**Clustering Analysis of Autumn Weather Regimes in the Northeast U.S.**

David Coe\*

*Department of Environmental, Earth, and Atmospheric Sciences,*

*University of Massachusetts Lowell, Lowell, United States*

Mathew Barlow

*Department of Environmental, Earth, and Atmospheric Sciences,*

*University of Massachusetts Lowell, Lowell, United States*

Laurie Agel

*Department of Environmental, Earth, and Atmospheric Sciences,*

*University of Massachusetts Lowell, Lowell, United States*

Frank Colby

*Department of Environmental, Earth, and Atmospheric Sciences,*

*University of Massachusetts Lowell, Lowell, United States*

Christopher Skinner

*Department of Environmental, Earth, and Atmospheric Sciences,*

*University of Massachusetts Lowell, Lowell, United States*

Jian-Hua Qian

*Department of Atmospheric Sciences,*

*National Taiwan University*

**\*Corresponding author address:** David W Coe, Department of Environmental, Earth, and Atmospheric Sciences, University of Massachusetts Lowell, One University Avenue, Lowell, MA 01854.

**E-mail:** David\_Coe@uml.edu

ABSTRACT

A k-means clustering method is applied to daily ERA5 500-hPa height, MSLP, and 850-hPa wind data to identify characteristic Weather Types (WTs) for Sep-Nov from 1979-2018 over the Northeast U.S. The resulting circulation clusters are analyzed in terms of their structure, frequency of occurrence throughout the season, typical progressions between clusters, precipitation and temperature characteristics, and relation to teleconnections. The clusters are then used to make a circulation-based distinction between early and late autumn.

Seven WTs are identified for the autumn season: Ridge-Trough pattern (WTs 1, 4, and 7), Maritime Trough pattern (WT2), Midwestern Trough (WTs 3 and 5), and Strong Ridge (WT6). The Midwestern Troughs are likely to have precipitation and it is likely to be extreme. The Strong Ridge is likely to coincide with a heat wave day. Two main sequences of typical progressions between the circulation clusters emerge: 1) A sequence in September, defined by Maritime Trough – Strong Ridge – Midwestern Trough, which can be further broken down into 3 subsequences; 2) A sequence that occurs most often from mid-October to November defined as Midwestern Trough – Ridge-Trough – Midwestern Trough. The seasonality of the sequences allows for a distinction between “Early Season” and “Late Season” WTs. A preliminary trend analysis comparing the first and last 20 years of the period indicates an increase in “Early Season” WTs later in the season and a decrease in “Late Season” WTs earlier in the season; that is, a shift toward warmer-season patterns occurring later in the season.

**1. Introduction**

The aim of Weather Type (WT) analysis is to identify a region’s characteristic weather patterns, which can then be used as a basis for further investigation of the dynamics, variability, impacts, and trends in the regional weather. There are numerous approaches to weather typing, as reviewed in, e.g., El-Kadi and Smithson (1992), Yarnal (1993), Huth et al. (2008), and Ramos et al. (2015). Here, we take an objective approach based on k-means clustering applied to daily regional height and circulation fields at the surface, low, and mid-tropospheric levels for the Northeast US in the autumn season (Sep-Nov). A similar typing approach for the Northeast U.S. has been previously applied by Roller et al. (2016) for the winter season. The aims of this study are to determine the WTs that describe the autumn weather in the Northeast U.S., identify the characteristic progression between WTs as weather systems evolve and move through the region, and examine the differences between early autumn WTs and late autumn WTs.

The WT analysis used in the current analysis is based on the use of the k-means clustering technique (Diday and Simon 1976; Ghil and Robertson 2002, Moron et al. 2010), with the Classifiability Index (CI) algorithm of Michaelangeli et al. (1995). Combining k-means with the CI provides a semi-objective approach to identifying the optimal clustering of the dataset (semi-objective in the sense that there is a subjective choice of domain, variables, and maximum number of clusters when running k-means). In the context of WTs, the k-means clustering method has so far only been performed for the winter season in the Northeast (Roller et al. 2016), however, this technique has been used in multiple other studies to isolate atmospheric circulation patterns on a timescale and regionalized basis (Lana and Fernandez-Mills 1994, Sheriden 2002, Coleman and Rogers 2007, Moron et al. 2008, Moron et al. 2015, Qian et al. 2013).

Previous analyses of the autumn season in the U.S. have focused on frost onset timing (Easterling 2002, Cooter and LeDuc 1995), definition of seasons (Trenberth 1984), circulation patterns across the Northern Hemisphere (Sheridan 2002, Fleming et al. 1986), and the effects of a changing climate (both natural and anthropogenic in nature) on seasonal length and onset/withdrawal timing (Allen and Sheridan 2016). Of these studies, only Sheridan (2002) and Allen and Sheridan (2016) used a clustering method (Spatial Synoptic Classification 2) to examine weather patterns during a season. Easterling (2002) and Cooter and LeDuc (1995)  noted changes to the growing season, with it extending on both ends due to changes to the onset of the winter season (with the first frost day being on average 0.5 days later per decade in the Northeast U.S. (Easterling 2002)). Circulation analysis across the Northern Hemisphere by Allen and Sheridan (2016) also noted changes similar to those in Easterling (2002); the summer season is extending on both ends due to a later start of the autumn season and an earlier start of the summer season itself.  While considerable previous work has been done on shifts in seasonality terms of temperature, here we want to lay the groundwork for examining shifts in seasonality in terms of daily circulation and weather patterns.

The purpose of this study is to examine the autumn season at a regional scale (for the Northeast U.S.) to understand the underlying seasonal circulation patterns (WTs) and the evolution of these WTs over both a seasonal and long-term climatological timescale. The autumn season is a transition season between the warmest and coldest time periods of the year in the Northeast and with summer’s length having been increasing due to a changing climate (Allen and Sheridan 2016), a WT analysis on the autumn season will provide further insight to changes in underlying circulation patterns on a regional basis. Accompanying this, an analysis into temperature and precipitation anomalies and extremes in relation to the underlying circulation patterns will help contextualize the local climatology of the region.

Roller et al. (2016) also applied the WTs and their frequency of occurrence to the frequency of occurrence of teleconnections. They looked at the Arctic Oscillation (AO), the North Atlantic Oscillation (NAO), the Pacific-North American Pattern (PNA), and the El Niño-Southern Oscillation (ENSO), finding a strong relationship between Northeast WT climatology and the AO and NAO, a weak relationship to the PNA, and no relationship with ENSO. Typically, the AO and NAO are talked about in relation to winter-time weather, as they play a more significant role due to their relative strength. However, studies such as Gong and Ho (2003) have shown that the AO does influence summer weather patterns in terms of influencing summer monsoon rainfall in Asia. Ogi et al. (2015) used a revised AO index to show that the AO influenced circulation anomalies that resulted in heat waves in Europe and Russia in 2003, which Hu and Feng (2010) surmise that since the AO describes anomalies in circumpolar circulation, the influences described in Europe and Russia can also be applied to North America. Since the NAO and AO will strengthen throughout the fall season as the Northern Hemisphere enters its cold period, they will increasingly play a role in the circulation anomalies for the region and can help explain variations in temperature and precipitation that occur.

The rest of the paper is organized as follows: In section 2, the k-means clustering method, datasets and variables used in this analysis are described. In section 3, the WTs described by the methodology in section 2 are presented and analyzed with respect to persistence, transition tendencies, and seasonal evolution. In section 4, the results are summarized and discussed.

**2. Methods**

***2a. Data***

The European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 daily 500-hPa height, 850-hPa zonal and meridional (u and v) component winds, and Mean Sea Level Pressure (MSLP) for September, October, and November (SON) from 1979–2018 are used in this study (European Centre for Medium-Range Weather Forecasts, 2017). These data are bound for the region of 30N-50N latitude and 60W-90W longitude centered around the Northeast U.S. (Figure 1). Since this study is concerned with circulation patterns, 500-hPa heights, 850-hPa meridional winds, and MSLP data are used in clustering to give a better understanding of how the circulation is influenced by surface pressures and mid-level heights. The winds were chosen at 850-hPa since this is the lowest level to the ground that is above the influence of the mountains of the Northeastern United States.

Gridded precipitation data are from the Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks-Climate Data Record (PERSIANN-CDR; Ashouri et al. 2015; NCDC 2015). PERSIANN-CDR is extracted from gridded satellite (GridSat-B1) infrared data using artificial neural networks and bias corrected using monthly GPCP 2.583 2.58 precipitation.

The teleconnection indices used include daily AO, NAO, and PNA indices from the Climate Prediction Center https://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily\_ao\_index/teleconnections.shtml), and the seasonal Niño-3.4 index (Reynolds et al. 2002).

***2b. k-means clustering***

K-means clustering seeks to sort data into k-number of clusters (pre-determined by the user) based on intra-cluster variance of the Squared Euclidean Distance between data points (Diday and Simon 1976, Roller et al. 2016). The lower the variance, the higher the confidence that a cluster closely represents the underlying daily fields, the significance of which is tested using the Classifiability Index (CI) (Michelangeli et al. 1995, Moron et al. 2002).

Since this study is using a multi-variate solution, the data must be standardized properly to prevent bias from any of the fields. Standardized anomaly fields are created at each grid point by removing the long-term seasonal mean, area weighting for latitude, dividing by the standard deviation and running through an empirical orthogonal function (EOF) filter to retain 95% of the variance in the dataset. This produces a grid of *nt by ns* data points to be used by the k-means algorithm.

The CI determines the minimum number of clusters, *K*, by which the data can be considered well-separated and depends on the mean anomaly correlation coefficients (ACC) between the clusters in various partitionings (where a partition represents a single clustering result for a given *K*, based on random initialization), with the number of partitionings for each *K* chosen by the user (Moron et al. 2002). This results in *K* distinct CI value*.* To determine the statistical significance of these values, the process is repeated with a red-noise dataset created from the input fields to establish a 90% confidence interval of CI values for each *K*. Any CI values from the data which fall outside of this 90% confidence interval indicate a clustering that is significantly more consistent and repeatable than that using random red-noise. For this study, 100 simulations and a max of 10 clusters (*K* = 10) are subjectively chosen. Figure 1 details the output, where the red line represents the CI values of the data and the gray shading indicates the 90% confidence interval. From this, a 7-cluster partitioning is chosen (based on best CI) as the best representation of the data for the season.

The same clustering performed using standardized anomalies made with the long-term daily mean removed left the conclusions of the study unchanged. We will be using the seasonal-mean removed data in the rest of this study as the patterns were more clearly defined and showed stronger correlations to days within the patterns than in the daily-mean removed data.

***2c. Monte Carlo Analysis***

To better understand significance in the data, a 95% confidence interval using the Monte Carlo method establishes the background frequency of WT occurrence. Each day for the period (1979–2018) is assigned a cluster value from the k-means algorithm. Then, every cluster value is randomly shuffled throughout all dates (while maintaining monthly and yearly integrity) and the calculation redone, 1000 times. After sorting, the 25th and 975th trials are used to determine the 2.5% and 97.5% confidence values, creating the background frequency (95% confidence interval). Data points that fall outside of the background frequency are significant compared to random chance.

***2d. Markov Chain Analysis***

Markov Chain analysis (as performed by Vautard et al. 1990) is used to better understand WT progressions between clusters. As found by Vautard et al. (1990), it is necessary to use Markov Chain analysis when clusters have large size discrepancies in number of cluster elements rather than using an equiprobability model. The equiprobability model assumes that each cluster is as likely to occur on the next day as any other cluster. This conditional probability does not account for the absolute probability of a cluster to occur, leading to bias in favor of clusters that have the most elements. In this study, WTs 1 and 6 both have > 650 days in their clusters, while WTs 2, 3, 4, and 5 have < 500 days in their clusters (Table 1), so an equiprobability model would be biased towards WTs 1 and 6.

To perform this analysis, first the transition matrix Tij is calculated for the original dataset. Then a Monte Carlo simulation is run 10,000 times, randomly shuffling the WTs each time while maintaining the integrity of the dataset. Each simulation, a new transition matrix is calculated, Bij, and is compared to Tij. The number of Cij times where Tij ≥ Bij and the number of Dij times where Tij ≤ Bij are recorded. At the end, values of Cij and Dij ≤ 500 are considered significant at the 95% level. Vautard et al. (1990) notes that the use of weak inequalities provides a more stringent criterion as it will not eliminate as not significant any transition to or from a cluster with only one element. ­­

**3. Results**

***3a. Weather Types***

Each WT is identifiable by unique circulation features at each of the three levels on which the clustering was performed, along with unique precipitation and 2-meter temperature anomalies (Figures 2 and 3). WT1 exhibits more zonal upper-level flow with weak troughing apparent. A surface low-pressure system is present over New Brunswick with low-level winds out of the northwest over the region. The New England and the East Coast exhibit warmer than average surface temperatures and high than average precipitation, while the Midwest is cooler than average with lower than average precipitation. In WT2, there is a weak upper level ridge over the western part of the region with a weak upper level trough over the eastern portion of the region. There is surface high-pressure over northern New England with an area of surface low-pressure over the ocean east of the Carolinas, resulting in higher than average precipitation over the ocean and lower than average surface temperatures over New England. Resulting low-level flow comes from the north/northeast over the region due to the influence of the low-pressure area. WT3 has an area of surface low pressure over northern New England accompanied by an upper-level trough over the Great Lakes region. Low-level winds over the region follow a counter-clockwise flow pattern centered in the middle of the upper-level trough and low-level surface pressure. This leads to an area of warmer than average surface temperatures and higher than average precipitation over the eastern half of the region, with the western half seeing colder than average temperatures and below average precipitation. WT4 presents an upper level trough over the Northeast with surface low-pressure and above average precipitation over New Brunswick. Low level winds are primarily out of the northwest, causing the region to experience below average surface temperatures. WT5 features an upper level ridge over the Atlantic Ocean and an upper level trough over the Midwest, with higher than average precipitation associated with the trough. There is an area of surface high-pressure over the Atlantic with winds flowing out of the southwest, leading to above average surface temperatures over the region. WT6 shows an area of surface high-pressure along with an upper level ridge over the Northeast, leading to below average precipitation over New England and above average temperatures over the whole of the region. Finally, WT7 features zonal upper level flow with a slight upper level ridge over the Midwest and slight upper level trough over the Atlantic Ocean, leading to below average precipitation over the whole of the region. Low-level flow is mainly out of the west/northwest with an area of high-pressure located at the surface over the Eastern Seaboard. Surface temperatures are below average in all regions except for the Midwest, where they are slightly above average.

***3b. Variations of WT Frequency Throughout the Season***

To measure monthly variability of WT frequency, the summation of WT frequency per month over all years was calculated (Figure 4) and from these, three groups were formed: 1.) WTs 1, 2, and 6 in September 2.) WTs 2 and 3 in October, and 3.) WTs 3, 4, 5, and 7 in November. WTs 1 and 6 feature the largest decrease in occurrence rate over the season from 37.9% and 35.7% of days in September to 4.1% and 4.3% of days in November, respectively. WTs 4 and 7 have the largest increase in occurrence rate over the season from 0.6% and 1.5% of days in September to 22.3% and 29.9% of days in November, respectively.

***3c. WT Precipitation and Temperature***

Each WT exhibited a unique precipitation and temperature pattern for the whole of the domain (Figures 2 and 3). WTs 3 and 5 (Midwestern troughs, MWT) see the highest precipitation rates (6.1 mm/day and 5.4 mm/day respectively), while WTs 4 and 7 (Midwestern ridges, MWR) see the lowest precipitation rates (2.7 mm/day and 2.5 mm/day respectively). The two MWT patterns contained the most extreme precipitation days that occurred during SON (26% and 19% of days respectively), and extreme precipitation days were more likely to occur during these two patterns (Figure 5a.). The two MWR and one strong Ridge (WT6) patterns saw the least amount of extreme precipitation days (7%, 8%, and 1% respectively) and extreme precipitation days were unlikely to occur during these patterns. WTs 1 and 3 (weak and strong upper-level trough) focused positive precipitation anomalies over New England and the Gulf of Maine, while WTs 2 and 4 (Maritime Troughs, MT) focused positive precipitation anomalies over the Atlantic Ocean, due to the positioning of the upper-level trough. WT5 had positive precipitation anomalies across the western two-thirds of the domain due to the strength of the MWT, while WT7 had predominately negative precipitation anomalies over the entirety of the region.

Positive and negative temperature anomalies in the domain were co-located to the positioning of the upper-level troughs and ridges. Parts of the region downstream from an upper-level trough experienced warmer than average temperatures and colder than average temperatures were experienced upstream from an upper-level trough, independent of calendar date in the season. Dependent of the calendar season, the mean temperatures for each WT varied from the warmest (17.8°C and 18.2°C, WTs 1 and 6 respectively) during WTs that occurred early in the season and the coldest (9.2°C and 10.0°C, WTs 4 and 7 respectively) during WTs that occurred late in the season. Heat wave days were seen during WTs 1, 2, and 6 (making up 98% of total heat wave days), due to their occurrence earlier in the season, with WTs 1 and 6 being the most likely to experience heat wave days (Figure 5b.).

***3d. Teleconnection Influences on WT Frequency***

The frequency of each WT is analyzed with respect to the daily phases of AO, NAO, and PNA and the monthly phases of ENSO, with significance determined through a Monte Carlo analysis (Figure 6). For all teleconnections during their neutral phase, there was no significance at the 95% level for any WT being likely or unlikely to occur.

WT frequency was least affected by the ENSO phase, with WTs 2 and 6 unlikely to occur (10% and 12% of days respectively) during a negative ENSO phase, WT7 likely to occur during either a positive or negative ENSO phase (20% and 22% of days respectively), and WT4 likely to occur during a negative ENSO phase (24% of days). It is not surprising as the Northeast weather and climate is not correlated directly to the ENSO phase. It is interesting that two of the WTs (2 and 6) that occur in the early season are unlikely to occur during a negative ENSO phase, as ENSO has a weaker influence during the beginning of autumn.

AO and NAO both feature similar patterns in WT frequency, with WTs 3 and 4 being likely to occur during a negative phase and unlikely during a positive phase and vice-versa for WTs 6 and 7. WTs 3 and 4 are classic negative AO and NAO circulation patterns, with upper-level troughing located over the Eastern U.S., accompanied by below average temperatures and above average precipitation. WTs 6 and 7 both feature upper-level ridging over the Midwest and Eastern U.S. accompanied by below average precipitation and above average temperatures, signatures of positive AO and NAO phases.

During a positive PNA phase, WT6 is likely to occur and WTs 5 and 7 are unlikely to occur, whereas during a negative PNA phase, WTs 1 and 5 are likely to occur and WTs 2, 4, and 7 are unlikely to occur. WT6 occurring during a positive PNA phase is surprising, as it features prominent ridging and warmer than average temperatures over the Northeast. It is also surprising to note that even with the prominent trough and below average temperatures in WTs 3 and 4, WT3 does not show any likelihood of occurring more or less often during either phase and WT4 is unlikely to occur during a negative PNA phase.

***3e. Sequences of Early and Late Season WTs***

Early and Late Season WTs for autumn are determined based on the monthly timing and on the progression of the WTs. To provide a more detailed look at the within-season evolution of the WTs, the average daily evolution throughout the season is shown in Figure 7. A 5-day moving mean is applied to show an approximate intersection point between the two data series, representing the transition date between the two (Fig 7c). October 16th represents this transition date; the day when there is a 50% chance that either an early or Late Season WT will occur. Early Season WTs typically occur during >50% of days before October 16th and drop to an average of 25% of days over the rest of October and <10% of days in the month of November. Late Season WT frequency trends show <10% of days experiencing a Late Season WT in September and >50% of days experiencing a Late Season WT after October 16th.

To further examine the sub-seasonality of the WTs, the monthly occurrence rates (Figure 4) are combined with an analysis of WT progression (Table 3 and Figure 7). First, the dataset of cluster numbers was split into a 91 x 40 matrix and then persistence was removed to be able to count patterns containing multiple consecutive cluster numbers (For example: “161” would be the same as “1166611” after persistence is removed). The number of sequences of lengths 2, 3, and 4 that occurred each year was counted and the yearly totals were summed together to get an overall count (Table 2; Note that only the top 7 results for each sequence are shown due to length and sequences greater than length 4 are not shown as the highest count for those sequences was less than 10 over the 40 year dataset). A Markov Chain analysis is performed on the dataset to produce a transition matrix, which is then run through a Monte Carlo process to find progressions that are significant at the 95% level. Initially, this was performed on the whole of the dataset, however the sequences “WT1-WT6” and “WT2-WT6” did not appear as significant, even though they occurred in the dataset (Table 2). Analyzing progression on a monthly timescale shows that these sequences are significant at the 95% level, however the month of October showed many sequences as significant, that the analysis was over-complicated. Instead, we opted to split the season in half and found similar sequences as in Table 2 that are significant at the 95% level (results of the Markov Chain analysis are found in Table 3).

Based on the Markov Chain analysis, four likely sequences of progression are found in the first half of the season and one in the second half of the season (Figure 8). The four sequences in the first half consist of one main sequence (WT1-WT2-WT6-WT5, hereafter E1, and three subsequences: 1. WT1-WT2-WT6, hereafter E2. 2. WT1-WT6-WT1, hereafter E3. 3. WT1-WT6-WT5 hereafter, E4. The one sequence in the second half is WT3-WT4-WT7-WT5, hereafter L1. There are subsequences of L1 that have occurred frequently (Table 2), however, none of them were significant at the 95% level. Overall, sequences E2-E4 occurred at least once during each calendar year of the study, with 1981 and 2000 seeing the least amount (2 occurrences of sequences) and 1983 seeing the most (7 occurrences of sequences) per year. Sequence E1 occurred 24 times over the whole study, not occurring at all in multiple years, while sequence L1 occurred 69 times, not occurring at all in only 6 years in the study (Figure 9). Subsequences of L1 that were not significant at the 95% level did occur in years where L1 did not, but were not shown.

***3f. Physical Characteristics of Early and Late Season Sequences***

Each of the sequences were characterized by distinct circulation, precipitation, and temperature features (Figures 10-14). In general, the sequences all involved a trough-ridge or ridge-trough (T-R or R-T) circulation flow through the region, varying in both strength and location of these features. Interestingly, the circulation, temperature, and precipitation features of the individual WT days within each sequence resemble the overall WTs described in Figures 2 and 3.

E1 features a MT-R-MWT circulation pattern. Most of the Northeast initially experiences above average precipitation and temperatures, with temperatures and precipitation becoming below average as the trough moves further east offshore, being replaced by a strong ridge. This strong ridge brings surface high pressure, with the eastern half of the region continuing to see below average temperatures and precipitation, while the western half sees both above average temperatures and precipitation. This is the leading edge of a MWT, which will bring above average precipitation and temperatures to the whole of the region.

Three subsequences of E1 can occur, depending on the location and strength of the MT and MWT. E2 starts out similarly to E1, with above average precipitation and temperatures occurring across the Northeast. Instead of the trough becoming a MT, it gets pushed northeast as ridge moves into the region along with surface high pressure. This ridge is weaker than that in E1, with the areas of above and below average temperatures and precipitation being weaker, albeit occurring in similar locations. This ridge is eventually replaced by a trough, weaker than the MWT seen in E1. E2 is the driest and coolest sequence, with the lowest precipitation and temperature anomalies. E3 begins similarly to E1, however the MT that develops is weaker and located further north than in E1. This leads to weaker areas of precipitation and temperatures anomalies and displaced northward compared to E1. The ridge that develops is of similar strength, but located further north than in E1, leading to above average temperature anomalies occurring over most of the region. However, precipitation anomalies are in similar locations and of similar strength to those in E1. In this case, the MWT is weak following the development of the ridge, leading to this weak trough replacing the ridge. E3 is overall our coolest pattern, with most of the region experiencing below average temperatures. E4 begins as a trough which is accompanied by positive temperature anomalies across most of the domain and positive precipitation anomalies across New England and the Midwest. A strong ridge builds into the region, with positive temperature anomalies across the whole of the region and a split between negative and positive precipitation anomalies for the eastern half and western half of the region, respectively. The surface high pressure accompanied with this ridge is weaker than seen in E1. A MWT then moves into the region, bringing positive precipitation anomalies everywhere except out over the Atlantic and positive temperature anomalies everywhere except for the Midwest. This is our wettest and warmest pattern, with the whole region seeing above average temperatures for the entire sequence, especially on the eastern seaboard where the highest anomalies occur.

L1 features a Hudson Bay trough (HBT) that crosses the northern half of the domain. This trough is accompanied by surface low pressure that brings positive precipitation anomalies to the whole of the eastern half of the region along with warmer than average temperatures. Below average temperatures follow upstream of the trough and overtake the entirety of the domain as the trough progresses eastward. Negative precipitation anomalies accompany the movement of this trough and overtake the whole of the region, linked with the cooler than average temperatures due to cold air advection from the northwest behind the surface low pressure system. Surface low pressure builds into the New England region as temperatures and precipitation remain below average. However, warmer temperatures and above average precipitation are building in the Midwest and move into the region as a strong MWT forms.

***3g. Shifts in the frequency of Early and Late Season WTs***

The categorization of autumn days into WTs can provide a basis for examining trends over time and shifts in seasonal timing in terms of daily circulation patterns. Here we provide a preliminary analysis of changes over time, considering the difference in frequency of occurrence of Early and Late Season WTs between the first twenty years of the record (1979-1998) and the last twenty years of the record (1999-2018).

The monthly averages of the difference between the periods for the Early and Late Season WTs are shown in Fig. 15. The thin gray bars indicate the 95% confidence interval obtained from Monte Carlo resampling, and the thick bars denote the difference between the periods, shaded red for significant increases, blue for significant decreases, and gray for changes that are not significant. Early Season WTs show a significant increase at the end of the season, in Nov, as well as a significant overall increase. Late Season WTs show a significant decrease at the beginning of the season, in Sep, as well as a significant overall decrease. That is, in addition to temperature shifts, daily circulation patterns are also indicating a seasonal shift in autumn in the Northeast U.S., with the early season weather continuing longer into the season as the late season weather retreats.

A preliminary look at extreme precipitation days and amounts and heat wave days was also performed, albeit without statistical significance accounted for. WTs 3 and 5 contain the most extreme precipitation days and are the most likely to have them occur, especially during the first half of the time period. However, during the second half of the time period both WTs saw a decrease in the number of extreme precipitation days, WT3 seeing a 1% decrease and WT5 seeing an 11% decrease. All other WTs saw an increase in extreme precipitation days during the latter half of the period with WTs 4 and 6 seeing the largest increase (4% and 5%, respectively). Even with seeing a decrease in extreme precipitation days between the two periods, WTs 3 and 5 experienced increased precipitation per event, while oppositely, WT4 experienced decreased precipitation per event. Heat wave days are most likely to occur during WTs 1 and 6, with WT6 seeing an increase in percent of heat wave days between the two periods (36% in the first half to 60% in the second half of the period). WTs 1 and 2 saw a decrease between the two periods (5% and 16%, respectively).

**4. Summary and Discussion**

Seven distinct WTs are identified for the Northeast U.S autumn season through use of the k-means clustering algorithm applied to ERA5 500-hPa height, MSLP, and 850-hPa wind data for the 1979-2018 time period. An analysis of WT frequency per month showed a clear change in frequency over the course of the season with 3 WTs (WTs 1, 2, and 6) likely to occur in September but not in November and 4 WTs (WTs 3,4,5, and 7) likely to occur in November but not in September. The WTs also showed preferred progression between one another. Two WT sequences emerged, of which one captures the progression of “Early Season” weather, able to be split into three subsequences, and one of which captures the progression of “Late Season” weather.

WT1 exhibits a weak T-R circulation, surface temperatures that are above average for most of the region, and New England and the Eastern U.S. experiencing above average precipitation. In WT2, there is a MT present coinciding with above average precipitation anomalies on the Atlantic Coast and over the Atlantic. Temperatures are generally average to slightly below average. WT3 has an area of surface low pressure over Northern New England accompanied by an HBT. This brings above average precipitation and temperatures to the eastern half of the region, while the western half is drier and cooler than average. WT4 exhibits an upper-level trough over the Northeast with a surface low-pressure over New Brunswick. Low level winds are primarily out of the northwest, bringing cooler than average temperatures and below average precipitation to the region. WT5 features an Atlantic ridge (AR) and MWT, coinciding with above average temperatures for the whole of the region and above average precipitation downstream of the MWT. There is an area of surface high-pressure over the Atlantic with winds flowing out of the southwest. WT6 shows an area of surface high-pressure along with an upper-level ridge over the Northeast, accompanied with above average temperatures and below average precipitation. Finally, WT7 features a more a slight R-T circulation pattern, with below average precipitation for the whole over the region and average to slightly below average temperatures. Surface high pressure dominates the entirety of the region.

Analysis of extreme precipitation and heat wave days was performed on the WTs, with the MT WTs (3 and 5) seeing the highest amount of extreme precipitation days and are more likely to see extreme precipitation occur when there is precipitation. Even though they were not all significant, every WT had at least 1 extreme precipitation day associated with it. Heat wave days were more divided, with WTs 1, 2, and 6 containing 98% of the days, with WTs 4 and 7 having no heat wave days occur. WTs 1 and 6 are likely to occur in conjunction with heat wave days (Figure 5).

In relation to teleconnections, since this is the transition between the warm and cold season, the NAO and AO are beginning to shift from weaker influence to stronger influence over circulation anomalies in the Northern Hemisphere. The WTs were weakly influenced by ENSO, with WTs 2 and 6 unlikely to occur and WTs 4 and 7 likely to occur during a negative phase. The AO and NAO showed the best relationship to the WTs, with WTs 6 and 7 likely to occur and WTs 3 and 4 unlikely to occur during a positive NAO and AO and the opposite during a negative AO and NAO.

WTs are further divided using monthly frequencies and progression between WTs was analyzed using Markov Chains, leading to the development of sequences of WT progression defined by their seasonality (Figures 4, 7, and 8 and Table 3). One “Early Season” sequence features 4 WTs (WTs 1, 2, 5, and 6), three of which are likely to occur in September and three of which are unlikely to occur in November (WTs 1 and 6). This sequence can be divided into three subsequences depending on the development of a maritime trough (MT) or the strength and positioning of the upper-level ridge. The “Late Season” sequence features 4 of the WTs (WTs 3, 4, 5, and 7), where all four WTs are significantly likely to occur in November and significantly unlikely to occur in September. These sequences each have their own distinct circulation features and associated temperature and precipitation anomalies, with the subsequences of the “Early Season” identifying both warm and cold patterns as well as which are likely to produce precipitation. While there are no subsequences identified for the “Late Season”, it can be noted that there are two subsequences that are likely to occur at the 87% and 84% levels (WT3-WT4-WT3 and WT3-WT4-WT7-WT3, Table 3). There are also two sequences that occurred in the dataset (WT5-WT4-WT7 and WT5-WT3-WT7, Table 2) that are not found to be significant by the Markov Chain Analysis. This can be attributed to whether the MWT that moves through the region during WT5 is strong or weak and how fast it moves through the area. If the trough moves through faster than over our daily synoptic timescale, our WT model will be unable to pick it up and misattribute it as a different WT. This could lead to WT3 or WT4 being skipped entirely, leading to the results in Table 2. It is also pertinent to note that each of the WTs are also statistically likely to transition to themselves on the next day (persist for 2 days), suggesting that these patterns are synoptic in nature.

There is evidence that the timing of the season is changing in terms of daily circulation. Over the past 20 years, the Early Season WTs have seen an increase in occurrence in November and the Late Season WTs have seen a decrease in occurrence in September, both of which are significant at the 95% confidence level. This idea of a delayed onset of the autumn season and thereby extension of the summer season has been noted before (Allen and Sheridan (2016), Easterling (2002), Park et al. 2018), most recently in Park et al. (2018). They noted that across all land masses in the Northern Hemisphere, the summer season has been getting longer due to both an earlier onset and longer duration, which is consistent with the results here. A few outstanding questions remain, including whether the circulation patterns themselves change substantially over time and how the observed regional temperature trends are apportioned between WTs. The planned extension of ERA5 back to 1950 will also allow for an examination of changes over a longer period. We plan to address these questions for the autumn season in future work, as well as extend our analysis to include both summer and winter, to capture the reflection of seasonal evolution more broadly in daily circulation.

**Data Availability Statement:**

Publicly available datasets were analyzed in this study. The ERA5 reanalysis data are available from the Copernicus data store at: <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels?tab=form> for pressure level data (500 hPa geopotential heights and 850 hPa u- and v- winds) and <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=form> for single level data (MSLP).

**Acknowledgements**

This work was supported by NASA NESSF grant NNX15AN91H S03 and NSF AGS-1623912, as well as by the use of EMCWF’s ERA5 data set

**References**

Agel, L., M. Barlow, F. Colby, H. Binder, J. L. Catto, A. Hoell, and J. Cohen, 2018: Dynamical analysis of extreme precipitation in the US northeast based on large-scale meteorological patterns. Climate Dynamics.

Allen, M.J. and Sheridan, S.C., 2016: Evaluating changes in season length, onset, and end dates across the United States (1948–2012). Int. J. Climatol., 36: 1268-1277. doi:[10.1002/joc.4422](https://doi.org/10.1002/joc.4422)

Ashouri, H., K.-L. Hsu, S. Sorooshian, D. K. Braithwaite, K. R. Knapp, L. D. Cecil, B. R. Nelson, and O. P. Prat, 2015: PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. Bull. Amer. Meteor. Soc., 96, 69-83, doi:10.1175/ BAMS-D-13-00068.1.

Chen, Y, Zhai, P, Liao, Z, Li, L.,2019: Persistent precipitation extremes in the Yangtze River Valley prolonged by opportune configuration among atmospheric teleconnections. Q J R Meteorol Soc, 145: 2603– 2626. https://doi.org/10.1002/qj.3581

Coleman, J.S.M., and J.C. Rogers, 2007: A Synoptic Climatology of the Central United States and Associations with Pacific Teleconnection Pattern Frequency. J. Climate, 20, 3485-3497.

Cooter, E., and S. LeDuc, 1995: Recent frost date trends in the northeastern United States. Int. J.

Climatol., 15, 65-75.

Diday, E., and J. J. Simon, 1976: Clustering analysis. Digital pattern recognition, K. S. Fu, Ed., Springer, 47–94.

Easterling, D.R., 2002: Recent changes in frost days and the frost-free season in the United States. *Bull. Amer. Meteor. Soc.,* **83**, 1327–1332,doi:10.1175/1520-0477-83.9.1327

El-Kadi, A. K. A., and P.A. Smithson, 1992: Atmospheric classifications and synoptic climatology. *Prog. Phys. Geo.,* **16**, 432-455.

European Centre for Medium-Range Weather Forecasts, 2017: ERA5 Reanalysis. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, Boulder, CO. [Available online at https://doi.org/10.5065/D6X34W69.]

Fleming, E.L., G. Lim, and J.M. Wallace, 1987: Differences between the Spring and Autumn Circulation of the Northern Hemisphere. J. Atmos. Sci., 44, 1266–1286, doi:10.1175/1520-0469

Ghil, M., and A. W. Robertson, 2002: “Waves” vs. “particles” in the atmosphere’s phase space: A pathway to long-range forecasting? Proc. Natl. Acad. Sci. USA, 99, 2493–2500, doi:10.1073/pnas.012580899

Gong, D.‐Y., and C.‐H. Ho (2003), Arctic oscillation signals in the East Asian summer monsoon, *J. Geophys. Res.*, 108(D2), 4066, doi:[10.1029/2002JD002193](https://doi.org/10.1029/2002JD002193" \o "Link to external resource: 10.1029/2002JD002193" \t "_blank).

Hu, Q., and Feng, S. (2010), Influence of the Arctic oscillation on central United States summer rainfall, *J. Geophys. Res.*, 115, D01102, doi:[10.1029/2009JD011805](https://doi.org/10.1029/2009JD011805" \o "Link to external resource: 10.1029/2009JD011805" \t "_blank).

Huth, R., C. Beck, A. Philipp, M. Demuzere, Z. Ustrnul, M. Cahynová, J. Kyselý, and O. E. Tveito, 2008: Classification of atmospheric circulation patterns – recent advances and applications. *Ann. New York Acad. Sci.,* **1146,** 105-152.

Lana, X., and G. Fernandez-Mills, 1994: Minimum sample size for synoptic weather type classification. Application to winter period data recorded on the Catalan coast (north-east Spain). Int. J. Climatol., 14, 1051–1060, doi:10.1002/joc.3370140909.

Legras, B., T. Desponts, and B. Piguet, 1987: Cluster analysis and weather regimes. Proc. Workshop on the Nature and Prediction of Extratropical Weather Systems. Vol. 2, Reading, Shinfield Park, UK, ECMWF, 123-149.

 Meseguer-Ruiz Oliver, Cortesi Nicola, Guijarro Jose A., Sarricolea Pablo, 2020: Weather regimes linked to daily precipitation anomalies in Northern Chile, Atmospheric Research, Volume 236, 104802, ISSN 0169-8095, <https://doi.org/10.1016/j.atmosres.2019.104802>.

Michelangeli, P.A., R. Vautard, and B. Legras, 1995: Weather regimes: recurrence and quasi-stationarity. J. Atmos. Sci., 52, 1237-1256.

Moron, V., A. W. Robertson, M. N. Ward, and O. Ndiaye, 2008: Weather types and rainautumn over Senegal. Part I: Observational analysis. J. Climate, 21, 266–287, doi:10.1175/2007JCLI1601.1

Moron, V., A.W. Robertson, and J.-H. Qian, 2010: Local versus regional-scale characteristics of monsoon onset and post-onset rainautumn over Indonesia. Climate Dynamics, 34, 281-299.

Moron V, Robertson AW, Qian J-H and Ghil M (2015) Weather types across the Maritime Continent: from the diurnal cycle to interannual variations. Front. Environ. Sci. 2:65. doi: 10.3389/fenvs.2014.00065

Muñoz, Á. G., X. Yang, G. A. Vecchi, A. W. Robertson, W. F. Cooke, 2017: A weather-type-based cross-time-scale diagnostic framework for coupled circulation models. *J. Climate*, **30**, 8951–8972, <https://doi.org/10.1175/JCLI-D-17-0115.1>

NCDC, 2015: PERSIANN data files. National Climatic Data Center, accessed 10 September 2020 [Available online at <https://www.ncei.noaa.gov/data/precipitation-persiann/access/>.]

Ogi, M., K. Yamazaki, and Y. Tachibana (2005), The summer northern annular mode and abnormal summer weather in 2003, *Geophys. Res. Lett.*, 32, L04706, doi:[10.1029/2004GL021528](https://doi.org/10.1029/2004GL021528" \o "Link to external resource: 10.1029/2004GL021528" \t "_blank).

Park, B., Y. Kim, S. Min, and E. Lim, 2018: Anthropogenic and Natural Contributions to the Lengthening of the Summer Season in the Northern Hemisphere. J. Climate, 31, 6803–6819, doi:10.1175/JCLI-D-17-0643.1

Ramos, A.M., D. Barriopedro, and E. Dutra, 2015. Circulation weather types as a tool in atmospheric, climate, and environmental research. *Frontiers Env. Sci.*, **3,** 44.

Roller, C.D., J.-H. Qian, L. Agel, M. Barlow, and V. Moron, 2016: Winter weather regimes in the Northeast United States. J. Climate, 29, 2963-2980.

Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, 2002: An improved in situ and satellite SST analysis for climate. J. Climate, 15, 1609-1625, doi:10.1175/ 1520-0442(2002)015,1609:AIISAS.2.0.CO;2

Sheridan, S. C., 2002: The redevelopment of a weather-type classification scheme for North America. Int. J. Climatol., 22, 51–68, doi:https://doi.org/10.1002/joc.709

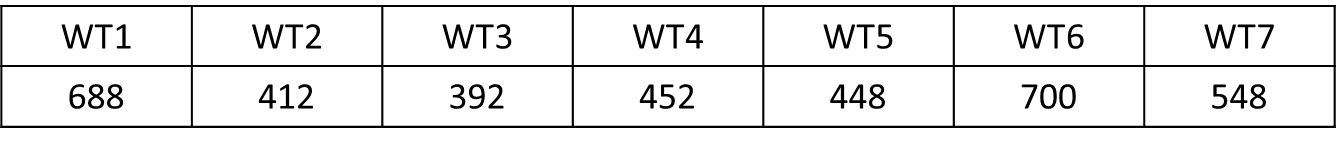
Sheridan, S. C., 2003: North American weather-type frequency and teleconnection indices. *Int. J. Climatol.*, **23**, 27–45, <https://doi.org/10.1002/joc.863>

Trenberth, K.E., 1983: What are the Seasons?. Bull. Amer. Meteor. Soc., 64, 1276–1282, doi: 10.1175/1520-0477

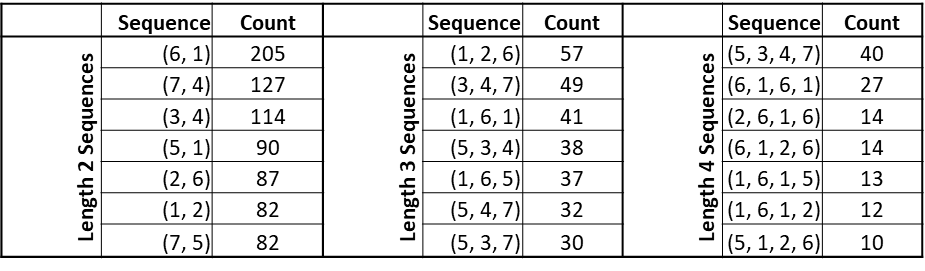
Qian, J.-H., A.W. Robertson, and V. Moron, 2013: Diurnal cycle in different weather regimes and rainfall variability over Borneo associated with ENSO. J. Climate, 26, 1772-1790.

Vautard, R., Mo, K. C., & Ghil, M. (1990). Statistical Significance Test for Transition Matrices of Atmospheric Markov Chains, Journal of Atmospheric Sciences, 47(15), 1926-1931.

Yarnal, B., 1993: Synoptic Climatology In Environmental Analysis. Belhaven Press, London, UK.



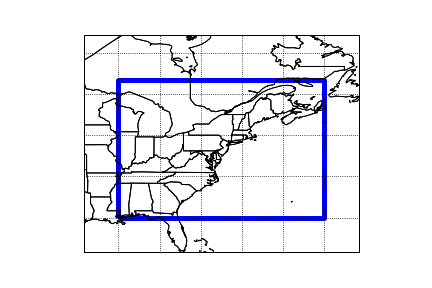
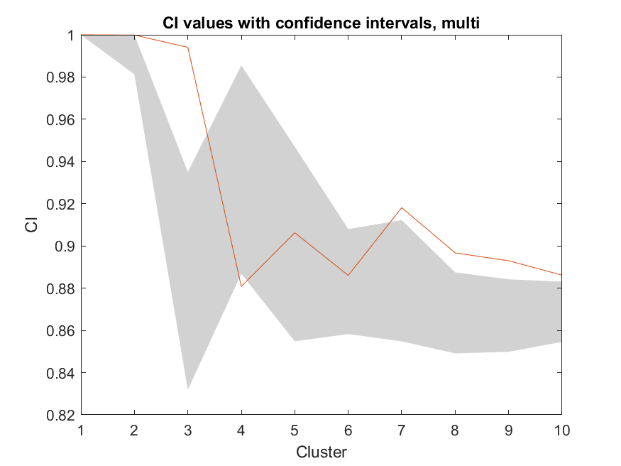
**Table 1.** Total number of days in each Weather Type.



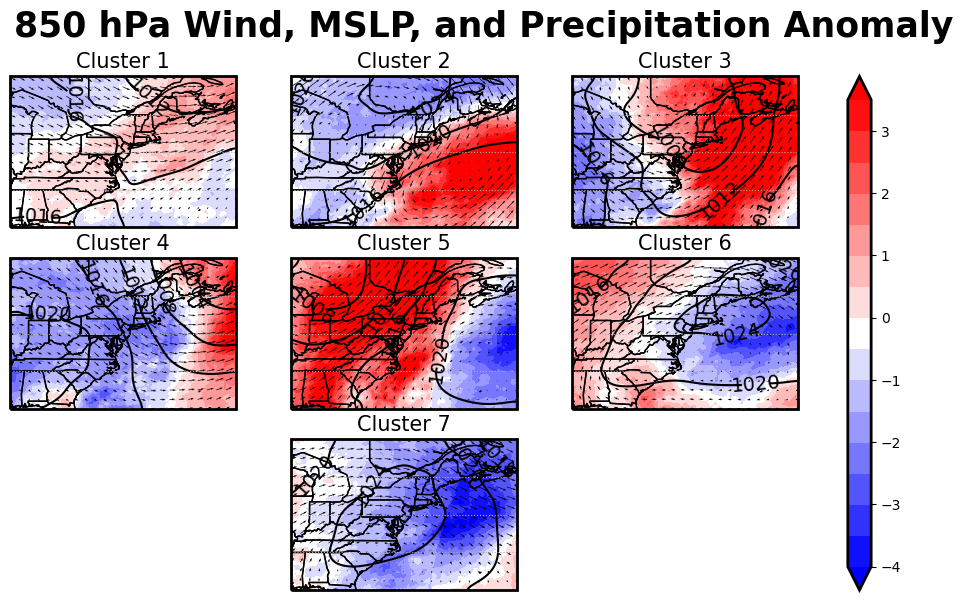
**Table 2.** Top 7 results for occurrences of sequences of length 2, 3, and 4 of WTs in the overall dataset.



**Table 3.** Markov Chain analysis (as performed in Vautard et al. 1990) indicating WT progression. Table C indicates WTs that are likely to transition between each other, with values below 500 being significant at the 95% level (bold and italicized). Table D indicates WTs that are unlikely to transition between each other, with values below 500 being significant at the 95% level (bold and italicized).

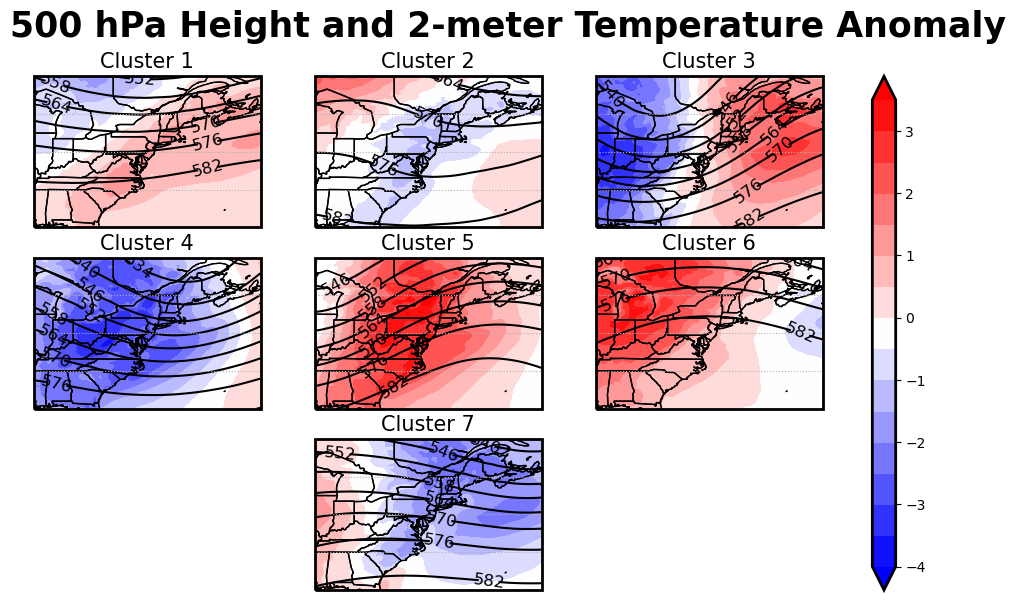


**Figure 1.** Map of the bounded area used in this study (left) and k-means clustering results (right). Classifiability Index (red line) shows the results of the k-means clustering for k=1...10. Red noise was made from the dataset and run through k-means to determine the 90% confidence interval of the CI (gray shading). The CI is significant for values where the red line lies above the gray shading.



**Figure 2.** **Autumn Season Weather Types Lower Atmosphere**

850 hPa u and v winds, Means Sea Level Pressure (contour), and daily mean removed average precipitation anomalies (mm/day, shaded) ERA5 data from 1979-2018 plotted for each WT.

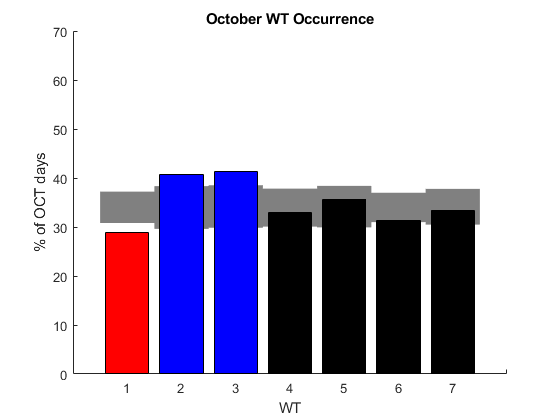


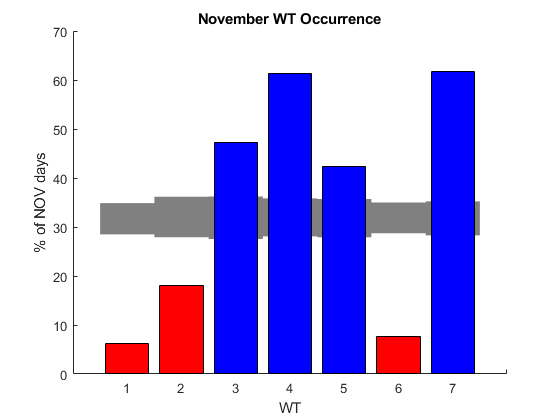
**Figure 3: Autumn Season Weather Types Upper Atmosphere**

500 hPa height (contour) and daily mean removed average 2-meter anomalies (°C, shaded) ERA5 data from 1979-2018 plotted for each WT.

Chart, bar chart

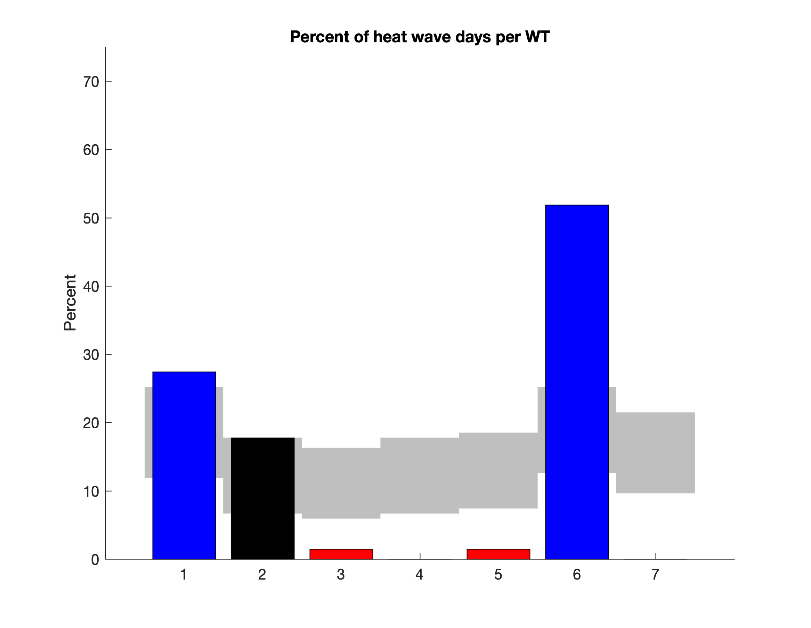
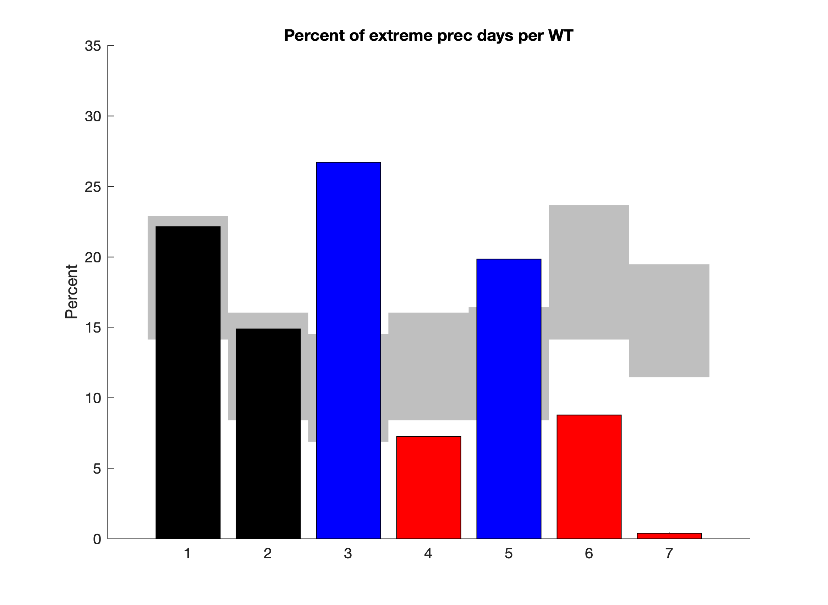
Description automatically generated





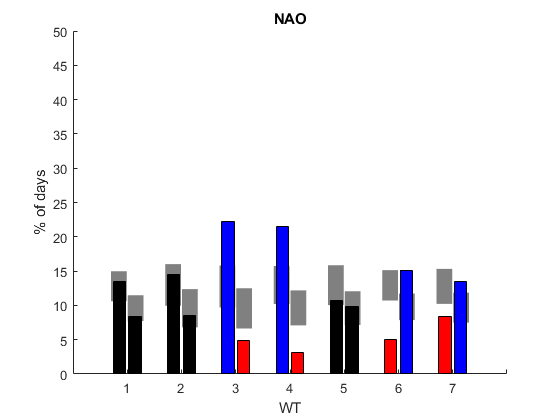
**Figure 4: Autumn WT Monthly Frequency**

The likelihood of occurrence during a month for each WT is displayed based on the number of days each WT occurs per month. A 95% confidence interval (grey bars) is found using the Monte Carlo method. Red bars indicate WTs that are more likely to occur than expected due to chance during a given month and blue bars indicate WTs that are less likely to occur during a given month.

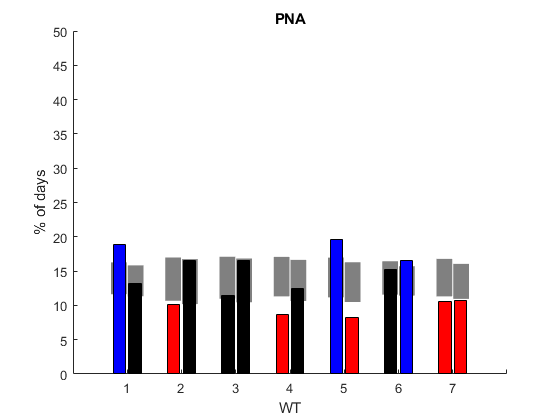
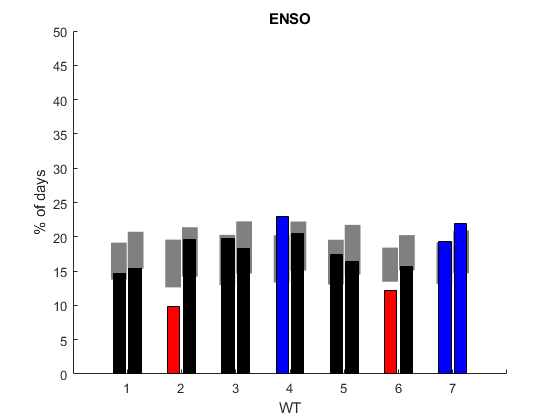


**Figure 5: Autumn Weather Extreme Precipitation and Heat Wave Days**

Percent of WT days featuring extreme precipitation (left) and heat waves (right). Grey shading indicates the Monte Carlo 95% confidence interval, with red shading indicating less likely to occur WTs and blue shading indicating more likely to occur WTs.

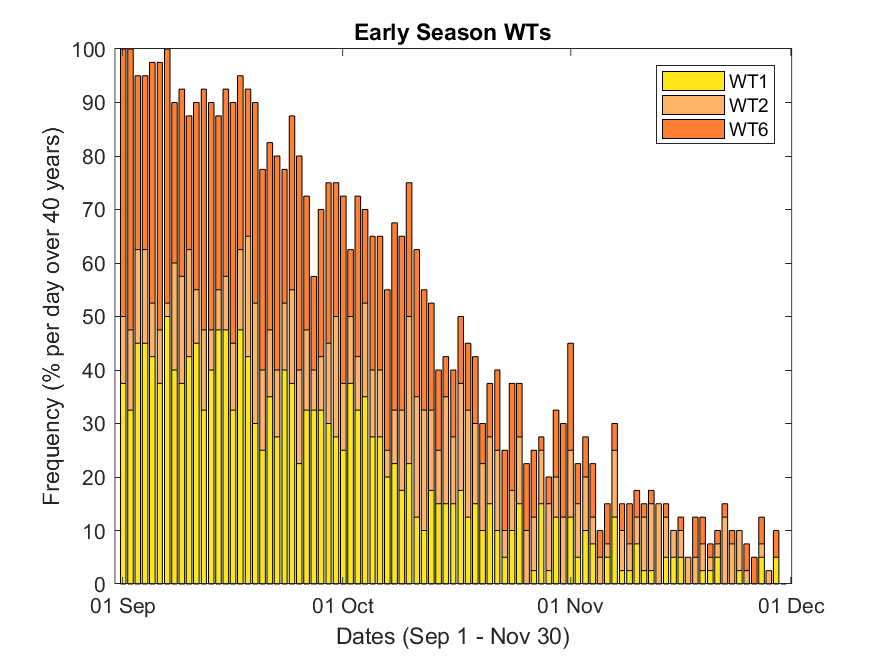
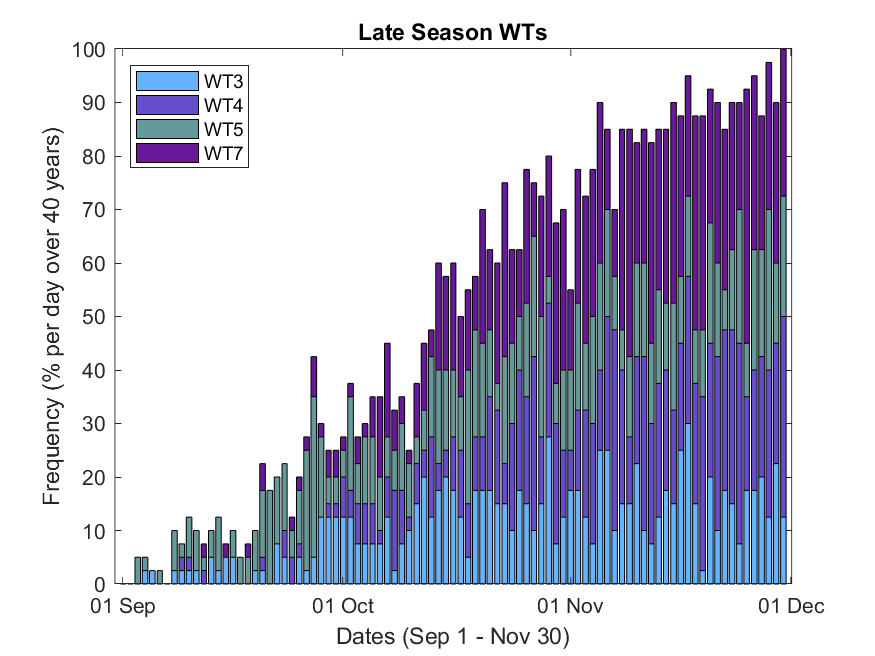
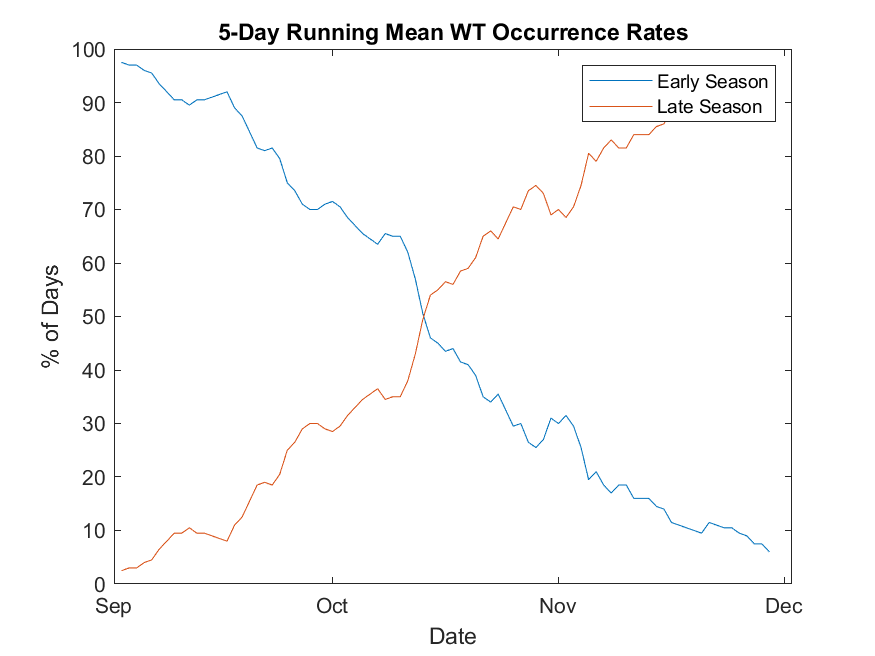
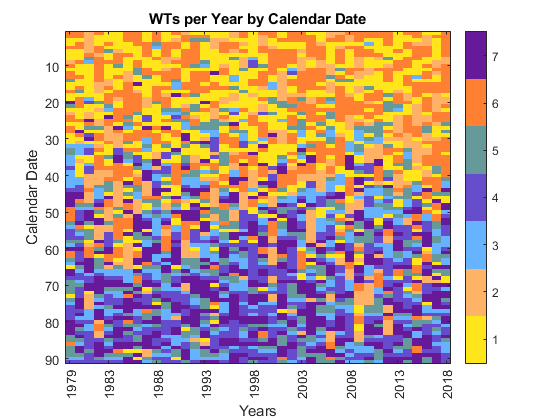
Chart, bar chart

Description automatically generated



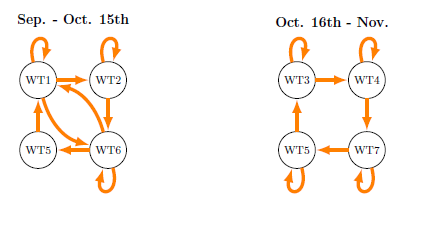
**Figure 6: Autumn Season WT Frequency During Teleconnections**

Teleconnection indices (-/+, neutral phases did not show any significance) in relation to each WT. Red shading indicates a WT that is less likely to occur and blue shading indicates a WT that is more likely to occur based on the 95% Monte Carlo confidence level (grey shaded bars).



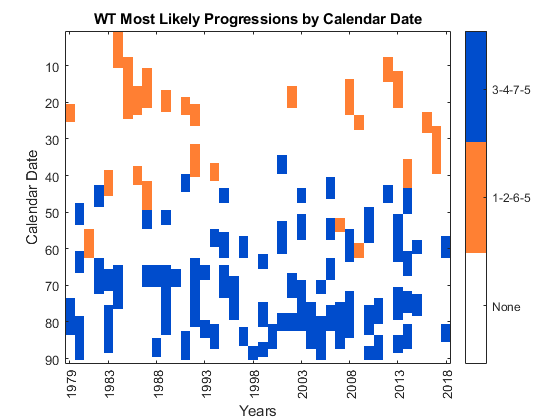
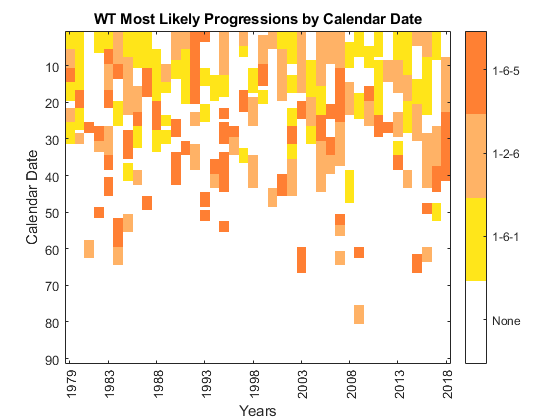
**Figure 7. Autumn Season Early and Late Season WT Daily Occurrence Rate** 

**(a,b.)** The frequency of occurrence of Early **(a)** and Late **(b)** Season WTs on each day of the season over all years. **(c)** The same as **(a)** and **(b)**, but smoothed using a 5-day running mean. Note the point of intersection at Oct 15th. **(d)** Klee diagram showing the WT occurrence per day over all 40 years. Y axis reads from top (Sep 1st) to bottom (Nov 30th)



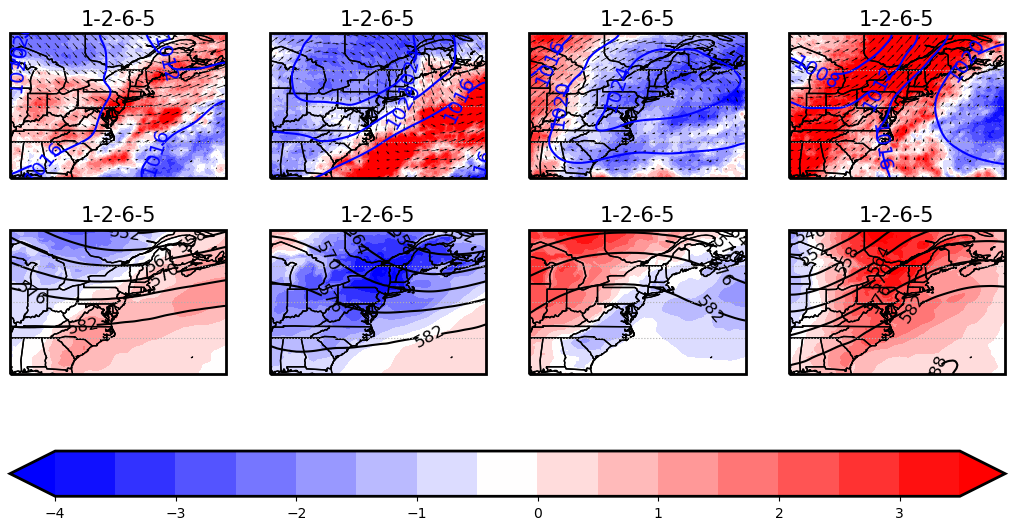
**Figure 8.** **Autumn season WT Progressions**

WT progression for the first and second halves of the season based on Markov chain analysis. Arrows indicate progressions that are significant at the 95% level.



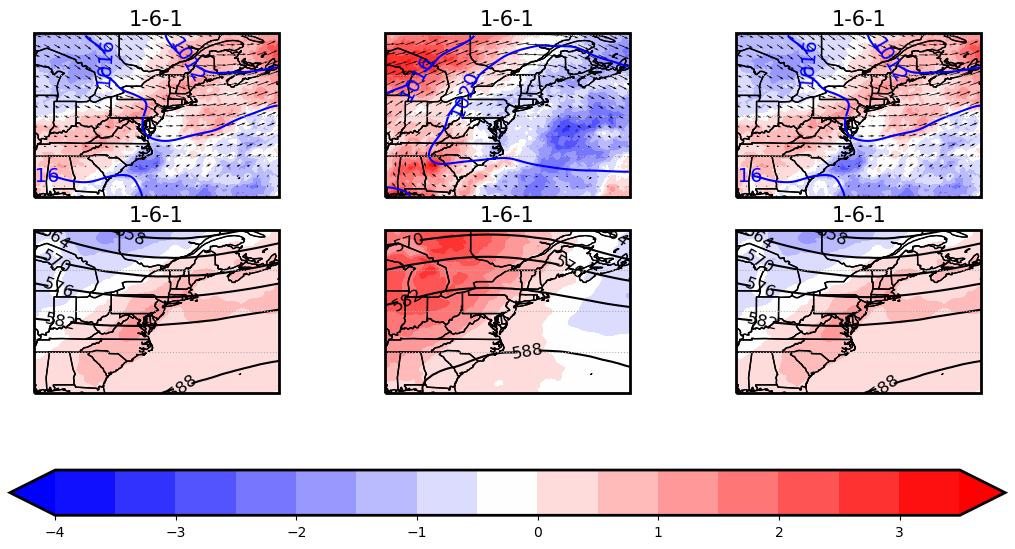
**Figure 9: Autumn Season Early and Late Season WT Daily Occurrence Rate**

**(a)** Klee diagram indicating the timing of the most likely WT sequences found using Markov Chain analysis. The Early Season was the only pattern which saw WTs that could progress in chains of 3 (e.g. 1-6-1,1-2-6). **(b)** Klee diagram indicating the timing of the most likely WT sequences found using Markov Chain analysis.



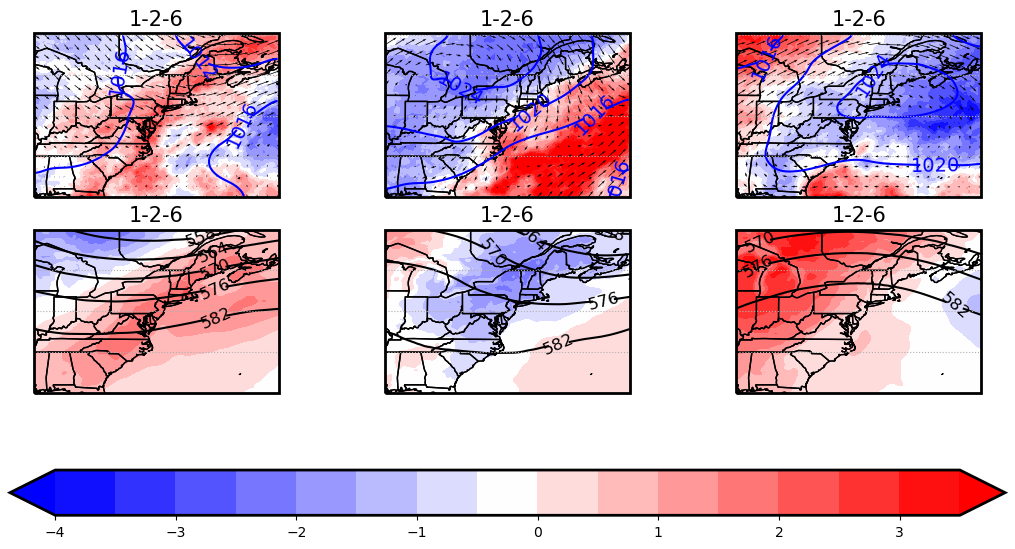
**Figure 10. Autumn Early Season Sequence (E1)**

**(a,b,c)** MSLP (blue), 850 hPa wind, and daily precipitation anomaly (mm/day, shaded); **(d,e,f)** 500 hPa height (black) and 2-meter temperature anomaly (°C, shaded) for WTs in the average 1-2-6-5 sequence.



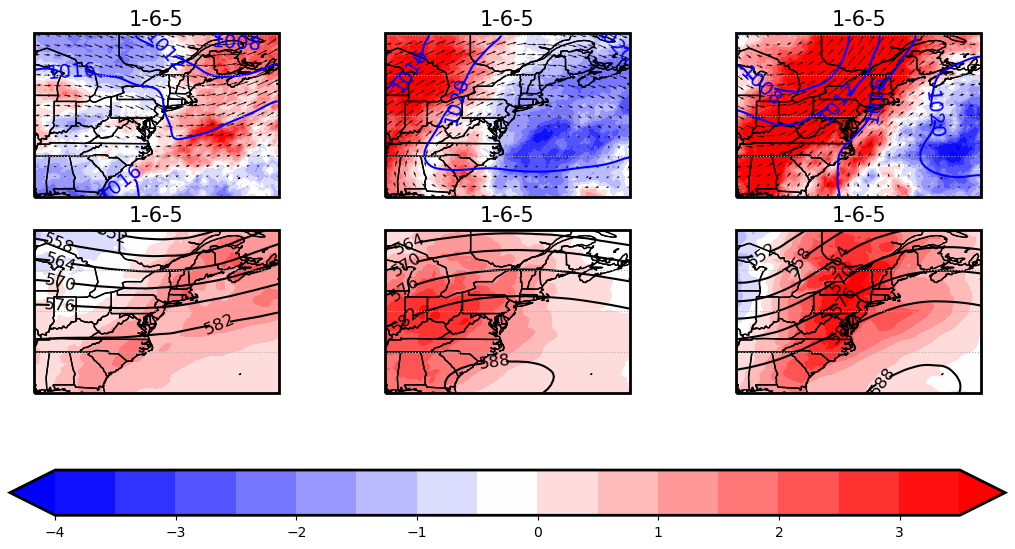
**Figure 11. Autumn Early Season Sequence (E2)**

**(a,b,c)** MSLP (blue) , 850 hPa wind, and daily precipitation anomaly (mm/day, shaded); **(d,e,f)** 500 hPa height (black) and 2-meter temperature anomaly (°C, shaded) for WTs in the average 1-6-1 sequence.



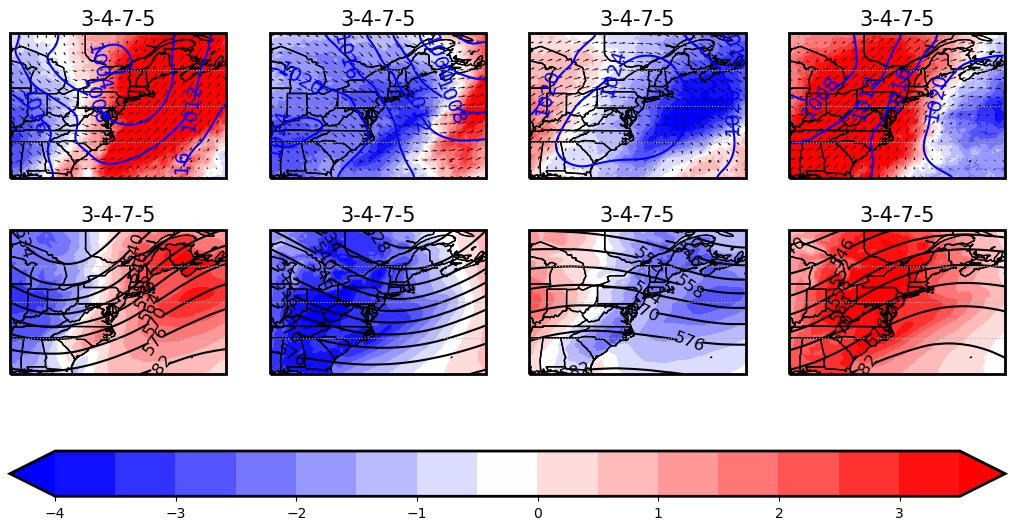
**Figure 12. Autumn Early Season Sequence (E3)**

**(a,b,c)** MSLP (blue) , 850 hPa wind, and daily precipitation anomaly (mm/day, shaded); **(d,e,f)** 500 hPa height (black) and 2-meter temperature anomaly (°C, shaded) for WTs in the average 1-2-6 sequence.



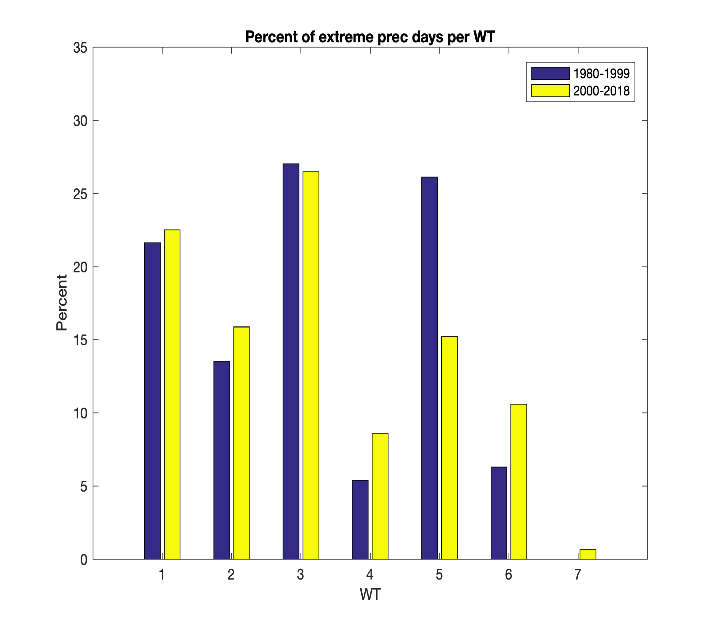
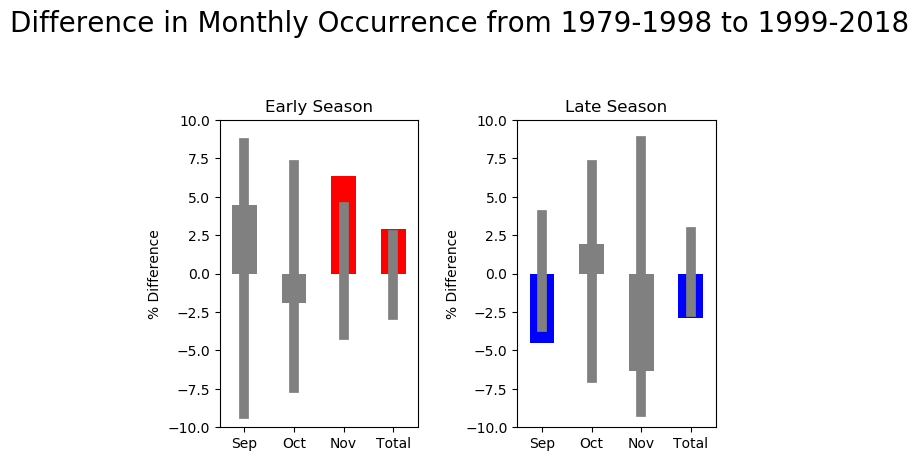
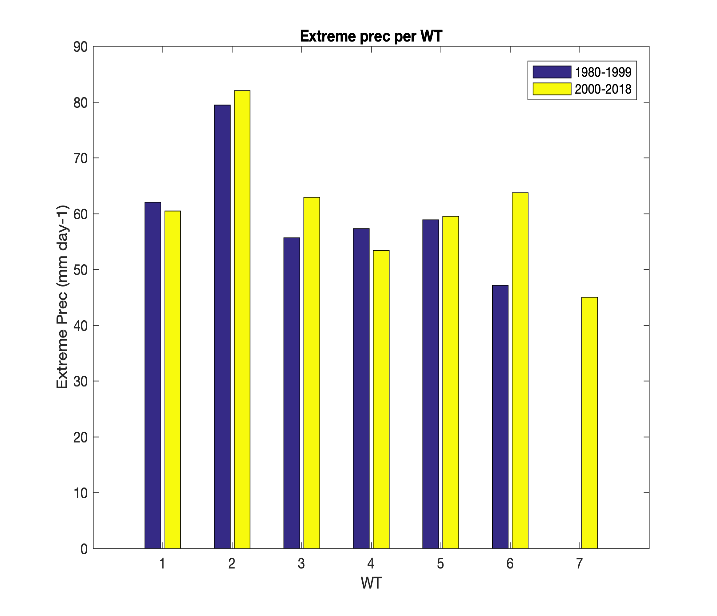
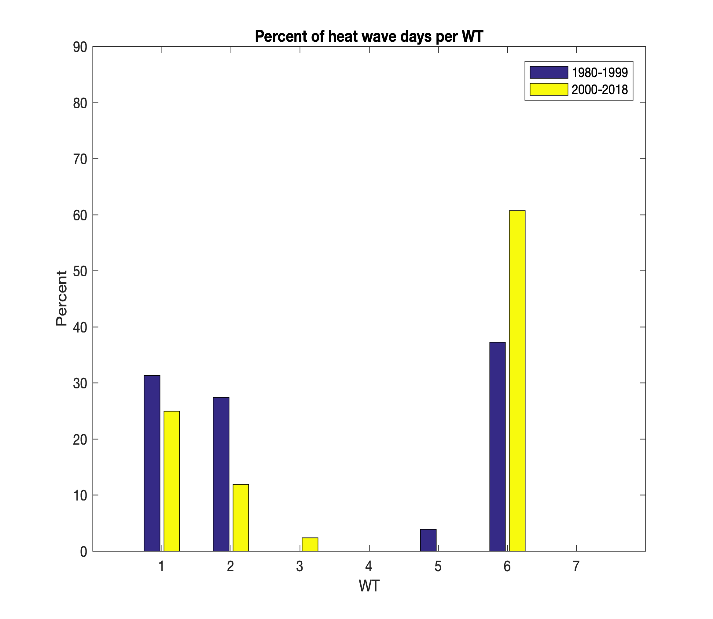
**Figure 13. Autumn Early Season Sequence (E4)**

**(a,b,c)** MSLP (blue) , 850 hPa wind, and daily precipitation anomaly (mm/day, shaded); **(d,e,f)** 500 hPa height (black) and 2-meter temperature anomaly (°C, shaded) for WTs in the average 1-6-5 sequence.



**Figure 14. Autumn Late Season Sequence (L1)**

**(a,b,c)** MSLP (blue) , 850 hPa wind, and daily precipitation anomaly (mm/day, shaded); **(d,e,f)** 500 hPa height (black) and 2-meter temperature anomaly (°C, shaded) for WTs in the average 3-4-7-5 sequence.



**Figure 15. Changes in Monthly Occurrence, Extreme Precipitation, and Heat Wave Days of Early and Late Season WTs**

Preliminary analysis between the first and second halves of the 1979-2018 period for differences in: **(a)** Frequency of occurrence of Early and Late season WTs, blue shaded bars show a decrease and red shaded bars show an increase that is significant at the 95% level . **(b)** Frequency of extreme precipitation days per WT. **(c)** Extreme precipitation amount per WT. **(d)** Frequency of occurrence of heat wave days per WT.