**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, and typical progressions between clusters. The clusters are then used to make a circulation-based distinction between early and late autumn.

Seven WTs are identified for the autumn season: WTs 1 and 2 feature a weak upper-level trough, WTs 3 and 4 have a strong upper-level trough, WTs 5 and 6 feature a strong upper-level ridge, and WT7 shows a weak upper-level ridge with surface high pressure over the region. Three main sequences of typical progressions between the circulation clusters emerge: 1) A sequence in September, defined by WT1-WT6-WT1; 2) A sequence that primarily occurs in October, defined by WT1-WT6-WT5-WT1; and 3) A sequence that occurs most often from mid-October to November defined as WT5-WT3-WT4-WT7-WT5. 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 2007a, Moron et al. 2008, Moron et al. 2015, Qian, J.-H, A.W. Robertson and V. Moron, 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 end 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 or mean fields, 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.

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.

***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. Correlation and Root-Mean-Squared-Error (RMSE) analysis of clusters***

An analysis of the spatial correlation and RMSE is performed on each pattern using all dates (Figure 2). In general, dates that fall within the clusters (blue bars) have high correlation coefficient and low RMSE values for the 500-hPa heights and MSLP, which suggests a successful clustering, comparatively; while, dates that fall outside of the clusters (gray bars) have low correlation coefficient and high RMSE. As with any pattern matching algorithm, there are a number of cluster days that correlate better with other clusters, since some days can have similar features to multiple clusters. These days make up only a small percentage (<10%) of the total days.

**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 (Figure 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. 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 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. WT4 presents an upper level trough over the Northeast with surface low-pressure over New Brunswick. Low level winds are primarily out of the northwest. WT5 features an upper level ridge over the Atlantic Ocean and an upper level trough over the Midwest. 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. 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. Low-level flow is mainly out of the west/northwest with an area of high-pressure located at the surface over the Eastern Seaboard.

A complementary way to describe the autumn season is by investigating the anomalies of each WT with respect to the seasonal mean (Figure 4). WTs 6 and 7 both feature higher than average surface pressure over the whole of the region (blue contour lines). In WT6, this is due to stronger than average upper-level heights, while in WT7 this is due to lower than average upper-level heights. WTs 1, 3, and 4 all feature lower than average surface pressure over most of the Northeast region. WTs 3 and 4 both have lower than average upper-level heights due to the strong troughing over the Northeast. WT1 sees slightly higher than average upper-level heights, but the portion of the region which experiences lower than average surface pressure exhibits average upper-level heights. WTs 2 and 5 both show areas of higher and lower than average surface pressure and upper-level heights, albeit in differing locations. WT5 has lower than average surface pressure and upper-level heights over the Midwest, while WT2 has them over the southern portion of the region extending east into the Atlantic Ocean. WT5 features higher than average pressure and upper-level heights in the eastern portion of the region, while WT2 sees them in the northern portion of the region.

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

To measure variability of WT frequency on a monthly basis, the summation of WT frequency per month over all years is calculated (Figure 5) and from this three groups are formed: 1.) WTs 1, 2, and 6 in September, 2.) WTs 2, 3, and 5 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. In the month of October, even though we expect WTs 2, 3, and 5 to be the most likely WTs to occur, they have a lower frequency of total days of occurrence over the study period compared to the WTs that are less likely to occur (WTs 1,6, and 7). This is most likely due to October being the middle of the season, when we would expect a slow decline in the occurrence of WTs that are likely to occur in September and a slow increase in the occurrence of WTs that are likely to occur in November.

To further examine the sub-seasonality of the WTs, the monthly occurrence rates (Figure 5) are combined with an analysis of WT progression (Figure 6). WT progression is determined by examining for each day the WT that is assigned to the day. This creates a record of the frequency of progression from each individual WT to the next. A Monte Carlo analysis, like the method used for monthly occurrence rates, is applied to determine statistical significance.  Individually, each WT, has a tendency (greater than expected from a random distribution) to persist; that is, to be followed on the next day by the same WT. Each WT is statistically likely to persist for at least 2 days over 85% of the time with WTs 2 and 6 likely to persist for up to 5 days 5% of the time (Figure 7). Additionally, all the WTs, except for WT2, tend to be followed by at least one other WT. From this, progression sequences between WTs can be identified and 2 sequences emerge that occur more often than would be expected by chance at the 0.05 level: 1) WT1 – WT6 – WT5 – WT1, and 2) WT5 – WT3 – WT4 – WT7 – WT5. WT1 can progress to WT2, but WT2 progresses to no other WTs other than itself, so it is disregarded as being part of any sequence.

The average frequency of occurrence for the two WT sequences varies notably over the course of the season and they are defined by when they occur during the season (as shown in Figure 8, along with percent frequencies showing progression tendency significant at the 0.05 level). This establishes WT1-WT6-WT5-WT1 as the “Early Season” sequence and WT5-WT3-WT4-WT7-WT5 as the “Late Season” sequence. The Early Season sequence breaks down further into two parts: 1.) WT1-WT6-WT1, and 2.) WT1-WT6-WT5-WT1. The primary sequence (1.) occurs throughout the entirety of September and October, while the secondary sequence (2.) occurs mainly in October, coinciding with an increase in the likelihood of WT5 to occur.

 The “Early Season” and “Late Season” sequences transition between one another depending both on the time of season and the WT5 progression. WT5 becomes more likely to occur in October, which is the same time that the Early Season WTs become less likely to occur and the Late Season WTs become more likely to occur. WT5 acts as a bridge between the early and Late Season sequences, eventually becoming part of the Late Season sequence in November (when Early Season WTs are statistically less likely to occur than expected by chance at the 0.05 level).

***3c. Early Season vs Late Season WTs***

From the previous sections, 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 9. 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 9c). 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.

***3d. 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. 10. 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.

**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. Three WT sequences emerged, of which two capture the progression of “Early Season” weather and one of which captures the progression of “Late Season” weather.

WT1 exhibits more zonal upper-level flow with weak troughing apparent. 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. WT3 has an area of surface low pressure over Northern New England accompanied by an upper-level trough over the Great Lakes region. 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. WT5 features an upper level ridge over the Atlantic Ocean and an upper level trough over the Midwest. 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. Finally, WT7 features a more zonal upper level flow with a slight upper level ridge over the Midwest and slight upper level trough over the Atlantic Ocean.

WTs are further divided using monthly frequencies and progression between WTs, leading to the development of sequences of WT progression defined by their seasonality. “Early Season” sequences feature 3 WTs (WTs 1, 5, and 6), two of which are significantly likely to occur in September and significantly unlikely to occur in November (WTs 1 and 6). 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. WT5 occurs in both sequences and is the WT which links them together. WT5 is significantly likely to transition to either itself, WT1 or WT3, and isn’t important to the progression of the Early Season until October, when it becomes more likely to occur, and helps to facilitate the development of the secondary “Early Season” sequence. 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 of time. 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 more broadly capture the reflection of seasonal evolution 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)

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

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>

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.

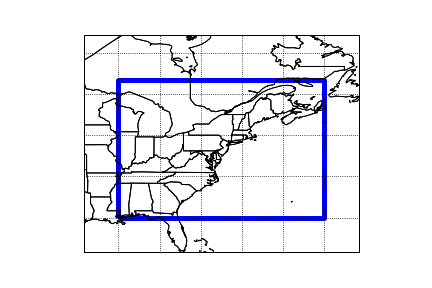
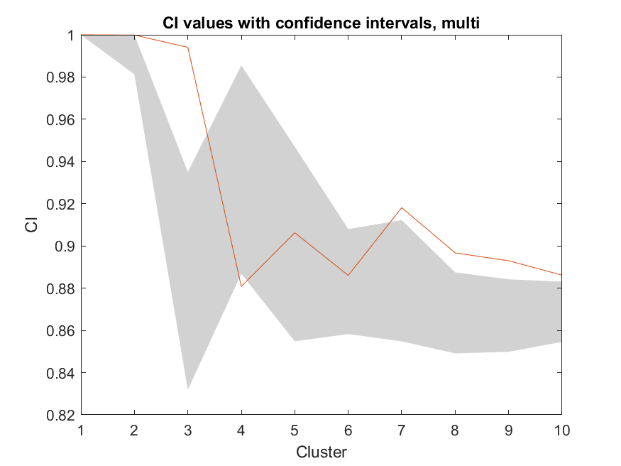
Sheridan, S. C., 2002: The redevelopment of a weather-type classification scheme for North America. Int. J. Climatol., 22, 51–68, doi:https://doi.org/10.1002/joc.709

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>

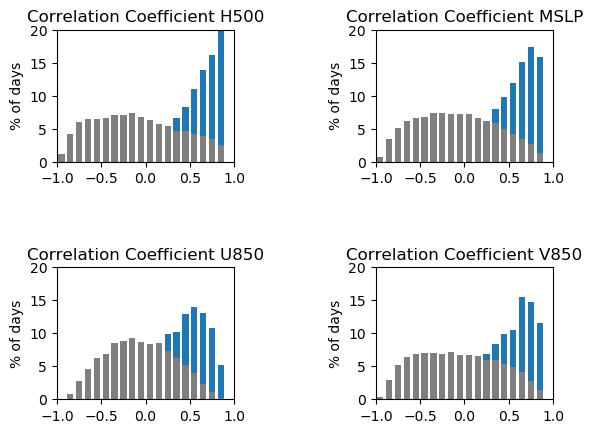
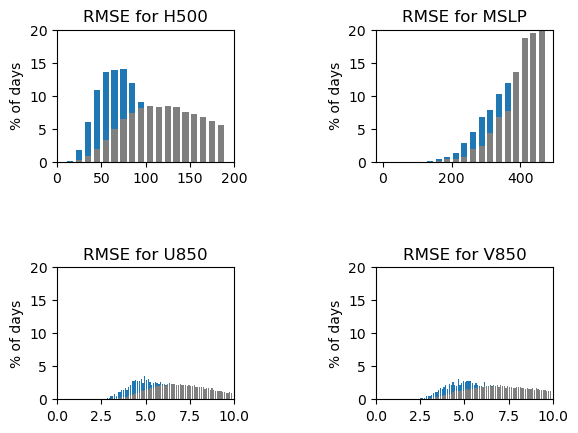
Trenberth, K.E., 1983: What are the Seasons?. Bull. Amer. Meteor. Soc., 64, 1276–1282, doi: 10.1175/1520-0477

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.

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

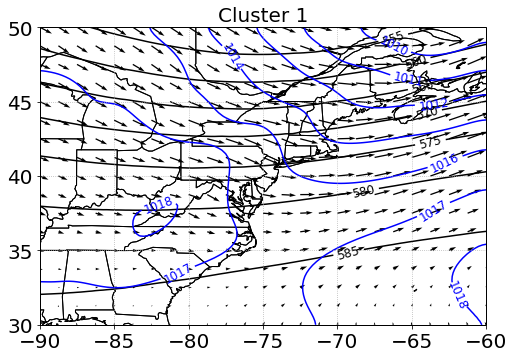
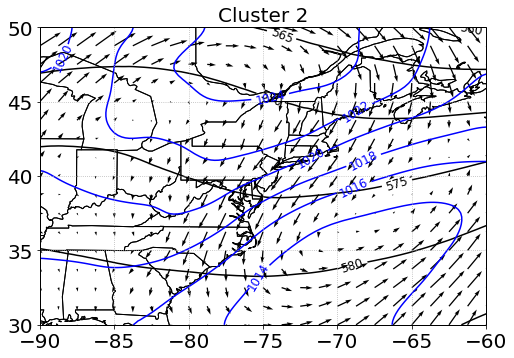
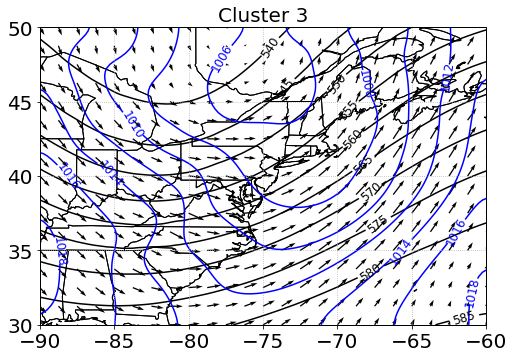
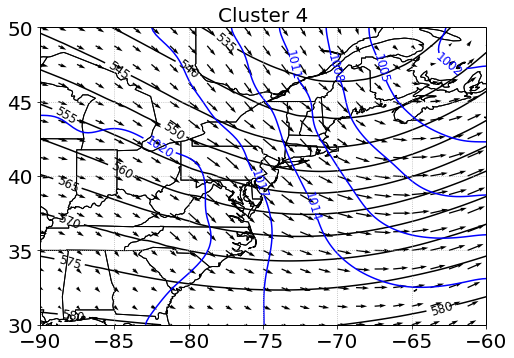
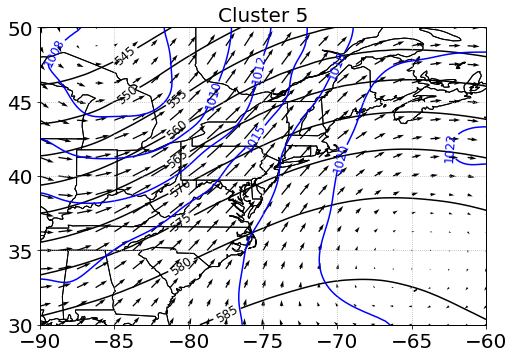
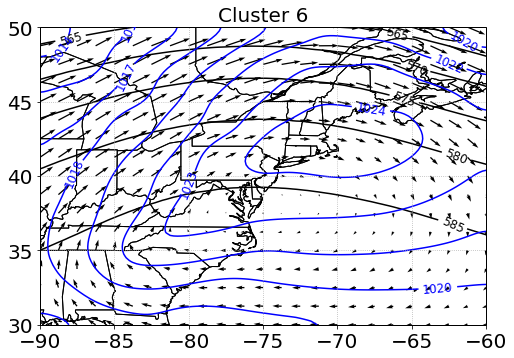
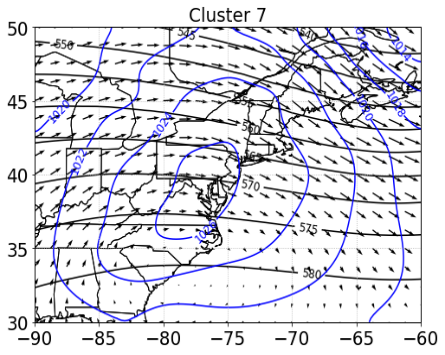


**Figure 1.** Map of the bounded area used in this study (left) and k-means clustering results (right). Classifiability Index (red line) shows the results of the k-means clustering for k=1...10. Red noise was made from the dataset and run through k-means to determine the 90% confidence interval of the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



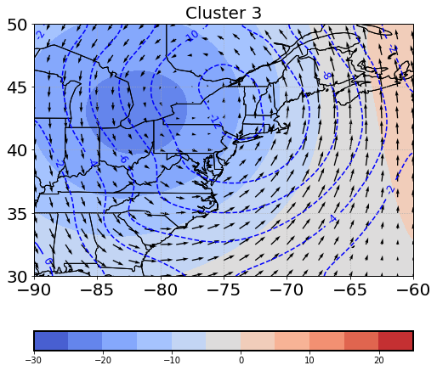
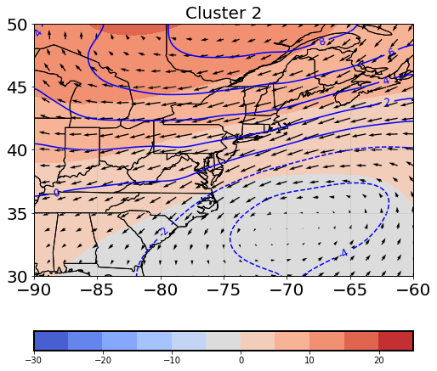
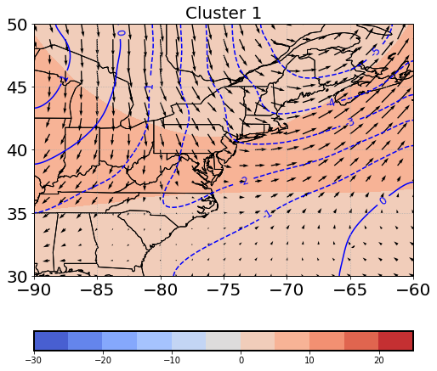
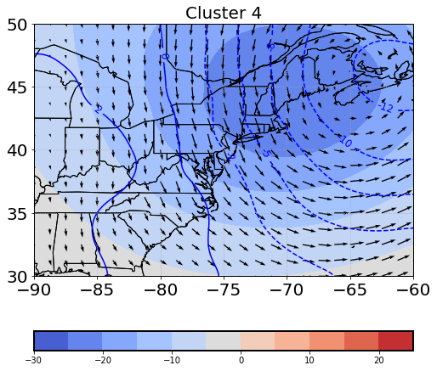
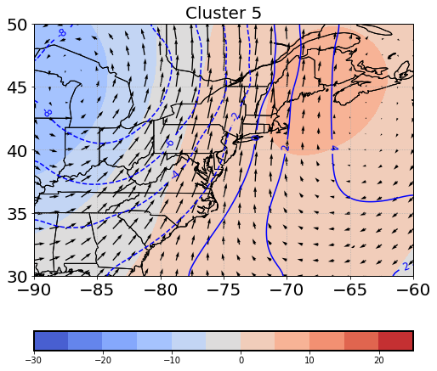
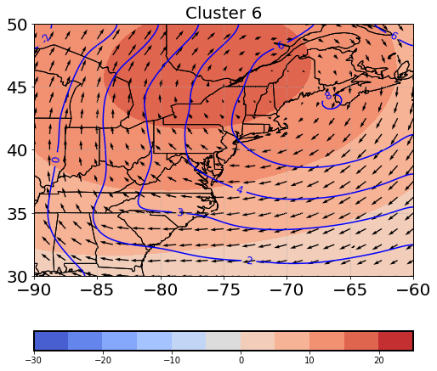
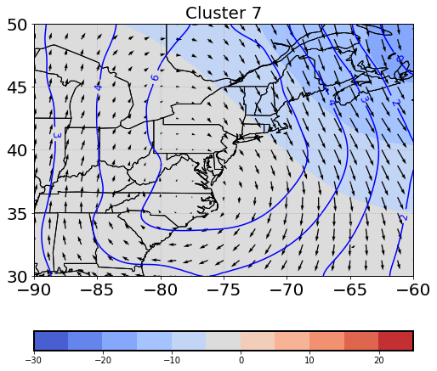
**Figure 2.**

Correlation Coefficient (left) and RMSE (right) for daily fields compared to each Weather Type with regards to 500-hPa height (top left panel), MSLP (top right panel), 850-hPa U-wind component (bottom left panel), and 850-hPa V-wind component (bottom right panel). Blue shaded bars represent values for daily fields compared to their assigned WT, and gray shaded bards represent daily fields compared to other WTs.



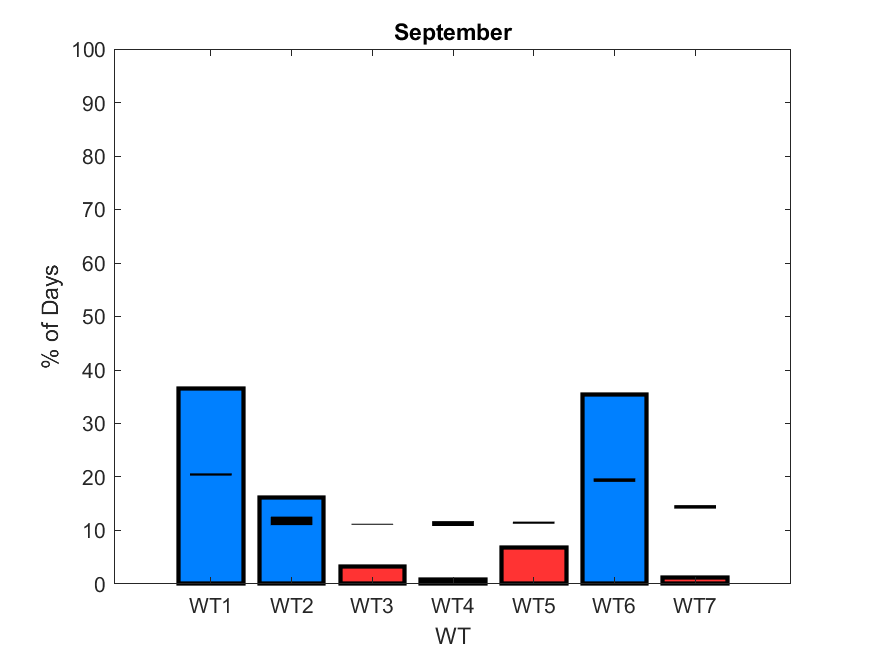
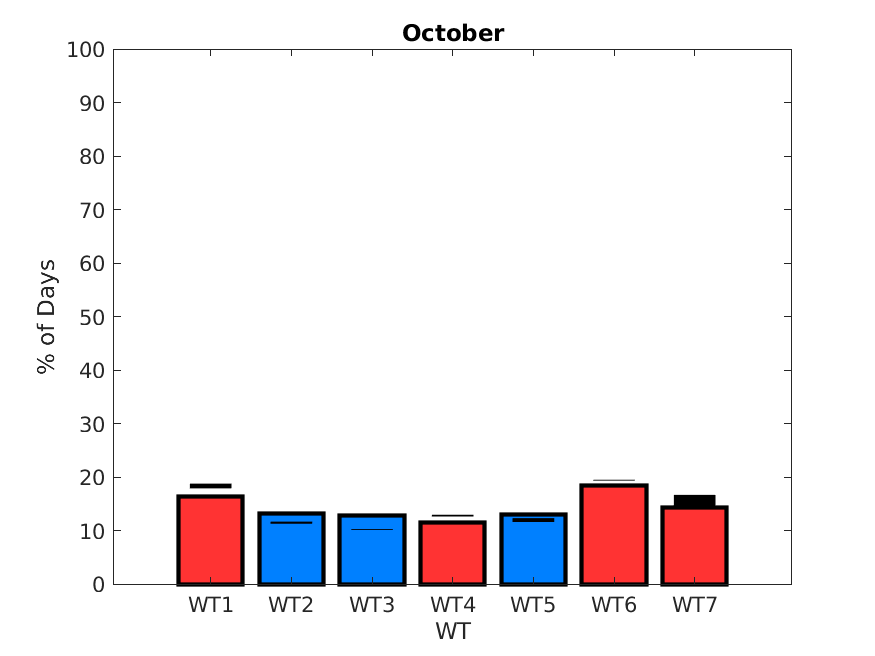
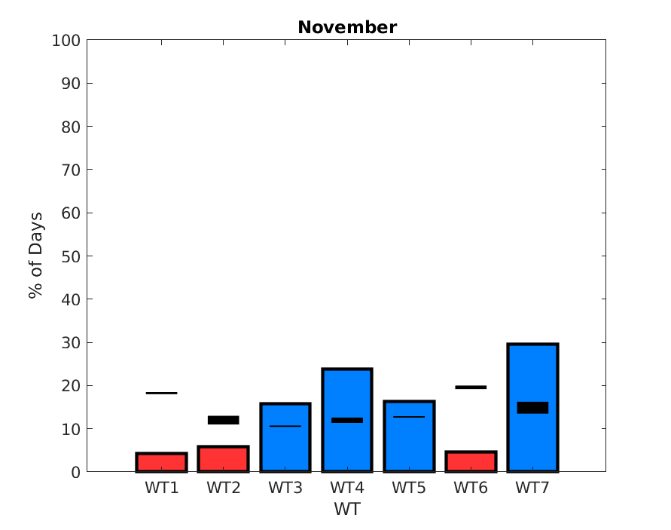
**Figure 3: Autumn Season Weather Types**

ERA-5 data from 1979-2018 split into 7 clusters using k-means analysis on the MSLP (blue), 500-hPa geopotential height (black), and 850-hPa wind fields (vectors).



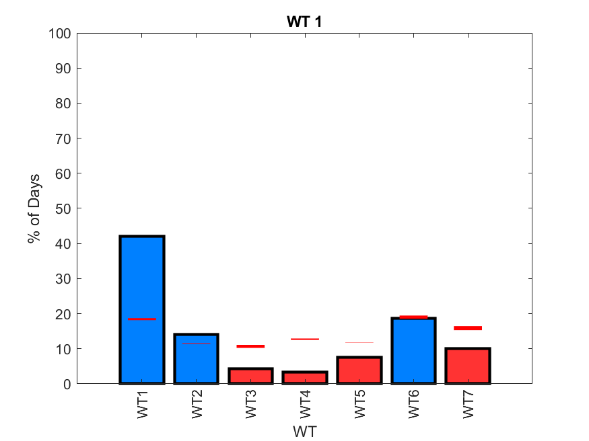
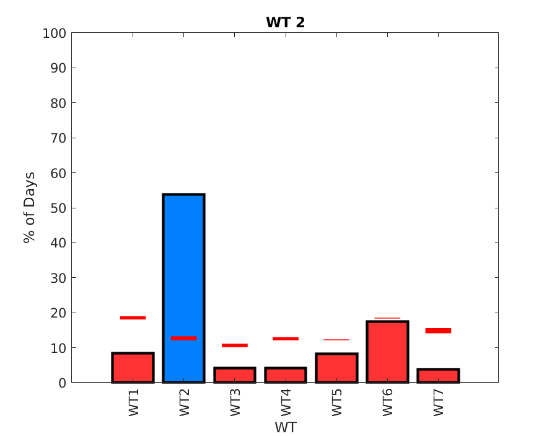
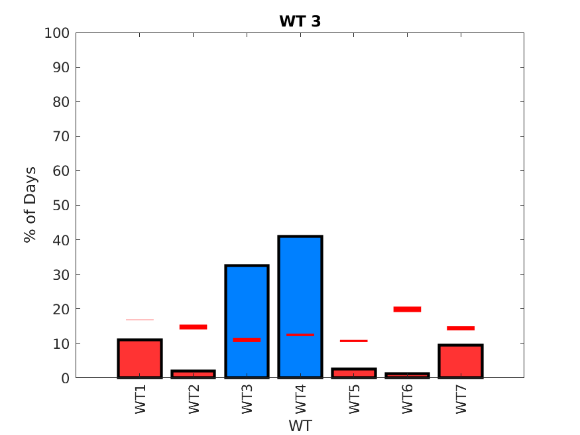
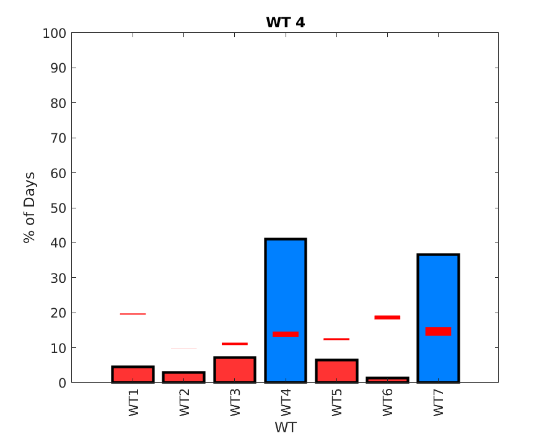
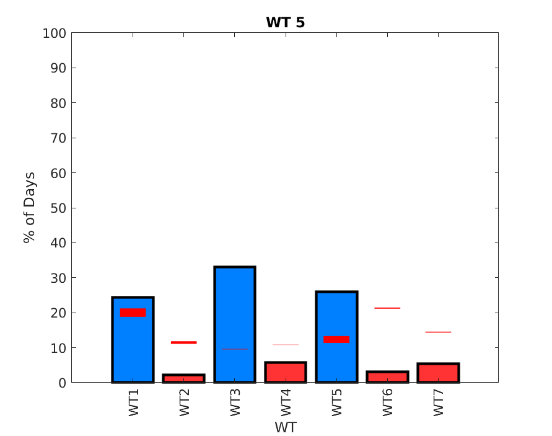
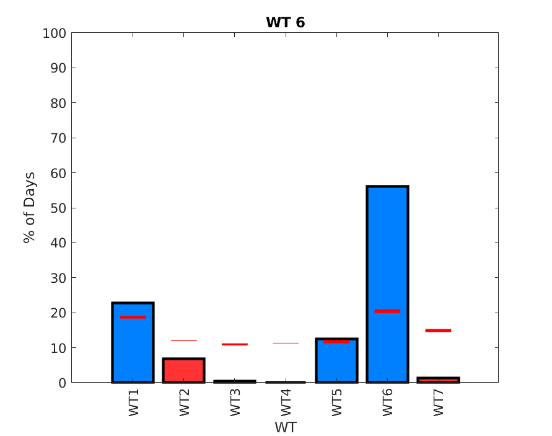
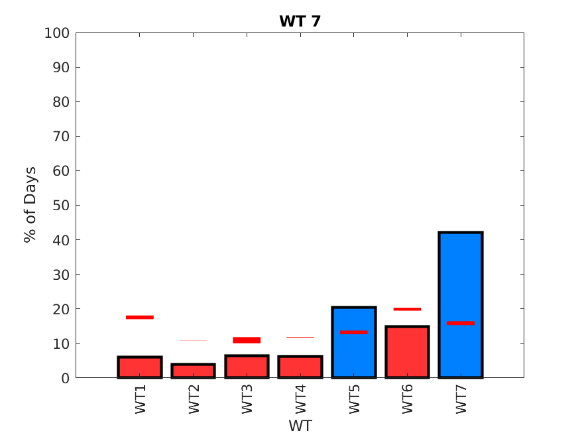
**Figure 4: Autumn Season Weather Type Anomalies**

ERA-5 MSLP, 500-hPa geopotential heights, and 850-hPa winds from 1979-2018 split into 7 clusters with the long-term season mean removed. MSLP anomalies are plotted in blue, 500-hPa height anomalies are shaded, and 850-hPa wind anomalies are vectors.



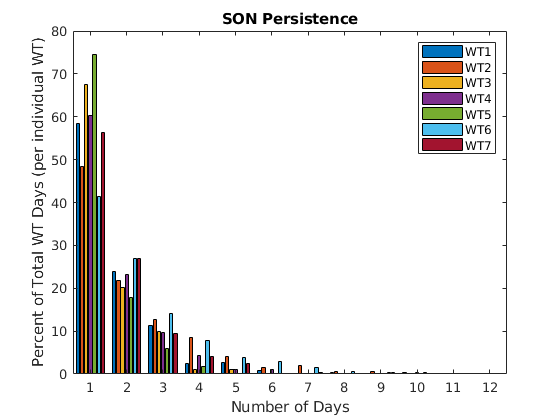
**Figure 5: Autumn Weather Type Monthly Occurrence**

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 (black bars) is found using the Monte Carlo method. Red bars indicate WTs that are less likely to occur than expected due to chance during a given month and blue bars indicate WTs that are more likely to occur during a given month.



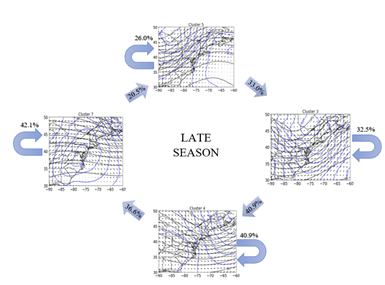
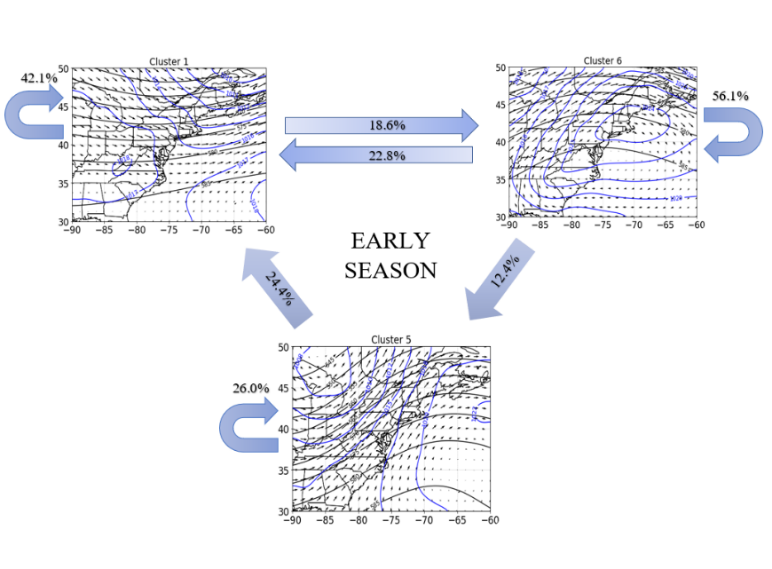
**Figure 6: Autumn Season Weather Type Progression**

The progression from each WT to every other WT as percent of total days of each WT. A Monte Carlo method is applied to create a 95% confidence interval (red bar). Blue shaded bars indicate a progression that is more likely between two WTs, while a red shaded bar indicates a progression that is less likely.



**Figure 7. Weather Type Persistence**

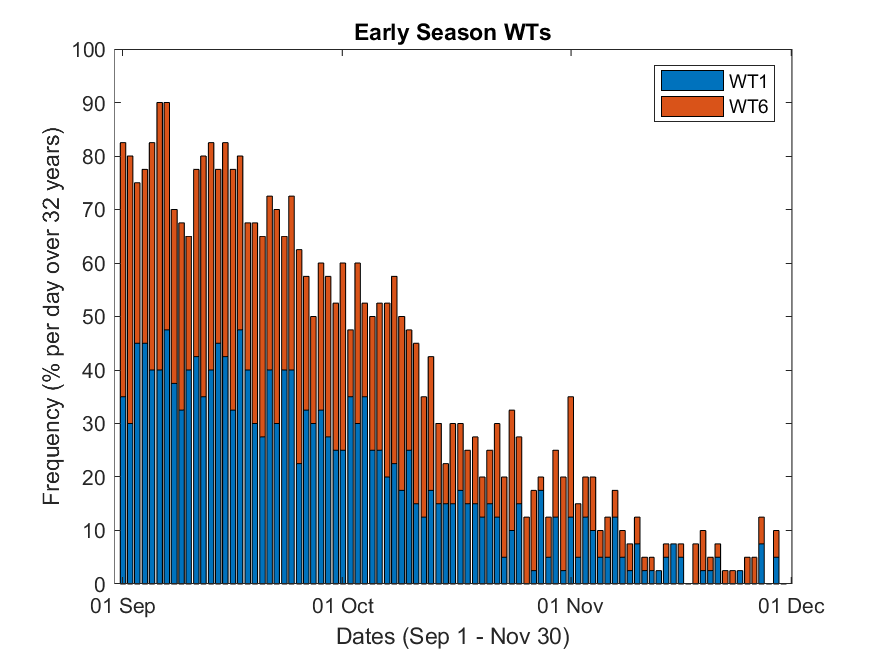
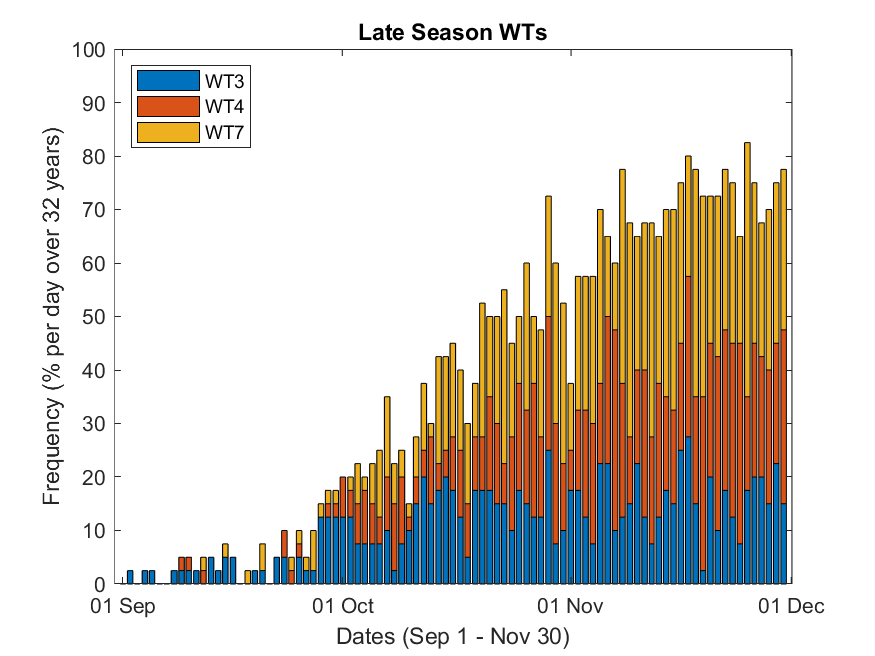
The persistence of each WT measured as percent of total days of each WT.

**Figure 8.** **Autumn season Early and Late Season WTs**

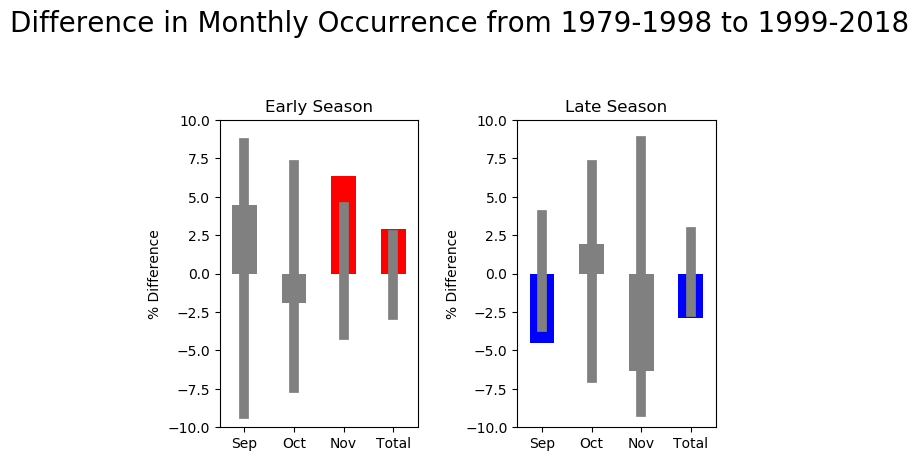
The Early (left) and Late (right) season WT patterns based on statistical significance of WT progression. Straight arrows show significant transitions (at the 0.05 level) between individual WTs with occurrence rates marked as percentages. Curved arrows signify the likelihood of each WT to persist.

A close up of a map

Description automatically generated

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

Early Season (left) and Late Season (right) WTs displayed as percent occurrence per day of the season over the 40-year period. The bottom image presents the “Early Season” (blue) and “Late Season” (red) percent occurrence per day of the season over the time period with a 5-day running mean applied.



**Figure 10. Change in Monthly Occurrence of Early and Late Season WTs**

The “Early Season” (left) and “Late Season” (right) WTs based on percent difference of WT occurrence rate between the last 20 years and the first 20 years of the study. Thin, gray bars show the 95% confidence interval found with the Monte Carlo method. Red bars indicate a significant increase, blue bars indicate a significant decrease and grey bars indicate no statistically significant change.