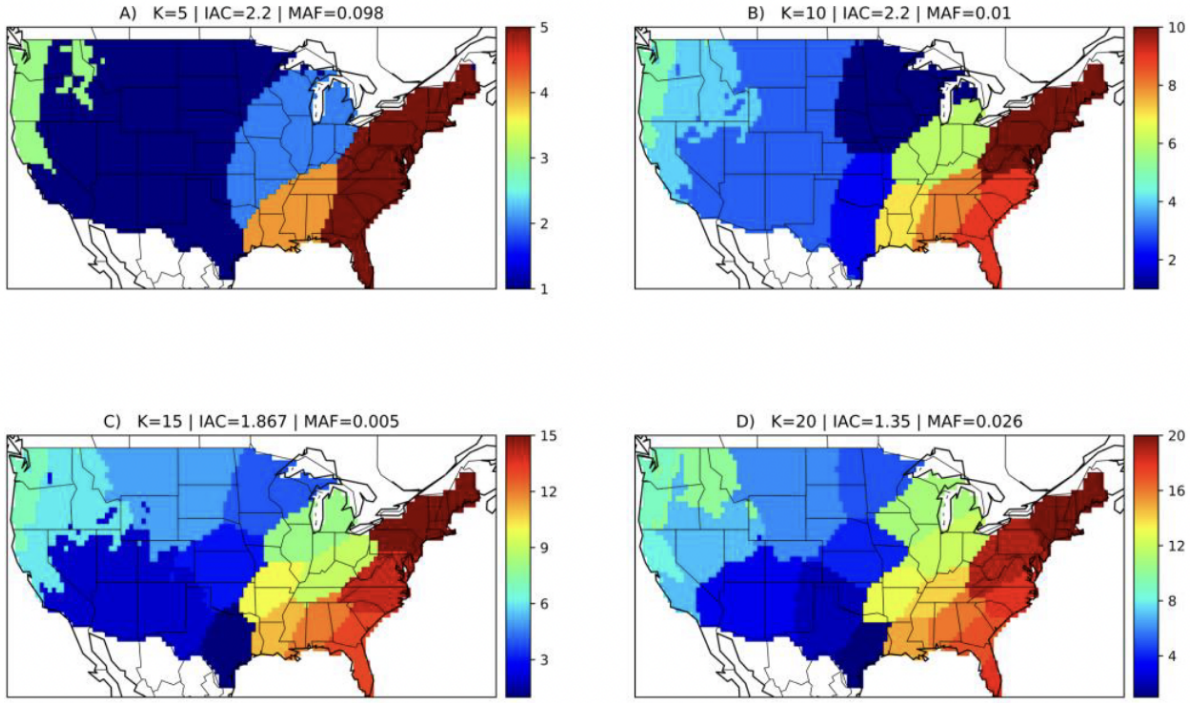

Figure Supplement1: SOMs regions from the full, unprocessed, CFSR daily precipitation record. A) K=5, B) K=10, C) K=15, and D) K=20. IAC is the mean value of the isolated area count and MAF is the mean value of the minor areas fraction.

Chapter 3 Simple Metrics to Identify the Process(es) Driving Extreme Precipitation

3.1 Introduction

In this chapter we will try using metrics based on the literature to test the central hypothesis that *The seasonality of mean precipitation is predictive of the distribution of extreme precipitation generating mechanisms.* The goal of this chapter is to develop an algorithm to identity the extreme precipitation generating mechanisms creating individual extremes. This algorithm will be based on previous such attempts and should be able to distinguish the primary process causing an extreme event. To assess this ability we will compare our algorithm to several individual cases from Kunkel et al. (2012) as well as the seasonal trends from Kunkel et al. (2012) and Dowdy and Catto (2017).

Section 2 provides a background of previous detection and classification algorithms. Section 3 discusses the methods used in this chapter. Section 4 presents the results, which are discussed in more detail in section 5, and section 6 presents conclusions.

3.2 Background

The importance of meteorological process that create extreme precipitation (e.g. fronts, convection, and extratropical cyclones) is known to vary seasonally and spatially (Kunkel et al., 2012). This knowledge is often linked to the seasonality of mean precipitation. Here we will test the hypothesis that *the seasonality of mean precipitation is predictive of the distribution of extreme precipitation generating mechanisms.* To do this we need to determine the primary meteorological cause of individual precipitation extremes. There have been numerous case studies where these

meteorological processes (fronts, convection, and extratropical cyclones) have been identified for individual extreme events. That procedure is done by hand and can be quite labor intensive. This labor can become a prohibitive barrier to create a climatology of the driving meteorological process behind precipitation events. There has been at least one notable attempt to do this from observed data (Kunkel et al., 2012), which required hand analysis of thousands of surface weather maps. This type of analysis is not feasible to repeat for even a single climate model run let alone a full ensemble. In this chapter our main objective is to create a set of algorithms to identify meteorological processes, which cause individual extreme precipitation events, from model data. The focus is on three main processes: fronts, convection, and extratropical cyclones.

3.2.1 Frontal Identification

Objective frontal analysis dates back at least as far as (Renard and Clarke, 1965) and debate on the precise definition of a front has persisted since the concept was introduced by Bjerknes (1919). Definitions include: gradient in air temperature, surface humidity gradient, leading edge of temperature advection, abrupt shift in wind direction, a change in the air's origin, and warm air side of the gradients in air temperature and low level humidity (Sanders et al., 2002). Objective frontal detection schemes typically focus on: a wind shift (Simmonds et al., 2012), a temperature gradient (Mills, 2005), a combination of moisture and temperature gradients (Catto et al., 2012; Clarke and Renard, 1966; Hewson, 1998) or a combination of temperature gradient and vorticity (Parfitt et al., 2017). The vertical levels these schemes are applied to are application dependent and vary from surface fields to around 850 hPa. Some schemes look at upper level fronts (600 hPa) but they are invariably secondary to the analysis. The schemes of each type are broadly similar but tuned to the specific needs of those using them. The specifics of one scheme of each type will be discussed in greater detail below.

Wind Shifts

Simmonds et al. (2012) looks for winds to shift from the northwest quadrant to from the southwest quadrant and for the change in meridional velocity to be greater than 2 m/s over a 6-hour interval.

This scheme is designed to look at southern hemisphere fronts so the direction of the wind shifts would need to be reversed for application to the northern hemisphere. A grid point is flagged as a front at the end of a 6-hour interval that meets these criteria. Adjacent frontal grid points are considered to be a frontal object and single unconnected frontal grid points are discarded. The front is then found to be the smoothed eastern edge of each frontal object. For each front the angle (relative to the meridian), center of gravity, length, and intensity is recorded. The intensity is the sum of all the changes in meridional velocity along the length of the front normalized for different spatial and temporal scales. This scheme cannot detect stationary or very slow moving fronts, both of which can create flash flood events (Maddox, 1979). Although useful in detecting and tracking mobile fronts, especially over water, and possessing a natural grid point based measure of intensity; the emphasis on 10m winds and inability to detect slow-moving or stationary fronts are severe drawbacks.

Temperature Gradient Only

Mills (2005) use the gradient of air temperature at 850 hPa to measure frontal strength with the goal of connecting strong pre and post frontal winds to major fire events in Australia. The author thought that if the temperature gradient was strong at 850 hPa that indicated a deeper tropospheric structure and strong associated winds. To connect this to precipitation, one would expect a strong thermal gradient and winds to create strong advection, generating ascent and precipitation (if the conditions are moist enough). This work differs significantly from many other frontal analyses in that it does not actually attempt to locate a front, although it can be adapted for that purpose (Hope et al., 2013). Instead, the goal is to find the area the front impacts with strong winds and measure the intensity of this impact. This is a very useful, impacts driven, frontal detection scheme. The primary limitation is that it cannot discern fronts arising from a sharp moisture gradient between two air masses, as commonly arises from the “Dry-line” in the central U.S.

Thermal Frontal Parameter

Catto and Pfahl (2013) use a thermal frontal parameter (TFP) defined by Renard and Clarke (1965) to identify fronts for the purpose of linking them to precipitation events. This TFP is defined, for some scaler variable τ, as the derivative of the magnitude of the gradient of τ in the direction of the gradient of τ (equation 3.1).

$$TFP = -\nabla|\nabla\tau| \cdot \frac{\nabla\tau}{|\nabla\tau|} \tag{3.1}$$

In Catto and Pfahl (2013) τ is the wet bulb potential temperature at 850 hPa. Points are identified where the TFP is negative and the gradient of the TFP is zero and those points are linked together to form fronts. This is done on a very coarse grid (2.5° resolution) to reduce the influence numerical noise can have on the detection of the location of a front (Simmonds et al., 2012). To associate precipitation (also on a 2. 5° resolution) with a front the grid box experiencing the precipitation and the surrounding 8 grid boxes are searched for an instance of a front at the beginning and end of the 6-hour period of precipitation accumulation. If a front is found then the precipitation event is associated with that front. Catto and Pfahl (2013) note that misallocation of precipitation is possible with all automated methods due to the area of influence of a front being substantially larger than the front itself. This method is also applied to daily precipitation data and misallocation is slightly higher due to the possibility of a short-lived rainstorm occurring when no front was present, however a front could pass through earlier or later in the day. In that case the precipitation would be flagged as frontal.

Vorticity And Temperature Gradient

Parfitt et al. (2017) develop a two variable method for frontal detection. This method combines a thermal variable, the magnitude of the temperature gradient, with the relative vorticity seen in equation form below (equation 3.2).

$$F = \eta_p|\nabla(T_p)| \tag{3.2}$$

This quantity is divided by the Coriolis parameter and a baseline temperature gradient of $\frac{0.45K}{100km}$ to normalize the metric. Grid points with values exceeding 1 at 900 hPa are considered frontal. This threshold must be found via case study analysis at each pressure level to be considered, which is a substantial drawback. Additionally, the authors only test their metric over oceans, an ideal place for automated detection because it completely removes topography. This is not an option afforded the analysis of fronts appearing over land.

3.2.2 Convective Identification

Thunderstorms and intense convection can often give rise to heavy and localized precipitation. Dowdy and Catto (2017) use a network of ground based lightning detectors (World Wide Lightning Location Network: WWLLN). Grid cells with at least two lightning strikes during the 6-hour precipitation time period of the study (Dowdy and Catto, 2017) are considered to be convective. This threshold provides an indication of the presence of a deep convective storm colocated with the precipitation.

Synoptic cloud observations are used to distinguish convective precipitation from stratiform precipitation by Berg et al. (2013). The authors classify precipitation as convective if observations find cumulus or cumulonimbus clouds during the period of precipitation.

In their large scale manual identification of meteorological causes of extreme precipitation, Kunkel et al. (2012) have two classes of convective event, Air Mass Convection (AMC) and Mesoscale Convective Systems (MCS). Each type was characterized by a convectively unstable vertical temperature profile. An event was classified as AMC if the precipitation is isolated (exists at an isolated grid cell or pair of cells). AMC events were also typically found in warm areas and times of year. The MCS type often needed to be separated from their frontal category because MCSs often spawn along frontal boundaries before separating. These MCSs are characterized by moderate southerly winds, sometimes lacking anomalously warm temperatures. Even so, MCS was often assigned as the category if no other category was appropriate.

The first two schemes rely on observational networks which limit the applicability of techniques to reanalysis data. But each of these schemes seeks signs of existing vertical instability associated

with an extreme event to classify the event as convective.

3.2.3 Vortical Identification

Vortical features are often investigated through the study of extratropical cyclones. Many detection schemes have been implemented to find and track these features, despite no universally accepted definition (Neu et al., 2013). These tracking schemes usually use one or more of these four variables: mean sea level pressure (MSLP) (Benestad and Chen, 2006; Hewson and Titley, 2010; Murray and Simmonds, 1991), upper level vorticity (Hewson and Titley, 2010; Murray and Simmonds, 1991; Serreze, 1995), lower level vorticity (Inatsu, 2009; Sinclair, 1994, 1997), and geopotential height at 850hPa (Inatsu, 2009; Sinclair, 1994, 1997). These methods use differing levels of terrain filtering depending on their geographic area of study or their method's reliability over topography. As with frontal identification, the misallocation of precipitation is possible by these automated methods due to the area of influence of an extratropical cyclone being substantially larger than the cyclone itself.

3.3 Methods

3.3.1 Data

This study uses the fifth version of the European Center for Medium-Range Weather Forecasting's (ECMWF) atmospheric reanalysis (ERA5) (Hersbach et al., 2020). All ERA5 data used in this study, was retrieved from the Climate Data Store (CDS) using their application programming interface (API) as detailed here. This interface allows us to regrid this data set to one degree resolution in the horizontal and three hours resolution in time. These resolutions are attainable by regridding and/or data are readily available at these or finer resolutions from many GCMs/reanalysis outputs, making them a good starting point for this analysis. We defined an extreme event to be the top 5% of nonzero 3 hour precipitation accumulation at each grid cell. This temporal resolution is short enough that we can match instantaneous measurements of process strength to the accumulated precipitation. We have chosen a period from 1980-2010 to diagnose the climatology of PEx drivers

in the current climate.

3.3.2 Metrics for the Identification of Processes

For each of the three processes of interest (fronts, convection, and cyclones) we create a simple and easy to calculate metric to measure its strength. These metrics take inspiration from previous work and attempt to build on their commonalities. A significant difference between this and the approaches outlined above is that we want to quantify the contribution of each process at the grid point level. This desire motivated our choice of metrics.

Frontal Intensity

To identify the presence and quantify the strength of *fronts* associated with the precipitation extreme event we use the absolute value of the horizontal gradient of equivalent potential temperature at 850 hPa (equation 3.3).

$$Frontal\ Metric = |\nabla_p \theta_e| \tag{3.3}$$

This metric accounts for both temperature and moisture gradients as precipitation can arise from sharp gradients of both kinds between air masses (Renard and Clarke, 1965). Because PEx can arise distant from the maximum thermal gradient we take the largest score from among neighboring grid points. In order to more easily the match the front to the 3-hour precipitation accumulation we also take the largest score from between the beginning and end of the accumulation period.

Convective Intensity

Convection arises from static instability in the atmosphere and can generate very high precipitation rates over a relatively small area. We use the change in Convective Available Potential Energy (CAPE; equation 3.4) during the precipitation event to measure the severity of the convection. Our metric uses CAPE measured for an air parcel lifted from the surface.

$$Convection\ Metric = \frac{\partial}{\partial t} CAPE \tag{3.4}$$

This metric measures the change in the stability of the column to convective motions during the event. Only negative values (decreases in CAPE) are considered an indication of convective activity. This was inspired by discussions on the nature of convection and CAPE in the context of statistical equilibrium by Rennó and Ingersoll (1996). Particularly the idea that the energy available for convection is proportional to the amount of CAPE that is present locally. It is also in line with the three schemes discussed previously in that it focuses on a sign of vertical instability, in this case CAPE.

Vortical Intensity

To assess the impact of cyclones and strong *vortical* features we calculate the vorticity advection at 500 hPa (equation 3.5).

$$Vorticity\ Metric = -U \cdot \nabla_p \zeta \tag{3.5}$$

Positive values of this metric are associated with increased upper level vorticity. This could be because the presence of a midlatitude cyclone. Alternatively, a less notable vortical feature may be present which could still drive ascending motion through the column during the precipitation extreme. Similar to the frontal metric we take the largest score from among neighboring grid points and from between the beginning and end of the precipitation accumulation period.

Scaling The Intensities

Because each of these metrics has different units, comparing their relative importance to a particular event carries some challenges. We considered several approaches to make the metrics intercomparable. The primary one being, placing each metric on a scale relative to the usual strength of the metric in a particular area. For this approach we created a 0-100 scale based on the local percentile rank of each metric. Above average frontal metric scores, negative convection metric scores, and positive vorticity metric scores are retained for each grid point. Other scores are set to 0 because they are judged to be either the wrong sign to directly generate precipitation (in the case of convection and vorticity) or unremarkable (in the case of below average frontal scores).

We judged that a below average thermal gradient (including non-precipitating time periods) was unlikely to be the direct cause of an extreme precipitation event. The retained scores are then ranked on a scale from 1-100 at each grid point.

The relative scales are created based on every value of the field in the data. This purposefully includes times with non-extreme, and even zero, precipitation. This choice allows the separation of the rarity of each process from whether or not large amounts of moisture were present. When regional and seasonal statistics are discussed in the following sections the events considered are every three hour period with precipitation accumulation greater than the local 95^{th} percentile of those periods with non-zero precipitation.

3.3.3 Self-Organizing Map Regionalization

Recall that our central hypothesis is that *The seasonality of mean precipitation is predictive of the distribution of extreme precipitation generating mechanisms.* To test this we use a self-organizing map (SOM) to identify regions that share the same mean precipitation seasonality (Swenson and Grotjahn, 2019). This is very useful because it allows us to match a set of extreme events (those occurring within a particular region) to the seasonal cycle of mean precipitation within that region. The raw output from the SOM is shown in figure 3.1, as well as it's isolated area count (the median number of isolated areas comprising a single region: IAC), minor areas fraction (the median fraction of a region's area that is not contained in the largest isolated area: MAF), and compactness ratio (the median ratio of the square root of a region's area to its perimeter) which all compare well to the scores in the results of Swenson and Grotjahn (2019). Additionally this set of regions passes the regional extremes ratio (RER) threshold of 20% set out in Swenson and Grotjahn (2019) using this chapter's three hour definition of a precipitation extreme. Because of the lower resolution used in this work we have to create maps using less regions than in Swenson and Grotjahn (2019). This results in Florida (FL) not being separated from the New Mexico / Texas border region naturally. This necessitates the use of the automatic intervention to separate large enough isolated areas grouped together by the SOM discussed in Swenson and Grotjahn (2019) to separate FL into a 7^{th} region. Further discussion of the link between the seasonality of precipitation in FL and the New

Mexico / Texas border region can be found in Swenson and Grotjahn (2019). We also remove the borders between regions from the analysis to reduce the uncertainty in the seasonal cycle in each region. The final regional map is shown in figure 3.2. This figure (3.2) does not display the scores because IAC and MAF are 1 and 0 respectively because of the processing while the compactness ratio increases. After processing, the RER is the only criterion likely to be negatively effected and the threshold is still passed for the processed regions.

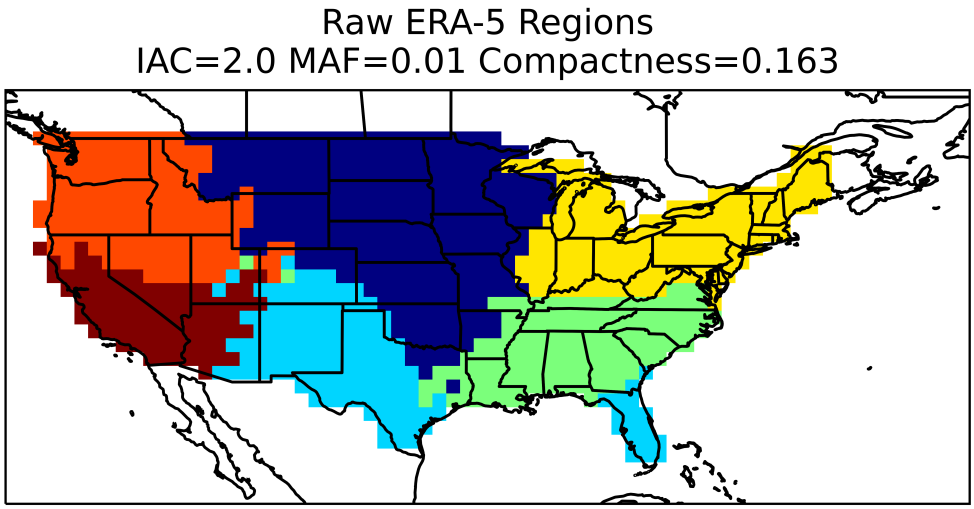

Figure 3.1: The raw output of the SOM trained on the normalized long term daily mean of the cube root of precipitation at each grid point as in Swenson and Grotjahn (2019). IAC and MAF refer to the map's median isolated area count and minor areas fraction respectively.

3.4 Results

3.4.1 Multivariate Metric Distributions

The simplest method to investigate the processes, as described previously, is to examine the various bivariate distributions of the metrics. An example of this is shown in figure 3.3. This example displays data from the Southeast region (figure 3.2) during summertime extreme events. There we see no extreme events with a score between between 5 and 50 in the convective metric (upper left panel). The overwhelming majority of the events in this 0-4 bin are events with a value of zero. This indicates that, for these extreme events, convection is either above average strength or not existent. This is in contrast with the other two metrics shown in this figure, each of which show a relatively flat distribution from 5-100 with the most frequent score between 0 and 4. The bivariate distributions show convection is about as likely to be "on" in the presence of a front as it is in the presence of vorticity (comparing the distributions of convection vs frontal to convection vs vorticity). Looking at the center bottom panel, when fronts and vorticity are mixed this occurs more frequently with higher vorticity scores and lower front scores than vice versa.

In contrast during wintertime the mix of processes (figure 3.4) associated with extreme precipitation is much different. Far fewer are associated with high convection scores but the gap in the convection score distribution remains (upper left panel of figure 3.4). The nonzero scores in the vorticity and frontal metrics are no longer flat, but have many more events at higher scores. The quadrant where Vorticity and Fronts are both relatively weak is the most sparsely populated with extreme events (middle right panel of figure 3.4).

As a comparison let's look at a region that experiences most of its precipitation during wintertime, like the Pacific Northwest (orange in figure 3.2). In summertime (figure 3.5) the distributions are very comparable to those of the southeast (figure 3.3) with an increased frequency of high scores in the vorticity and frontal metrics. But in wintertime there is a decrease in the relative strengths of the frontal metric scores compared to the vorticity metric scores (figure 3.6 that is not seen in the southeast. This is somewhat surprising because fronts are discussed regularly in relation to precipitation in this region. However previous studies disagree on whether the primary

driver of winter extreme precipitation here is extratropical cyclones (Kunkel et al., 2012) or fronts (Dowdy and Catto, 2017). Although the two processes are interrelated (extratropical cyclones commonly develop strong fronts (Bjerknes, 1919)) this study finds unusually high levels of vorticity advection (which points to cyclonic activity) more commonly associated with wintertime extreme precipitation than strong thermal gradients in the Pacific Northwest.

3.4.2 Categorical Comparison

For our three metrics we create 8 categories: Front, Vorticity, Convection, Front & Vorticity, Front & Convection, Vorticity & Convection, All, and None. Only about 1% of events fall into the "None" category. This is beneficial in that these metrics capture most extremes but can also be interpreted as a sign that the metrics are too broad. A process is excluded as a factor for an event if it has either a score of 0 or a score that is 20 less than the process with the highest score. Events fall into the "None" category if all three metrics have a score of zero and into the "All" category if each metric is within 20 points of the other two. This 20 point threshold is a subjective choice, which was made to balance the desire to find the most important process and leave some room for the possibility of mixed type events.

To examine the geographic and seasonal variations in the relative strengths and frequencies of each process associated with extreme precipitation the percentage of events that fall into each category is plotted for each region during the entire year, summer, and winter (figures 3.7-3.9 respectively). The most noticeable trend here is the increased fraction of events that are in the Convection category during summer for all regions. This is true for regions with strong precipitation seasonality (the Pacific Northwest, in orange), as well as a region with a flatter seasonal cycle (the Northeast, in yellow). In the annual plot (figure 3.7) four regions; the Great Plains, Desert Southwest, Southeast, and Florida have Convection as the category most associated with extreme precipitation. The two west coast regions share Vorticity as their most frequent category. The Northeast is the only region with a mixed category (Vorticity and Frontal) as its single most common category during extreme precipitation. Unsurprisingly Florida is the region with the largest gap between its most and second most common categories. Florida is also the region with

the most extreme seasonal shift. It goes from 60% convection in summer to 35% vorticity in winter. Both of these are large percentages for a single category to occupy.

Case Study - Front

As a way to check this analysis against previous work we consider a single event from the previously mentioned Kunkel et al. (2012) paper. This event was placed into the frontal type based on the manual analysis of a limited set of meteorological fields. The results of this work's analysis are shown in figure 3.10 at 8 times during the day in question. The colored areas are where the precipitation threshold of 12.5 mm from Kunkel et al. (2012) is met. The comparison focuses on the largest connected extreme, running SW-NE from Ohio to Maine. The categories including fronts in figure 3.10 are prevalent throughout the breadth and duration of this event which shows good agreement with the previous study. Figure 3.10 also shows definite changes in the mixtures of processes that are featured throughout event in time. There is a shift in secondary process whereby the Front & Convection category is present in the western part of the event early and the eastern part of the event later in the day. This is coupled with the reverse trend in the Vorticity & Front category. Convection & Vorticity appear together, without as notable thermal gradients, in isolated pockets that fall mostly around the edges of the precipitation extreme. This case study largely agrees with the finding that this event is "Frontal" but is also reveals that this front is being detected by all three metrics. Indeed, no event that was examined in this way was a truly "pure" case. The identified categories are quite varied in space and time. Some grid points belong to five different categories during the 24 hour period plotted in figure 3.10. Notably this precipitation occurs during the last day of Hurricane Katrina's life cycle. The then tropical depression Katrina is near the southwestern edge of our precipitation event during this day.

Case Study - Extratropical Cyclone

Figure 3.11 shows how the processes driving the extreme precipitation event of June 17, 1996 change throughout the course of the day. We will focus on the area from Kansas up to southern Michigan but it is briefly worth noting the other large area of strong precipitation east of Florida. This area

lies in the path of tropical storm Arthur and is dominated by the vorticity and convection metrics. The event in the area of focus was classified as being primarily driven by an extratropical cyclone in Kunkel et al. (2012). The metrics presented in this paper don't find unusually strong vorticity advection, compared to the strength of the thermal gradients for much of this event. The only place strong vorticity advection shows up is in the southern "tail" of the event and only around hour 12. For the rest of the event the frontal and convective metrics show up much more strongly, and are often mixed. This event underlines both the ability for convection to be embedded in, or near, frontal features and the difficulty of teasing apart the influences of extratropical cyclones and associated fronts. We again see all three metrics responding to this one event and generate varied results in space and time. It is interesting to note the way the categories with the frontal metric active begin as dominant over the region, while categories with convection active grow to dominate the area later.

Case Study - Front & Tropical Cyclone

The final case study (figure 3.12) contains two separate events and event types, as classified by Kunkel et al. (2012). The first event is tropical storm Erin creating precipitation over Oklahoma. This storm moves through north Texas into Oklahoma during the day of this precipitation event. In the metrics this event shows up very strongly in terms of local thermal gradients with the yellow, green, and cyan of the frontal categories well represented. This event looks very different than the area associated with tropical storm Arthur in the previous case (figure 3.11). That event had almost no grid points that fell into any of the three frontal categories whereas this one is dominated by them. Both events are associated with tropical storms, Arthur in the previous example and Erin in the current example. The primary differences are in the stage of development and the presence of land or water underneath the storm which causes them to appear very differently in these metrics.

The other event was classified as a front by Kunkel et al. (2012). While the frontal categories are very present here there are also a few areas that fall outside their scope, primarily in the convection and vorticity category. There are also significant areas where all three metrics are closely matched in score. While a front is definitely present, this event serves as an example of the mixed nature

of extreme precipitation events, which makes distinguishing the primary cause difficult for these metrics. As in the previous two examples all three metrics return high scores during this one event. generating varied results in space and time. For example, some areas along the southern edge of the front take up three different categories in the first 12 hours of the event.

3.5 Discussion

This work has presented an approach to identifying the process or processes responsible for an extreme precipitation event. The three processes were chosen to be commonly found processes present in the vast majority of events. This aspect was successful as only 1% of events were not captured by at least one of our metrics. This does have a downside in that there is a large overlap between each of the processes, which can be seen clearly in the fractured nature of the case studies (figures 3.10, 3.11, and 3.12). This fracturing happens in both space (the mosaic like appearance of the plots) and time as seen by the rapid changing of categories during the course of each event at some grid points. Because of this overlap between our metrics none of the cases we analyzed fell into spatially or temporally consistent categories.

Some examples of how these metrics overlap include mesoscale convective systems (MCSs), which are often born along frontal boundaries before separating (Kunkel et al., 2012). These events would score strongly in both our convective and frontal metrics. Depending on the presence and position of a trough the vortical metric could also score highly in some of these MCSs. Squall lines are a convective system even more closely tied to frontal activity. These type of events could contribute to the temporal inconsistencies by reaching an area just before the passage of a front (Newton, 1950). Additionally, CAPE is simply a measure of the current state of the atmosphere. CAPE can be decreased without convection by the advection of less buoyant air into the column at low levels. It is also possible for CAPE to decrease due to the heating generated by precipitation whether or not the precipitation was generated by convection. This contributes to the fractured nature of these case study events sometimes increasing the convection metric.

One of the ways fronts are sometimes identified is by observing shifts in the wind direction along the frontal boundary. This can lead to increased vorticity and vorticity advection in the vicinity

of the front. A link between the frontal and vortical metrics also arises because both fronts (areas of strong thermal gradients) and troughs (areas with largely cyclonic flow) offten occur toghether. The frontal and vortical metrics are also both related to the detection of frontal cyclones. These systems will have areas of strong vorticity advection that overlap with frontal features giving both metrics high scores. If, for instance, a squall line develops near a front in a frontal cyclone all three of our metrics will score highly.

In some ways, these results are enlightening. During the case studies we could observe very general trends in the importance of certain metrics to the extreme event, either changing in time or moving in space. For our purpose however, the interrelation between the processes do not permit us to cleanly identify the atmospheric process most associated with a particular precipitation extreme.

Climatologically, these metrics show increased strength of vorticity advection to wintertime extreme precipitation on the west coast relative to frontal strength (figure 3.8). Though perhaps initially surprising, previous studies (Dowdy and Catto, 2017; Kunkel et al., 2012) have disagreed on how to disentangle extratropical cyclones from their associated fronts as the cause of extreme precipitation. This method finds a stronger relationship between the vortical field and precipitation extremes rather than the thermodynamic field. In the Northeast, the three frontal categories narrowly edge the three vortical categories over the course of the year (figure 3.7) where Kunkel et al. (2012) find that fronts are more common than extratropical cyclones by greater than a 2 to 1 margin. In the Desert Southwest we find a majority of events are convectively influenced. This is in agreement with the findings of Dowdy and Catto (2017) and disagreement with Kunkel et al. (2012) who find more than half of events to be frontal in nature. These direct comparisons are somewhat difficult given the different data, methods, and regional aggregations at play. But, the disagreements between the previous studies and between this work and the previous studies highlights the uncertainty still remaining in this kind of endeavor.

3.6 Conclusions

The three processes of focus were Convection, Vorticity Advection, and Fronts which form a analogous approach to the work in Dowdy and Catto (2017). Each process was given a metric that

had a physical link to the mechanism used to create extreme precipitation. Thermal gradients for fronts, vertical instability for convection, and upper level vorticity advection for cyclones.

For convection the metric is based on amount of CAPE consumed during the event, for vorticity the metric is based on the amount of positive vorticity advection, and for fronts the metric is based on the strength of the local thermal gradient. These metrics emphasize the strength of each process relative to strengths found at each location. This avoids the creation of thresholds for each process that correspond to similar strength or importance between processes. This would be an extremely problematic process because these processes are often linked, as in the earlier discussion of extratropical cyclones and fronts. This work clearly shows that these physically relevant metrics are not strictly related to a single type of atmospheric process and, therefore, do not separate extreme events cleanly. This means that we are unable to test our central hypothesis that *the seasonality of mean precipitation is predictive of the distribution of extreme precipitation generating mechanisms.* However, we showed that these physically reasonable metrics are unable to determine the main cause of extreme precipitation.

The metrics are created using simple calculations that can be done at the grid point level to investigate the spatial change in process mix within an event (figures 3.10-3.12). This comparison showed broad agreement with Kunkel et al. (2012) when considering the influence of fronts. The metrics helped us identify areas where convection and vorticity each played the primary supporting role and how those areas changed during the course of each event. In the first and third event the "front" was somewhat close to a tropical storm. The remnants of Hurricane Katrina in the first case and Tropical Storm Erin in the second. Disentangling a probable moisture source from the dynamical process is a challenge for any process attribution framework. Sometimes a choice needs to be made between the source of anomalously high moisture and the dynamical process responsible for the lifting and condensation. This emphasizes a takeaway from this work, that synoptic scale weather systems can create extreme precipitation through a variety of processes and that the primary cause cannot be identified from these physically based, grid point level metrics. The following chapter will present descriptive, rather than prescriptive, approach to identifying the primary meteorological process behind extreme events.

Acknowledgements This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research program under Award DE-SC0016605 "A framework for improving analysis and modeling of Earth system and intersectoral dynamics at regional scales." This project took advantage of netCDF software developed by UCAR/Unidata (http://doi.org/10.5065/D6H70CW6).

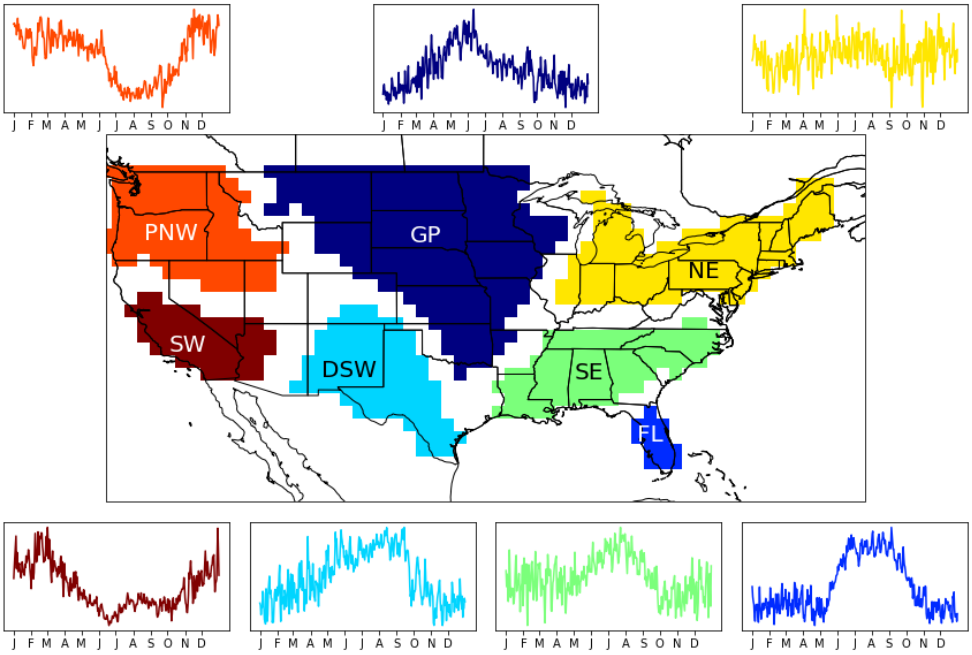

Figure 3.2: The 7 regions of similar annual cycle by which our results may be grouped. These regions result from the approach in Swenson and Grotjahn (2019) for ERA5 daily precipitation data. Abbreviations for the 7 regions are as follows: Great Plains (GP), Florida (FL), Desert Southwest (DSW), Southeast (SE), Northeast (NE), Pacific Northwest (PNW), Southwest (SW). The mean annual cycle of precipitation averaged over all grid points within each region is displayed in the subplot matching the color of the region. The y-axis of these subplots is precipitation amount normalized to the region, with the limits chosen separately to span the normalized data for that region. The x-axis is day of the year with tick marks at the first day of every month.

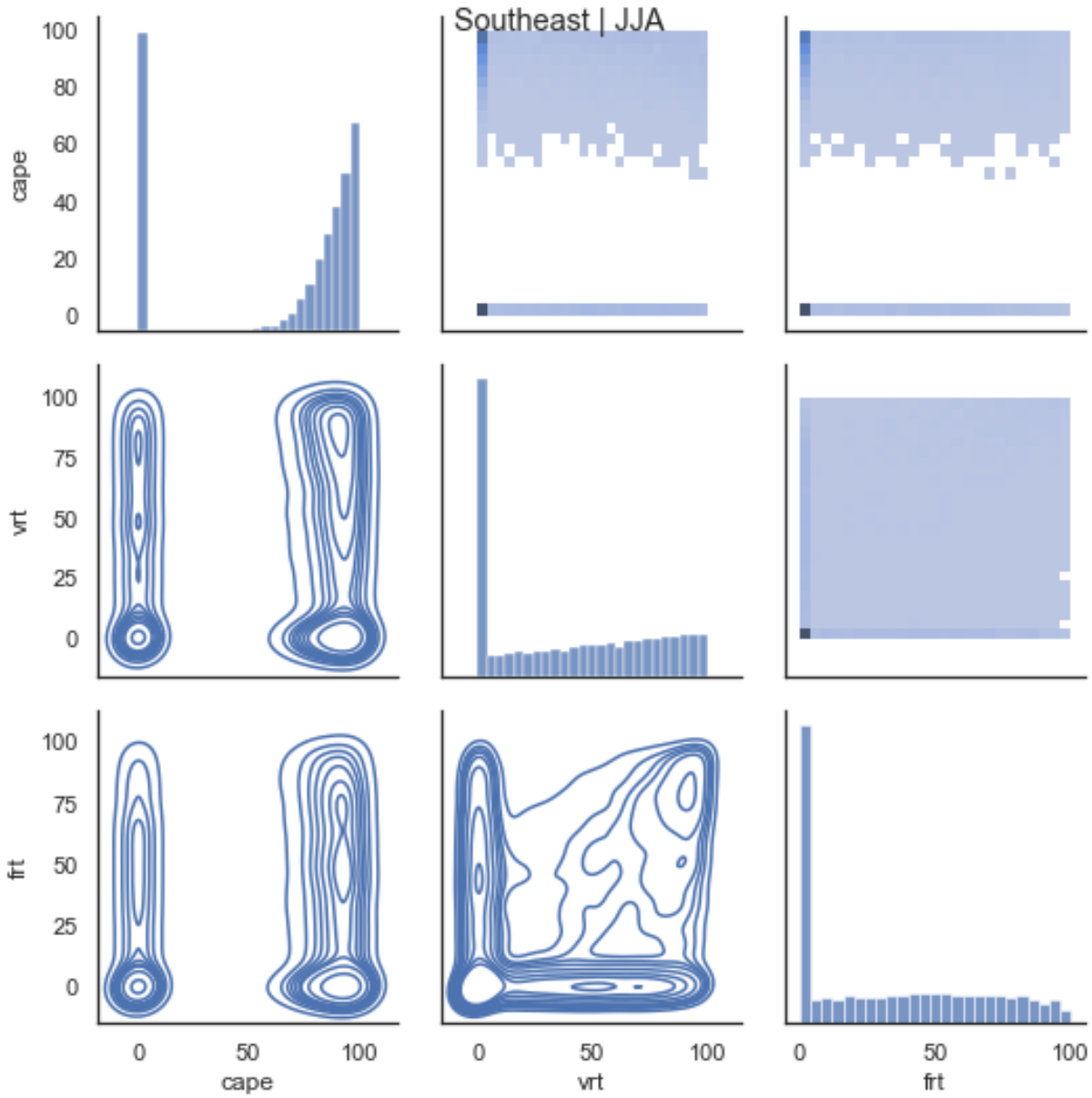

Figure 3.3: The plots along the diagonal are the univariate distribution of each of the three processes (Convection, Vorticity, and Frontal). The x-axis is the rank of the process (broken into 25 bins) and the y-axis the number of events in that bin (the y axis labels do not apply to the panels on the diagonal). The lower triangular panels show a kernel density plot of the two processes indicated on the corresponding x and y axis labels. The upper triangular panels show the same data as a bivariate histogram.

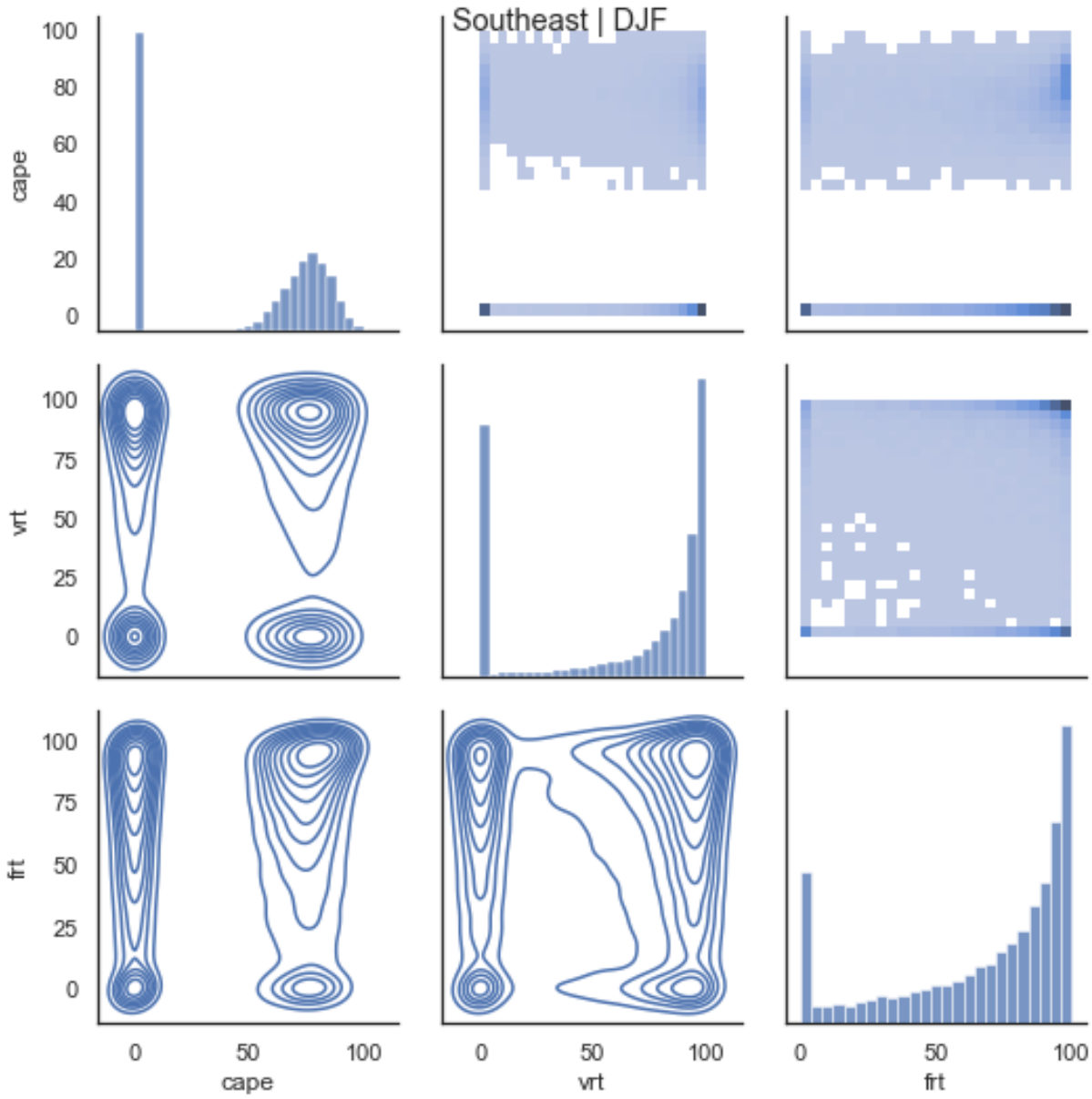

Figure 3.4: The plots along the diagonal are the univariate distribution of each of the three processes (Convection, Vorticity, and Frontal). The x-axis is the rank of the process and the y-axis is a density function over all events in the named region and season. The lower triangular panels show a kernel density plot of the two processes indicated on the corresponding x and y axis labels. The upper triangular panels show the same data as a bivariate histogram.

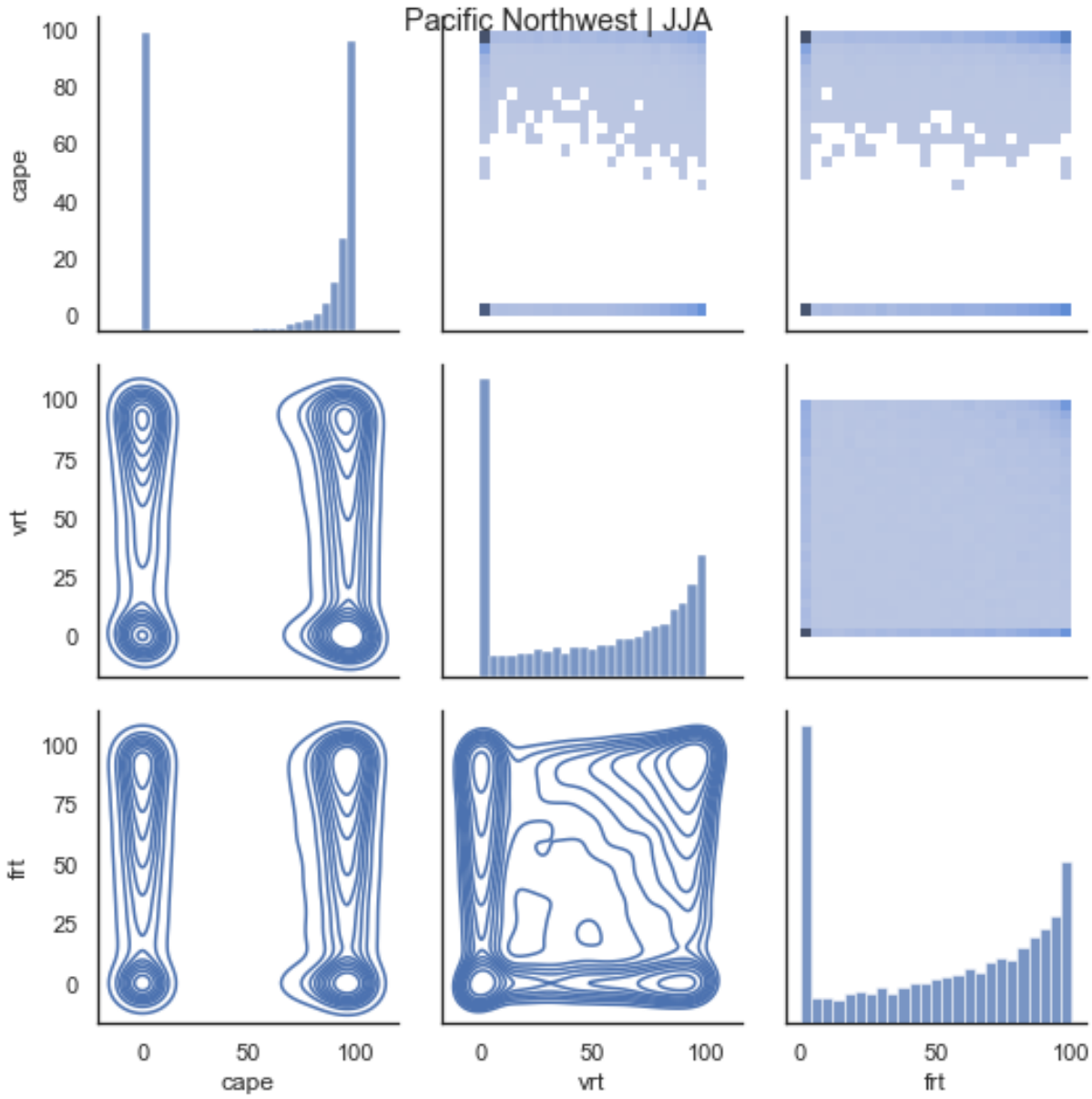

Figure 3.5: The plots along the diagonal are the univariate distribution of each of the three processes (Convection, Vorticity, and Frontal). The x-axis is the rank of the process and the y-axis is a density function over all events in the named region and season. The lower triangular panels show a kernel density plot of the two processes indicated on the corresponding x and y axis labels. The upper triangular panels show the same data as a bivariate histogram.

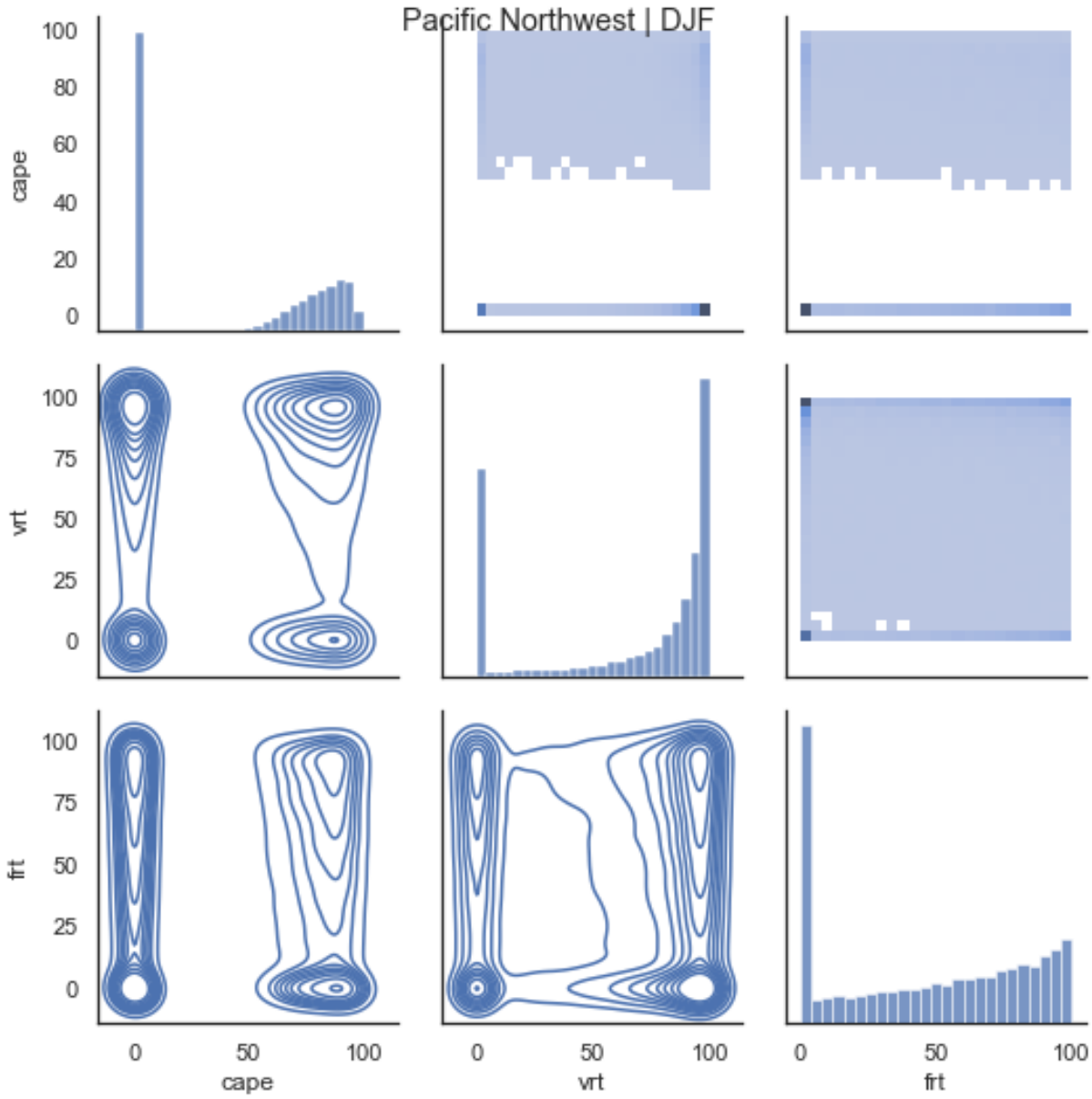

Figure 3.6: The plots along the diagonal are the univariate distribution of each of the three processes (Convection, Vorticity, and Frontal). The x-axis is the rank of the process and the y-axis is a density function over all events in the named region and season. The lower triangular panels show a kernel density plot of the two processes indicated on the corresponding x and y axis labels. The upper triangular panels show the same data as a bivariate histogram.

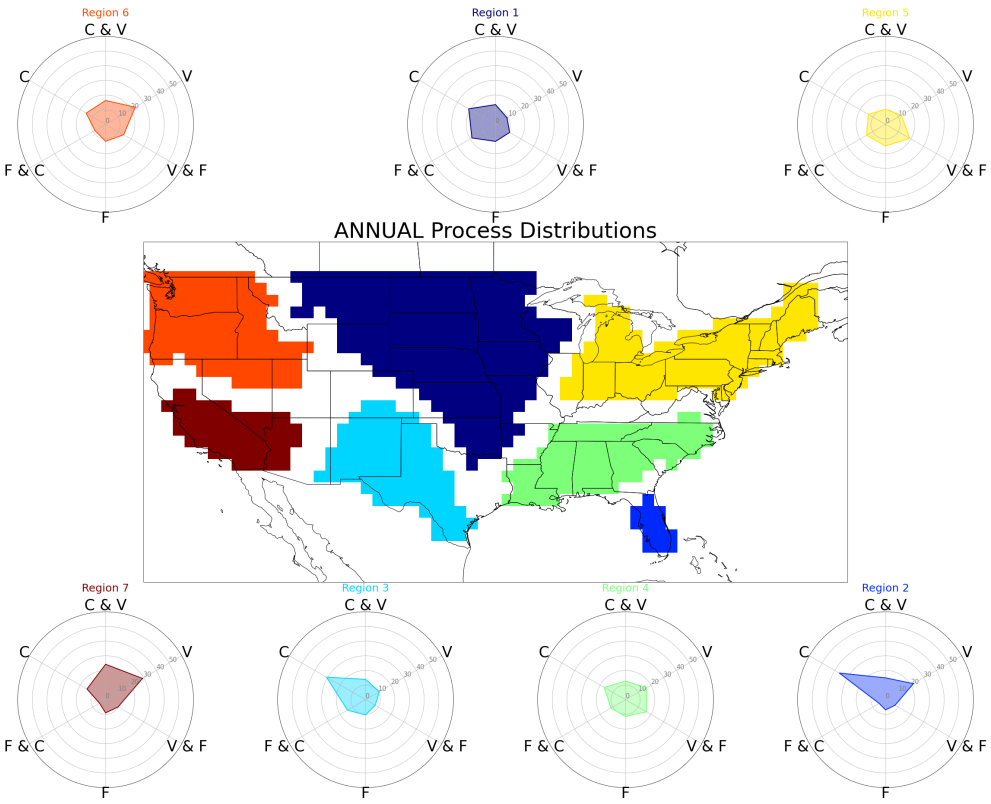

Figure 3.7: Central plot shows the regions over which extreme precipitation was aggregated. Around this are plotted the distributions of the various categories for extreme precipitation within each region. Concentric rings are in increments of 10%. This includes every event in the record used bar those in the "ALL" and "None" categories. Those were omitted to aid the visualization. "None" type events make up no more than 4% of the events of any combination of region and season.

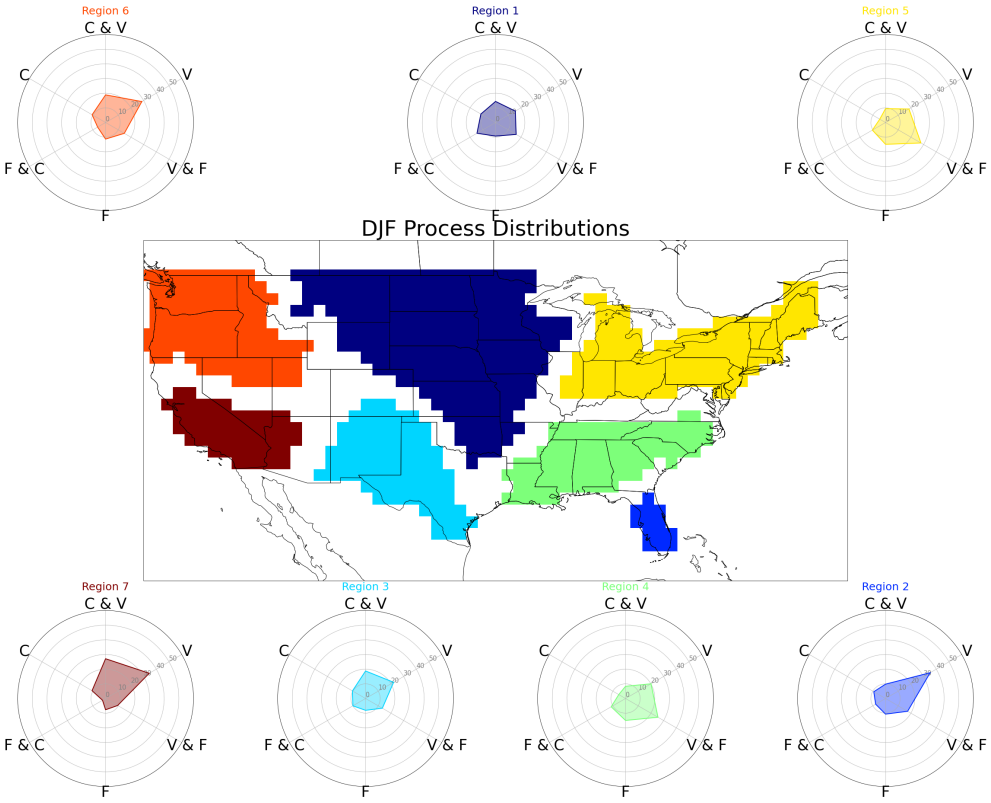

Figure 3.8: Same as figure 3.7 but for only the winter months (DJF)

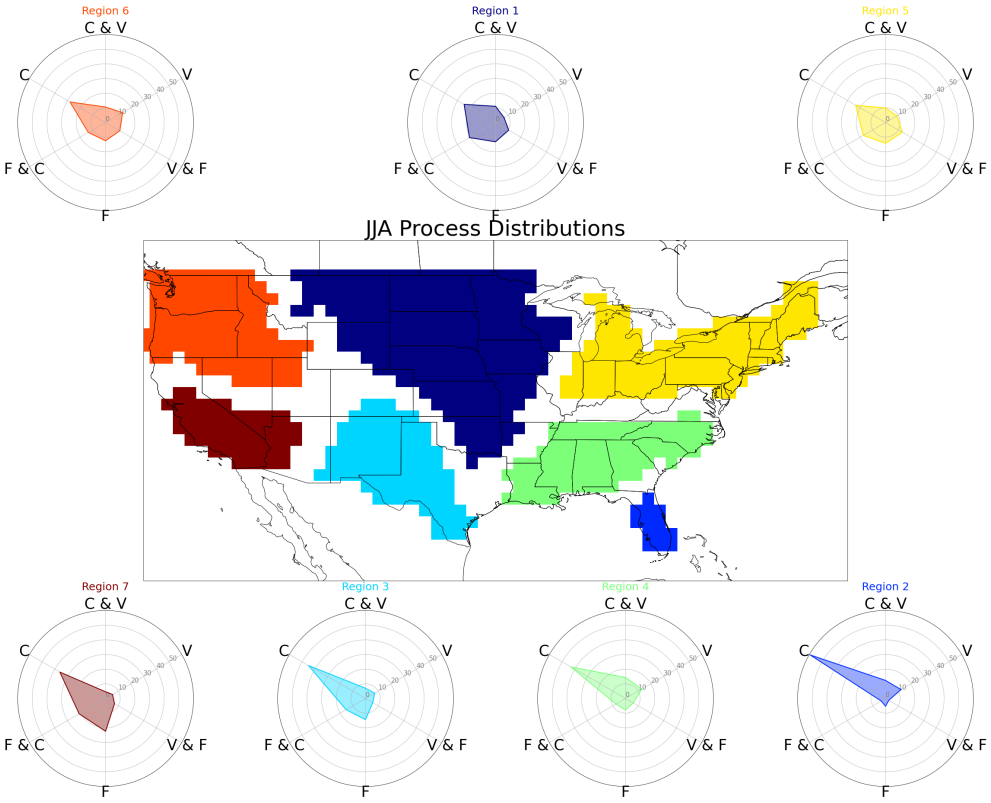

Figure 3.9: Same as figure 3.7 but for only the summer months (JJA)

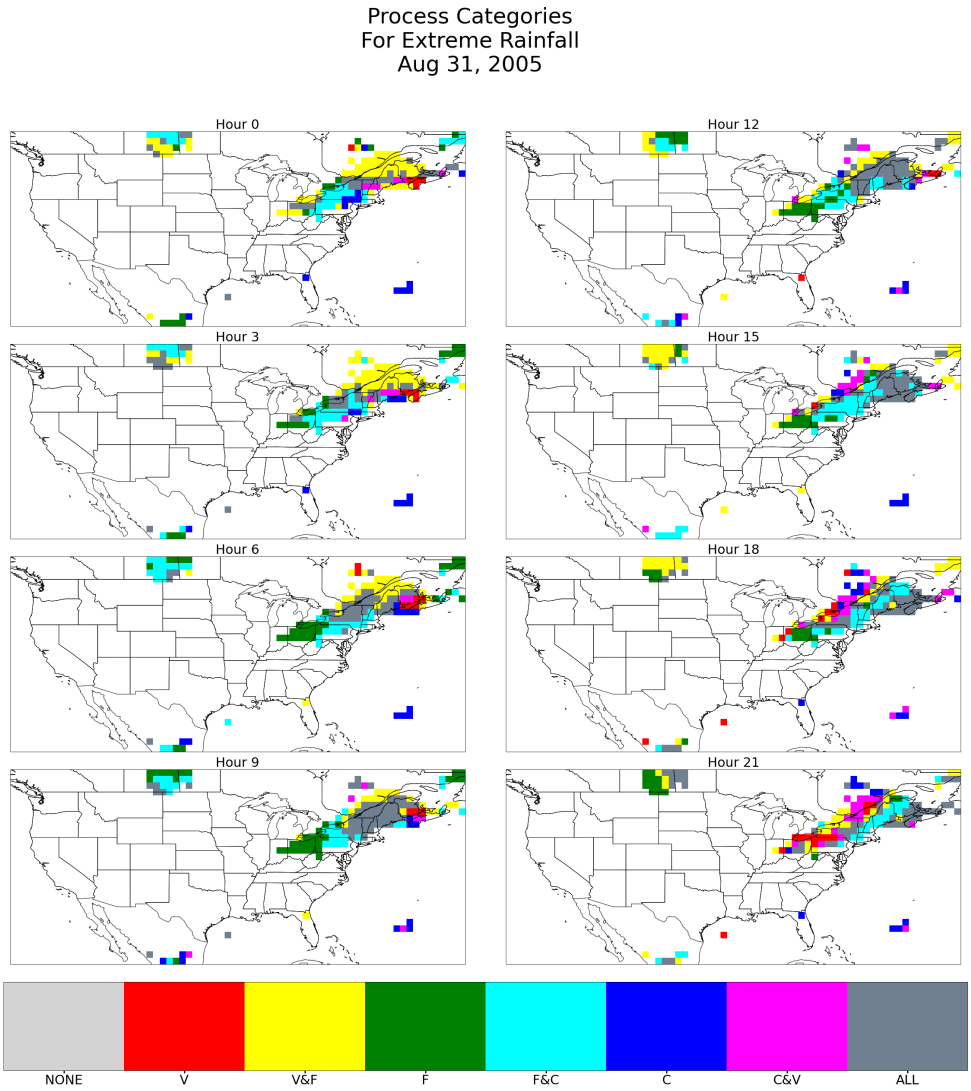

Figure 3.10: Plot of the process category for each grid point with precipitation $> 12.5\frac{mm}{day}$ 8 times during the day of Aug 31, 2005. The categories are labeled as: V - Vorticity, F - Frontal, C - Convection. The area of interest is the largest connected area of colored dots, classified as a frontal extreme by Kunkel et al. (2012). All times are UTC.

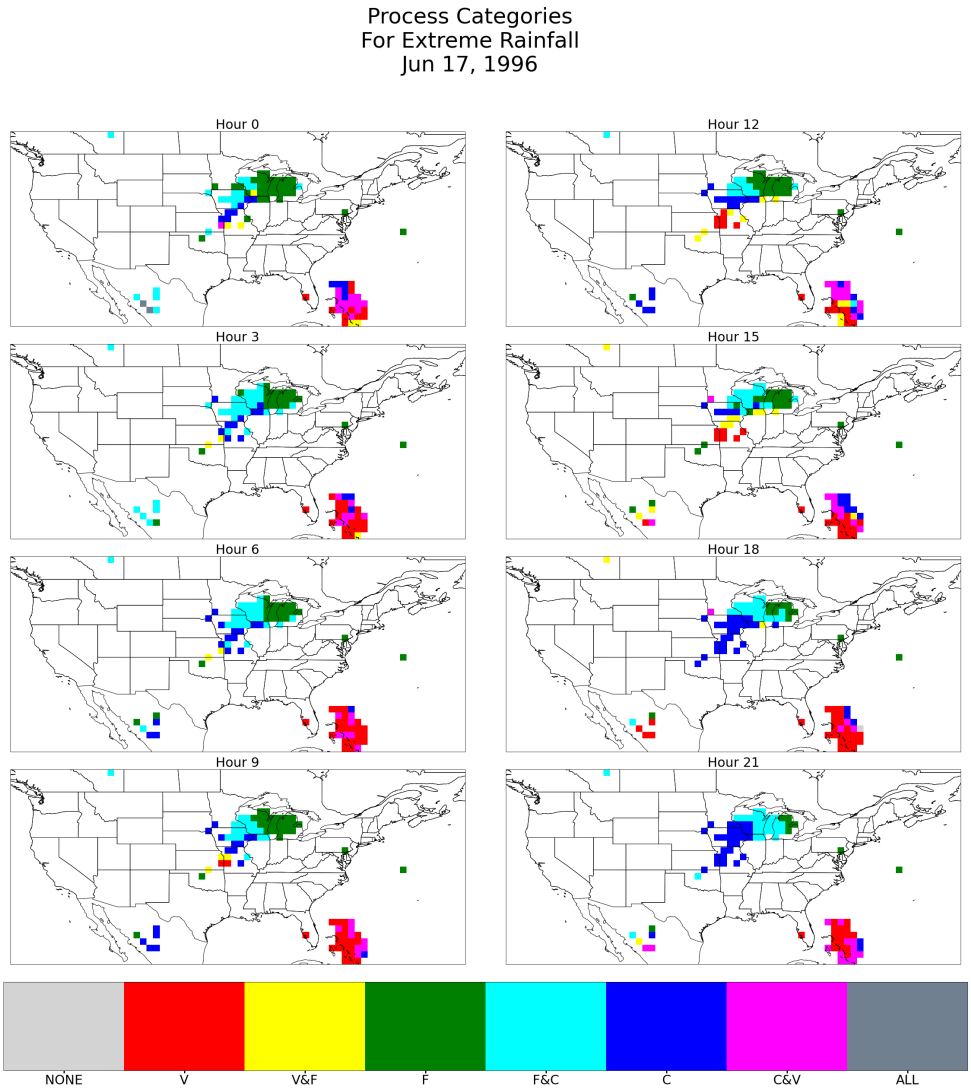

Figure 3.11: Plot of the process category for each grid point with precipitation $> 12.5 \frac{mm}{day}$ 8 times during the day of June 17, 1996. The categories are labeled as: V - Vorticity, F - Frontal, C - Convection. The area of interest is the connected area of colored dots surrounding Wisconsin, labeled an extratropical cyclone by Kunkel et al. (2012). All times are UTC.

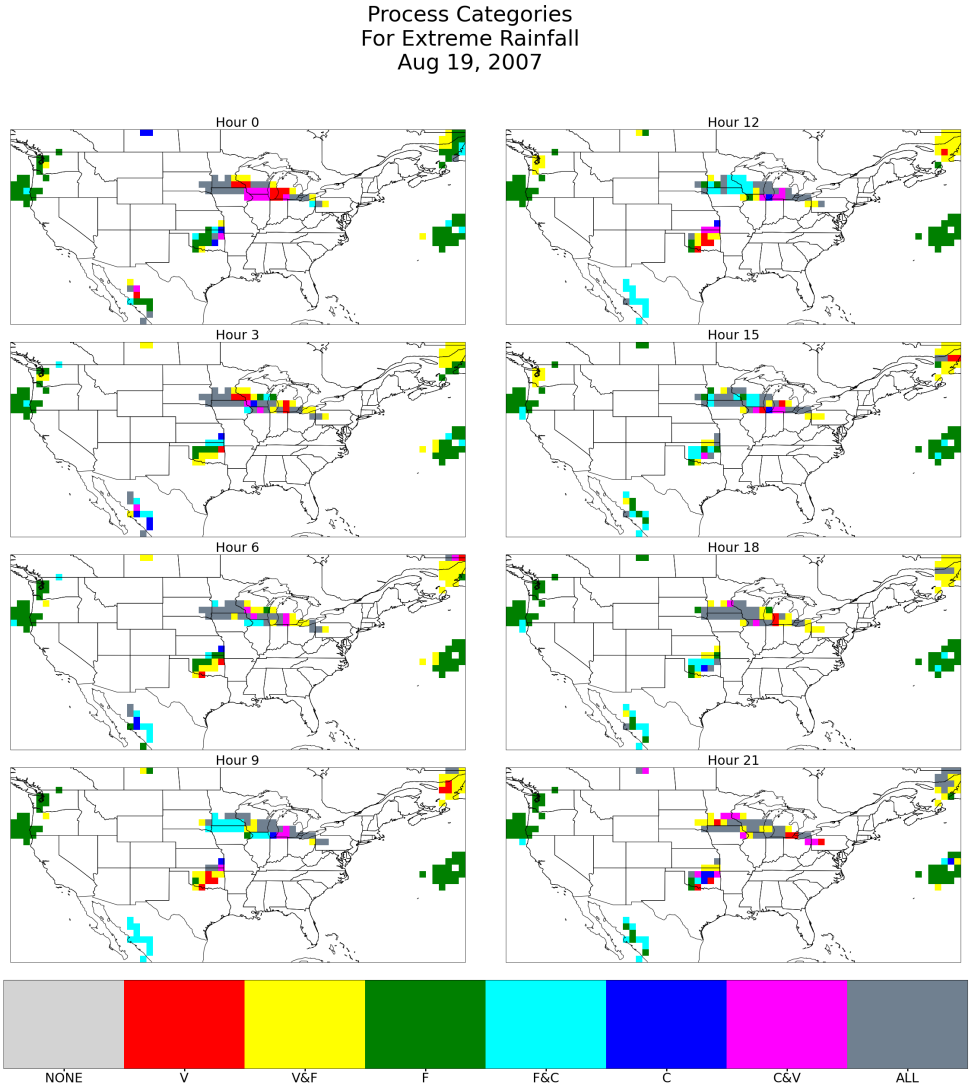

Figure 3.12: Plot of the process category for each grid point with precipitation $> 12.5 \frac{mm}{day}$ 8 times during the day of Aug 19, 2007. The categories are labeled as: V - Vorticity, F - Frontal, C - Convection. The areas of interest are Oklahoma (classified as a tropical cyclone by Kunkel et al. (2012)) and the long line stretching from South Dakota to Pennsylvania (classified as a front by Kunkel et al. (2012)). All times are UTC.

Chapter 4 Fundamental Weather Patterns During Extreme Precipitation: Identified From The QG Omega Equation

4.1 Introduction

Understanding how changes to dynamical processes will affect precipitation extremes (PEx) must start with understanding the links between those processes and PEx in current climate. O'Gorman and Schneider (2009) showed that a combination of vertical velocity and precipitable water (PW) can be used to scale PEx intensity. The relationship between precipitable water and precipitation intensity is straightforward (as PW increases so does precipitation intensity) and has been investigated (Kunkel et al., 2020). Matching instantaneous vertical velocity to precipitation accumulation is not as straightforward. Vertical profiles at a single time are not sufficient since PEx accumulates over time. A single vertical level over several times is also insufficient since the vertical distribution varies for different PEx events. Vertical velocities sometimes peak hours before, or indeed after, the onset of peak precipitation rates. Previous studies have focused on the relationship between omega and precipitation in one or two specific PEx events (Nie et al., 2016; Pauley and Nieman, 1992). This level of depth is not feasible for for developing a climatology of PEx events over large and diverse geographic regions. We will group the PEx events objectively into a small number of categories by applying the self-organizing map (SOM) as in Kohonen (1982). This automatic grouping will also allow us to examine the types of weather patterns that explain the variations in PEx as done by Kunkel et al. (2012) without necessitating a huge time investment

in analysing each individual event by hand. This will, in conjunction with the results of Swenson and Grotjahn (2019), allow us to test our central hypothesis: *The seasonality of mean precipitation is predictive of the distribution of extreme precipitation generating mechanisms.* To do this we will compare the distribution of mean and extreme precipitation throughout the year, throughout the CONUS, and within each region. We will then analyze the number and relative importance of each SOM identified weather pattern throughout the year within regions.

This work considers two methods of forcing vertical motion: differential (in pressure) advection of vorticity and advection of temperature. Increasing positive vorticity advection over a layer of the atmosphere causes a larger height decrease at the top of the layer than the bottom. This leads to a decrease in thickness which corresponds to a cooling of the layer. Hydrostatically, this is caused by adiabatic cooling from upwards vertical motion. Positive temperature advection through a layer of atmosphere increases the thickness of the layer. The expansion implies upward motion but also higher pressure aloft that supports upper level divergence of mass so that the atmospheric column could have a net loss of mass and thus lowered pressure at bottom. The lower pressure supports convergence below. The divergence aloft and convergence below cause upward motion.

These two most dominant dynamical process in the QG system (Holton and Hakim, 2013) are where we focus our effort to identify the weather systems associated with PEx. Though there exists some cancellation between the two terms (Trenberth, 1978) we use both individually in our method to identify the large scale weather pattern active during the PEx event. Similar to precipitable water, diabatic heating is not incorporated into the SOMs because it is not prescriptive of the different weather phenomena of interest here. When diabatic heating is present the upwards motion must be present to facilitate cooling to keep the atmosphere in hydrostatic balance.

This paper is divided into the following sections. Section 2 describes the data sets that were used. Section 3 describes the methods used in this work and is subdivided into subsections dedicated to numerical methods and self-organizing maps. Section 4 describes the patterns that result from our method. In Section 5 we discuss the results while section 6 provides the conclusions.

4.2 Data

For this study we primarily use the fifth version of the European Center for Medium-Range Weather Forecasting's (ECMWF) atmospheric reanalysis (ERA5) (Hersbach et al., 2020). This data, and all ERA5 data used in this study, was retrieved from the Climate Data Store (CDS) using their application programming interface (API) as detailed here on their website. We regrid this data set to one degree resolution in the horizontal, 50 hectopascal resolution in the vertical, and three hours resolution in time. These resolutions are are attainable by regridding and/or data are readily available at these or finer resolutions from many GCMs/reanalysis outputs. All regridding operations were done via the CDS API. The horizontal resolution is as fine as is practical for the quasi-geographic (QG) omega equation (Battalio and Dyer, 2017). The vertical resolution is fine enough to well sample the vertical structure of the troposphere as well as yielding similar results to a Q-vector formulation of the omega equation (Pauley and Nieman, 1992). The temporal resolution allows us to catch short duration events (e.g., small scale convection) and sample over the course of longer duration events (e.g., stationary fronts or slow-moving mesoscale convection). We have chosen a period from 1980-2010 to diagnose the climatology of PEx drivers in the current climate.

ERA5 is our source of gridded precipitation accumulations on the same one degree resolution horizontal grid. PEx events are identified when the 24 hour precipitation accumulation at a grid cell is in the top 5% of all 24 hour accumulations (0 UTC to 0 UTC) at that grid cell. We center the time series of omega from each event on the 3 hour period with the highest total precipitation during each event. Making each event satisfy a 24 hour threshold focuses on events with enough total rainfall to achieve severe impacts. Focusing on the peak 3 hour precipitation within each event lets us capture the dynamical processes and instantaneous weather patterns at play near the most intense precipitation rate.

4.3 Methods

4.3.1 Self-Organizing Maps

A Self-Organizing Map (SOM) is a type of artificial neural network first introduced by Kohonen (1982). SOMs utilize a competitive and unsupervised learning algorithm to produce a lower dimensional representation of the input data that can be more readily analyzed. A SOM produces a user defined number of patterns that span the input data. The competitive learning balances each pattern between being dissimilar to the other patterns and being similar to each input that is best matched to itself. A valid set of patterns must pass a field significance test based upon the False Discovery Rate (FDR) (Johnson, 2013) of 1%.

Self-Organizing Maps: Regions

Recall that our central hypothesis is that *The seasonality of mean precipitation is predictive of the distribution of extreme precipitation generating mechanisms.* To test this we use a self-organizing map (SOM) to identify regions that share the same mean precipitation seasonality (Swenson and Grotjahn, 2019). This is very useful because it allows us to match a set of extreme events (those occurring within a particular region) to the seasonal cycle of mean precipitation within that region. The raw output from the SOM is shown in figure 4.1, as well as it's isolated area count (the median number of isolated areas comprising a single region: IAC), minor areas fraction (the median fraction of a region's area that is not contained in the largest isolated area: MAF), and compactness ratio (the median ratio of the square root of a region's area to its perimeter) which all compare well to the scores in the results of Swenson and Grotjahn (2019). Additionally this set of regions passes the regional extremes ratio (RER) threshold of 20% set out in Swenson and Grotjahn (2019) using this chapter's 24 hour definition of a precipitation extreme. Because of the lower resolution used in this work we have to create maps using less regions than in Swenson and Grotjahn (2019). This results in Florida (FL) not being separated from the New Mexico / Texas border region naturally. This necessitates the use of the automatic intervention to separate large enough isolated areas grouped together by the SOM discussed in Swenson and Grotjahn (2019) to separate FL into a 7^{th} region.

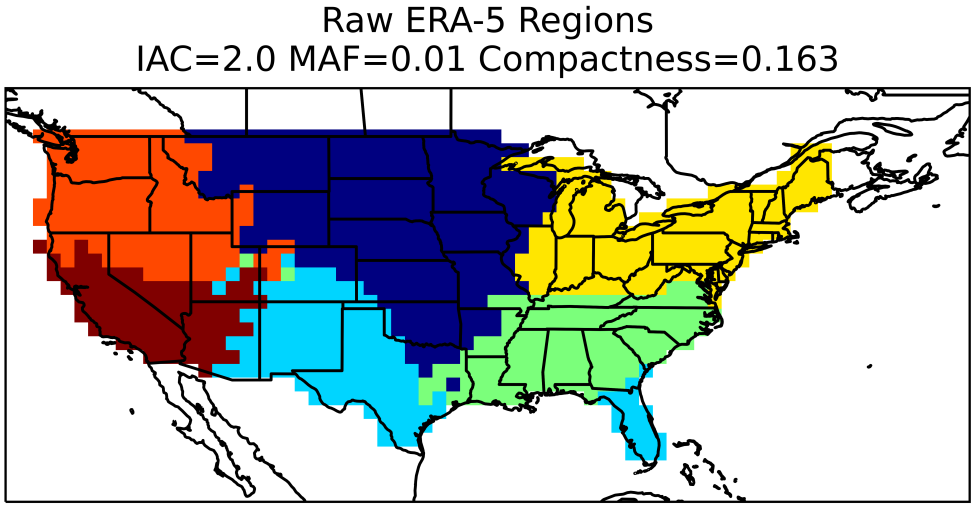

Figure 4.1: The raw output of the SOM trained on the normalized long term daily mean of the cube root of precipitation at each grid cell as in Swenson and Grotjahn (2019). IAC and MAF refer to the map's median isolated area count and minor areas fraction respectively.

Further discussion of the link between the seasonality of precipitation in FL and the New Mexico / Texas border region can be found in Swenson and Grotjahn (2019). We also remove the borders between regions from the analysis to reduce the uncertainty in the seasonal cycle in each region. The final regional map is shown in figure 4.2. Figure 4.2 does not display the scores because IAC and MAF are 1 and 0 respectively because of the processing while the compactness ratio increases. After processing, the RER is the only criterion likely to be negatively effected and the threshold is still passed for the processed regions.

Self-Organizing Maps: Omega Patterns

Our SOM analysis of processes (hereafter "SOM") is trained on the time versus pressure pattern of omega created from each of the advective terms of the QG omega equation centered on each PEx event. PEx events are chosen following the procedure outlined in the "Data" section. This analysis is applied between 1980 and 2010 for every PEx event at each of the colored grid cells in figure 4.2. For each PEx event we create a pressure vs time matrix of omega values forced by each of two forcing terms. These two matrices are combined to be one input data element to the SOM. This results in each SOM pattern being a set of two Hovmöller diagrams featuring both components of the QG omega equation with time on the x-axis and pressure descending on the y-axis. We use five repetitions of a two-phase learning cycle with 30 rough training iterations and 300 fine tuning iterations. The lattice of nodes is chosen to have a shape which minimizes the difference between the number of rows and columns. The initial neighborhood radius is set to span the lattice of patterns. The final neighborhood radius is set to one to prevent over fitting. In every case the neighborhood function, which weights how much each pattern adjusts with a particular input, is an Epanechikov function. The displayed results (figure 4.3) come from a SOM with nine nodes in a 3x3 arrangement. The duration of the training data was chosen to be 36 hours after testing different length time series as inputs to the SOM and finding only small differences between the different sets of output patterns. Nine nodes were chosen as the largest square number of statistically distinguishable nodes for all lengths of time series that were tested. One of the benefits of the SOM as a method is that the results are sorted by similarity with adjacent patterns being the most similar to one another. This helps us organize our results into smaller groups of similar patterns. In the Results section the discussion of the 9 patterns is collected into groups of similar patterns.

4.3.2 Numerical

The QG omega equation is a central tool of this work and can be expressed as equation (4.1).

$$\underbrace{\sigma_0 {\nabla_h}^2 \omega + f^2 \frac{\partial^2 \omega}{\partial p^2}}_{LHS} = \underbrace{\underbrace{f \frac{\partial}{\partial p}[V_g \cdot \nabla_h(\zeta_g + f)]}_{F_V} + \underbrace{\frac{R}{p} {\nabla_h}^2 (V_g \cdot \nabla_h T)}_{F_T}}_{RHS} \tag{4.1}$$

Where the subscript "g" means the quantities take their geostrophic form and the subscript "h" indicates the operation is done only in the horizontal. The square of the coriolis parameter (f^2) is fixed to a value of 1e-8 s-2 on the left-hand side (LHS) but f is allowed to vary freely with latitude on the right-hand side (RHS) as in Räisänen (1995). σ_0 is the domain and time average of the profile of static stability and a function of pressure only. The operator on the (LHS) is inverted over a limited domain chosen separately for each region. We found that a rectangular domain that encompassed the region with at least a buffer of one grid cell in each direction was sufficient for every region. These choices of spatial boundaries result in the need to preform 2D matrix operations on arrays ranging from 630 to 7410 points per side. The inversion is done using the python package NumPy (Harris et al., 2020), which is quite suitable for matrix manipulations of the sizes required. For each day with extreme precipitation somewhere in a region the operator is inverted for each of the advective forcings separately at enough time steps to create a 36-hour window around the event at each grid cell (13 inversions per event for the time resolution of our data). The finite differencing used to calculate the derivatives in equation (4.1) were done using centered 2nd order differences at two time levels and taking the average. Before the inversion operation we smooth each forcing term using a 3x3 Gaussian filter in the horizontal. We use homogeneous boundary conditions (values of omega outside the domain are set to zero) at all six boundaries to isolate the contributions to vertical motion from each of the advective forcing terms. This is possible because the LHS operator is linear in omega (Krishnamurti, 1968).

The green contours on figure 4.4 are our proxy for detecting frontal locations in the horizontal map composites. This field is a thermal frontal parameter (TFP) used by Renard and Clarke

(1965) as well as Catto et al. (2012). This parameter is defined as the directional derivative of the horizontal gradient of a scalar thermodynamic variable along its gradient (equation 4.2). We chose equivalent potential temperature (θ_e) at 850 hPa as our thermodynamic scalar. This TFP was selected from among many existing options (many of which may be found in Hewson (1998)) because it takes its maximum values along the warm air side of frontal boundaries, as detailed by Renard and Clarke (1965). Because figure 4.4 displays composite events, which somewhat smooth sharp features like fronts, we interpret the areas inside green contours of figure 4.4 as regions with increased frontal gradient.

$$TFP = -\nabla|\nabla\tau| \cdot \frac{\nabla\tau}{|\nabla\tau|} \tag{4.2}$$

4.4 Results

Our method of organizing the results is as follows. Our choice of 9 total patterns was made, as described earlier, without foreknowledge of the types of patterns we might see. This was done to create a process that was open to novel insights that an unsupervised machine learning process can provide. Upon reviewing the 9 patterns we saw that one of the key distinctions the SOM made was in the timing of extreme precipitation relative to the passage of a midlatitude cyclone. We find a cyclone very close to the PEx grid cell in 6 of the 9 patterns; 4 are best described as mature frontal cyclones and 2 SOMs are cyclones with apparent occlusions. We group these 4 and 2 patterns together respectively, for a more concise discussion because of their similarities. During testing of the SOM with fewer patterns we noticed that for a SOM with 5 patterns we see close analogues for each of the 5 groups. This reinforces our assessment of the intra-group similarity between the frontal cyclone and occluded cyclone patterns.

4.4.1 Convection: Pattern 0

Pattern 0 (upper left panel of figure 4.3) has extremely weak upwards motion from both QG terms. The members of this pattern have by far the weakest agreement on the sign of omega (green contours of figure 4.3) with only small areas reaching 2/3 agreement. This level of agreement only

occurs in the vorticity driven omega panel. Because the agreement is so low there is the possibility that large magnitudes of omega are canceling each other to produce the pattern shown in figure 4.3. To eliminate this possibility we looked at the average of the absolute value of each member event and found that the members of pattern 0 do indeed have the weakest QG forced vertical velocity of the nine patterns. The precipitation field is narrow in time (upper left panel of figure 4.5), evidenced by the large drop in both directions from the middle bar. Figures 4.4 and 4.6 depict a large horizontal domain (30 degrees latitude by 60 degrees longitude) composite of the events that make up each pattern. The synoptic elements of figure 4.4 are consistent with convection as it is sometimes isolated (as in a convection category used by Kunkel et al. (2012)). There is also a lack of any strong upper level anomaly, the area encompassed by notable precipitation (dotted contours) is smallest, and CAPE values are the largest of any pattern (figure 4.4). Additionally; the low level temperature anomalies are weak, pattern 0 is one of only two patterns without a closed low level circulation in the anomaly wind field, and the PEx grid cell is centered within the total column water (TCW) anomaly (figure 4.6). These factors lead us to categorize this pattern as convective, possibly triggered by surface fluxes. We do see a frontal zone (green contours in figure 4.4) so some slow moving or stationary fronts with weak advection may also be incorporated into this pattern. A pseudo-frontal zone may also appear because the frontal gradient parameter can be large due to a strong moisture gradient even where the temperature gradient is small. Tropical cyclone (TC) related PEx events are also very likely to be found in this pattern due to a weaker QG omega field compared to our other patterns (Fischer et al., 2017).

4.4.2 Atmospheric Rivers: Pattern 2

Pattern 2, the upper right panel of figure (4.3, shows very strong sinking motion from vorticity advection as well as strong rising motion (constrained below 400 hPa) from temperature advection. Both motions exist throughout the duration of the pattern in time, but the strongest upwards values are found in the temperature omega field at low levels between 6 and 3 hours before the event. This pattern has very strong sign agreement between the members with $> 75\%$ agreeing on: downwards motion from vorticity advection below 500hPa and upwards motion from temperature advection

below 700hPa. These motions sum to positive (downwards) values of omega throughout the time pressure domain in figure 4.3. This is the only one of the nine patterns to do this and indicates the presence of some uncaptured factor causing upwards motion during these types of PEx events. In figure 4.4 the PEx grid cell is northeast of the midpoint between the upper level low and high (along the the zero line for 300hPa height anomalies). Also, the surface low is northwest of the event and a frontal zone is located southwest of the extreme. Pattern 2 is commonly seen in the Pacific Northwest wherein the surface and upper level low are near the Gulf of Alaska with a strong ridge ahead over the Rockies and a frontal zone trailing to the southwest. Such a pattern during heavy rain is likely to occur with an atmospheric river (Ralph et al. (2017), Zhang et al. (2019), and Collow et al. (2020) among others). Additionally we see an elongated area of anomalous moisture and temperature that extends far to the southwest of the PEx grid cell (figure 4.6). This is accompanied by strong low level southwesterly flow along the anomalous areas. This is also consistent with a phenomena called 'training', where a line of heavy precipitation cells moves slowly eastward while the individual cells in that line move along the line (typically SW to NE). Training can be found in Atmospheric River (AR) events and can cause severe flooding and damage (Nash and Carvalho, 2020). This pattern is found in the western part of the CONUS more often than the east and a plurality of events take place in the Pacific Northwest Region (figure 4.7) They are more common there during winter (figure 4.8) than summer (figure 4.9) as one would expect for strong ARs. The strong southwesterly warm air advection and the orientation of the high and low anomalies discussed earlier, strongly indicates that this group of patterns corresponds to Pacific coast AR driven events. Because AR driven precipitation often occurs in conjunction with upslope flow and the prevalence of this type of event along the western slopes of the Cascade and Rocky Mountain ranges in the Pacific Northwest (upper right panel of figure 4.7) we believe that the uncaptured factor causing upwards motion is upslope flow. This could be verified in subsequent work by setting omega at the lower boundary equal to the vertical velocity forced by the interaction between topography and the low level horizontal winds while setting the right-hand side forcings to zero.

Pattern 2 is also a common pattern in the northern part of the Desert Southwest region (South-

ern Colorado, see the upper right panel of figure 4.7). This occurs within the path of the North American Monsoon which, as opposed to ARs in the Pacific Northwest, is a summertime phenomena in the Desert Southwest (compare panel 2 between figures 4.8 and 4.9). Convection (Pattern 0) remains the most common during summer in the DSW, but the prevalence of patterns 2 and 1 indicates that PEx there can also be associated with increased low level moisture transport of the North American Monsoon (NAM). For both AR and NAM events in this SOM, the PEx grid cell is not co-located with an approaching upper level extratropical cyclone (ETC) center (no upwards omega from vorticity advection) but both show definite low level warm, moist air advection.

4.4.3 A Transitional Pattern, Convection Enhanced by Horizontal Moisture Flux: Pattern 1

We label pattern 1 as a transition pattern for both technical and meteorological reasons. On The technical side; the SOM organizes patterns by similarity, placing pattern 1 as *most* like patterns 0, 2, and 4. Meteorologically, pattern 1 has visibly more in common with patterns 0 and 2 than it does with pattern 4 (figures 4.3, 4.4, and 4.6). The omega pattern itself bears a striking resemblance to pattern 2, though of lower magnitude (and less sign agreement). Members of this pattern show very consistent upwards motion from temperature advection but have less agreement in omega from vorticity than all other patterns except pattern 0. This is despite having similar composite absolute values of the two omega fields (not shown). Patterns 1 and 2 look fairly different when comparing their upper and surface highs and lows (figure 4.4). The only point of similarity is that in both patterns 1 and 2 the PEx grid cell is located around the zero contour of the 300 hPa height anomaly field. By contrast pattern 1 strongly resembles pattern 0 in terms of upper and lower level anomaly fields, CAPE amount, shape of frontal zone, and precipitation area (figure 4.4). Pattern 0 and 1 also appear similar in the fields depicted in figure 4.6 but the increased size of the warm anomaly leads to the presence of warm air advection through the PEx grid cell. This warm air advection has a more southerly flow in pattern 1 compared to a more southwesterly flow in pattern 2.

4.4.4 Frontal Cyclones: Patterns 3, 6, 7, and 8

The members of this "frontal cyclones" group are all characterized by two things in the SOM omega patterns: a transition from rising to sinking motion forced by temperature advection near time 0 and rising motion during and after the event forced by vorticity advection centered near 500 hPa (figure 4.3). The differences between the patterns in this category are primarily in the timing of the peak precipitation rate relative the passage of the frontal cyclone. The associated cold front shows up distinctly in the omega patterns as the transition from rising to sinking motion in the temperature advection subpanels mentioned previously. These differences in timing are captured, not only in the composite values of omega, but are also reflected in the areas of maximum sign agreement among the various members. The key features in the omega fields (upper level rising motion due to vorticity advection and sinking motion due to temperature advection) each have at least 75% sign agreement. In fact, pattern 3 is the only frontal cyclone pattern without 90% sign agreement over some part of these key features. In patterns 3 and 6 the peak precipitation rate occurs prior to the cold front as judged by upward omega during the peak period in figure 4.3. However, in pattern 7 the peak precipitation is found overlapping the transition from rising to sinking motion and in pattern 8 the peak precipitation rate happens behind the cold frontal passage (sinking in the temperature advection subpanel). Unsurprisingly the patterns with later peak precipitation rates correspondingly have more left skewed precipitation time series (figure 4.5). One other difference to note is the much larger amount of CAPE present in the horizontal composite (figure 4.4) indicating that pattern 3 is substantially enhanced by convection compared with the other patterns in this group. Each pattern's PEx grid cell is approximately coincident with the center of the surface low with the center of the upper level low to the west (figure 4.4).

4.4.5 Occluded Cyclones: Patterns 4 and 5

These patterns both show a distinct maxima in rising motion forced by temperature advection around 700 hPa followed in time by one around 300 hPa, as well as strong rising motion from vorticity advection associated with a cyclone passage (figure 4.3). The peak precipitation in pattern

4 happens earlier, relative to the described features, than in pattern 5. This conclusion matches with the precipitation rate data in figure 4.5, which shows significantly less precipitation in the hours after the peak in pattern 5 compared to pattern 4. In both patterns peak warm air advection (WAA) shifts from lower to upper levels during the course of the event, which is characteristic of the motion of warm sector air forming an occlusion. That tongue of warm sector air moves cyclonically around the trough axis while rising; as the whole system moves across the PEx grid cell the elevation of the warm air advection driven omega increases and includes a later separate tropopause level component (Hirschberg and Fritsch, 1991). Both the upwards shift in WAA and the peak positive vorticity advection (PVA) are well agreed upon by the member events with $> 75\%$ for the WAA shift and $> 90\%$ for the location of the PVA. The horizontal plots of figure 4.4 show a transition between pattern 4 and pattern 5: for pattern 5 the PEx grid cell is further from the identified frontal zone, west of the surface low center, and northeast of the 300 hPa trough anomaly. A PEx location at an occluded front would have weak or no thermal frontal parameter (Catto and Pfahl, 2013). This strengthens our conclusion that the PEx events in pattern 5 are likely later in the life-cycle of the frontal cyclone than the events in pattern 4 because the frontal activity is further from the center of the event. Patterns 4 and 5 also differentiate themselves from the previous cyclonic category by having a slightly southward tilt with height in the low pressure anomaly. These patterns may also include some closed-low type events, characterized by strong PVA aloft with weaker (than average for this pattern) temperature advection. Upper level closed lows without a clear frontal zone also have upper level PVA occurring simultaneously with low level moisture/temperature advection would be present in these patterns (especially pattern 4). These are also the only two patterns with the PEx grid cell inside the positive relative vorticity anomaly (not shown). Figure 4.6 shows the distinctive comma shape in precipitable water which is also suggestive of an occluded stage not present in other patterns associated with midlatitude cyclones.

4.4.6 Does The Mean Seasonality Predict The Mixture of Omega Patterns?

There are two aspects to assessing the validity of this common assumption. First, we need to check if the mixtures of extreme processes differ between regions. Swenson and Grotjahn (2019) showed

that the mean seasonality of the regions used in this work are statistically distinguishable from one another (except between SGP and FL as already mentioned) so if we cannot tell the mixtures of omega patterns apart between regions then we cannot say that mean seasonality is strongly linked to the population of PEx causing patterns. Second, we need to assess the extent that the regional distributions of extreme processes are representative of the grid cells within each region. The methodologies to make these two assessments are as follows.

The first method is to plot the average number of events that belong to each group of patterns (Convection, Convection Enhanced by Horizontal Moisture Flux, Atmospheric Rivers, Frontal Cyclones, and Occluded Cyclones) during each month of the year (figure 4.10). This is used to assess the inter-regional variations in the distribution of extreme precipitation generating processes. In figure 4.10 we see that in almost every region the relative number of extreme events corresponds to the relative amount of mean precipitation, with the exception being the Southeast. The Southeast's mean and extreme precipitation are almost completely out of phase with one another, as can be seen in figure 4.10 by comparing the black curve to the total stacked area. Each region's stacked area is readily differentiable from each other region; either by shape, size, or color. To quantify these observations we consider the monthly number of events for each process as an individual time series. When we compute the correlation between the same omega pattern across regions we find that every pair of regions has at least two processes with a correlation coefficient below 0.4, including at least one process with a negative correlation. This is clear evidence that even regions with similar seasonal cycles of mean and extreme precipitation (i.e. regions that both have most of their precipitation in the same season) have different mixtures of PEx generating processes. A complete tabulation of these correlations between regions can be found in figure 4.11. While figure 4.11 could be investigated in depth and contains many useful insights about the spatial relationships of extreme precipitation generating processes, it's primary function in this paper is to quantify the way in which each region's mixture of extreme precipitation generating processes differ from one another.

The second method uses the same data as the first to investigate intra-regional variations in the distribution of extreme precipitation generating processes. We start with the number of occurrences

of each omega pattern at each grid cell, separated by month. This gives us a vector with $9 * 12$ elements representing each grid cell. It's straightforward to construct such a vector for each region by averaging all the vectors within the region. We then plot the correlation coefficient between the vector at each grid cell and the regional average (figure 4.12A). This shows strong correlation over much of the CONUS with notable exceptions in the Great Plains, Desert Southwest, Pacific Northwest, and Southwest. These areas of low correlation all are characterized by mountainous terrain. This result is consistent with precipitation being difficult to analyze in mountainous areas (Basist et al., 1994; Daly et al., 1994; Marquınez et al., 2003). Turning to figure 4.12B we see that the areas of weaker correlation from figure 4.12A correspond to areas where the best matching PEx pattern seasonality is from a different region. These areas are small and figure 4.12B is very similar to figure 4.2. These low-correlation areas also correspond to the yellow areas of figure 4.12C which compare the mixture of extreme processes at each grid cell to the average over many random samples of grid cells. This is done for each grid cell by creating 100 samples of X grid cells, where X matches the number of grid cells in the original grid cell's region. These samples are created without replacement from among shaded grid cells in figure 4.2. The percentage of these samples which were a better match that the regional PEx generating pattern seasonality is plotted in figure 4.12C. The yellow areas line up well with the areas of disagreement in figure 4.12B but cover a smaller area in most regions. The areas where this is not the case involve the Northeast. This region's mean and extreme seasonality are both relatively flat over the year compared to that of the other regions. This is a trait also shared by the vast majority of the random samples, meaning that there is increased similarity between a typical random sample and the Northeast region. This is not to say that the Northeast region is a collection of grid cells with vastly different extreme seasonalities because the correlation values in figure 4.12A are too large. Instead this says that for a random sample of grid cells the seasonality is largely smoothed out in the aggregate.

A main area of disagreement in the Pacific Northwest falls between the eastern slope of the Cascade mountain range and the western slope of the Rocky mountains. There we see that the best matching regions are the Southeast and Southwest. Both these regions have a similar overall shape to PEx seasonality to the Pacific Northwest (figure 4.10) but with many fewer of the AR type

of event. This indicates that there is an area with a different mixture of PEx generating processes within the Pacific Northwest, primarily due to the relative lack of AR type events between the mountain ranges. It should be noted that the Southeast and Southwest are the closest to the Pacific Northwest in terms of their mixtures of PEx generating patterns so the disagreement is not large. It should also be noted that the area of disagreement in the Pacific Northwest shrinks from figure 4.12B to 4.12C. This indicates that the winter dominated seasonal cycle provides predictive value above that from a random sample.

Another area of disagreement lies along the eastern border of the Great Plains. There the grid points predominantly are a better match for the extreme process seasonality of the Northeast (figure 4.12B) and are dissimilar enough from the seasonality of the Great Plains to best match a random sample of grid points a majority of the time (figure 4.12C). Figure 4.11 shows a strong correlation between the Great Plains and Northeast for the first six patterns and a weak correlation patterns 6, 7, and 8. This is corroborated by the bottom 3 panels of figure 4.8 where we see the complete lack of those patterns throughout the Great Plains except for the eastern border. Those patterns are also prevalent throughout the Northeast during winter (figure 4.8).

These two examples show that the regions created from mean precipitation seasonality do not perfectly predict the mixture of PEx generating processes. However, they show that for most areas, a consistently similar mean precipitation cycle leads to a consistent mix of PEx processes.

4.5 Discussion

4.5.1 Omega Patterns

Our method of focusing on the QG or large-scale forcing of vertical motions associated with PEx finds more geographic variation in the patterns during winter than summer. This is due to the high proportion of 95th percentile daily precipitation accumulations that are in our "convection" pattern instead of patterns with strong quasi-geostrophically forced vertical motion. This method is very sensitive to distinguishing precipitation events that occur near cyclones (six patterns) and insensitive to different types of triggered convection (one pattern). At the broadest level this method

separates events with notable QG influences from those without (pattern 0). Of the remaining eight patterns, six (patterns 3-8) have evidence of the presence of an extratropical cyclone in the omega field directly. Another (pattern 2) shows a strong cyclonic feature in the horizontal composite which is farther from the event than the previously mentioned patterns. The remaining pattern (1) does not have a strong upper level height anomaly and its most notable feature is the steady upwards motion driven by positive temperature advection throughout the event, especially at low levels, which corresponds with moisture advection from the south inferred from figure 4.6. The six patterns (3-8) with direct evidence of an extratropical cyclone are differentiated by the nature of the frontal feature most active during the peak of precipitation in time. Patterns 3 and 6 peak in precipitation at least three hours before the shift from positive to negative temperature advection (the telltale sign of a cold front) occurs locally. This marks patterns 3 and 6 as including PEx at squall lines ahead of the front in the warm sector (the area between the warm air edges of the warm and cold fronts in an ETC). Patterns 7 and 8 both shift from positive to negative temperature advection between t=0 and t=3 hours indicating that the PEx is occurring near the surface low center or along a cold front. When compared to the "warm sector" patterns 3 and 6, we see that significant precipitation cuts off 6-9 hours earlier in the "cold frontal" patterns 7 and 8. The last cyclonic patterns (4 and 5) exhibit a shift in positive temperature advection from lower to upper levels and show frontal zones away from the center of the event. This supports the label "occlusion". The almost total shutdown of temperature advection after t=3 hours in pattern 5 indicates that the occlusion is coincident with the PEx. Whereas, the shutdown is not captured in pattern 4. Additionally the positive temperature advection is stronger and the frontal zone extends closer to the PEx grid cell in pattern 4 than 5. We say that occlusion happens in both cases but is more clear for pattern 5.

The frequency of each pattern exhibits significant seasonal dependence within the CONUS. This can be seen by comparing figures 4.8 and 4.9. The seasonal dependence differs between patterns. The convective pattern 0 is more prevalent everywhere but the west coast during summer. Patterns 1 and 2 are most active during winter in the west coast regions. While pattern 2 happens very rarely in any of the three east coast regions at all, pattern 1 can be found during both seasons

with some regularity in the NE region and during winter in the SE region. Pattern 1 type events are a predominantly summertime phenomena in the Great Plains and DSW regions. Regardless of season, areas with pattern 2 as the *most frequent* pattern lie amongst areas with pattern 1 as the most frequent (figure 4.13). The frontal cyclone patterns (3, 6-8) occur more often in winter (figure 4.8) than summer (figure 4.9) in the NW, SW, DSW, SE and FL regions. Patterns 3, 6, 7, 8 occur more frequently during summer in the GP region. Pattern 3 also occurs more frequently during summer in the NE region. The remaining two patterns, 4 and 5 occur more commonly during winter except the northern parts of the GP and NE regions.

4.5.2 PEx Regional Climatology

Of the seven regions used (figure 4.2) three experience a plurality of PEx during summertime: the Great Plains (GP), Desert Southwest (DSW), and Florida (FL). The annual cycle of precipitation, figure 4.2, is also larger during those months. Only the Northeast (NE) has the most in Fall (not shown), albeit with a relatively even spread of PEx events among the four seasons (though the annual cycle of precipitation, figure 4.2, is relatively less during the fall there). The remaining three regions each have the largest number of events in the winter. These three regions are the Southeast (SE), Pacific Northwest (PNW), and Southwest (SW). The last of which has more than half of its events during the winter months. The annual cycle of precipitation in the PNW and SW is also larger during those winter months, but not so for the SE region. We will classify the GP, DSW, FL, and NE regions as warm season dominated in terms of extreme precipitation. The remaining regions (SE, PNW, and SW) are cold season dominated for PEx. Only the SE has a peak in PEx events that does not match its peak in mean precipitation seasonality (figure 4.10).

Convection is the most common single pattern of PEx annually over the CONUS and with at least one grid cell in every region (figure 4.13, top panel) and season. During summer, convection is the most common event type, of the 9, over the vast majority of the CONUS, with a few notable exceptions: annually and during summer, portions of the GP and NE regions have convectively-modified frontal cyclone (pattern 3) as most common. But as noted above, several of the patterns are logically grouped: 4 with 5, and 3 with 6, 7, and 8. When the frequencies of these groups are

examined, figure 4.14, the apparent dominance of convection (pattern 0) is much lessened in favor of the four SOMs comprising the frontal cyclone group. Large areas in both the GP and NE regions have the frontal cyclones group as their most frequent driver of PEx events overall, and they retain their status over a subset of these area even during summer.

The PNW is heavily influenced by events with notable horizontal moisture fluxes (patterns 1 and 2), especially during fall and winter. The coastal area is particularly dominated by pattern 2, which produces an upper level height field (figure 4.4) very analogous to the composites made by Gao et al. (2014) for PEx events along the Northern Pacific Coast. In their other regions the comparison is less clear either because we find a mix of event types to be influential or because their horizontal domain differs substantially from ours. During summer, only a scattering of the PNW grid cells have PEx most frequently by cyclones and moisture flux while at most locations PEx events are by convection.

The SW experiences more than half of its extreme events during winter. During winter the most frequent type of event varies zonally: the west edge is primarily pattern 1, moving east, inland and southern California PEx events are most frequently driven by frontal cyclones, and further east, southern Nevada has a plurality of occluded cyclones (figure 4.14). Our method appears to be capturing different timings of the extreme precipitation relative to the evolution of a weather system. From west to east the system might advect moisture into the coastal area, develop a cold front over the central valley and, finally, occlude and decay over southern Nevada. Much of central California during summer has the most PEx events from occluding or occluded cyclones whether all 9 patterns or only the 5 groups are viewed. As mentioned above, these are upper level closed lows from cold air aloft which creates potential instability soundings and triggers thunderstorms beneath.

The GP region sees a clear seasonal cycle to the type of event which brings PEx. In summer frontal cyclones dominate the upper middle of the region, surrounded by areas where convective events are most frequent. Spring and fall see a relative shift and expansion east and south in the area dominated by frontal cyclones and the introduction of an area heavily influenced by occluded cyclones. During winter much of the region is most frequently the occlusion group with much of

the rest being the frontal cyclone group. Finally, there is some apparent spill over from the PNW of atmospheric river (pattern 2) and convection / moisture fluxes (pattern 1) in the far northwest corner of the region.

The northern tip of the DSW is most influenced by the atmospheric river and convection / moisture flux patterns. The southern part of the DSW is dominated by purely convective PEx events except during winter which accounts for only 15% of the annual total of PEx events there. During summer, when most PEx events occur there, the northern part of the DSW is most heavily influenced by patterns with substantial moisture flux. These are mostly pattern 1 with a small area dominated by pattern 2. During the other three seasons grid cells in this area have a mixture of frontal cyclone and occluded type events as the most frequent type.

The NE region is strongly influenced by the frontal cyclone group nearly year round (figure 4.10), with the exception being convection in the western half during the summer (figure 4.14). The exact make up of these cyclones changes seasonally, with winter having the most diverse group of most frequent patterns scattered over the grid cells (figure 4.13).

The SE is a bit of an outlier in terms of seasonality of extreme precipitation patterns. The region experiences the highest average daily precipitation in summer (inset of figure 4.2) yet winter has the most extreme events of any season (figure 4.10), which matches with the PEx seasonality found by Gao et al. (2014). This seems to be because warm season convection occurs frequently but increased frontal cyclone activity during winter and spring months generates the larger fraction of PEx events. However, the annual cycle of precipitation, figure 4.2, is lowest during winter and spring. PEx events during the SE summer and fall are most frequently caused by pattern 0: convection. It is also the only region that shows a substantial difference in the most frequent type of event associated with PEx between spring and fall (figure 4.14).

The FL region is dominated by the convection pattern annually and in every season. This result is easily anticipated from figures 4.7, 4.8, and 4.9. As mentioned previously tropical cyclones will be predominantly found as members of the convective pattern. The work of Fischer et al. (2017) indicates that the upper level QG omega field of a tropical cyclones is highly asymmetric and somewhat weaker in magnitude relative to our patterns. This indicates that depending on the

location of the PEx relative to the tropical cyclone center, some additional tropical cyclone related PEx events may be found in patterns 1 and 3. The presence of those patterns (0, 1, and 3) in FL during summer, as well as the near total lack of other patterns, lends credence to this idea.

4.6 Conclusions

In this paper we have captured and described the nine representative time vs pressure patterns of vertical velocity as they relate to PEx events over the CONUS. A SOM analysis was made by focusing on QG-forced vertical velocity. The SOM analysis finds multiple different varieties of frontal and cyclonically driven events while grouping primarily convective events all together. Our "convective" events have little signature in the QG forced omega fields but are characterized by the strongest local CAPE. The other 8 patterns all have notable QG forced omega signatures, of which the key features discussed are well agreed upon by the members of those patterns. The 9 patterns can be further reduced to 5 groups based on the overall synoptics where some of the SOMs are analogous but differ in the timing of the QG forcing. From our analysis we learn where extreme precipitation arises from strong QG forcings and where these forcings are usually absent. Only in two regions (FL and DSW) are PEx events predominantly generated *without* strong QG forcings, but even this is only true seasonally in DSW. In four other regions (GP, NE, PNW, and SW) the fraction of PEx events generated without strong QG forcings is around one quarter; most PEx events have strong QG forcing. The Southeast region sits between these two collections of regions as around 40% of PEx events occur without strong QG forcings.

In addition to creating a descriptive method of allowing vertical velocity to show us the primary meteorological process driving extreme precipitation, we also examined the idea that *the seasonality of mean precipitation is predictive of the distribution of extreme precipitation generating mechanisms*. Figure 4.12 shows that for most of the CONUS there is a definite relationship between the mean and extreme process seasonalities. In figure 4.12C, we see that there are areas with a stronger correlation to a set of random grid points than to the region they belong to. These areas are by no means the majority but they do show that the relationship between mean seasonality and the distribution of extreme precipitation generating mechanisms is not always one-to-one,

especially near topography. While mean seasonality is somewhat predictive of the distribution of extreme precipitation mechanisms, further study will be required to definitively reject or confirm our hypothesis.

Acknowledgements This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research program under Award DE-SC0016605 "A framework for improving analysis and modeling of Earth system and intersectoral dynamics at regional scales." This project took advantage of netCDF software developed by UCAR/Unidata (http://doi.org/10.5065/D6H70CW6). The authors would also like to thank Drs. Da Yang, Matthew Igel, Paul Ullrich, Michael Wehner, and Xiaoye Li for the time they took to provide feedback during the duration of this project.

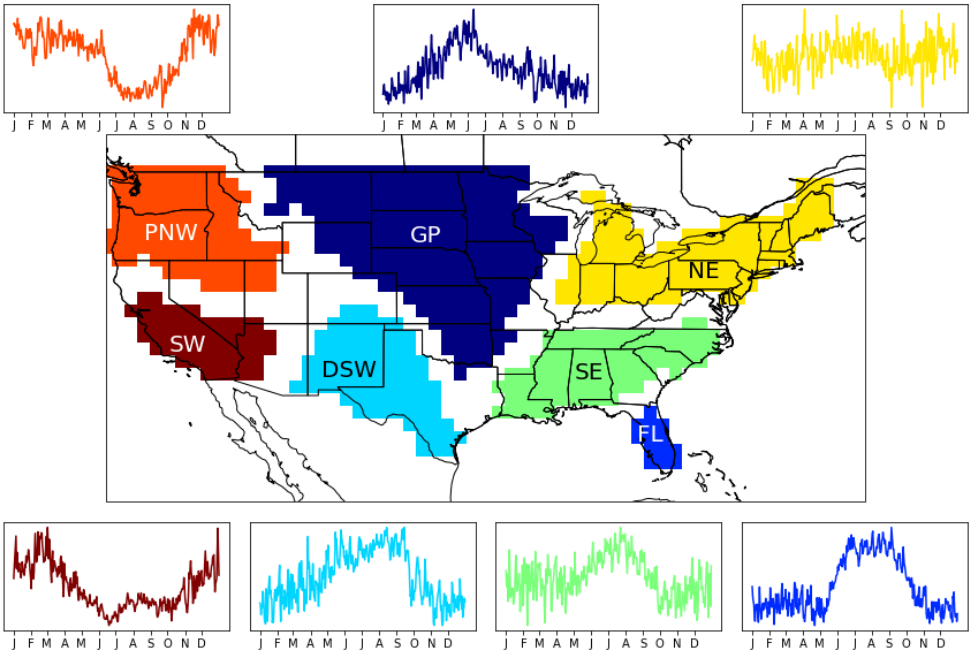

Figure 4.2: The 7 regions of similar annual cycle by which our results may be grouped. These regions result from the approach in Swenson and Grotjahn (2019) for ERA5 daily precipitation data. Abbreviations for the 7 regions are as follows: Great Plains (GP), Florida (FL), Desert Southwest (DSW), Southeast (SE), Northeast (NE), Pacific Northwest (PNW), Southwest (SW). The mean annual cycle of precipitation averaged over all grid cells within each region is displayed in the subplot matching the color of the region. The y-axis of these subplots is precipitation amount normalized to the region, with the limits chosen separately to span the normalized data for that region. The x-axis is day of the year with tick marks at the first day of every month.

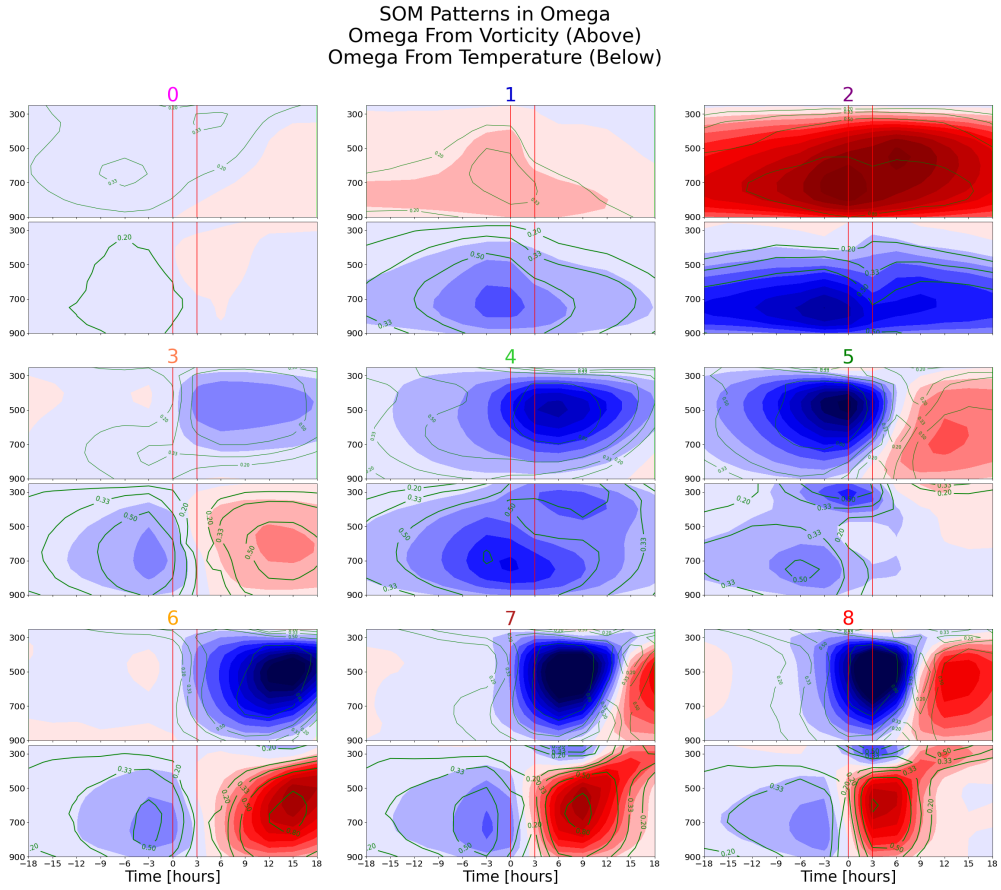

Figure 4.3: These are the nine SOM patterns. In each of the 9 panels the upper sub-panel is the omega forced by differential vorticity advection. The lower sub-panel is the omega forced by temperature advection. Blue filled contours are negative (upwards) omega (vertical velocity), red filled contours are positive (downwards) omega. The green contours indicate the degree to which the individual member events of each pattern share the sign of the displayed pattern. The contour values are calculated by subtracting the number of disagreements from agreements and dividing by the total number of events. The contours are at 0.2, 0.33, 0.5, and 0.8. These correspond to 50, 66, 75, and 90 percent agreement. The vertical red lines at times 0 and 3 bound the time period the PEx event is centered on. The color of the title text corresponds to the color bar on figure 4.13.

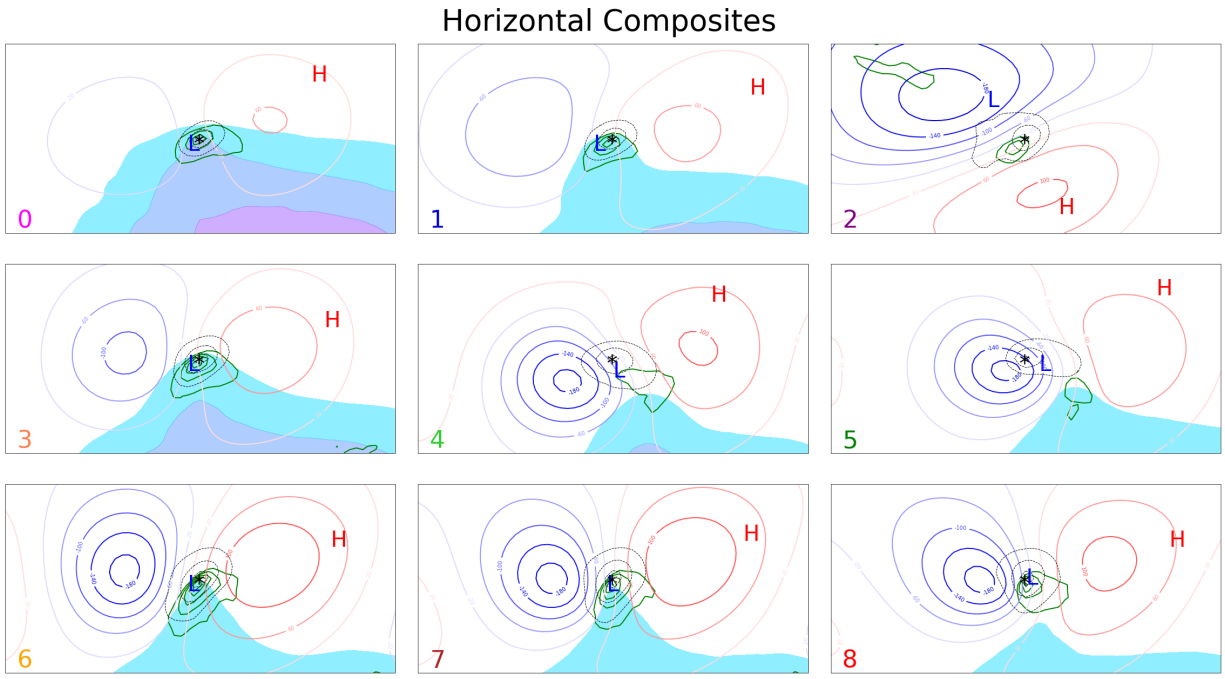

Figure 4.4: Horizontal composites (30 degrees latitude by 60 degrees longitude) at time 0 for the member events of each of the omega patterns. Red (blue) lines correspond to high (low) 300hPa height anomalies. The blue "L" is the center of the sea level pressure negative anomaly field. The red "H" is the center of the sea level pressure positive anomaly field. The green contours are of the thermal frontal parameter (Catto and Pfahl, 2013). The axis of relative maximum corresponds to a frontal feature. The dotted contours are 3 hour precipitation accumulation with contour intervals of 1.5, 3, 6, and 12mm. The shaded contours are Convective Available Potential Energy (CAPE) in intervals of 250 J kg-1. The "*" indicates the location of the grid cell experiencing the extreme event. The color of the title text corresponds to the color bar on figure 4.13.

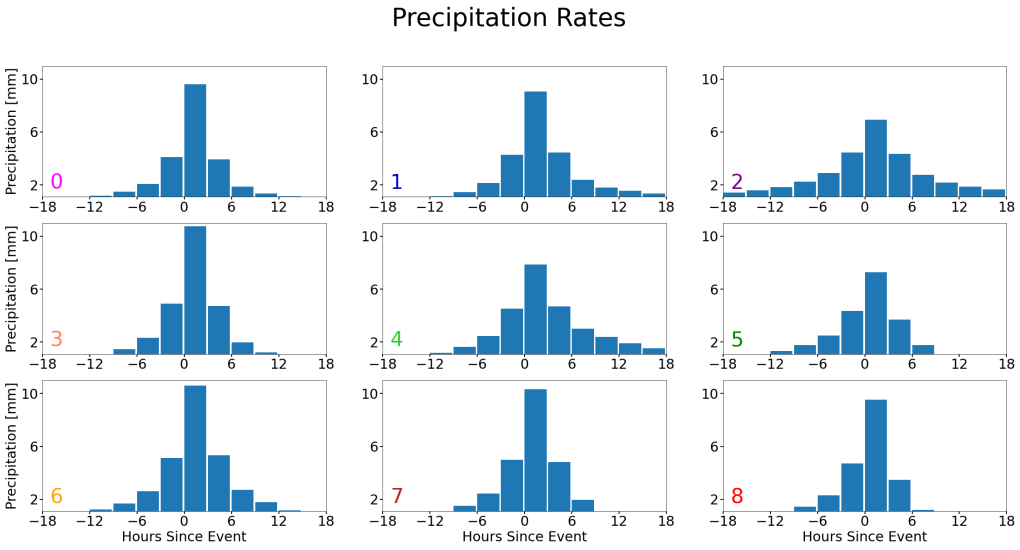

Figure 4.5: Composite time series of three hour precipitation for each pattern. Precipitation values below 1mm are not shown. Vertical red lines bound the time period the PEx event is centered on. The color of the title text corresponds to the color bar on figure 4.13.

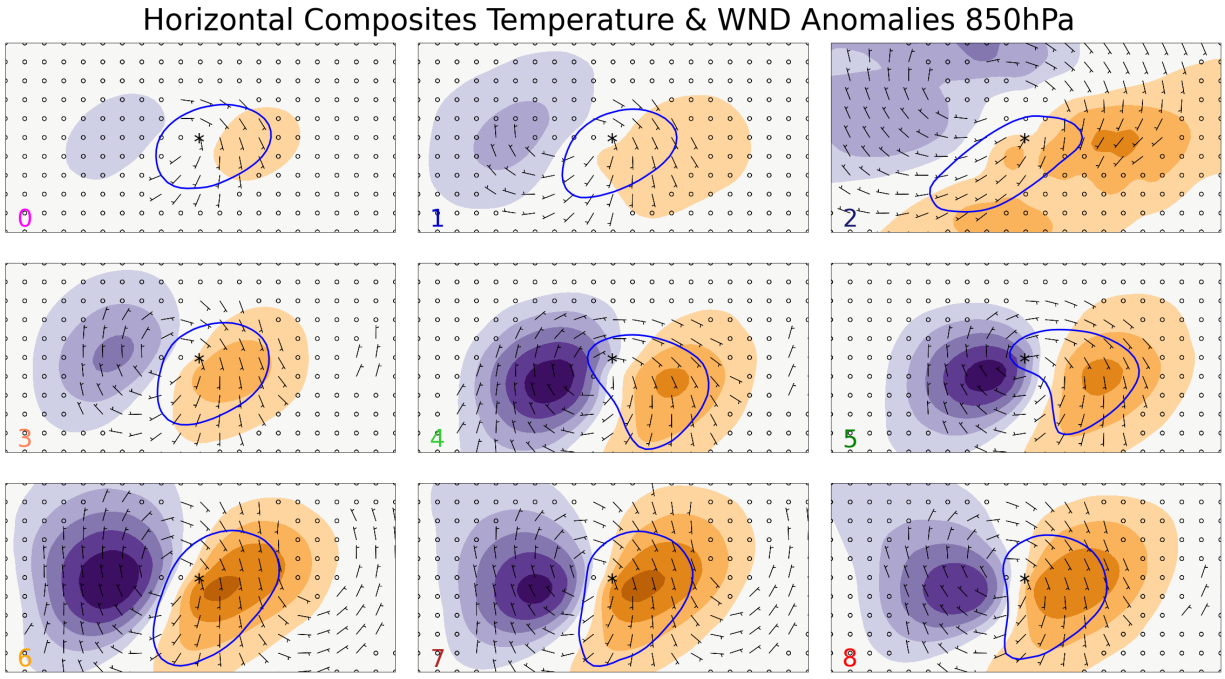

Figure 4.6: Horizontal composites (30 degrees latitude by 60 degrees longitude) at time 0 for each of the omega patterns. Orange (purple) filled contours correspond to positive (negative) 850 hPa temperature anomalies. The contour interval is 1 Kelvin with the 0 contour skipped. Wind barbs show the direction of the 850 hPa horizontal wind anomalies for areas where the anomalous velocity is greater than 5 m/s. The blue contour indicates that the total column water (TCW) anomaly is greater than 3mm. The color of the title text corresponds to the color bar on figure 4.13.

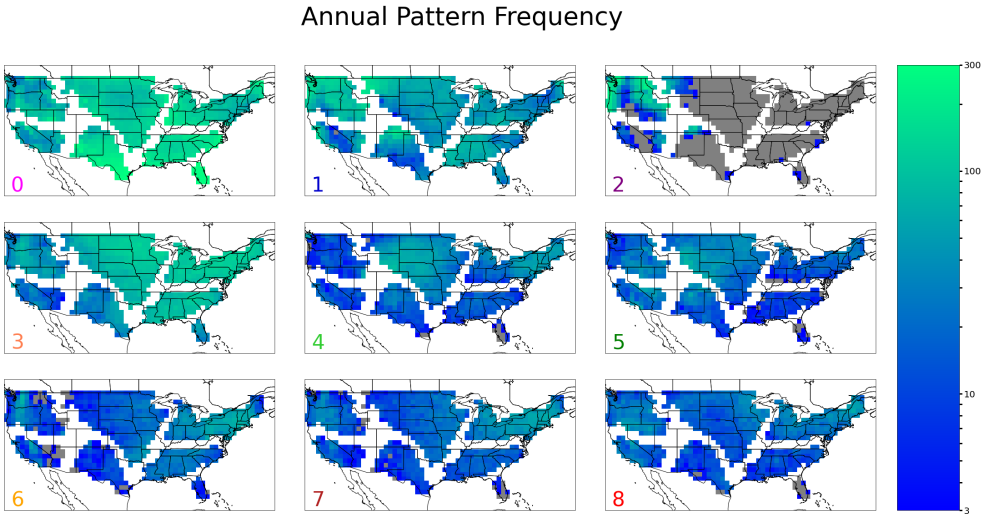

Figure 4.7: The number of times each pattern occurs at each grid cell considered. The color scale is logarithmic and grid cells with less than 3 total events of a given pattern are displayed as gray. The color of the title text corresponds to the color bar on figure 4.13

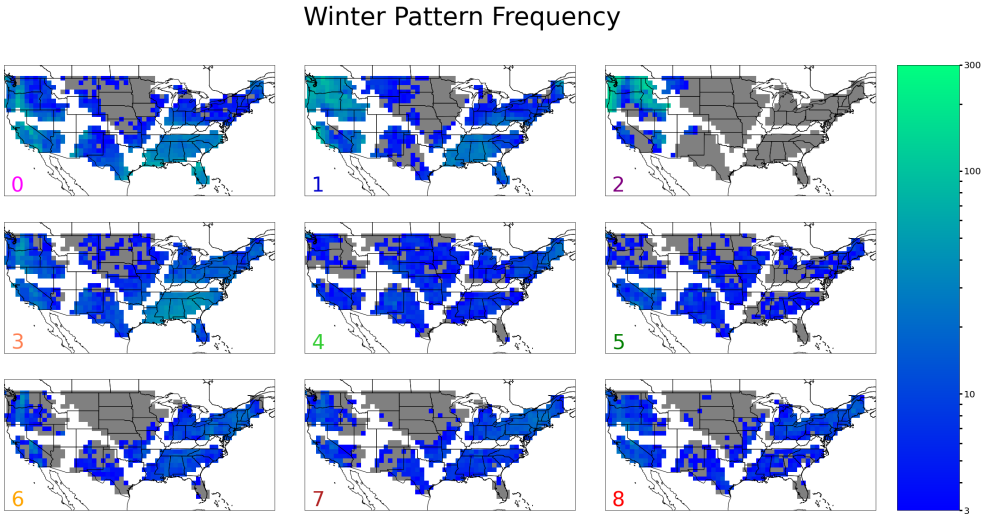

Figure 4.8: Same as figure 4.7 but for the winter months December-February (DJF).

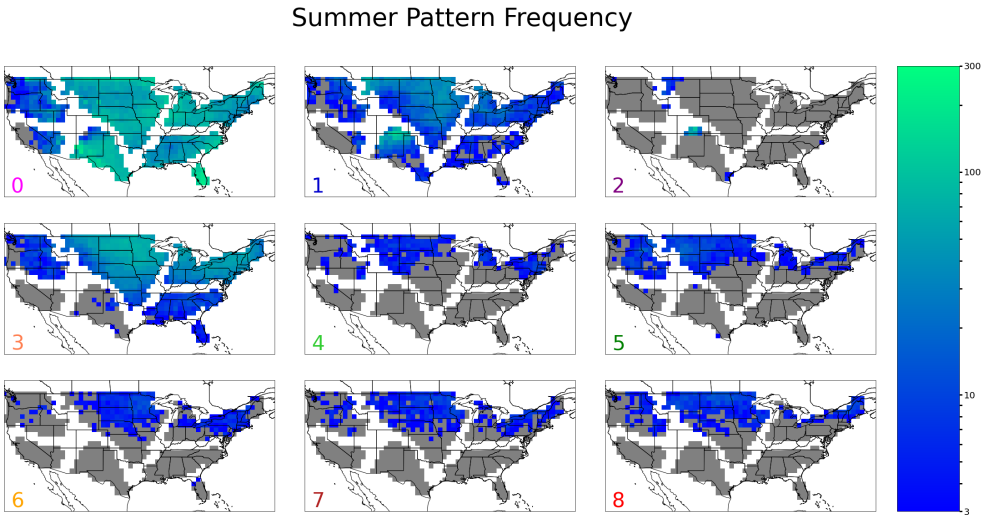

Figure 4.9: Same as figure 4.7 but for the summer months June-August (JJA).

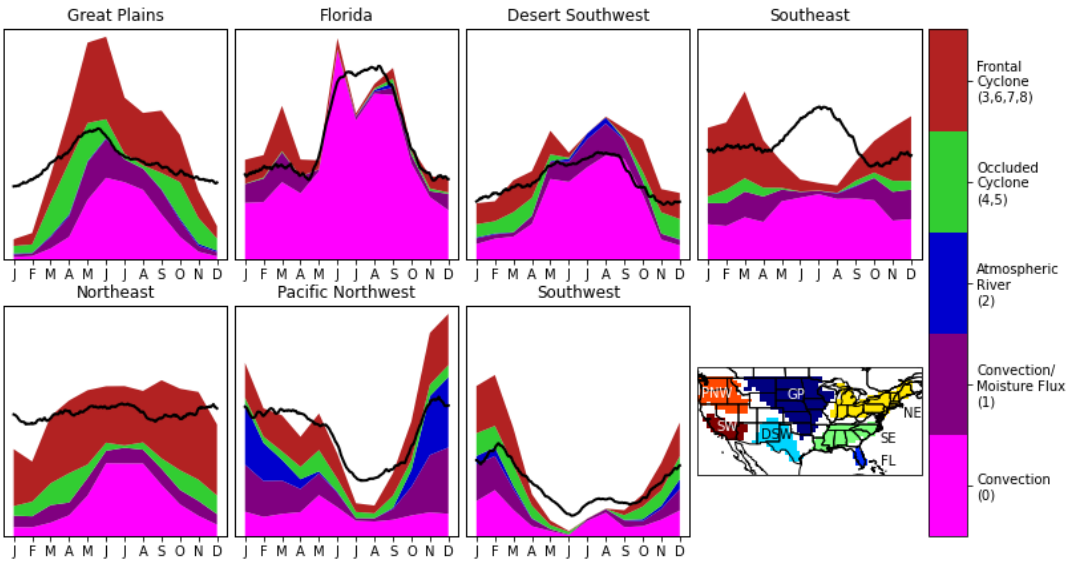

Figure 4.10: Each subplot is a stacked area graph with the average number of events of each type on the y-axis for each month of the year along the x-axis. The number of events of each type is represented by the height of the corresponding color at each month and the y-axis associated with this feature ranges from 0 to 65 events / grid cell. The overlayed black line is a smoothed version of the mean seasonality from figure 4.2 and the y-axis associated with this feature ranges from 0 to 0.33 normalized precipitation units.

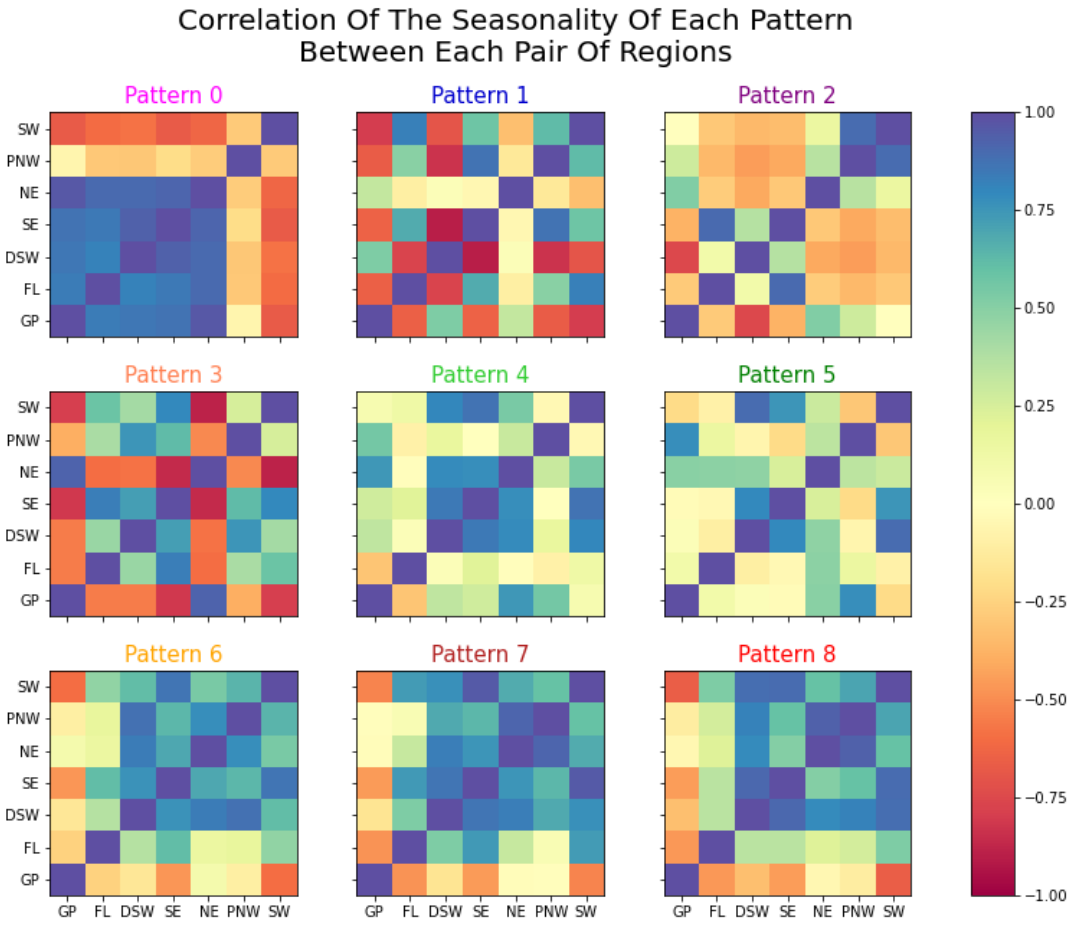

Figure 4.11: This plot displays the correlation coefficient between the seasonality of each omega pattern between each pair of regions. Each panel displays the correlation between each pair of regions for the omega pattern indicated in the panel's title. The x and y axis of these panels both show the region's short names from figure 4.2. The seasonality of each omega pattern in a region is calculated by finding the total number of occurrences in that region in each month and dividing by the number of grid cells in the region. This gives it a time series of length 12 for each region and omega pattern

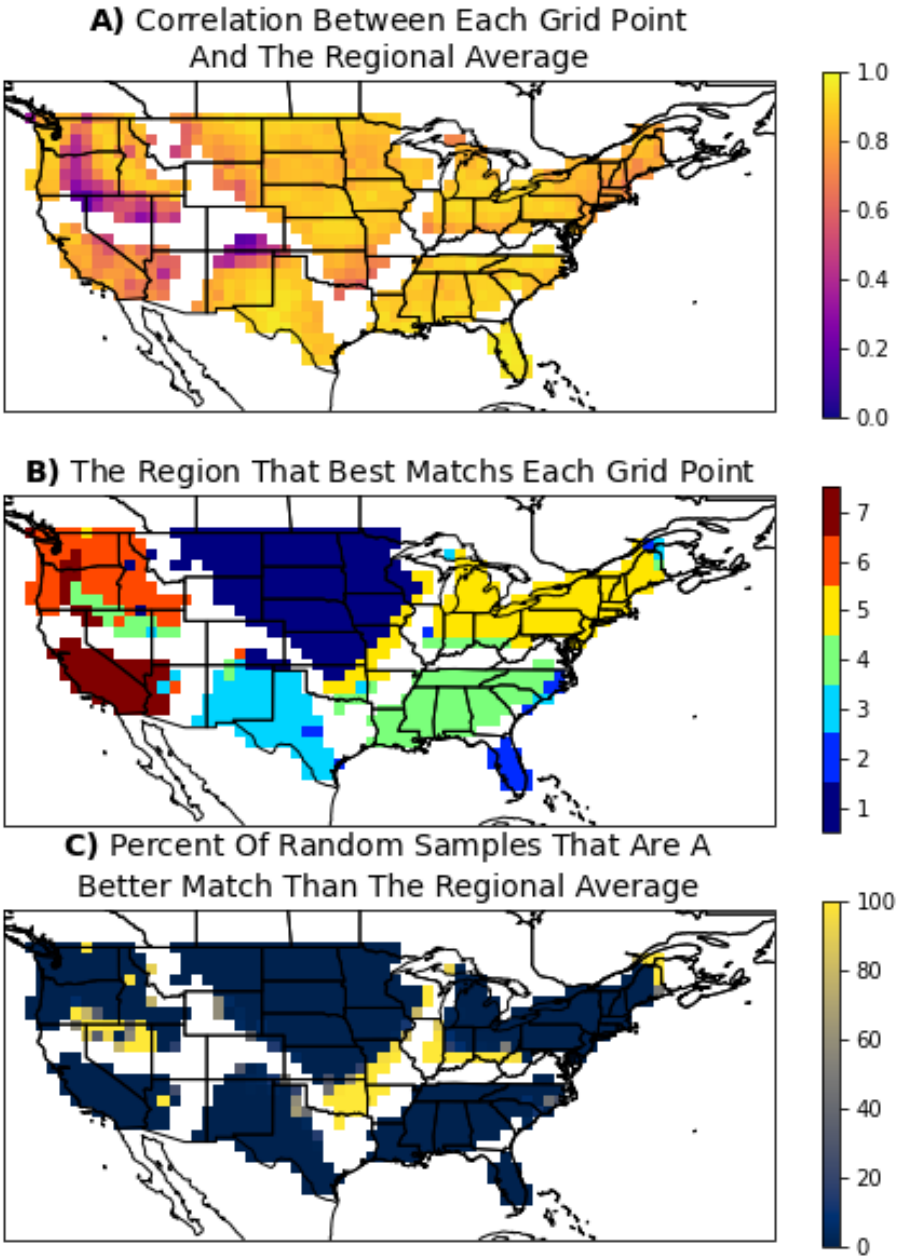

Figure 4.12: We create a vector containing the number of extreme events of each type during each month for each grid cell. **A)** Shows the correlation coefficient between the vector from each grid cell and the average of the vectors from all grid cells within that region. **B)** Shows which regional average vector has the highest correlation coefficient with the vector from each grid cell. **C)** Shows the percentage of random samples of grid cells with an average vector that has a higher correlation coefficient with the vector from grid cell than its own region's average vector.

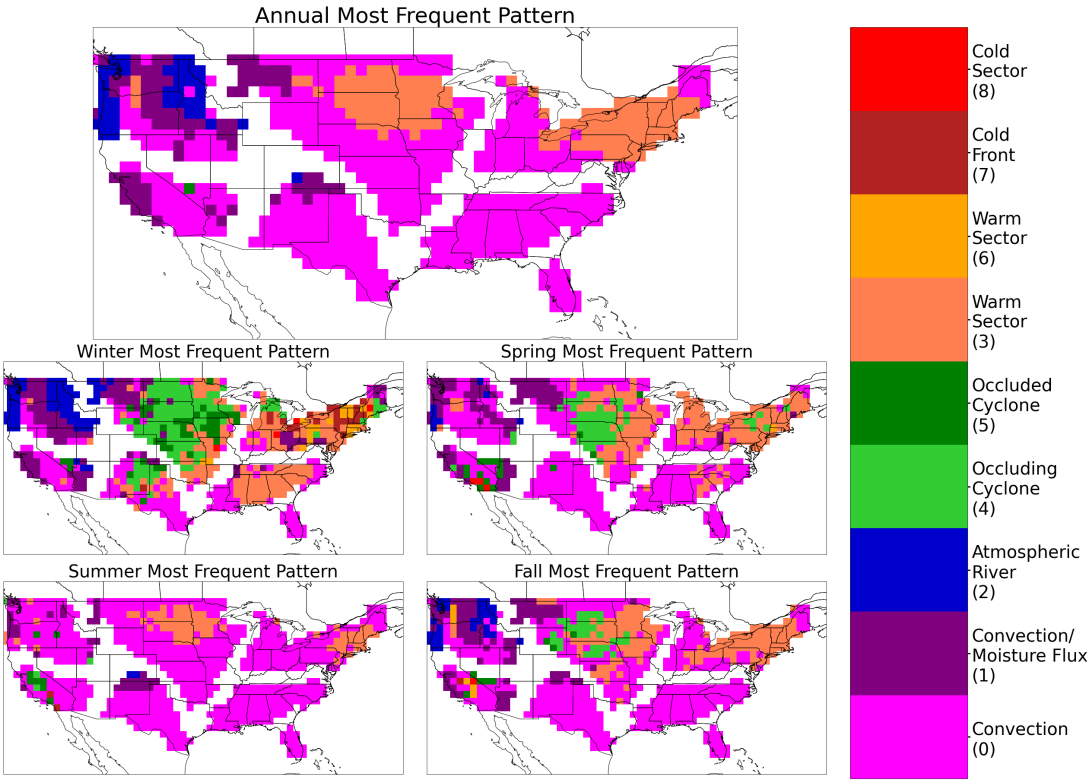

Figure 4.13: Most frequently occurring pattern at each grid cell over the entire time period (center top). The other panels reflect the same information but for a single season, as indicated in the subtitle. Winter is again months DJF, spring is MAM, summer is JJA and fall is SON. Pattern number is indicated by color, which matches the titles on figures 4.3-4.9, and the label on the color bar.

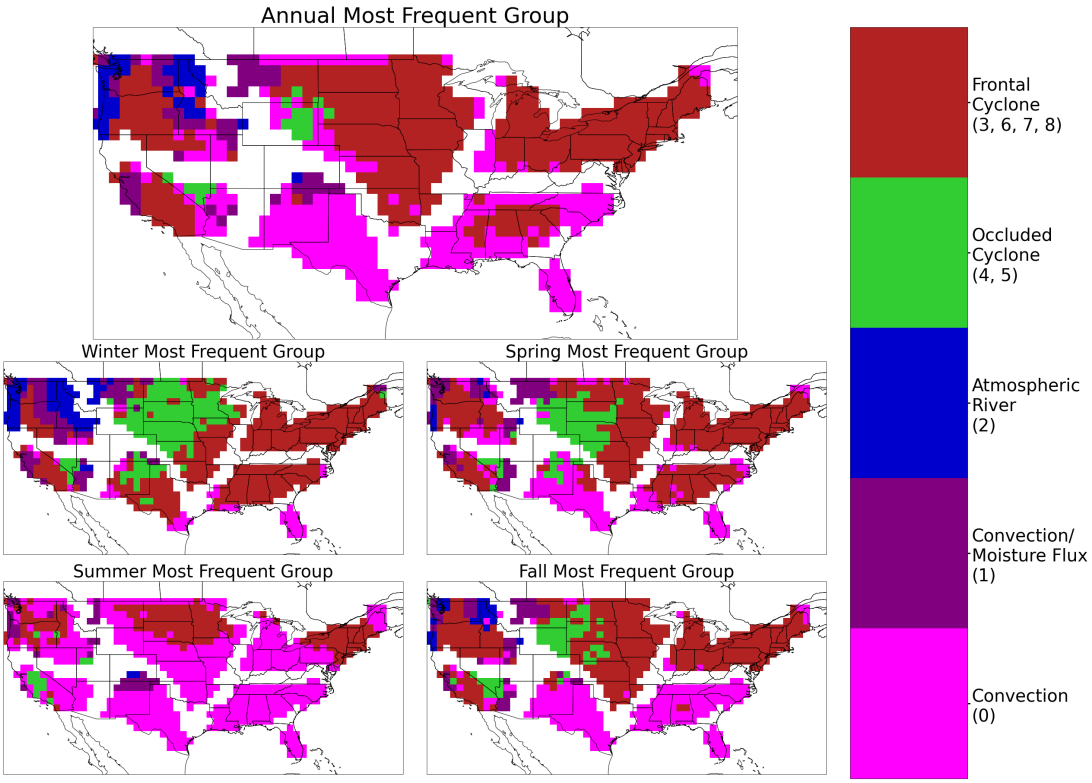

Figure 4.14: Similar to figure 4.13 except for the most frequently occurring *group* of patterns at each grid cell over the entire time period (center top). The other panels reflect the same information but for a single season, as indicated in the subtitle. Groups are as follows: Convection - pattern 0, Convection/Moisture Flux - pattern 1, Atmospheric River - pattern 2, Occluded Cyclone - pattern 4 and 5, Frontal Cyclone- patterns 3, 6, 7, and 8

Chapter 5 Conclusions

This work was motivated by the desire to create an automatic scheme to identify the meteorological process most responsible an extreme precipitation event. Studying extreme events often necessitates aggregation over a larger area to build statistics that one can be confident in. We found that commonly used regions based on political boundaries (Karl and Knight, 1998) were less than fully satisfactory because the atmosphere doesn't care about those boundaries. Other (non-political) regionalizations have been made but they were tailored for a different task and therefor didn't meet our needs for several reasons. Because it is often assumed precipitation and the processes that drive it can be highly seasonal we created regions based on the long term seasonality of precipitation (Karl and Knight, 1998; Kunkel et al., 2012). This groups areas that experience their rainy seasons at the same time together. We also introduced several novel criteria for selecting the appropriate number of regions for any gridded precipitation data set. These regions also allowed us to test the hypothesis that *the seasonality of mean precipitation is predictive of the distribution of extreme precipitation generating mechanisms* in chapter 4.

With regions that we are confident in aggregating within, work continued on an automatic scheme to identify the meteorological process most responsible an event. We selected metrics for three key atmospheric processes: fronts, vorticity, and convection. Along the way we showed that these simple and physically motivated metrics are very interdependent so these metrics are unable to isolate specific processes. This was shown in chapter 3 with extratropical cyclones and their associated fronts, as well as, for fronts and convective storms like squall lines. Consistently choosing the "correct" process is fraught with challenges, not least of which is certain fronts may not have existed without the presence of their associated extratropical cyclone. An additional wrinkle occurs when a process like a tropical cyclone moistens a distant environment. This can enhance existing precipitation, making it extreme when it might not have been without the distant cyclone. If the increased moisture runs up against an unusually cold and/or dry air mass one can see heavy

precipitation arising just from this moisture front. The front itself would not have existed without the moisture supplied by the tropical cyclone.

This interdependence led us to use vertical velocity to describe the primary meteorological process creating each extreme event. This is a natural choice of variable because extreme precipitation can be thought of, and scaled, as a combination of vertical velocity and atmospheric moisture O'Gorman and Schneider (2009). Our interest is in extreme precipitation events so every event we look at has a lot of moisture by definition. This leaves us to consider the vertical velocity field, which nicely represents dynamical contributions to extreme rainfall. We found that the two main forcings of the omega equation should be considered separately, rather than in sum, to display the shifts in time of one's importance relative to another. This approach led to our identification of the set of 9 most representative time vs pressure patterns in both omega forcings. These 9 patterns map onto familiar large scale meteorological patterns which were described and analyzed in chapter 4. This work also let us test our hypothesis that *the seasonality of mean precipitation is predictive of the distribution of extreme precipitation generating mechanisms* in conjunction with the regions created in chapter 2. We found that, *while not perfectly predictive, mean seasonality is clearly related to the distribution of extreme precipitation generating mechanisms over most of the CONUS.*

In the future I hope for the opportunity to take this framework of regionalization and omega equation analysis into the realm of climate models. This type of analysis could be used to investigate and quantify shifts in precipitation seasonality due to changes in climate. Differences in the number, type, and distribution of the patterns could serve as a useful tool for model evaluation while describing the dynamical factors giving rise to extreme precipitation in various models. Examining how these patterns change in future climate could advantageously change our understanding of which weather patterns are more likely to be associated with heavy rainfall and aid in describing how these dynamical process are affected by the climate state. Additionally, defining regions based on seasonality elsewhere in the world may may give key insights into the physical mechanisms that define precipitation seasonality.

Reference

Adams, D. K. and Comrie, A. C.: The North American Monsoon, Bulletin of the American Meteorological Society, 78, 2197–2213, https://doi.org/10.1175/1520-0477(1997)078⟨2197:TNAM⟩2.0.CO;2.

Basist, A., Bell, G. D., and Meentemeyer, V.: Statistical Relationships between Topography and Precipitation Patterns, Journal of Climate, 7, 1305 – 1315, https://doi.org/10.1175/1520-0442(1994)007⟨1305:SRBTAP⟩2.0.CO;2, 1994.

Battalio, M. and Dyer, J.: The Minimum Length Scale for Evaluating QG Omega Using High-Resolution Model Data, Monthly Weather Review, 145, 1659–1678, https://doi.org/10.1175/MWR-D-16-0241.1, 2017.

Benestad, R. and Chen, D.: The use of a calculus-based cyclone identification method for generating storm statistics, Tellus A: Dynamic Meteorology and Oceanography, 58, 473–486, https://doi.org/10.1111/j.1600-0870.2006.00191.x, 2006.

Berg, P., Moseley, C., and Haerter, J. O.: Strong Increase in Convective Precipitation in Rsponse to Higher Temperatures, Nature Geoscience, 6, 181 – 185, https://doi.org/10.1038/ngeo1731, 2013.

Bjerknes, J.: ON THE STRUCTURE OF MOVING CYCLONES, Monthly Weather Review, 47, 95–99, https://doi.org/10.1175/1520-0493(1919)47⟨95:OTSOMC⟩2.0.CO;2, 1919.

Bukovsky, M.: Masks for the Bukovsky regionalization of North America, Regional Integrated Sciences Collective, Institute for Mathematics Applied to Geosciences, National Center for Atmospheric Research, Boulder, CO. Downloaded 2018-12-24. [http://www.narccap.ucar.edu/contrib/bukovsky/], 2011.

Catto, J. L. and Pfahl, S.: The importance of fronts for extreme precipitation, Journal of Geophysical Research Atmospheres, 118, 10 791–10 801, https://doi.org/10.1002/jgrd.50852, 2013.

Catto, J. L., Jakob, C., Berry, G., and Nicholls, N.: Relating global precipitation to atmospheric fronts, Geophysical Research Letters, 39, 1–6, https://doi.org/10.1029/2012GL051736, 2012.

Chen, M., Shi, W., Xie, P., Silva, V. B. S., Kousky, V. E., Higgins, R. W., and Janowiak, J. E.: CPC Unified Gauge-BasedAnalysis of Daily Precipitation over CONUS, V1.0. EarthScience Research Lab Physical Sciences Division, accessed1 October 2016, https://psl.noaa.gov/data/gridded/data.cpc.globalprecip.html, 2008a.

Chen, M., Shi, W., Xie, P., Silva, V. B. S., Kousky, V. E., Higgins, R. W., and Janowiak, J. E.: Assessing objective techniques for gauge-based analyses of global daily precipitation, Journal of Geophysical Research, 113, D04 110, https://doi.org/10.1029/2007JD009132, 2008b.

Clarke, L. C. and Renard, R. J.: The U.S. Navy Numerical Frontal Analysis Scheme: Further Development and a Limited Evaluation, 1966.

Collow, A. B., Mersiovsky, H., and Bosilovich, M. G.: Large-scale influences on atmospheric river–induced extreme precipitation events along the coast of Washington State, Journal of Hydrometeorology, 21, https://doi.org/10.1175/JHM-D-19-0272.1, 2020.

Daly, C., Neilson, R. P., and Phillips, D. L.: A Statistical-Topographic Model for Mapping Climatological Precipitation over Mountainous Terrain, Journal of Applied Meteorology and Climatology, 33, 140 – 158, https://doi.org/10.1175/1520-0450(1994)033⟨0140:ASTMFM⟩2.0.CO;2, 1994.

Dowdy, A. J. and Catto, J. L.: Extreme weather caused by concurrent cyclone, front and thunderstorm occurrences, Scientific Reports, 7, 1–8, https://doi.org/10.1038/srep40359, 2017.

Fischer, M. S., Tang, B. H., and Corbosiero, K. L.: Assessing the Influence of Upper-Tropospheric Troughs on Tropical Cyclone Intensification Rates after Genesis, Monthly Weather Review, 145, 1295 – 1313, https://doi.org/10.1175/MWR-D-16-0275.1, 2017.

Gao, X., Schlosser, C. A., Xie, P., Monier, E., Entekhabi, D., Gao, X., Schlosser, C. A., Xie, P., Monier, E., and Entekhabi, D.: An analogue approach to identify heavy precipitation events: Evaluation and application to CMIP5 climate models in the United States, Journal of Climate, 27, 5941–5963, https://doi.org/10.1175/JCLI-D-13-00598.1, 2014.

Grotjahn, R.: Identifying extreme hottest days from large scale upper air data: A pilot scheme to find California Central Valley summertime maximum surface temperatures, Climate Dynamics, 37, 587–604, https://doi.org/10.1007/s00382-011-0999-z, 2011.

Grotjahn, R. and Faure, G.: Composite Predictor Maps of Extraordinary Weather Events in the Sacramento, California, Region*, Weather and Forecasting, 23, https://doi.org/10.1175/2007WAF2006055.1, 2008.

Grotjahn, R. and Zhang, R.: Synoptic Analysis of Cold Air Outbreaks over the California Central Valley, Journal of Climate, 30, 9417–9433, https://doi.org/10.1175/JCLI-D-17-0167.1, 2017.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E.: Array programming with NumPy, Nature, 585, 357–362, https://doi.org/10.1038/s41586-020-2649-2, 2020.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/https://doi.org/10.1002/qj.3803, 2020.

Hewitson, B. C. and Crane, R. G.: Gridded Area-Averaged Daily Precipitation via Conditional Interpolation, Journal of Climate, 18, 41–57, https://doi.org/10.1175/JCLI3246.1, 2005.

Hewson, T. D.: Objective fronts, Meteorological Applications, 5, 37–65, https://doi.org/10.1017/S1350482798000553, 1998.

Hewson, T. D. and Titley, H. A.: Objective identification, typing and tracking of the complete life-cycles of cyclonic features at high spatial resolution, Meteorological Applications, 17, 355–381, https://doi.org/https://doi.org/10.1002/met.204, 2010.

Hirschberg, P. A. and Fritsch, J. M.: Tropopause undulations and the development of extratropical cyclones. Part II: Diagnostic analysis and conceptual model, Monthly weather review, 119, 518–550, https://doi.org/https://doi.org/10.1175/1520-0493(1991)119⟨0518:TUATDO⟩2.0.CO;2, 1991.

Holton, J. R. and Hakim, G. J.: Chapter 6 - Quasi-geostrophic Analysis, in: An Introduction to Dynamic Meteorology (Fifth Edition), edited by Holton, J. R. and Hakim, G. J., pp. 171–211, Academic Press, Boston, fifth edition edn., https://doi.org/10.1016/B978-0-12-384866-6.00006-4, 2013.

Hope, P., Keay, K., Pook, M., Catto, J., Simmonds, I., Mills, G., McIntosh, P., Risbey, J., and Berry, G.: A Comparison of Automated Methods of Front Recognition for Climate Studies: A Case Study in Southwest Western Australia, Monthly Weather Review, 142, 343–363, https://doi.org/10.1175/mwr-d-12-00252.1, 2013.

Inatsu, M.: The neighbor enclosed area tracking algorithm for extratropical wintertime cyclones, Atmospheric Science Letters, 10, 267–272, https://doi.org/https://doi.org/10.1002/asl.238, 2009.

Ison, N. T., Feyerherm, A. M., Bark, L. D., Ison, N. T., Feyerherm, A. M., and Bark, L. D.: Wet Period Precipitation and the Gamma Distribution, Journal of Applied Meteorology, 10, 658–665, https://doi.org/10.1175/1520-0450(1971)010⟨0658:WPPATG⟩2.0.CO;2.

Johnson, N. C.: How Many ENSO Flavors Can We Distinguish?*, Journal of Climate, 26, 4816–4827, https://doi.org/10.1175/JCLI-D-12-00649.1, 2013.

Kampe, T. U., Johnson, B. R., Kuester, M. A., and Keller, M.: NEON: the first continental-scale ecological observatory with airborne remote sensing of vegetation canopy biochemistry and structure, Journal of Applied Remote Sensing, 4, 043 510, https://doi.org/10.1117/1.3361375, 2010.

Karl, T. R. and Knight, R. W.: Secular Trends of Precipitation Amount, Frequency, and Intensity in the United States, Bulletin of the American Meteorological Society, 79, 231–241, https://doi.org/10.1175/1520-0477(1998)079⟨0231:STOPAF⟩2.0.CO;2, 1998.

Kohonen, T.: Self-organized formation of topologically correct feature maps, Biological Cybernetics, 43, 59–69, https://doi.org/10.1007/BF00337288, 1982.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F.: World Map of the Köppen-Geiger climate classification updated, Meteorologische Zeitschrift, 15, 259–263, https://doi.org/10.1127/0941-2948/2006/0130, 2006.

Krishnamurti, T. N.: a Diagnostic Balance Model for Studies of Weather Systems of Low and High Latitudes, Rossby Number Less Than 1, Monthly Weather Review, 96, 197–207, https://doi.org/10.1175/1520-0493(1968)096⟨0197:adbmfs⟩2.0.co;2, 1968.

Kunkel, K. E., Changnon, S. A., and Shealy, R. T.: Temporal and Spatial Characteristics of Heavy-Precipitation Events in the Midwest, Monthly Weather Review, 121, 858–866, https://doi.org/10.1175/1520-0493(1993)121⟨0858:tascoh⟩2.0.co;2, 1993.

Kunkel, K. E., Easterling, D. R., Kristovich, D. A. R., Gleason, B., Stoecker, L., and Smith, R.: Meteorological Causes of the Secular Variations in Observed Extreme Precipitation Events for the Conterminous United States, Journal of Hydrometeorology, 13, 1131–1141, https://doi.org/10.1175/JHM-D-11-0108.1, 2012.

Kunkel, K. E., Stevens, S. E., Stevens, L. E., and Karl, T. R.: Observed Climatological Relationships of Extreme Daily Precipitation Events With Precipitable Water and Vertical Velocity in the Contiguous United States, Geophysical Research Letters, 47, https://doi.org/10.1029/2019GL086721, 2020.

Maddox, R. A.: A Methodology For Forecasting Heavy Convective Precipitation and Flash Flooding, National Weather Digest, 4, 30–42, URL `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.663.8634&rep=rep1&type=pdf`, 1979.

Marquınez, J., Lastra, J., and Garcıa, P.: Estimation models for precipitation in mountainous regions: the use of GIS and multivariate analysis, Journal of Hydrology, 270, 1–11, https://doi.org/https://doi.org/10.1016/S0022-1694(02)00110-5, 2003.

Mearns, L. O., Arritt, R., Biner, S., Bukovsky, M. S., McGinnis, S., Sain, S., Caya, D., Correia, J., Flory, D., Gutowski, W., Takle, E. S., Jones, R., Leung, R., Moufouma-Okia, W., McDaniel, L., Nunes, A. M. B., Qian, Y., Roads, J., Sloan, L., Snyder, M., Mearns, L. O., Arritt, R., Biner, S., Bukovsky, M. S., McGinnis, S., Sain, S., Caya, D., Jr., J. C., Flory, D., Gutowski, W., Takle, E. S., Jones, R., Leung, R., Moufouma-Okia, W., McDaniel, L., Nunes, A. M. B., Qian, Y., Roads, J., Sloan, L., and Snyder, M.: The North American Regional Climate Change Assessment Program: Overview of Phase I Results, Bulletin of the American Meteorological Society, 93, 1337–1362, https://doi.org/10.1175/BAMS-D-11-00223.1, 2012.

Mills, G. A.: A re-examination of the synoptic and mesoscale meteorology of Ash Wednesday 1983, Australian Meteorological Magazine, 54, 35–55, fire events!!, 2005.

Murray, R. J. and Simmonds, I.: A numerical scheme for tracking cyclone centres from digital data. Part I: development and operation of the scheme, Australian Meteorological Magazine, 39, 1991.

Nash, D. and Carvalho, L.: Brief Communication: An electrifying atmospheric river–understanding the thunderstorm event in Santa Barbara County during March 2019, Natural Hazards and Earth System Sciences, 20, 1931–1940, 2020.

Neu, U., Akperov, M. G., Bellenbaum, N., Benestad, R., Blender, R., Caballero, R., Cocozza, A., Dacre, H. F., Feng, Y., Fraedrich, K., Grieger, J., Gulev, S., Hanley, J., Hewson, T., Inatsu, M., Keay, K., Kew, S. F., Kindem, I., Leckebusch, G. C., Liberato, M. L., Lionello, P., Mokhov, I. I., Pinto, J. G., Raible, C. C., Reale, M., Rudeva, I., Schuster, M., Simmonds, I., Sinclair, M., Sprenger, M., Tilinina, N. D., Trigo, I. F., Ulbrich, S., Ulbrich, U., Wang, X. L., and Wernli, H.: Imilast: A community effort to intercompare extratropical cyclone detection and tracking algorithms, Bulletin of the American Meteorological Society, 94, 529–547, https://doi.org/10.1175/BAMS-D-11-00154.1, 2013.

Newton, C. W.: STRUCTURE AND MECHANISM OF THE PREFRONTAL SQUALL LINE, Journal of Atmospheric Sciences, 7, 210 – 222, https://doi.org/10.1175/1520-0469(1950)007⟨0210:SAMOTP⟩2.0.CO;2, 1950.

Nie, J., Shaevitz, D. A., and Sobel, A. H.: Forcings and feedbacks on convection in the 2010 Pakistan flood: Modeling extreme precipitation with interactive large-scale ascent, Journal of Advances in Modeling Earth Systems, 8, https://doi.org/10.1002/2016MS000663, 2016.

O'Gorman, P. A. and Schneider, T.: The physical basis for increases in precipitation extremes in simulations of 21st-century climate change., Proceedings of the National Academy of Sciences of the United States of America, 106, 14 773–7, https://doi.org/10.1073/pnas.0907610106, 2009.

Parfitt, R., Czaja, A., and Seo, H.: A simple diagnostic for the detection of atmospheric fronts, Geophysical Research Letters, 44, 4351–4358, https://doi.org/10.1002/2017GL073662, 2017.

Pauley, P. M. and Nieman, S. J.: A Comparison of QG and Non-QG Vertical Motions for a Model-simulated Rapidly Intensifying Marine Extratropical Cyclone, Monthly Weather Review, 120, 1108–1134, 1992.

Ralph, F. M., Iacobellis, S. F., Neiman, P. J., Cordeira, J. M., Spackman, J. R., Waliser, D. E., Wick, G. A., White, A. B., and Fairall, C.: Dropsonde Observations of Total Integrated Water Vapor Transport within North Pacific Atmospheric Rivers, Journal of Hydrometeorology, 18, 2577 – 2596, https://doi.org/10.1175/JHM-D-17-0036.1, 2017.

Renard, R. J. and Clarke, L. C.: Experiments in Numerical Objective Frontal Analysis 1, Monthly Weather Review, 93, 547–556, https://doi.org/10.1175/1520-0493(1965)093⟨0547:einofa⟩2.3.co;2, 1965.

Rennó, N. O. and Ingersoll, A. P.: Natural convection as a heat engine: A theory for CAPE, Journal of the Atmospheric Sciences, 53, 572–585, https://doi.org/10.1175/1520-0469(1996)053⟨0572:NCAAHE⟩2.0.CO;2, 1996.

Räisänen, J.: Factors Affecting Synoptic-Scale Vertical Motions: A Statistical Study Using a Generalized Omega Equation, Monthly Weather Review, 123, 2447–2460, https://doi.org/10.1175/1520-0493(1995)123⟨2447:fassvm⟩2.0.co;2, 1995.

Saha, S. and Coauthors: NCEP Climate Forecast System Reanalysis (CFSR) Selected Hourly Time-Series Products,January 1979 to December 2010. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, accessed 13 November 2016,https://doi.org/10.5065/D6513W89, 2010a.

Saha, S. and Coauthors: The NCEP Climate Forecast System Reanalysis, Bulletin of the American Meteorological Society, 91, 1015–1058, https://doi.org/10.1175/2010BAMS3001.1, 2010b.

Sanders, F., Hoffman, E. G., Sanders, F., and Hoffman, E. G.: A Climatology of Surface Baroclinic Zones, Weather and Forecasting, 17, 774–782, https://doi.org/10.1175/1520-0434(2002)017⟨0774:ACOSBZ⟩2.0.CO;2, 2002.

Serreze, M. C.: Climatological aspects of cyclone development and decay in the Arctic, Atmosphere-Ocean, 33, 1–23, https://doi.org/10.1080/07055900.1995.9649522, 1995.

Simmonds, I., Keay, K., Bye, J. A. T., and Bye, J. A. T.: Identification and climatology of Southern Hemisphere mobile fronts in a modern reanalysis, Journal of Climate, 25, 1945–1962, https://doi.org/10.1175/JCLI-D-11-00100.1, 2012.

Sinclair, M. R.: An Objective Cyclone Climatology for the Southern Hemisphere, Monthly Weather

Review, 122, 2239 – 2256, https://doi.org/10.1175/1520-0493(1994)122⟨2239:AOCCFT⟩2.0.CO;2, 1994.

Sinclair, M. R.: Objective Identification of Cyclones and Their Circulation Intensity, and Climatology, Weather and Forecasting, 12, 595 – 612, https://doi.org/10.1175/1520-0434(1997)012⟨0595:OIOCAT⟩2.0.CO;2, 1997.

Stidd, C. K.: Cube-root-normal precipitation distributions, Transactions, American Geophysical Union, 34, 31, https://doi.org/10.1029/TR034i001p00031, 1953.

Swenson, L. M.: Simple Metrics to Determine the Mix of Processes Responsible for Extreme Precipitation, Climate, 2023, In Preparation.

Swenson, L. M. and Grotjahn, R.: Using Self-Organizing Maps to Identify Coherent CONUS Precipitation Regions, Journal of Climate, 32, 7747–7761, https://doi.org/10.1175/JCLI-D-19-0352.1, 2019.

Swenson, L. M. and Grotjahn, R.: Seasonal and Geographical Variations in Fundamental Weather Patterns During Extreme Precipitation as Identified from Omega Equation Forcing, Journal of Climate, 2023, In Review.

Trenberth, K. E.: On the Interpretation of the Diagnostic Quasi-Geostrophic Omega Equation, Monthly Weather Review, 106, 131 – 137, https://doi.org/10.1175/1520-0493(1978)106⟨0131:OTIOTD⟩2.0.CO;2, 1978.

Wang, B. and LinHo: Rainy Season of the Asian–Pacific Summer Monsoon*, Journal of Climate, 15, 386–398, https://doi.org/10.1175/1520-0442(2002)015⟨0386:RSOTAP⟩2.0.CO;2.

Wang, B. and Wang, B.: Climatic Regimes of Tropical Convection and Rainfall, Journal of Climate, 7, 1109–1118, https://doi.org/10.1175/1520-0442(1994)007⟨1109:CROTCA⟩2.0.CO;2.

Xie, P., Chen, M., Yang, S., Yatagai, A., Hayasaka, T., Fukushima, Y., Liu, C., Xie, P., Chen, M., Yang, S., Yatagai, A., Hayasaka, T., Fukushima, Y., and Liu, C.: A Gauge-Based Analysis of

Daily Precipitation over East Asia, Journal of Hydrometeorology, 8, 607–626, https://doi.org/10.1175/JHM583.1, 2007.

Zhang, Z., Ralph, F. M., and Zheng, M.: The Relationship Between Extratropical Cyclone Strength and Atmospheric River Intensity and Position, Geophysical Research Letters, 46, https://doi.org/10.1029/2018GL079071, 2019.