Clustering Analysis of Autumn Weather Regimes in the Northeast United States

DAVID COE,[a] MATHEW BARLOW,[a] LAURIE AGEL,[a] FRANK COLBY,[a], CHRISTOPHER SKINNER,[a], AND JIAN-HUA QIAN[b]

[a] *Department of Environmental, Earth, and Atmospheric Sciences, University of Massachusetts Lowell, Lowell*
[b] *Savannah River National Laboratory, Aiken, South Carolina*

(Manuscript received 5 June 2020, in final form 27 May 2021)

ABSTRACT: A *k*-means clustering method is applied to daily ERA5 500-hPa heights, sea level pressure, and 850-hPa winds, 1979–2008, to identify characteristic weather types (WTs) for September–November for the northeast United States. The resulting WTs are analyzed in terms of structure, frequency of occurrence, typical progressions, precipitation and temperature characteristics, and relation to teleconnections. The WTs are used to make a daily circulation-based distinction between early and late autumn and consider shifts in seasonality. Seven WTs are identified for the autumn season, representing a range of trough and ridge patterns. The largest average values of precipitation and greatest likelihood of extremes occur in the Midwestern Trough and Atlantic Ridge patterns. The greatest likelihood of extreme temperatures occurs in the Northeast Ridge. Some WTs are strongly associated with the phase of the North Atlantic Oscillation and Pacific–North America pattern, with frequency of occurrence for several WTs changing by more than a factor of 2. The two most common progressions between the WTs are one most frequent in September, Mid-Atlantic Trough to Northeast Ridge to Mid-Atlantic Trough, and one most frequent in mid-October–November, Midwestern Trough to Northeast Trough to Midwestern Trough. This seasonality allows for a daily WT-based distinction between early and late season. A preliminary trend analysis 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 a longer period of warm season patterns and a shorter, delayed period of cold season patterns.

KEYWORDS: Annual variations; Climate variability; Multidecadal variability; Seasonal cycle; Seasonal variability

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, for example, Legras et al. (1987), El-Kadi and Smithson (1992), Yarnal (1993), Huth et al. (2008), Ramos et al. (2015), and Muñoz et al. (2017). Here, we take an objective approach based on *k*-means clustering applied to daily regional height and circulation fields at the surface, low, and midtropospheric levels for the northeast United States in the autumn season (September–November). A similar typing approach for the northeast United States has been previously applied by Roller et al. (2016) for the winter season. The goals of this study are to determine the WTs that describe the autumn weather in the northeast United States, to examine their basic properties, to identify the characteristic progression between WTs as weather systems evolve and move through the region, and to analyze the differences between the early and late autumn season as described by the WTs. The focus on the autumn transition season allows for a test of whether daily circulation patterns can be used to characterize seasonal evolution at the regional level.

The WT analysis approach used here is the *k*-means clustering technique (Diday and Simon 1976; Ghil and Robertson 2002; Moron et al. 2010), with the classifiability index (CI) algorithm of Michelangeli et al. (1995). Combining *k*-means with the CI provides a semiobjective approach to identifying the optimal clustering of the dataset (semiobjective 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 only been applied to the Northeast for analysis of the winter season (Roller et al. 2016). However, this technique has been used in multiple other studies to isolate atmospheric circulation patterns on a time-scale and continental and regionalized basis (Lana and Fernandez-Mills 1994; Sheridan 2002; Coleman and Rogers 2007; Moron et al. 2008, 2015; Qian et al. 2013; Allen and Sheridan 2016; Meseguer-Ruiz et al. 2020).

In their analysis of Northeast winter circulation, Roller et al. (2016) found clustering analysis useful for identifying the evolution of characteristic daily weather patterns and their association with precipitation and large-scale teleconnections. They identified five distinct WTs, with a typical progression between them capturing the movement of a trough and associated precipitation across the region. They found a strong relationship with the Arctic Oscillation (AO) and North Atlantic Oscillation (NAO), a weak relationship with the Pacific–North American Pattern (PNA), and little significance in relationship with El Niño–Southern Oscillation (ENSO), consistent with the previous research on the influence of teleconnections on the region. However, the daily WT perspective adds considerable nuance, as the influence of the teleconnections varies strongly by WT.

While winter variability has been the focus of considerable analysis, there has been much less attention to the autumn season. At the continental scale, previous analyses of autumn in the United States have focused on frost onset timing (Easterling 2002;

Denotes content that is immediately available upon publication as open access.

Corresponding author: David Coe, david_coe@uml.edu

DOI: 10.1175/JCLI-D-20-0243.1

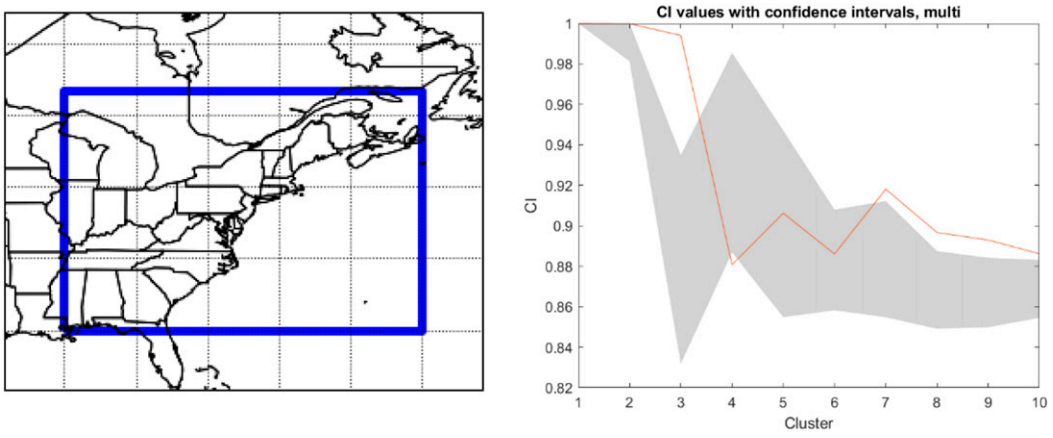

FIG. 1. (left) Map of the bounded area used in this study and (right) *k*-means clustering results. CI (red line) shows the results of the *k*-means clustering for $k = 1, \ldots, 10$. Red noise was made from the dataset and run through *k*-means to determine the 90% confidence interval of the CI (gray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

Cooter and LeDuc 1995), definition of seasons (Trenberth 1983), circulation patterns across the Northern Hemisphere (Sheridan 2002; Fleming et al. 1987), and the effects of a changing climate (both natural and anthropogenic in nature) on seasonal length and onset/withdrawal timing (Allen and Sheridan 2016; Park et al. 2018). 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 and focused on patterns indicating the broad continental-scale changes in circulation. 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 United States; 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 becoming longer due to both an earlier start of the summer season and a later start of the autumn season. How these changes are expressed in regional daily circulations is not yet clear.

The analysis here considers daily WTs during the autumn season, to provide more information on the warm season to cold season transition from a daily circulation perspective, and focuses regionally on the Northeast, to resolve a detailed view of the circulation. 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 are presented and analyzed with respect to precipitation and temperature, persistence, and progression between WTs, influence of teleconnections, and seasonal evolution. In section 4, the results are summarized and discussed.

2. Methods

a. 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 to 2018 are used in this study (Hersbach et al. 2020). These data are bound for the region of 30°–50°N latitude and 60°–90°W longitude centered on the northeast United States (Fig. 1). Since this study is concerned with circulation patterns, 500-hPa heights, 850-hPa meridional winds, and MSLP data are used when clustering to give a better understanding of how the circulation is influenced by surface pressures and midlevel 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 northeast 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).

b. *k*-means clustering

The *k*-means clustering method sorts data into k clusters (predetermined by the user) based on intracluster variance of the squared Euclidean distance between data points (Diday and Simon 1976; Roller et al. 2016). The number of clusters is chosen here based on the CI (Michelangeli et al. 1995; Moron et al. 2008), discussed further below. Each of the resulting clusters is considered to be a weather type.

Since this study uses a multivariate analysis, the data must be standardized to prevent undue weight from any individual field. Standardized anomaly fields are created at each grid

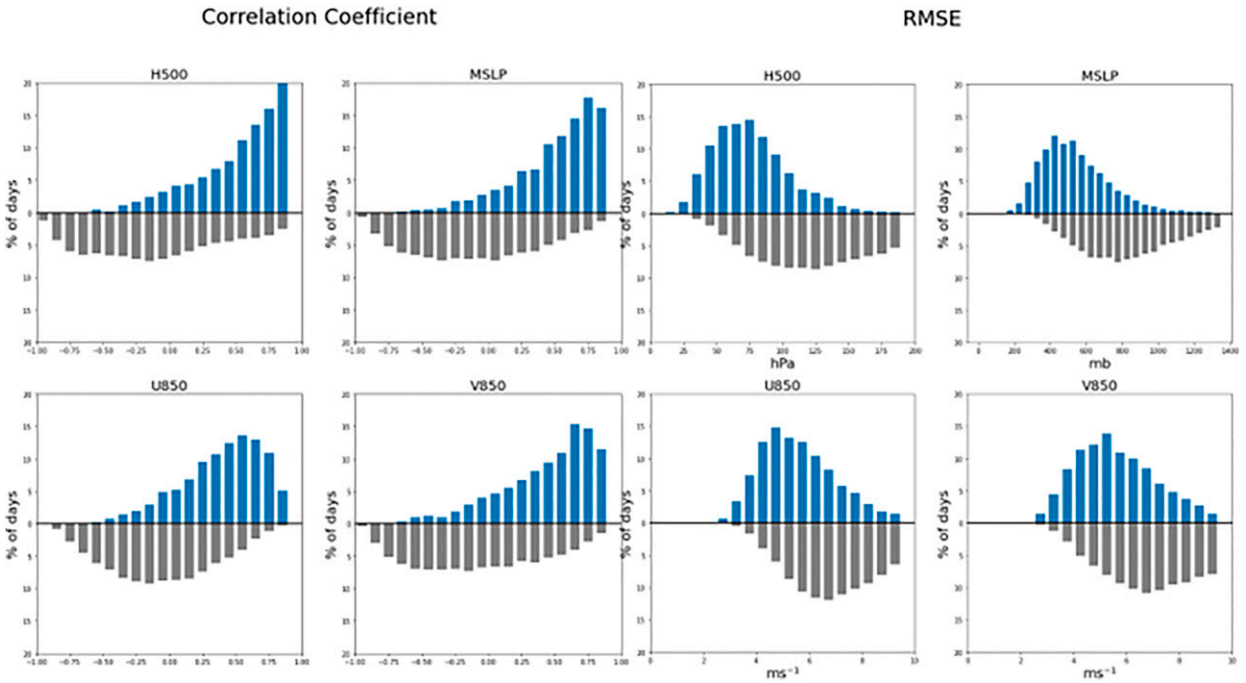

FIG. 2. Analysis of spatial correlation and RMSE between individual days and WT patterns, presented as a mirrored histogram for individual fields. The blue bars are the histograms of values between individual days and the WT pattern to which they are assigned, while the gray bars are the histograms of values between individual days and the WT patterns they are not assigned to. The histograms are mirrored in the *y* axis at zero, shown as a horizontal line; that is, values increase upward for the blue bars and downward for the gray bars. (left) Histograms of spatial correlations for the individual fields (H500, MSLP, U850, and V850) and (right) histograms of RMSE for the same fields.

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 $nt \times ns$ 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. 2008). This results in k distinct CI values. 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 that 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 chosen. Figure 1 shows the results, where the red line represents the CI values of the data, and the gray shading indicates the 90% confidence interval. Based on the CI results, there are several significant k that may be considered (3, 7, 8, 9, and 10). The choice of cluster number within this set is based on a trade-off between sample size and cluster number as well as an assessment of the sensitivity of the results. A smaller number of clusters will be better sampled but may conflate distinct processes, while a larger number of clusters may provide more distinction between patterns but will not be as well sampled in the data. Using only three clusters is not sufficient to capture multiple weather processes in a transition season, so that solution is not used. Comparison of the 7-, 8-, 9-, and 10-cluster solutions shows that using more than 7 clusters does not provide substantial new information and so the 7-cluster solution is used here. These seven clusters define the seven weather types used in the remainder of the analysis.

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.

To further examine the validity of the clustering algorithm, a spatial correlation and root-mean-square error (RMSE) analysis is performed between the spatial patterns of the seven clusters and every individual day in the dataset for each of the four fields considered (500-hPa heights, 850-hPa *U* wind, 850-hPa *V* wind, and mean sea level pressure). This provides an assessment of how representative the clusters are of individual days. Figure 2 presents the results as mirrored histograms, with the values between individual days and the pattern of the cluster to which they are assigned shown in blue (values

increasing upward from the zero line), while the values between individual days and the pattern of the cluster to which they are *not* assigned are shown in gray (values increasing downward from the zero line). As every day is assigned to a cluster and there are only seven clusters, we expect there may be considerable differences between the spatial pattern of the four fields for an individual day and the spatial patterns of the cluster it is assigned to. This calculation provides a quantitative evaluation of the relationships and also compares them to the out-of-cluster relationships for context. In each of the four fields, a wide range in spatial correlations and RMSE between individual days and their associated cluster is observed, as expected, including some negative correlations. Approximately 60% of days have correlation coefficients greater than 0.5 to their assigned clusters, showing that many of the days are well represented by the clusters. A small number of noncluster days have high correlations to individual fields but the peaks of noncluster days on the histograms are generally located between −0.25 and 0. Similarly, the RMSE showed the lowest values for assigned-cluster days and the highest values for noncluster days. Overall, correlations are highest and RMSE is lowest for the 500-hPa height field (H500), which is the field with the least spatial variability, and correlations are lowest and RMSE is highest for the two horizontal components of the 850-hPa wind field (U850 and V850), which have more spatial structure. Although not pursued further here, this approach could be generally applied to any cluster analysis to provide a quantitative measure of goodness of fit.

c. Monte Carlo analysis

To estimate significance, that is, how likely a given WT is likely to occur under a given condition (a given month, a given teleconnection phase, etc.), a 95% confidence interval based on the Monte Carlo random resampling method is used to compare the results with the null hypothesis that the WTs are occurring randomly. 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 the observed monthly and yearly frequencies) and the calculation is repeated 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 considered to be significant compared to random chance.

d. Markov chain analysis

Markov chain analysis, as in Vautard et al. (1990), is used to examine WT progressions, the likelihood that the WTs tend to occur in specific sequences of order. Vautard et al. (1990) found that it is necessary to use Markov chain analysis when clusters (here, the WTs) 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 (each cluster is equally likely to occur the next day) 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, 5, and 7 have <600 days in their clusters, so an equiprobability model would be biased toward WTs 1 and 6.

To perform this analysis, first the transition matrix T_{ij} 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, B_{ij}, and is compared to T_{ij}. The number of C_{ij} times where $T_{ij} \geq B_{ij}$ and the number of D_{ij} times where $T_{ij} \leq B_{ij}$ are recorded, with C_{ij} representing the likely to occur transitions and D_{ij} representing the unlikely to occur transitions. At the end, values of C_{ij} and $D_{ij} \leq 500$ are considered significant at the 95% level. Vautard et al. (1990) note 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

a. Weather types

The circulation fields of the WTs at each of the three levels on which the clustering was performed, along with associated precipitation and 2-m temperature anomalies, are shown in Figs. 3 and 4. In terms of circulation fields, WT1 (Shallow Northeast Trough; 688 days) exhibits weak troughing with a surface low pressure system present over New Brunswick with low-level winds out of the northwest. New England and the East Coast exhibit warmer-than-average surface temperatures and higher-than-average precipitation, while the Midwest is cooler than average with lower-than-average precipitation. In WT2 (Mid-Atlantic Trough; 412 days), there is weak ridging over the western part of the region with a maritime 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. WT3 (Midwestern Trough; 392 days) has an area of surface low pressure over northern New England accompanied by a trough over the Great Lakes region. Low-level winds over the region follow a counterclockwise flow pattern centered in the middle of the 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 having colder-than-average temperatures and below-average precipitation. WT4 (Deep Northeast Trough; 452 days) presents a trough over the Northeast with surface low pressure and above-average precipitation over New Brunswick. Low-level winds are primarily out of the northwest, bringing below-average surface temperatures to the region. WT5 (Atlantic Ridge; 448 days) features a ridge over the Atlantic Ocean and a trough over the Midwest, with higher-than-average precipitation downstream of the trough. There is an area of surface high pressure over the Atlantic with winds flowing out of the

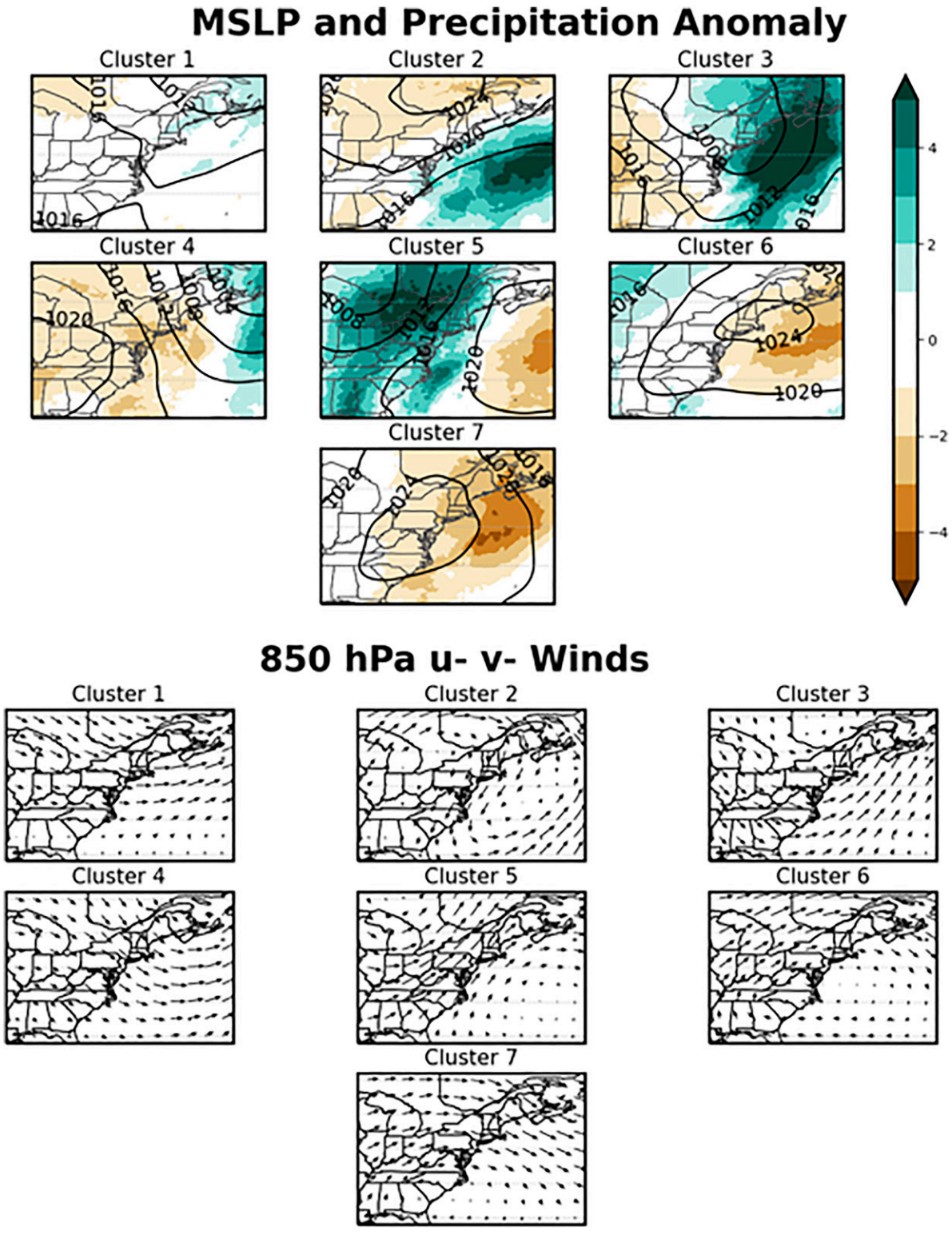

FIG. 3. ERA5 data from 1979 to 2018 plotted for (top) mean sea level pressure (contour) and daily mean removed average precipitation anomalies (mm day^{-1}; shaded) and (bottom) 850-hPa u and v wind components.

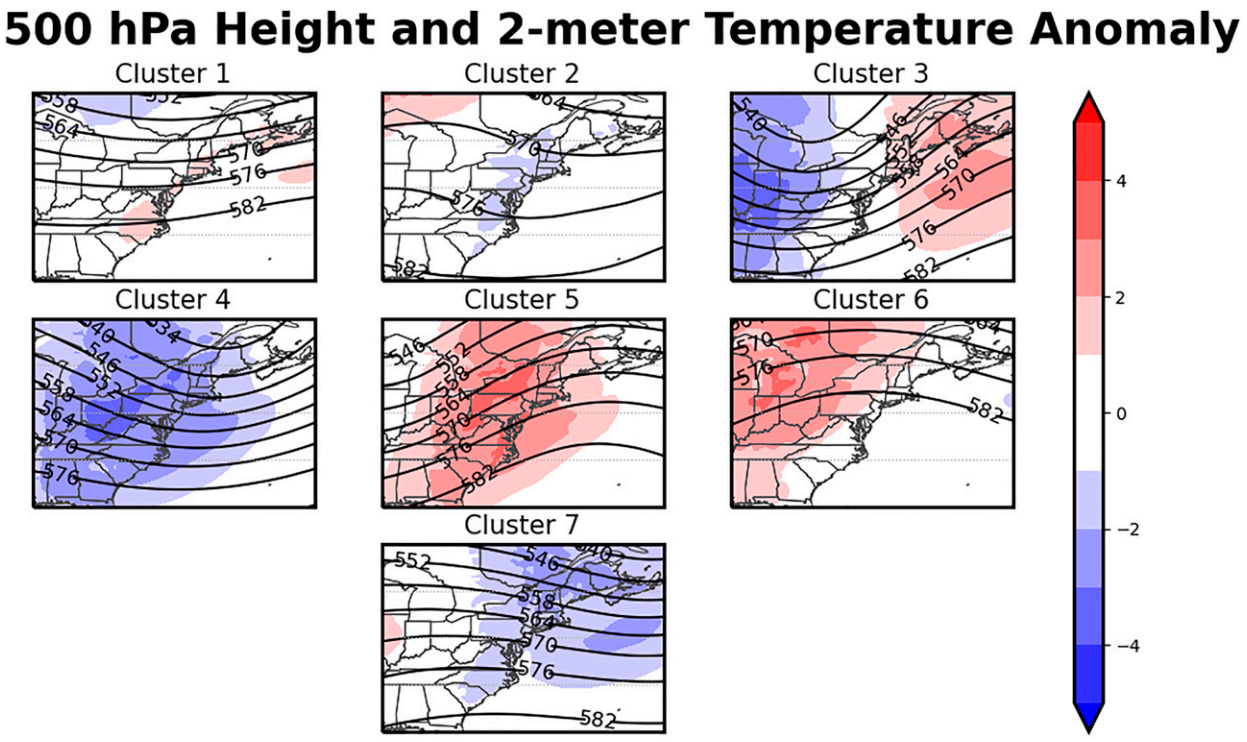

FIG. 4. Autumn season weather types: midlevel atmosphere and temperature. 500-hPa height (contour) and daily mean removed average 2-m anomalies (°C; shaded) ERA5 data from 1979 to 2018 plotted for each WT.

southwest, leading to above-average surface temperatures over the region. WT6 (Northeast Ridge; 700 days) shows an area of surface high pressure along with a ridge over the Northeast, leading to below-average precipitation over New England and above-average temperatures over the whole of the region. Finally, WT7 (Midwestern Ridge; 548 days in cluster) features slight ridging over the Midwest and slight troughing over the Atlantic Ocean with below-average precipitation across the whole 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.

b. *WT precipitation and temperature*

Precipitation and temperature are considered both in terms of average daily anomalies, shown in Figs. 3 and 4, and extremes, shown in Fig. 5. Extreme precipitation is defined as any day with station precipitation greater than the 99th percentile for that day, following Agel et al. (2015), and extreme temperature is analyzed in terms of heat-wave days, which are defined as any day with a maximum station temperature greater than the 95th percentile for that day, following Agel et al. (2021).

In terms of average precipitation anomalies (Fig. 3a), WTs 3 and 5 (Midwestern Trough and Atlantic Ridge) have the highest precipitation rates (6.1 and 5.4 mm day^{-1}, respectively), while WTs 4 and 7 (Deep Northeast Trough and Midwestern Ridge) have the lowest precipitation rates (2.7 and 2.5 mm day^{-1}, respectively). WTs 1 and 3 (Shallow Northeast Trough and Midwestern Trough) have the largest positive precipitation anomalies over New England and the Gulf of Maine, while WTs 2 and 4 (Mid-Atlantic and Deep Northeast Trough) have the largest positive precipitation anomalies over the Atlantic Ocean. WT5 has positive precipitation anomalies across the western two-thirds of the domain due to the strength of the trough over the Midwest, while WT7 has predominately negative precipitation anomalies over the entirety of the region.

For extreme precipitation days, the Midwestern Trough and Atlantic Ridge patterns (WTs 3 and 5) contain the most (26% and 19% of days, respectively), and extreme precipitation days are more likely to occur during these two patterns (Fig. 5a). During these two patterns, extreme precipitation is most likely in inland stations, across New York State in WT5 and across New York State and northern New England in WT3. The Deep Northeast Trough, the Northeast Ridge, and the Midwestern Ridge patterns (WTs 4, 6, and 7) have the fewest extreme precipitation days (7%, 8%, and 1%, respectively) and extreme precipitation days are unlikely to occur during these patterns. Similarly, extreme precipitation occurs mainly at inland stations in these WTs, in New York State during WTs 6 and 7, influenced by precipitation from the great lakes due to the positioning of the ridge, and in northern New England during WT4. WT2 is the only pattern in which extreme precipitation days are more likely in coastal stations, due to the location of the Mid-Atlantic Trough.

In terms of average temperature anomalies (Fig. 4), positive and negative temperature anomalies in the region are collocated

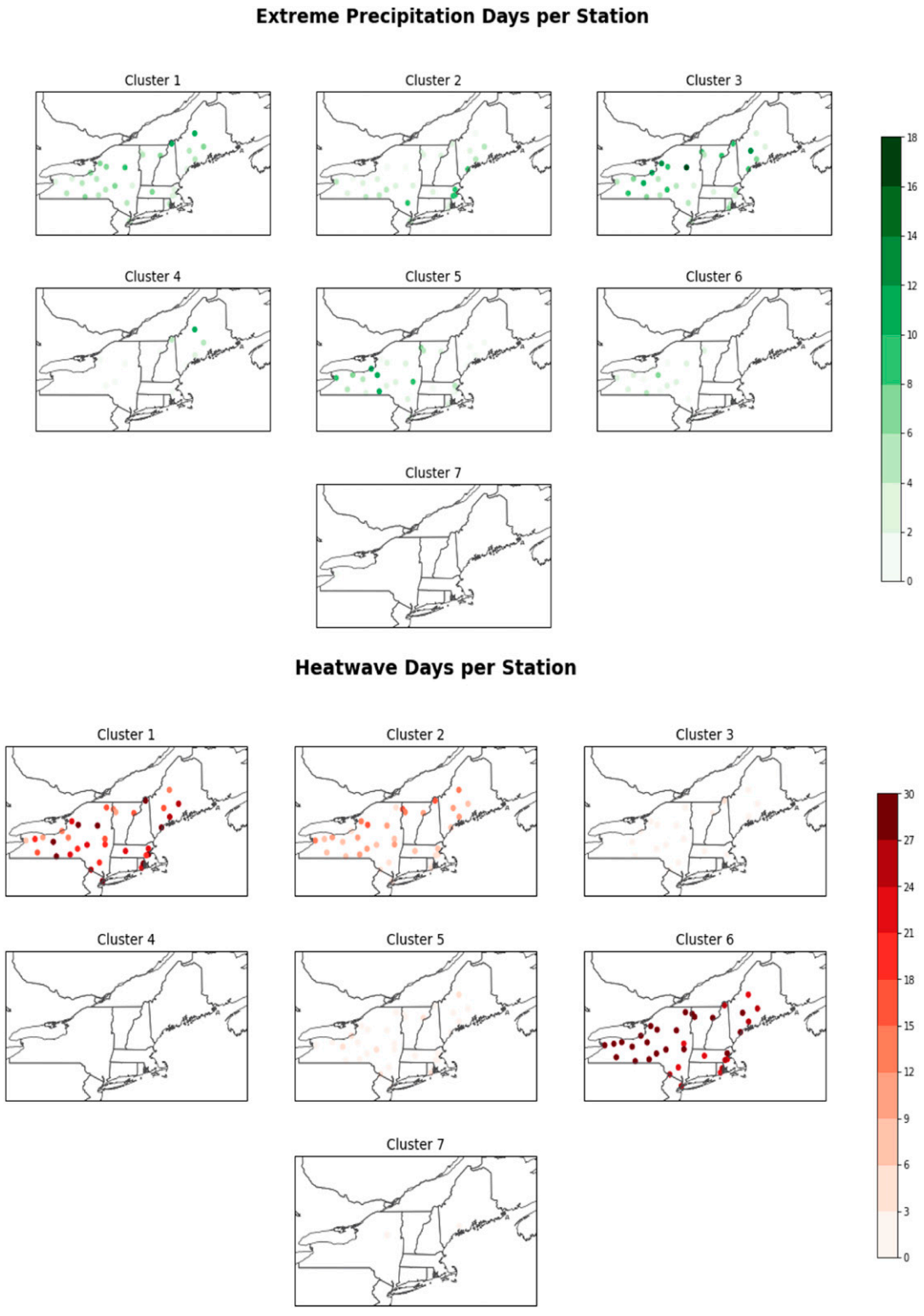

FIG. 5. Extreme precipitation and heat-wave days by WT analysis of (top) extreme precipitation days and (bottom) heat-wave days using U.S. Historical Climatology Network (USHCN) stations across New England and New York that had no more than 1% of their daily data missing. Extreme precipitation days is defined as any day with precipitation > 99th percentile of daily precipitation at a station. A heat-wave day is defined as any day with Tmax > 95th percentile at a station.

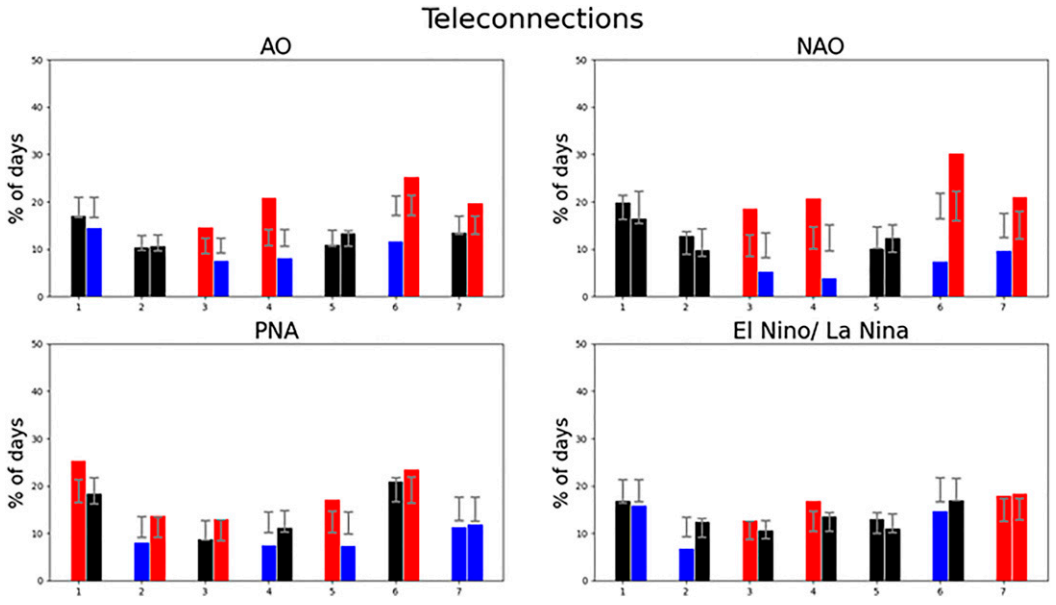

FIG. 6. Autumn season WT frequency during teleconnections. Teleconnection indices (±, neutral phases did not show any significance) in relation to each WT. In every panel, there are two columns for each WT: frequency of occurrence during the negative phase of the teleconnection (left column) and frequency during the positive phase (right column). Blue shading indicates a WT that is less likely to occur, and red shading indicates a WT that is more likely to occur based on the 95% Monte Carlo confidence level (gray lines).

to the positioning of the troughs and ridges. Parts of the region downstream from a trough experience warmer-than-average temperatures and colder-than-average temperatures are experienced upstream from a trough, independent of calendar date in the season. The mean temperatures for each WT are the warmest (17.8° and 18.2°C, WTs 1 and 6, respectively) for WTs that occur early in the season and the coldest (9.2° and 10.0°C, WTs 4 and 7, respectively) for WTs that occur late in the season.

Heat-wave days occur mostly in WTs 1, 2, and 6 (making up 98% of total heat-wave days), with WTs 1 and 6 being the most likely to experience heat-wave days (Fig. 5b). During a WT1 or WT6 day, heat-wave days occur across the majority of the region, but with WT1 having the lowest heat-wave day stations in western and northern portions of the domain, which are areas downstream of the upper-level trough. During WT2, heat-wave days are more prevalent in central and northern parts of the region, aligning with an area upstream of the upper-level trough. Heat-wave days during WTs 3, 5, and 7 are rare. It is interesting that the WT perspective is able to capture a traditional heat-wave pattern for the Northeast (WT6 with a strong ridge) without considering temperatures when clustering. This provides additional confirmation of the regional relevance of the circulation-based WT approach.

c. *Teleconnection relationship to WT frequency*

The frequency of each WT with respect to the daily phases of the AO, NAO, and PNA and the monthly phases of ENSO is shown in Fig. 6, with significance determined through Monte Carlo analysis. For all teleconnections during their neutral phase, there is no significance at the 95% level for frequency change in any of the WTs, and so the neutral phase results are not shown.

The AO and NAO have similar relationships to WT frequency, as expected due to the close relationship between the two indices (Deser 2000). WTs 3 and 4 are much more likely to occur during the negative phase of either (22% and 21% of WT days) and much less likely to occur during the positive phase (4% and 3% of WT days) while the opposite holds true for WTs 6 and 7 (5% and 8% of WT days during a negative phase and 15% and 13% of days during a positive phase). WTs 3 and 4 are classic negative AO and NAO circulation patterns, with troughing and lower-than-average heights over the eastern United States, accompanied by below-average temperatures and above-average precipitation. WTs 6 and 7 both feature upper-level ridging over the Midwest and eastern United States accompanied by below-average precipitation and above-average temperatures, signatures of positive AO and NAO phases. While the results are similar between the AO and NAO, the largest differences are obtained with the NAO index, consistent with its regionally based definition. These results are consistent with those from Sheridan (2003), where polar airmass WTs (our WTs 3 and 4, which feature strong flow from the northwest and below-average temperatures) have an inverse relationship with tropical and moderate airmass WTs (our WTs 6 and 7) based on the NAO phase (with polar airmass WTs unlikely to occur during +NAO and likely to occur during −NAO).

The PNA has a weaker association than the AO and NAO, although it is still substantial. Statistically significant changes are observed for 6 of the WTs and large differences are

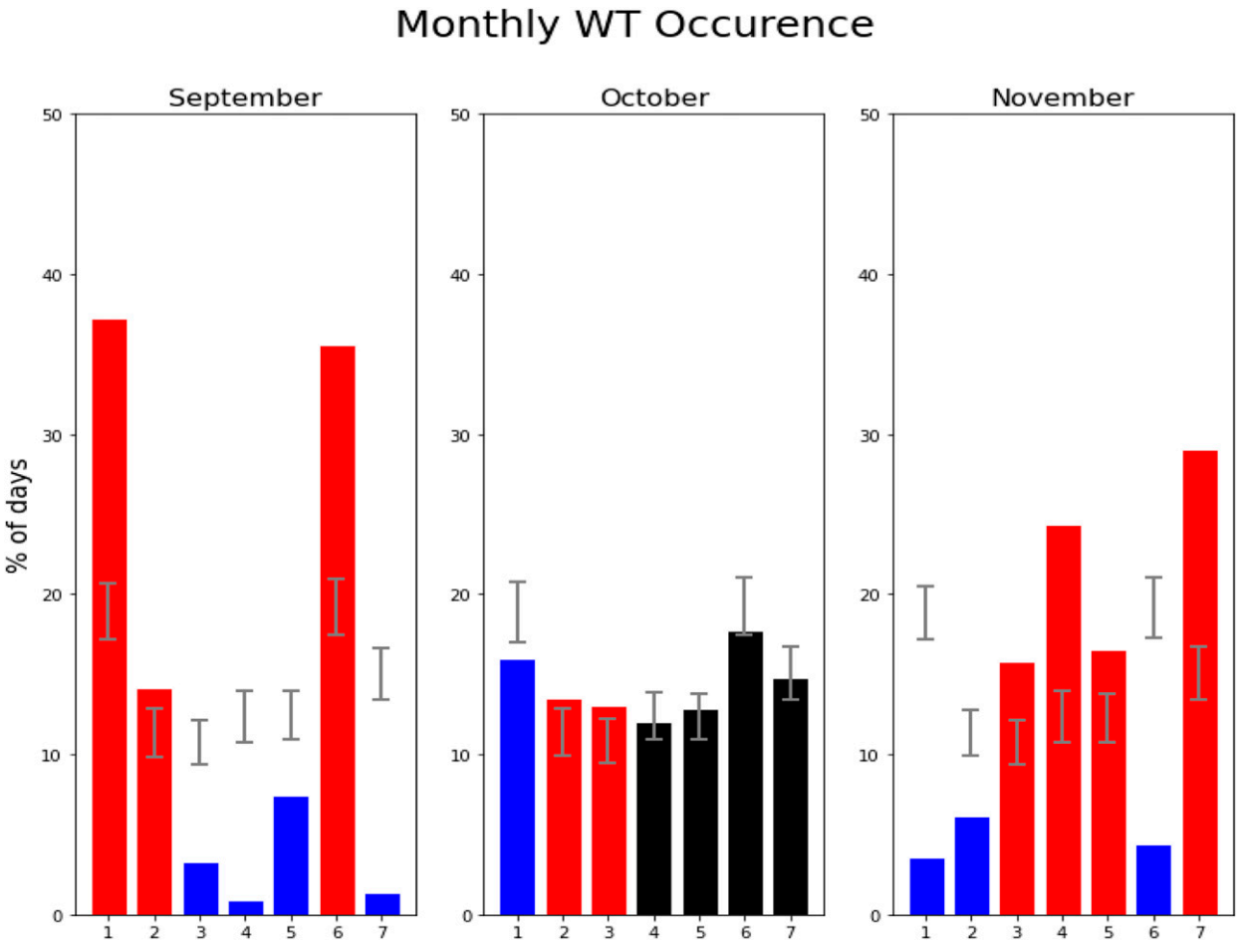

FIG. 7. Autumn WT monthly frequency. The likelihood of occurrence during a month for each WT is displayed based on the number of days each WT occurs per month. A 95% confidence interval (gray lines) is found using the Monte Carlo method. Blue bars indicate WTs that are less likely to occur than expected due to chance during a given month, and red bars indicate WTs that are more likely to occur during a given month.

observed for WT5, whose frequency doubles during the negative phase of the PNA relative to the positive phase. While many studies have looked at the PNA over the whole of North America (Sheridan 2003; Coleman and Rogers 2007), there are only a handful that have looked at its relationship to weather in the northeast United States. Notaro et al. (2006), being one of the first and major studies on PNA associations with wintertime weather in the northeast United States, found that a positive PNA coincided with colder and drier than average conditions over the region. That association is also found in our autumn WTs here, with WTs 2 and 6 (both dry and cooler than average) being likely to occur during a +PNA phase, although they are WTs that most commonly occur in the early season rather than during the late season. Only two of the WTs that most commonly occur in the late fall season have a significant relationship to the PNA phase, with WT7 being unlikely to occur during both +PNA and −PNA phases and WT 5 being likely during a −PNA phase. WTs 2 and 5 best resemble the +PNA (WT2) and −PNA (WT5) patterns in the northeast United States as described in Notaro et al. (2006), with WT2 featuring predominately dry and cool conditions with the heaviest precipitation offshore and WT5 featuring warmer and wetter conditions most likely related to the increased frontal passages through New York noted by Notaro et al. (2006).

WT frequency is less associated with ENSO than the other teleconnections. Although five WTs have a statistically significant change in frequency related to at least one ENSO phase, only one, WT2, has a substantial change. WT2 is approximately half as likely to occur during a negative ENSO phase, La Niña, then during a positive ENSO phase, El Niño. Given that ENSO does not have a strong influence over the northeast United States (e.g., Trenberth et al. 1998), the modest association with WT frequency is unsurprising. It is interesting that two of the WTs (2 and 6) that typically occur in the early season are unlikely to occur during a negative ENSO phase and WT7 is likely to occur during both a positive and negative ENSO phase.

Although Northern Hemisphere teleconnections are generally strongest in winter, the autumn association considered here is quite large for individual WTs. In particular, the likelihood of WTs 3, 4, and 6 changes by more than a factor

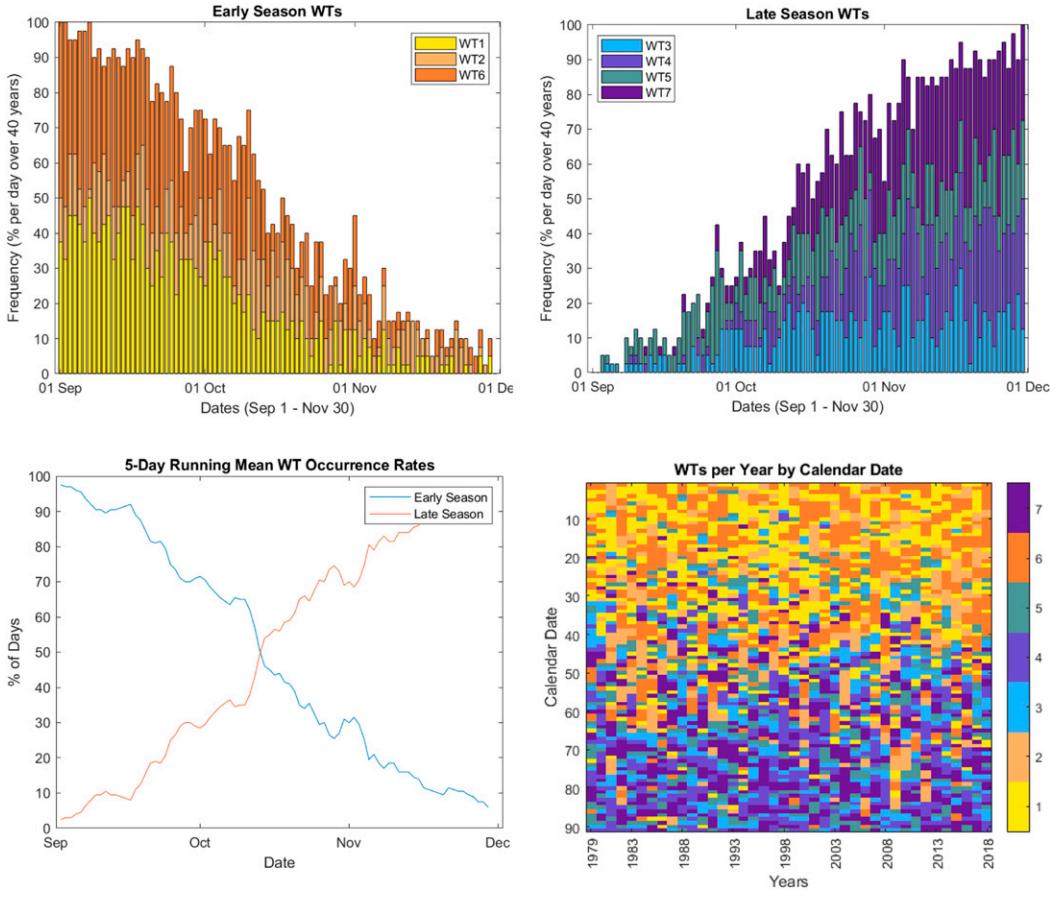

FIG. 8. Autumn season early and late season WT daily occurrence rate. The frequency of occurrence of (a) early season WTs and (b) late season WTs on each day of the season over all years. (c) As in (a) and (b), but smoothed using a 5-day running mean. Note the point of intersection at 15 Oct. (d) Klee diagram showing the WT occurrence per day over all 40 yr. The *y* axis is time, increasing down from the top (1 Sep) to bottom (30 Nov).

of 2 depending on the phase of the AO/NAO, and WT5 for the PNA. That such a strong association is found outside the peak winter season suggests that the WT perspective is providing additional information complementary to traditional correlation and composite analysis, which recover a more modest signal.

d. Progressions of early and late season WTs

The difference in the occurrence of the WTs over the course of the autumn season is examined in terms of monthly frequency, average daily evolution, and typical progressions between WTs. Early and late season WTs are determined based on changes in the frequency of occurrence over the 3 months of the season for individual WTs and for the characteristic progressions between WTs.

To measure monthly variability of WT frequency, the sum of WT frequency per month over all years is calculated, as shown in Fig. 7. Based on this, three groups are identified: 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 have 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

TABLE 1. Top seven results for occurrences of progressions of length 2, 3, and 4 of WTs in the overall dataset.

	Sequence	Count		Sequence	Count		Sequence	Count
Length 2 sequences	(6, 1)	205	Length 3 sequences	(1, 2, 6)	57	Length 4 sequences	(5, 3, 4, 7)	40
	(7, 4)	127		(3, 4, 7)	49		(6, 1, 6, 1)	27
	(3, 4)	114		(1, 6, 1)	41		(2, 6, 1, 6)	14
	(5, 1)	90		(5, 3, 4)	38		(6, 1, 2, 6)	14
	(2, 6)	87		(1, 6, 5)	37		(1, 6, 1, 5)	13
	(1, 2)	82		(5, 4, 7)	32		(1, 6, 1,2)	12
	(7, 5)	82		(5, 3, 7)	30		(5, 1, 2, 6)	10

TABLE 2. Markov chain analysis [as performed in Vautard et al. (1990)] indicating WT progression. The top portion of the table indicates WTs that are likely to transition between each other, with values below 500 being significant at the 95% level (bold and italicized). The bottom portion indicates WTs that are unlikely to transition between each other, with values below 500 being significant at the 95% level (bold and italicized).

		1	2	3	4	5	6	7
				Likelihood of transition				
1 Sep–15 Oct	1	***0***	***0***	9984	10 000	7206	***0***	9994
	2	9059	***0***	10 000	10 000	9991	***1***	10 000
	3	8328	10 000	***1***	8098	10 000	10 000	10 000
	4	10 000	10 000	10 000	9750	10 000	10 000	10 000
	5	***0***	10 000	***31***	10 000	623	10 000	10 000
	6	***0***	4825	10 000	10 000	***475***	***0***	10 000
	7	10 000	10 000	10 000	10 000	10 000	10 000	10 000
16 Oct–30 Nov	1	10 000	10 000	10 000	10 000	10 000	10 000	9993
	2	10 000	***0***	9953	9967	9721	10 000	10 000
	3	10 000	9997	***0***	***0***	10000	10 000	5493
	4	10 000	10 000	2294	***0***	4847	10 000	***0***
	5	8574	10 000	***0***	7291	***0***	10 000	9036
	6	10 000	10 000	10 000	10 000	8024	9343	10 000
	7	10 000	9997	2541	4321	***0***	7608	***0***
				Likelihood of no transition				
1 Sep–15 Oct	1	10 000	10 000	***27***	***0***	3433	10 000	***17***
	2	1351	10 000	***0***	***0***	***26***	10 000	***0***
	3	2170	***0***	10 000	2576	***0***	***0***	***0***
	4	***0***	***0***	***0***	*410*	***0***	***0***	***0***
	5	10 000	***0***	9985	***0***	9571	***0***	***0***
	6	10 000	5912	***0***	***0***	9649	10 000	***0***
	7	***0***	***0***	***0***	***0***	***0***	***1***	***0***
16 Oct–30 Nov	1	***0***	***1***	***0***	***0***	**0**	***0***	***9***
	2	***0***	10 000	***106***	***61***	***485***	***0***	***0***
	3	***0***	***5***	10 000	10 000	***0***	***0***	5309
	4	***0***	***0***	8319	10 000	5908	***0***	10 000
	5	1866	***0***	10 000	3501	10 000	***0***	1348
	6	***0***	***0***	***0***	***0***	2504	842	***0***
	7	***0***	***9***	8080	6399	10 000	*2883*	10 000

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.

To provide a more detailed look at the within-season evolution of the WTs, the average daily evolution throughout the season is shown in Fig. 8. A 5-day moving mean is applied to show an approximate intersection point between the two data series and is considered the midpoint of the autumn season (Fig. 8c); 16 October 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 16 October 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 in September and >50% of days experiencing a late season WT after 16 October.

To further examine the within-season evolution of the WTs, the monthly occurrence rates (Fig. 7) are combined with an analysis of WT progression (Table 2 and Fig. 8). First, the dataset of cluster numbers is split into a 91 × 40 matrix and then persistence is removed to be able to count patterns containing multiple consecutive cluster numbers (e.g., ''161'' is the same as ''1166611''). The number of progressions of lengths 2, 3, and 4 that occur each year are counted, and the yearly totals summed together to get an overall count (Table 1; note that only the top seven results for each progression are shown due to length, and progressions greater than length 4 are not shown as the highest count for those progressions is less than 10 occurrences over the 40-yr 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 is performed on the whole of the dataset; however, the progressions ''WT1–WT6'' and ''WT2–WT6'' do not appear as significant, even though they occurred frequently (Table 1). Analyzing the progressions on a monthly basis is also attempted. This produces similar results to what actually occurred (Table 1); however, the month of October is indiscernible as there are too many patterns that occur at various significance levels (ranging from 50% to 85%). Since part of the intention of this study is to understand the transition from the early autumn season to late autumn season in terms of circulation patterns, a monthly analysis is unable to fully capture and comprehend this. To try and provide more detail within the season, the significance analysis is applied separately to the two halves of the season (1 September–15 October and 16 October–30 November); similar progressions as in Table 1 are found that are significant at the 95% level (results of the Markov chain analysis are found in Table 2). Since 16 October was previously established as the midpoint of the autumn season (where early and late season WTs both occur 50% of the time, Fig. 7c), splitting the season into two halves is justifiable.

Based on the Markov chain analysis, four frequently occurring progressions are found in the first half of the season and one in the second half of the season (Fig. 9). In the early half of the season, there is one longer progression and three shorter progressions (consisting of different combinations of types

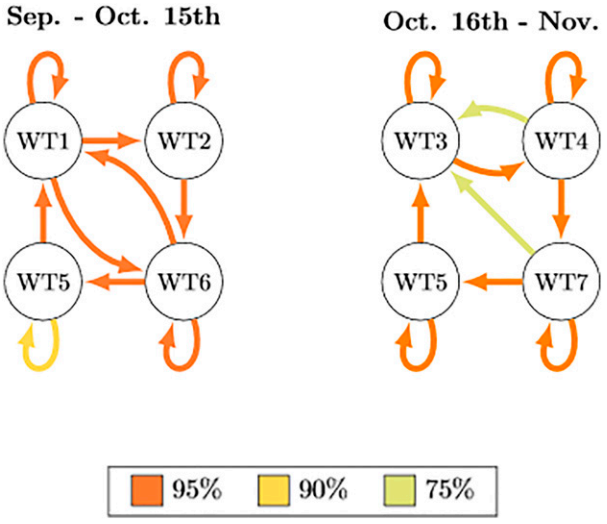

FIG. 9. Autumn season WT progressions. WT progression for the first and second halves of the season based on Markov chain analysis. Arrows indicate progressions that are significant, colored based on the percent level. Note that there are no progressions that were significant at the 85% or 80% levels, and only progressions at or above the 75% level were considered significant.

from the longer progression). The early season progressions are WT1–WT2–WT6–WT5 (herein, E1), WT1–WT6–WT1 (E2), WT1–WT2–WT6 (E3), and WT1–WT6–WT5 (E4). The one late season progression is WT3–WT4–WT7–WT5 (L1). There are subsets of L1 that also occur frequently (Table 1); however, none of them were significant at the 95% level. Overall, progressions E2–E4 occurred at least once during each calendar year of the study, with the fewest in 1981 and 2000 (two occurrences of progressions) and the most in 1983 (seven occurrences of progressions) per year. Progression E1 occurred 24 times over the whole study, not occurring at all in multiple years, while progression L1 occurred 69 times, not occurring at all in only 6 years in the study (Fig. 10). Subprogressions of L1 that are not significant at the 95% level did occur in years where L1 did not but are not shown.

e. *Physical characteristics of early and late season progressions*

The physical characteristics of the progressions identified in the previous section are considered in terms of the evolution of the circulation and sensible weather fields. The WTs shown in the progressions (Figs. 11–15) are composited based only days occurring during those progressions but are quite similar to the original WTs (Figs. 3 and 4), further demonstrating the stability and representativeness of the WTs. As the weather types are fixed in space, the progressions between weather types reflect the movement of synoptic systems across the region, with greater persistence (a weather type followed by the same weather type) corresponding to slower movement. In general, the progressions all involved a propagating trough–ridge or ridge–trough circulation flow, varying in both strength and location.

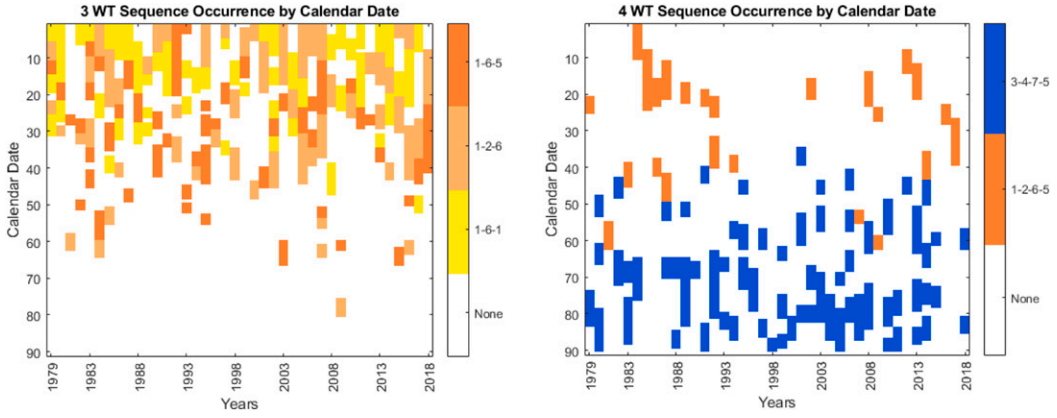

FIG. 10. Autumn season early and late season WT daily occurrence rate. Klee diagram indicating the seasonal timing of (a) the three-WT progressions (all early season) and (b) the four-WT progressions (1–2–6–5 occurs in early season and 3–4–7–5 occurs in late season).

E1 features a Mid-Atlantic Trough–Northeast Ridge–Midwestern Trough progression, as a trough moves off the eastern side of the domain and another trough moves in from the west (Fig. 11). Most of the Northeast initially experiences above-average precipitation and temperatures, with temperatures and precipitation becoming below average as the trough moves farther east offshore, being replaced by a strong ridge. This strong ridge brings surface high pressure, with the eastern half of the region continuing to have below-average temperatures and precipitation, while the western half has both above-average temperatures and precipitation. This is the leading edge of a Midwestern Trough, which will bring above-average precipitation and temperatures to the whole of the region.

Three subsets of the E1 progression can occur, which are variations on the location and strength of the troughs (Fig. 12). E2 starts out similarly to E1, with above-average precipitation and temperatures occurring across the Northeast. Instead of the trough becoming a Mid-Atlantic Trough, however, it gets

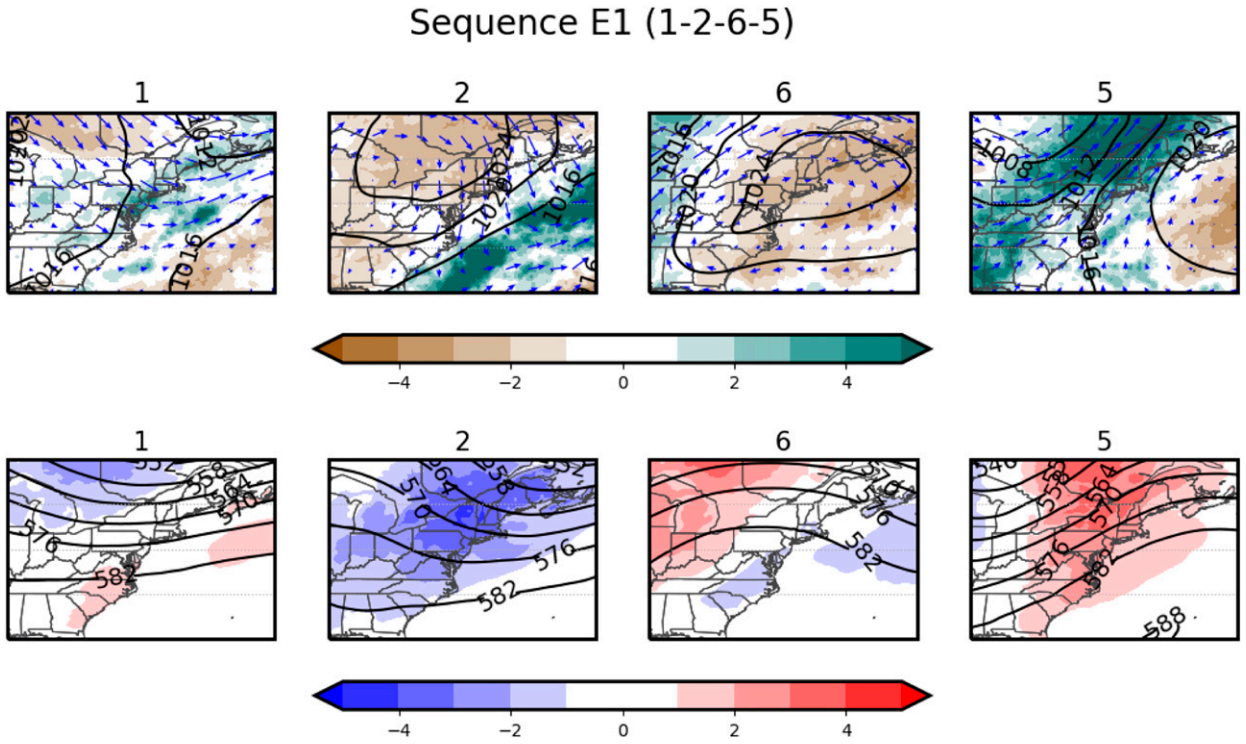

FIG. 11. Autumn early season progression (E1). (top) MSLP (black), 850-hPa wind, and daily precipitation anomaly (mm day^{-1}; shaded); (bottom) 500-hPa height (black) and 2-m temperature anomaly (°C; shaded) for WTs in the average 1–2–6–5 progression.

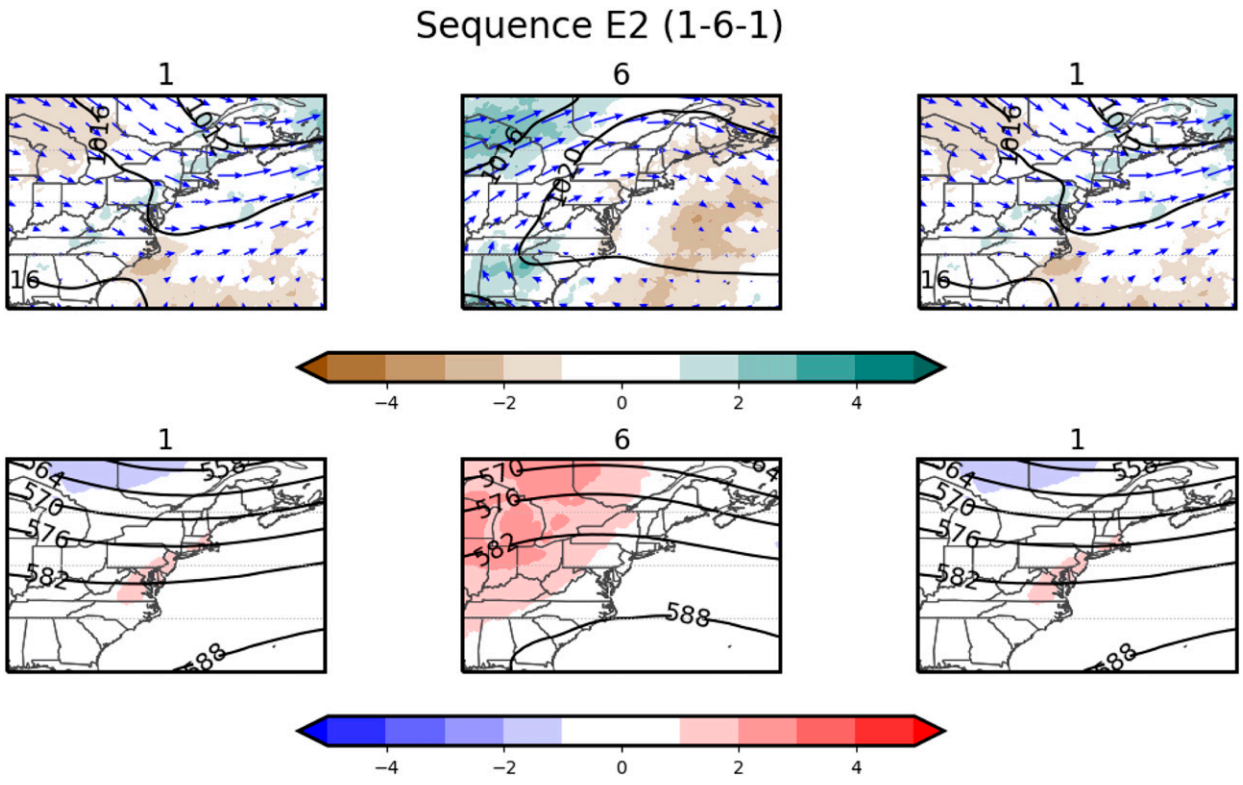

FIG. 12. Autumn early season progression (E2). (top) MSLP (black), 850-hPa wind, and daily precipitation anomaly (mm day^{-1}; shaded); (bottom) 500-hPa height (black) and 2-m temperature anomaly (°C; shaded) for WTs in the average 1–6–1 progression.

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 less positive, albeit occurring in similar locations. This ridge is eventually replaced by a shallow Northeast Trough. E2 is the driest and coolest progression, with the lowest precipitation and temperature anomalies.

E3 begins similarly to E1; however, the Mid-Atlantic Trough that develops is weaker and located farther north than in E1 (Fig. 13). This leads to areas of negative precipitation and temperature anomalies displaced northward compared to E1. The ridge that develops is of similar strength but located farther north than in E1, leading to above-average temperature anomalies occurring over most of the region. Precipitation anomalies are in similar locations and of similar strength to those in E1. In this case, the Midwestern Trough is weak following the development of the ridge. E3 is the coolest progression, with most of the region experiencing below-average temperatures.

E4 begins as a trough that is accompanied by positive temperature anomalies across most of the domain and positive precipitation anomalies across New England and the Midwest (Fig. 14). 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 in E1. A Midwestern Trough then moves into the region, bringing positive precipitation anomalies everywhere except over the Atlantic and positive temperature anomalies everywhere except for the Midwest. This is the wettest and warmest pattern, with the whole region experiencing above-average temperatures for the entire progression, especially on the Eastern Seaboard where the highest anomalies occur.

L1 features a Hudson Bay trough that crosses the northern half of the domain (Fig. 15). 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 in the Midwest move into the region as a deep Midwestern Trough forms. This is similar to the typical winter WT progression identified in Roller et al. (2016), and so this pattern represents a shift to winterlike circulation patterns and weather.

f. 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

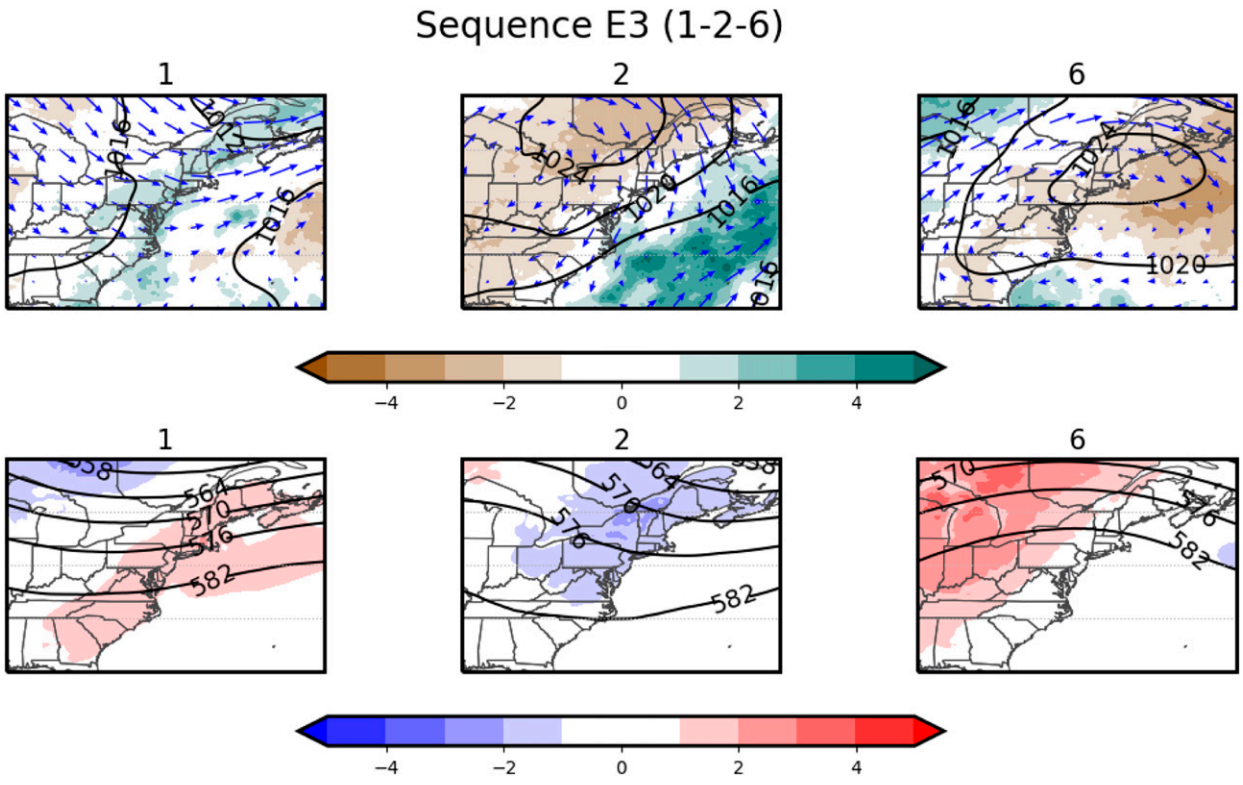

FIG. 13. Autumn early season progression (E3). (top) MSLP (black), 850-hPa wind, and daily precipitation anomaly (mm day^{-1}; shaded); (bottom) 500-hPa height (black) and 2-m temperature anomaly (°C; shaded) for WTs in the average 1–2–6 progression.

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 20 years of the record (1979–98) and the last 20 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. 16. Early season WTs show an increase at the end of the season, in Nov, as well as an overall increase and late season WTs show a decrease at the beginning of the season, in September and the end of the season, in November, as well as an overall decrease between the two periods, all being significant at the 95% level. That is, in addition to temperature shifts, daily circulation patterns are also shifting, with early season weather patterns continuing longer into the season and late season weather patterns occurring later.

4. Summary and discussion

Seven distinct daily 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 the occurrence of the WTs by calendar month showed a clear change in frequency over the course of the season with three WTs (WTs 1, 2, and 6) likely to occur in September but not in November and four WTs (WTs 3, 4, 5, and 7) likely to occur in November but not in September. The WTs showed preferred progressions between one another. Two primary WT progressions were identified via Markov chain analysis, one capturing the evolution of an early season weather system, which can occur in multiple variations, and one capturing the evolution of a late season weather system that is similar to the winter pattern identified in Roller et al. (2016). The early season weather is characterized by shorter sequences of WTs with weaker gradients and shallower troughs and ridges, while the late season is characterized mainly by a single longer-lasting, stronger system. The WTs were also used as a new perspective to examine the regional occurrence of extreme precipitation and temperature, the relationship to standard teleconnection indices, and changes in seasonality over the course of the period, which are discussed further below. The clear seasonality in daily weather types demonstrated here for autumn suggests that there may be additional value in considering seasonality not only in terms of average temperature (e.g., Trenberth 1983), sometimes referred to as ''climatological seasonality,'' but also in terms of daily weather patterns, which may be termed ''meteorological seasonality.''

In terms of circulation and sensible weather, WT1 exhibits a weak trough–ridge circulation and surface temperatures that are above average for most of the region, and New England and the eastern United States experience above-average precipitation. In WT2, there is a trough present with above-average precipitation anomalies on the Atlantic Coast. Temperatures are generally average to slightly below average. WT3 has an area of surface low pressure over northern New England

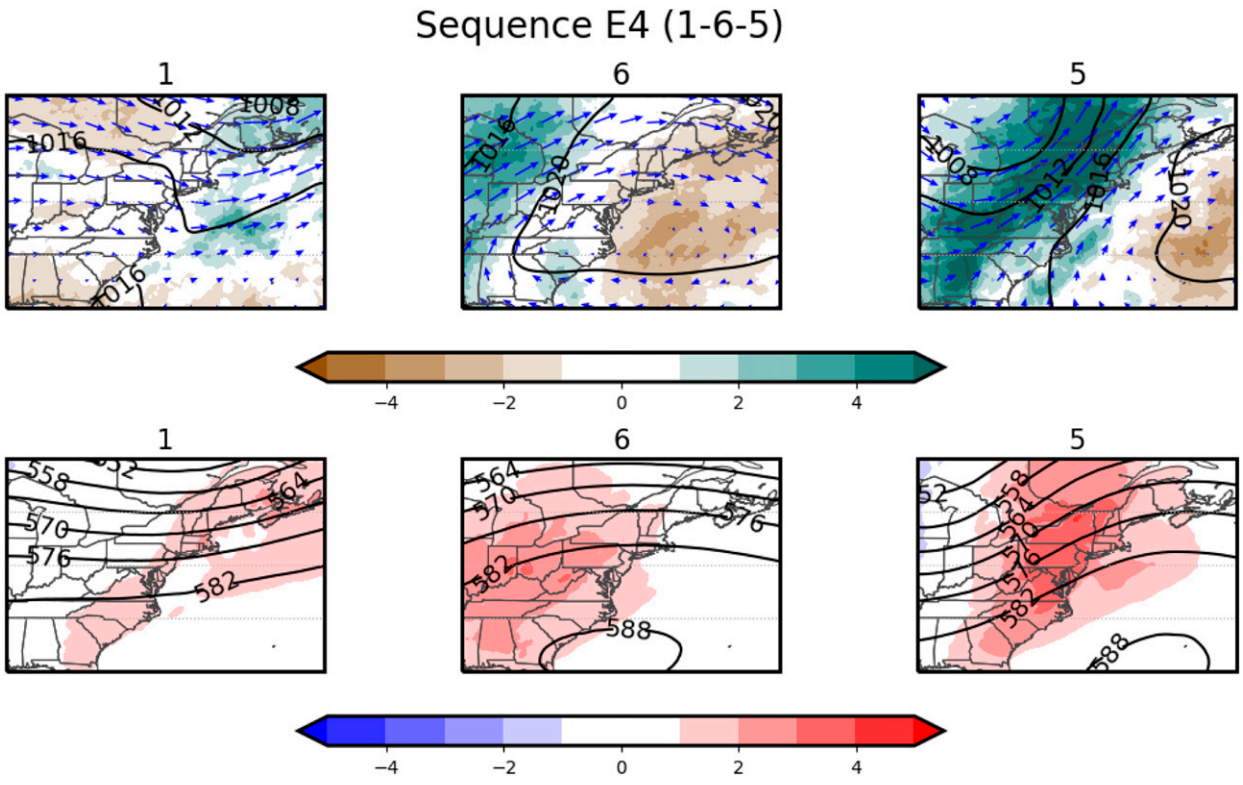

FIG. 14. Autumn early season progression (E4). (top) MSLP (black), 850-hPa wind, and daily precipitation anomaly (mm day^{-1}; shaded); (bottom) 500-hPa height (black) and 2-m temperature anomaly (°C; shaded) for WTs in the average 1–6–5 progression.

accompanied by a Midwestern Trough. 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 and Midwestern Trough, coinciding with above-average temperatures for the whole of the region and above-average precipitation downstream of the trough. 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 Midwestern Ridge with below-average precipitation over much of the region, and average to slightly below-average temperatures, and high pressure at the surface.

The occurrence of extreme precipitation and heat-wave days was analyzed relative to the WTs, with the highest amount of extreme precipitation days (with extreme precipitation days more likely to occur than due to random chance) during the Midwestern Trough and Atlantic Trough (WTs 3 and 5), although extreme precipitation occurred at least once during every WT. Heat-wave days were more divided, with WTs 1, 2, and 6 containing 98% of the days, and WTs 4 and 7 having no heat-wave days. Heat waves are most likely to occur in WTs 1 and 6, which is consistent with their greatest prevalence in the early season, and WTs 1, 2, and 6 are similar to three of the four characteristic heat-wave patterns identified for the region (Agel et al. 2021).

The relationship between the AO, NAO, PNA, and ENSO teleconnections and the frequency of occurrence of the WTs was examined. The NAO has the strongest relationship and ENSO has the weakest. The likelihood of WTs 3, 4, and 6 changes by more than a factor of 2 depending on the phase of the NAO, and WT5 changes by more than a factor of 2 based on PNA phase. While previous teleconnection analysis for the Northeast has primarily focused on the winter season (e.g., Notaro et al. 2006; Ning and Bradley 2014; Roller et al. 2016), the WT perspective here shows that there are some strong relationships in autumn as well, for individual weather types. Additionally, since the NAO influence does not occur uniformly across all days but is concentrated in a few WTs, standard correlation or composite analysis that includes all days or uses monthly averages will dilute the maximum strength of the relationship.

Markov chain analysis was used to analyze the progression between WTs, leading to the development of progressions defined by their seasonality (Figs. 7, 8, and 9 and Table 2). One ''early season'' progression features four 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, 2, and 6). This progression can be divided into three sub-progressions depending on the development of a maritime trough and the strength and positioning of the upper-level

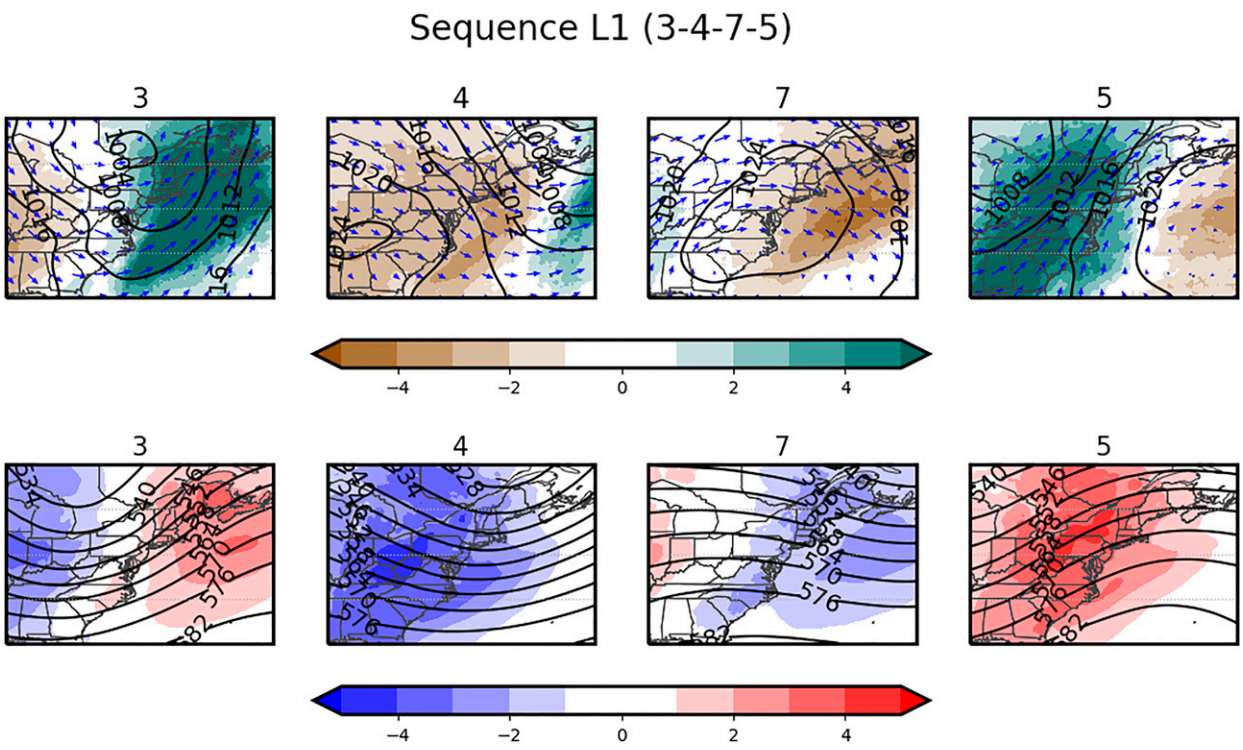

FIG. 15. Autumn late season progression (L1). (top) MSLP (black), 850-hPa wind, and daily precipitation anomaly (mm day^{-1}; shaded); (bottom) 500-hPa height (black) and 2-m temperature anomaly (°C; shaded) for WTs in the average 3–4–7–5 progression.

ridge. The ''late season'' progression features four 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. It is also important to note that each of the WTs are also likely to transition to themselves on the next day, that is, to persist for 2 days. This both validates the use of daily data as being generally sufficient to resolve the evolution of the WTs in most cases and highlights the synoptic time scales of the WTs.

The seasonality of the WTs is also relevant to the observed shifting of the seasons, which has primarily been previously investigated in terms of temperature. Comparison of WT occurrence for the most recent half of the period compared to the first half shows that the occurrence of early season WTs has

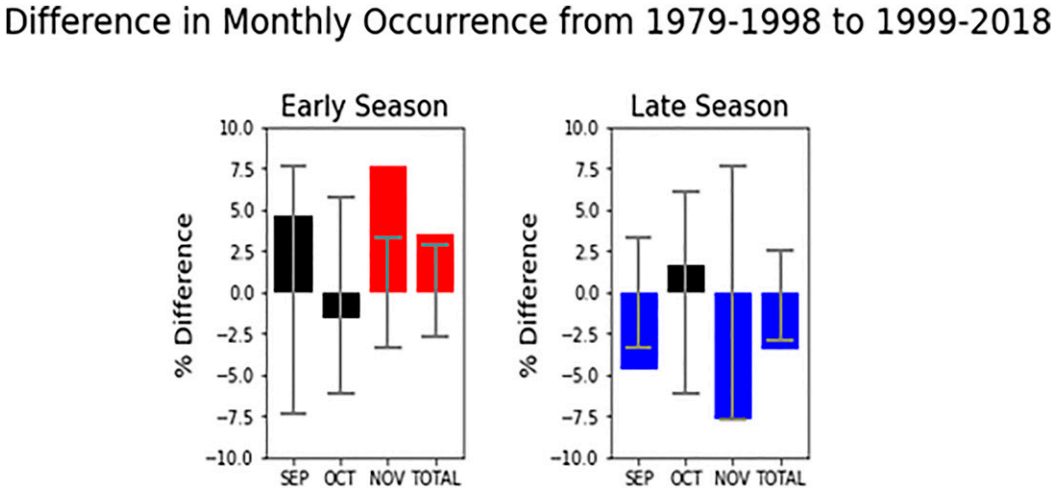

FIG. 16. Changes in monthly occurrence. Preliminary analysis between the first and second halves of the 1979–2018 period for differences in frequency of occurrence of early and late season WTs; blue shaded bars show a decrease and red shaded bars show an increase that is significant at the 95% level.

increased in November and late season WTs have decreased in September and November, both of which are significant at the 95% confidence level. That is, early season WTs are occurring later into the season and late season WTs are also delayed later into the season, demonstrating that the observed changes in regional seasonality in terms of temperature are also associated with clear changes in daily circulation, and offering a new perspective for considering seasonal shifts.

Several outstanding questions remain. In terms of trend behavior, it would be useful to also examine whether the structure of the WT patterns change over time in addition to the changes in frequency of occurrence considered here, and whether some WTs warm more than others. The extension of ERA5 back to 1950, at the preliminary stage at the time of this writing, will also allow for an examination of changes over a longer period. It would also be useful to examine how these changes relate to both climate change and decadal variability, through use of datasets that go back further in time and consideration of how the regional circulation patterns relate to large-scale circulation changes. We plan to address these questions for the autumn season in future work, as well as extend our analysis to the spring season, to consider the characteristic regional WTs in the context of the transition from the cold season to the warm season. We also plan to consider the evolution of WTs over the full year, to provide a full picture of how the annual cycle is reflected in changes to daily weather. Daily circulation patterns also provide a useful context for dynamical analysis (Agel et al. 2018), model assessment (Agel et al. 2020; Agel and Barlow 2020), and we also expect to utilize them for predictability studies. Finally, the approach here is easily extended to other regional domains and we expect the methodology to be widely applicable and useful.

Acknowledgments. This work was supported by NOAA NA20OAR4310424, NSF AGS-1623912, and NASA NESSF Grant NNX15AN91H S03, as well as by the use of EMCWF's ERA5 data set. We would also like to thank the editor and anonymous reviewers for their comments and suggestions that helped improve this manuscript.

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).

REFERENCES

Agel, L., and M. Barlow, 2020: How well do CMIP6 historical runs match observed northeast U.S. precipitation and extreme precipitation–related circulation? *J. Climate*, **33**, 9835–9848, https://doi.org/10.1175/JCLI-D-19-1025.1.

——, ——, J. Qian, F. Colby, E. Douglas, and T. Eichler, 2015: Climatology of daily precipitation and extreme precipitation events in the northeast United States. *J. Hydrometeor.*, **16**, 2537–2557, https://doi.org/10.1175/JHM-D-14-0147.1.

——, ——, 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 Dyn.*, **50**, 1819–1839, https://doi.org/10.1007/s00382-017-3724-8.

——, ——, J. Polonia, and D. Coe, 2020: Simulation of northeast U.S. extreme precipitation and its associated circulation by CMIP5 models. *J. Climate*, **33**, 9817–9834, https://doi.org/10.1175/JCLI-D-19-0757.1.

——, ——, C. Skinner, F. Colby, and J. Cohen, 2021: Four distinct northeast US heat wave circulation patterns and associated mechanisms, trends, and electric usage. *npj Climate Atmos. Sci.*, **4**, 31, https://doi.org/10.1038/s41612-021-00186-7.

Allen, M. J., and S. C. Sheridan, 2016: Evaluating changes in season length, onset, and end dates across the United States (1948–2012). *Int. J. Climatol.*, **36**, 1268–1277, 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, https://doi.org/10.1175/BAMS-D-13-00068.1.

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, https://doi.org/10.1175/JCLI4201.1.

Cooter, E., and S. LeDuc, 1995: Recent frost date trends in the northeastern United States. *Int. J. Climatol.*, **15**, 65–75, https://doi.org/10.1002/joc.3370150108.

Deser, C., 2000: On the teleconnectivity of the "Arctic Oscillation." *Geophys. Res. Lett.*, **27**, 779–782, https://doi.org/10.1029/1999GL010945.

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, https://doi.org/10.1175/1520-0477-83.9.1327.

El-Kadi, A. K. A., and P. A. Smithson, 1992: Atmospheric classifications and synoptic climatology. *Prog. Phys. Geogr.*, **16**, 432–455, https://doi.org/10.1177/030913339201600403.

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, https://doi.org/10.1175/1520-0469(1987)044<1266:DBTSAA>2.0.CO;2.

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** (Suppl. 1), 2493–2500, https://doi.org/10.1073/pnas.012580899.

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

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. N. Y. Acad. Sci.*, **1146**, 105–152, https://doi.org/10.1196/annals.1446.019.

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, https://doi.org/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, Shinfield Park, United Kingdom, ECMWF, 123–149.

Meseguer-Ruiz, O., N. Cortesi, J. A. Guijarro, and P. Sarricolea, 2020: Weather regimes linked to daily precipitation anomalies in northern Chile. *Atmos. Res.*, **236**, 104802, 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, https://doi.org/10.1175/1520-0469(1995)052<1237:WRRAQS>2.0.CO;2.

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

——, ——, and J.-H. Qian, 2010: Local versus regional-scale characteristics of monsoon onset and post-onset rainfall over Indonesia. *Climate Dyn.*, **34**, 281–299, https://doi.org/10.1007/s00382-009-0547-2.

——, ——, ——, and M. Ghil, 2015: Weather types across the Maritime Continent: From the diurnal cycle to interannual variations. *Front. Environ. Sci.*, **2**, 65, https://doi.org/10.3389/fenvs.2014.00065.

Muñoz, Á. G., X. Yang, G. A. Vecchi, A. W. Robertson, and 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. NOAA, accessed 10 September 2020, https://www.ncei.noaa.gov/data/precipitation-persiann/access/.

Ning, L., and R. S. Bradley, 2014: Winter precipitation variability and corresponding teleconnections over the northeastern United States. *J. Geophys. Res.*, **119**, 7931–7945, https://doi.org/10.1002/2014JD021591.

Notaro, M., W.-C. Wang, and W. Gong, 2006: Model and observational analysis of the northeast U.S. regional climate and its relationship to the PNA and NAO patterns during early winter. *Mon. Wea. Rev.*, **134**, 3479–3505, https://doi.org/10.1175/MWR3234.1.

Park, B.-J., Y.-H. Kim, S.-K. Min, and E.-P. Lim, 2018: Anthropogenic and natural contributions to the lengthening of the summer season in the Northern Hemisphere. *J. Climate*, **31**, 6803–6819, https://doi.org/10.1175/JCLI-D-17-0643.1.

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, https://doi.org/10.1175/JCLI-D-12-00178.1.

Ramos, A. M., D. Barriopedro, and E. Dutra, 2015: Circulation weather types as a tool in atmospheric, climate, and environmental research. *Front. Environ. Sci.*, **3**, 44, https://doi.org/10.3389/fenvs.2015.00044.

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, https://doi.org/10.1175/1520-0442(2002)015<1609:AIISAS>2.0.CO;2.

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, https://doi.org/10.1175/JCLI-D-15-0274.1.

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

——, 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, https://doi.org/10.1175/1520-0477(1983)064<1276:WATS>2.0.CO;2.

——, G. W. Branstator, D. Karoly, A. Kumar, N. C. Lau, and C. Ropelewski, 1998: Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. *J. Geophys. Res.*, **103**, 14 291–14 324, https://doi.org/10.1029/97JC01444.

Vautard, R., K. C. Mo, and M. Ghil, 1990: Statistical significance test for transition matrices of atmospheric Markov chains. *J. Atmos. Sci.*, **47**, 1926–1931, https://doi.org/10.1175/1520-0469(1990)047<1926:SSTFTM>2.0.CO;2.

Yarnal, B., 1993: *Synoptic Climatology in Environmental Analysis*. Belhaven Press, 195 pp.