**Cover Letter**

Here we submit the revised version of our manuscript “*Clustering Analysis of Autumn Weather Regimes in the Northeast U.S.”* by *Coe et al*. We thank the editor and reviewers for their careful attention to our work. We have clarified what we feel is a key new contribution of the analysis – analyzing seasonal shifts in terms of regional daily weather patterns – and also added a substantial amount of new work: the precipitation and temperature fields associated with the weather types are now also analyzed, including a consideration of extreme events; the relationship with standard teleconnections is now assessed, and a detailed examination of progressions between weather types through a Markov Chain analysis has been added. Finally, we have also extensively edited the text and figures for clarity. We look forward to the response to our revisions.

**Response to Reviewers**

The reviewer’s comments are given in blue and the responses are given in black.

**Response to Reviewer 1**

Thank you for the review, especially during the pandemic.

Reviewer #1: The manuscript presents an observational analysis of fall synoptic weather patterns in the Northeastern US. The authors use clustering analysis to identify seven different weather patterns with different trough and ridge axes across the region. The authors then identified favored pattern transitions and found two main sequences early- and late-season pattern transitions. Trend analysis shows that the frequency of early season pattern transition is occurring more frequently, and the late season pattern transition is occurring less frequently in the most recent two decades. This is consistent with an expanding summer season across the Northern Hemisphere.

This paper has had a difficult time finding a loving home so far and it finds its way to the latest review with lots of baggage. At least some of the criticism seems to be that it is not sufficiently novel from Roller et al. (2016) to justify publication. Not sure that I agree with that critique, Roller et al. analyzed the winter season and this manuscript provides analysis for the fall, to me this is sufficiently different to justify publication; even as unsatisfying as that may be. However, what does puzzle me is that Roller et al. provided more expansive analysis including precipitation and storm track density that this manuscript avoids. But maybe the more comprehensive analysis was included in an earlier version of the manuscript that has now been removed.

As far as we know, our analysis provides the first use of regional daily weather types to characterize shifts in seasonal timing (although Allen and Sheridan (2016) have used daily height clusters to analyze seasonal evolution at the continental scale). We felt that was an interesting perspective on an important topic and, given the prior history with the review process, thought it would provide a clearer presentation to just focus on that aspect.

However, we have now added several additional parts to the analysis: the precipitation and temperature fields associated with the weather types are now provided, including an analysis of extremes; the relationship with standard teleconnections is now assessed, and a detailed analysis of progressions between weather types through a Markov Chain analysis has been added.

This is not an exciting paper and I think many if not all the analysis is intuitive. The early season transition pattern seems consistent with a general summer circulation and the late season transition pattern consistent with a late fall circulation pattern of vigorous troughs moving across the Northeastern US from west to east. Also, the increasing trend of the summer-like pattern at the expense of the late fall transition pattern is consistent with our ideas of climate change and the existing scientific literature. Still intuition is not sufficient, and it is important and necessary to document trends and I thought using clustering analysis to identify pattern progressions that matches our own experience was of interest. And though the analysis is what I would consider modest in scope, it seemed well confined and straightforward.

While we agree that both physics and synoptic experience dictate a shift between warm season and cold season weather patterns during autumn outside of using surface temperatures and continental upper-level circulation patterns (Easterling 2002, Cooter and LeDuc 1995, and Allen and Sheridan 2016), we have not found any previous work that has managed to quantify this shift and so we suggest that the analysis may not be as obvious as it appears. Additionally, as noted above, we have added several new parts to the analysis to make the scope more substantial.

As I note below Figure 3 is too busy and I recommended that 500 hPa analysis be removed from the analysis of MSLP and 850 hPa. There is no analysis of temperature in the manuscript and I recommend that at a minimum the surface temperature anomalies of each weather type be included with the separated plots of 500 hPa. A temperature trend analysis would be nice as well, but I will take the authors at their word that this is planned future work.

Thank you for the recommendation. We have split the original Figure 3 into two figures. The first shows an analysis of MSLP, 850 hPa winds, and precipitation anomalies, while the second features 500 hPa heights and 2-meter temperature anomalies. We agree that a temperature trend analysis would be interesting but feel that is better suited for a separate paper due to length considerations.

Line 178/Figure 2 - I was confused what is being shown exactly in Figure 2 and I think more explanation is needed. I assumed that all dates feel within at least one of the seven clusters but are there many dates that are not included in any of the clusters? Shouldn't there be seven blue bars in all the panels? I count six in most.

This figure is no longer in the manuscript.

Figure 3 - Each panel is quite busy, and I am having a hard time making out the trough and ridge features. I suggest including SLP and 850hPa wind vectors in one plot and 500 hPa in a separate plot

Thank you for the suggestion. We have split this into two figures, one showing the MSLP, 850 hPa winds, and precipitation anomalies and the other showing 500 hPa heights and 2-meter temperatures anomalies.

Figure 4 - In order to see the contouring levels for MSLP, I really needed to blow up the figure but is the contouring interval irregular in the plot?

This figure is no longer in the manuscript. In other figures, we have adjusted the text size and contour intervals to make them easier to read.

Lines 212-214 - does this really makes sense that high surface pressure is due sometimes to higher heights in the mid-troposphere and sometimes due to higher heights in the lower-troposphere? Seems to me that higher surface pressure is always going to be associated with higher heights in the lower troposphere due to subsidence. The mid-tropospheric heights are then a function of the atmospheric column temperature and surface pressure.

These lines are no longer in the manuscript.

Figure 8 - I think a description of the percentages shown in the arrows could be provided at least in the caption, is it percentage of days when WT(X) occurs it will transition to WT(Y)?

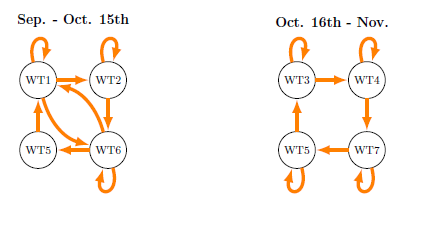


Figure R1: *WT progression for the first and second halves of the season based on Markov chain analysis. Arrows indicate progressions that are significant at the 95% level.*

This figure has been redone to be more easily read (copied above as Figure R1). Now, the arrows show transitions that are likely at the 95% level between WTs including persistence (curved arrows). The percentage of days when the transition between WTs occurs is no longer included since we are using the 95% level of likelihood of transition.

Figure 9 - how come the frequency of WT5 is not shown?

The frequency of WTs 2 and 5 have been added in to their respective Early and Late season appearances.

**Response to Reviewer 2**

Thank you for the review, especially during the pandemic.

Section 3b (and elsewhere): The discussion of weather types (WT) always by number has some drawbacks. The number designations are convenient and provide a succinct way of summarizing the data but for first time readers, the meaning of the WTs and the circulation pattern changes transpiring is often not quite so clear. For example, insert information clarifying the meaning of a WT1-WT6-WT5-WT1 (Line 250) sequence. What is happening with the circulation? Provide these insights throughout the results so the reader can connect arbitrary WT numbers with meaningful conclusions.

Thank you very much for your feedback and suggestions. We agree that referring to the WT that way has both advantages and disadvantages. We have added labels to the sequences to describe their circulation pattern (either Trough-Ridge, Ridge-Trough, Midwestern Trough, Maritime Trough, or Strong Ridge) and tried to strike a balance between referring to the WTs by circulation pattern versus number. The progressions have also been labelled with numbers (E1-4 and L1) based on when in the season they occur (“E” for “Early” and “L” for “Late) to make it easier for the reader to remember their context.

Summary and Discussion: The utility of this information needs further discussion. How are is the knowledge of these WT useful? What are the regional climate implications? Any comparison to other regions?

Thank you for the suggestion. We have added an analysis and discussion of how the WTs relate to surface sensible weather, extreme events, and teleconnections; clarified the link to seasonal shifts; and expanded the discussion of how the results can be used as a basis for further work. We have also added more information on the relevance and motivation of the analysis to the introduction.

Section 1, Lines 90-92: Extra "a" after Rogers 2007. Shorten to Qian et al. 2013.

Done.

Section 1, Line 104: Pluralize end.

Done.

Section 1, Lines 107: What is "mean fields"?

Removed “mean fields”.

Section 1, Lines 116-119: Delete without information loss.

We feel this is an important part of the motivation for the paper.

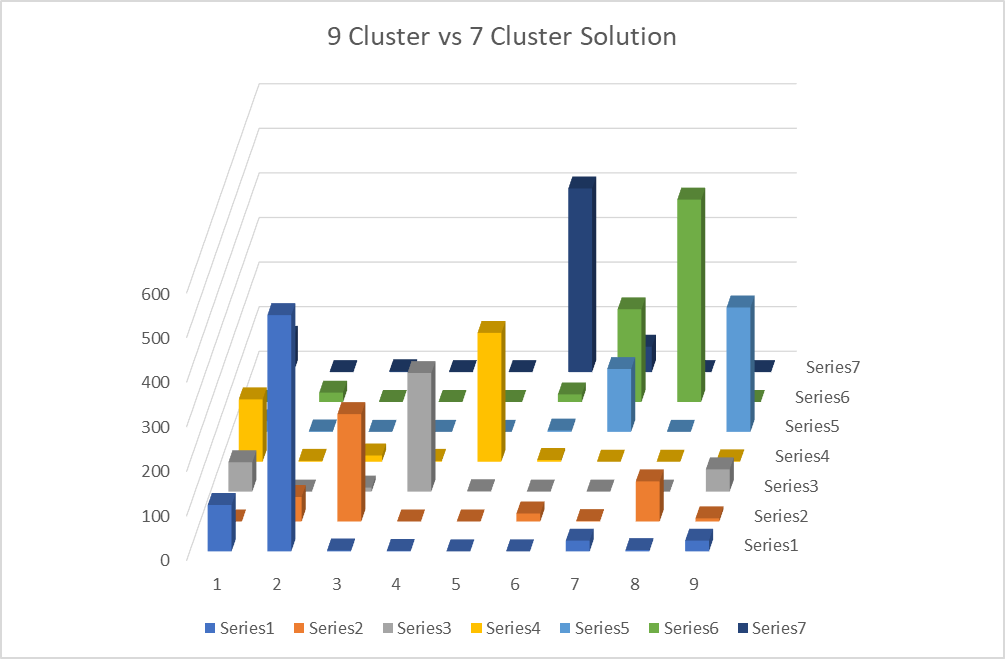
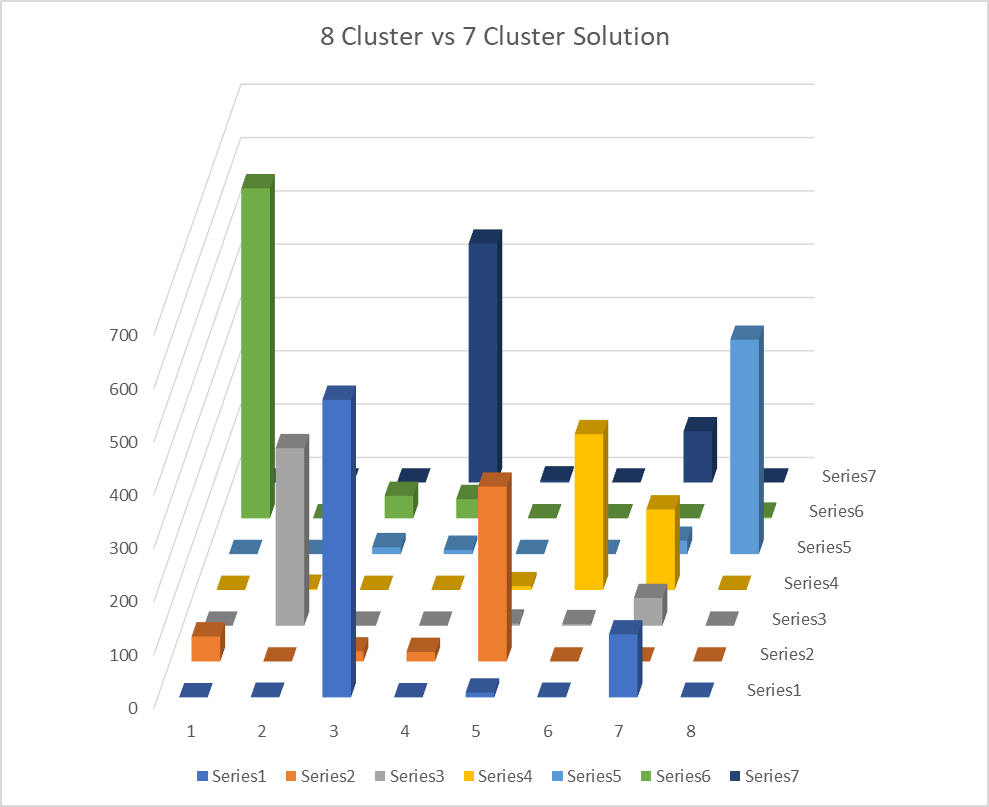
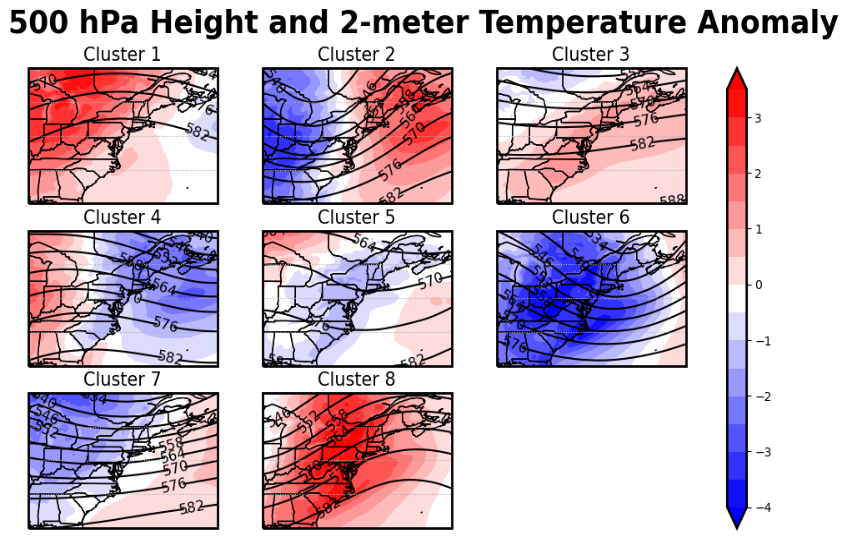
Section 2, Lines 155-156: As noted, the maximum number of clusters was subjectively chosen. Were other maximum cluster numbers investigated? What were the impacts on the results? Provide a rationale for the subjective numbers selected.

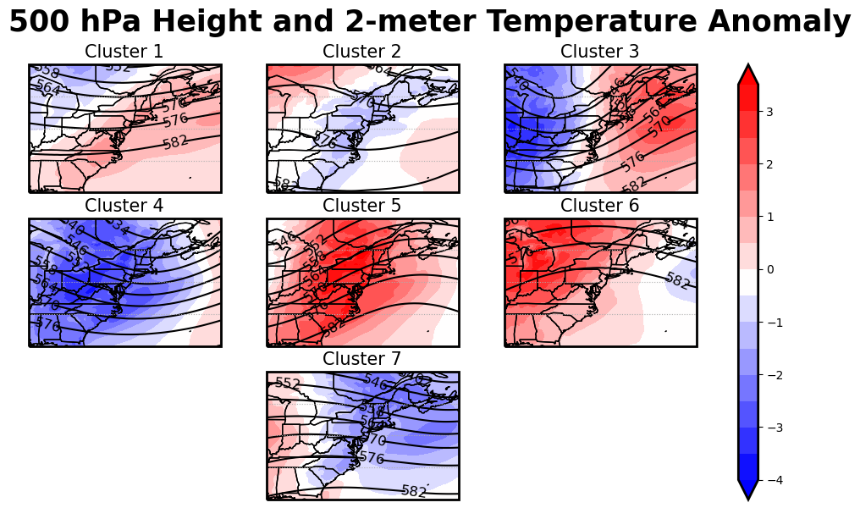
We regret the confusion. The number of clusters is chosen based on the analysis of the classifiability index, which is discussed later in that section. We have now clarified this in the text.

The classifiability results (Fig. 1b) seek to 1) identify a robust and repeatable partitioning of the data into k clusters utilizing multiple k-means runs, and assigning a value (the CI) to this partitioning and 2) comparing this CI value to values produced by repeating the process for k using red noise from the data. Based on CI results, there are several potential k to consider (3, 7, 8, 9, and 10). A small number of clusters may conflate distinct processes and a large number of clusters may not be well sampled by the data. 3 clusters are not sufficient to capture multiple weather processes in a transition season. We focus then on 7-10 clusters, with the aim of finding the fewest number of clusters, for robustness and simplicity, that capture the main patterns. We compared the 7-cluster solution to the 8 and 9-cluster solutions for cluster similarity and physical interpretation of the atmospheric circulation patterns, shown in Figure R2. We did not include the 10 cluster solution as it has the lowest CI value of the four clusters and we felt that 10 patterns for a transition season would be too many.

The 8-cluster solution had two clusters which best matched one cluster from the 7-cluster solution and the 9-cluster solution had three sets of two clusters that best matched one cluster from the 7-cluster solution. That is, the 8 and 9-cluster solutions do not identify distinct new patterns but rather subdivide existing patterns from the 7-cluster solution. Does this additional level of detail provide a clear advantage for the expense of greater complexity and smaller sample sizes? Comparing the 8 and 7-cluster solutions, Cluster 4 of the 7-cluster solution best matched clusters 6 and 7 from the 8-cluster solution. All three feature troughs and below average temperatures for the region. The trough is stronger in Cluster 6 than Cluster 4, however splitting Cluster 4 into these two clusters does not clearly provide new or useful information/features. Accordingly, the 8-cluster solution is not chosen. A similar comparison done between the 7-cluster and 9-cluster solutions also provided no clearly new or useful information when splitting clusters from the 7-cluster solution, and so that solution is not preferred either. In summary, the 7-cluster solution captures the primary patterns with the fewest number of clusters, and hence the greatest sample size.

Additionally, we have not found the results to be sensitive to the specific definition of the domain.





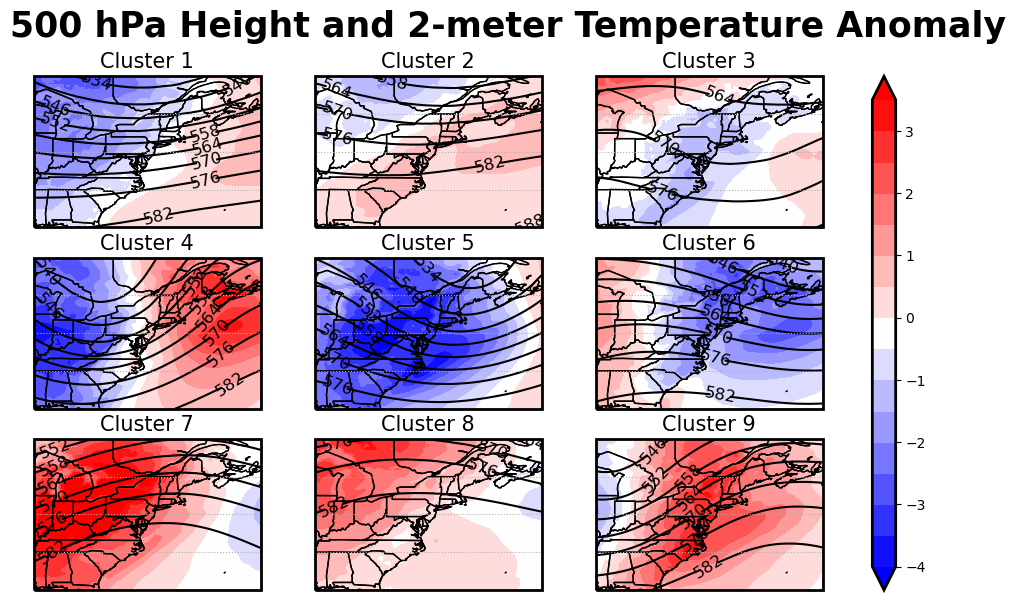


Figure R2: Comparison of chosen 7 cluster solution to 8 and 9 cluster solutions. (a,b) The results of comparing which clusters in the 8 and 9 cluster solution (x axis) occur on days in the 7 cluster solution (y axis). (c-e) 500 hPa heights and 2-meter temperature anomalies for the 7, 8, and 9 cluster solutions for comparison.

Figure 2: Last sentence in caption should read "…gray shaded bars…".

This figure is no longer in the manuscript.

Figures 3 and Figure 4: To avoid confusion with shaded blue height anomalies for Figure 4, identify MSLP anomalies as blue solid lines. What is the contour/anomaly interval? The geopotential height intervals in Figure 3 are virtually unreadable. What is the scaling on the wind vectors or at the very least, value range?

Figure 4 with the anomalies is no longer in the manuscript. Contour intervals have been standardized using 6 hPa intervals for 500 hPa heights, 4 hPa intervals for MSLP, and 0.5 (°C and mm/day) for the temperature and precipitation anomalies. The text has also been made larger for readability.

Figure 7: The color scheme in combination with the narrow bars make this figure difficult to read. Given the percentages are so low past 5 days, suggest truncating the figure to 5 days to spread the bars; then either explain in the caption and/or text the information past 5 days or create a Figure 7b that shows beyond 5 days with a different vertical scaling.

This figure is no longer in the manuscript.

References: Inconsistent AMS reference formatting throughout. Obvious errors with Angel et al; Chen et. al.; Moron et al. 2015:

These have been fixed, thank you.

**4. Response to Reviewer 3**

Thank you for the review, especially during the pandemic.

This paper applies k-means cluster analysis to ERA5 reanalysis data for the north-eastern United States to derive a set of weather types for the autumn season (Sep/Oct/Nov). Different clusters are found to occur more frequently in different months, transitions between different weather types are examined, and changes between the first and last 2 decades of the data are examined.

I find the revised paper to still be lacking in motivation and substance - the authors have applied a statistical method to some reanalysis data and describe the statistical properties of the results, but there's little sense of why they're doing this, any dynamical basis for the results, or what important impacts (eg in terms of local temperature or precipitation) the results have.

We regret the lack of clarity. As submitted, our two key results were: 1) identifying and investigating the primary autumn weather patterns in the Northeast, and 2) capturing the seasonal transition and its shift over time based on regional daily weather patterns. Neither analysis has been done yet for the region and, to our knowledge, this is the first use of daily weather types to characterize shifts in seasonal timing for any region (although Allen and Sheridan (2016) have used daily upper-level height clusters to analyze seasonal evolution at the continental scale). Therefore, the results improve our understanding of both the characteristic weather of the Northeast in an understudied season, as well as how that weather is changing in a changing climate.

Additionally, in this revision, we have now added consideration of the associated precipitation and surface temperature fields, including analysis of extreme events; analysis of the relationship between the weather types and standard teleconnections; a more in-depth examination of progressions between weather types through a Markov Chain analysis; and more discussion of the synoptic context of the weather types.

Presentation of the results could also be improved - some figures are rather cluttered, and panels and annotating text are rather small.

Thank you for the comment. We have extensively revised the figures.

The authors' findings do not seem particularly surprising either - for example, since autumn is a transition season, it would be expected that the atmospheric circulation types at the start of September are rather different from those in late November.

To find a difference is surely not surprising, but determining how that difference is actually realized in changes to individual weather patterns and the degree to which those changes can be captured in a small number of patterns would seem to be a considerably less obvious enterprise. Moreover, we have not found any previous analysis that has managed to quantify the seasonal shift in terms of regional weather patterns and so we suggest that conducting this analysis successfully may not be quite as simple as it appears. Finally, as noted above, we have added several new parts to the analysis which we feel considerably broadens the significance of the work, as well as more physically-based discussion of the weather types, themselves.

The analysis presented is almost entirely in terms of the statistical properties of the weather type time series, with no exploration of any underlying physical reasons behind the results. There might be an interesting story in terms of longer-term changes in the frequency of the atmospheric circulation types hinting at a change to the seasonal cycle, but again this idea is not explored in anything like enough depth.

We have added more physically-based discussion of the weather types and considerably expanded the scope of analysis to consider the associated changes in precipitation and temperature, as well as the connection to large-scale teleconnections. We feel that the story of seasonal cycle changes here is quite interesting – that the seasonal shift is not just in terms of monthly changes to average temperature and the position of climatological features as already shown in previous work but is also evident in changes to the frequencies of individual weather patterns. Additionally, we feel that the methodology, developed in this analysis, to resolve aspects of the seasonal shift at the daily timescale, also offers many exciting prospects for future work.

In summary, I feel that the paper is still not suitable for publication in Journal of Climate and recommend rejection.

We appreciate your critique and have adding several new parts to the analysis and extensively revised the text and figures in response, which we feel has considerably improved the manuscript.

"To better understand significance in the data," - the significance of what quantity in the data? Cluster frequency?

We are using the Monte Carlo method to find the 95% confidence interval of likelihood of occurrence of multiple variables in this study (WT frequency per month, WT seasonal frequency, WT progression, etc.). This has now been explained in more detail.

"WT2 progresses to no other WTs other than itself, so it is disregarded as being part of any sequence." This isn't strictly true though - there are still transitions to other clusters even though they might not be significantly different from chance.

You are correct, that there are transitions to other clusters that may not be significant. We have reworded this in the manuscript and now explain that we are looking at transitions which are significant at the 95% level.

How often do the preferred multi-type sequences actually occur in the cluster time series? It looks like the sequences are just based on the preferred transition from one state to a second state that is more likely than chance, so is it really justified to extend this to a whole sequence of states - eg WT5-WT3-WT4-WT7-WT5?

Thank you for your comment. We have gone back and analyzed the data for the occurrence of progressions of lengths 2, 3, 4, and 5 with persistence removed (Table R1). We do not show progressions of length 5 here as it occurred on average only once every 10 years. We have determined the progressions seen in Figure R1 using a Markov Chain Analysis combined with a Monte Carlo method, which is now described in the paper. Comparing the sequences that appear in our dataset to those in the Markov Chain analysis, we find the extension of preferred state transitions for the following progressions: 1-2-6-5, 3-4-7-5, 1-6-5, 1-6-2, and 1-6-1. The progressions 5-4-7 and 5-3-7 do appear 30 and 32 times; however, based on the Markov Chain analysis, they are only significant below the 80% level, so we do not include them in this analysis.

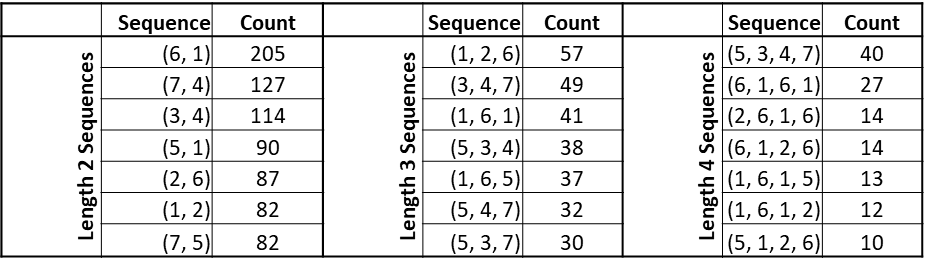


Table R1: Top 7 progressions that occurred in the SON dataset. Only the top 7 results are shown for space reasons.

Are there any dynamical reasons behind the preferred sequences of weather types, or their distribution across the autumn season? These seem like quite basic questions, but there's no discussion of these issues in the paper.

We regret the lack of clarity. 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. We have clarified this in the text. We have also added figures for the individual progressions, to make the interpretation of them more apparent.

How do we know whether these changes are due to climate change or just decadal variability? There are other reanalyses available that go back much further in time than ERA5 currently does, so why not use these to examine this question?

The relationship of the seasonal shift in this period to climate change has already been examined in the previous literature, which is discussed in the introduction. ERA5 is desirable for a regional circulation analysis because it is recently released, and so takes advantage of improvements in reanalysis methodology and data processing, and is also of comparatively high resolution. Analysis of satellite-era reanalysis products has been very popular, even for some aspects of climate change, due to the greater data availability in the post-1979 period, and the usefulness of establishing a baseline of analysis in a more data-rich period. As of this writing, ERA5 is in the process of being extended back to 1950 (the data is currently available only in a preliminary form), so we expect to be able to build off of, and extend, this analysis in the future. Additionally, we plan to use the observed weather types established in the current analysis as a baseline to both assess climate models and to examine how weather types change in the modeled warming projections, which will give further insights into the underlying dynamics.

Larger text for labelling is needed throughout - it's a struggle to make out a lot of the text around the plots.

Done.

Figure 2 - label the x axis. Presumably the figure showing histograms of correlation coefficients and RMSE values, but it's not immediately clear. I don't really understand the plots. If the y axis is percentage of days, shouldn't the bars in each plot take up the same area? Why are values for the u850 and v850 RMSE panels so much smaller than the others? Are there more blue bars that we can't see because they're hidden behind the gray bars, or have the blue bars just been stacked on top of the gray bars?

This figure is no longer in the manuscript.

Figure 3 has 3 separate fields all plotted in the same panels, plus boundaries between American states - it's a bit overloaded with information. The geopotential height might be more informative if the zonal mean values were removed? There's a strong meridional gradient in height which tends to obscure the features of each cluster.

*Diagram

Description automatically generated*Thank you for the suggestion. We have incorporated other suggestions from the other reviewers and split Figure 3 into two new figures. One figure shows the MSLP, 850 hPa wind, and precipitation anomaly, while the other shows 500 hPa heights and 2-meter temperature anomaly. This allows for the readability issue to be addressed. We have done an analysis with the zonal mean removed (Figure R3), however, we feel that the full version in the manuscript is more easily understood now that it has been moved to its own figure.

Figure R3: 500 hPa heights for each WT with the zonal mean removed.

I presume the point of figure 7 is to show that some clusters are more persistent than others, but it is difficult to make this out from the way the authors have plotted the data. I suggest having a separate panel for each cluster might be clearer.

This figure is no longer in the manuscript.

Figure 9 - presenting the data as stacked bar graphs makes it harder to see the trends in the individual types. Again, I suggest showing the frequency for each cluster separately.

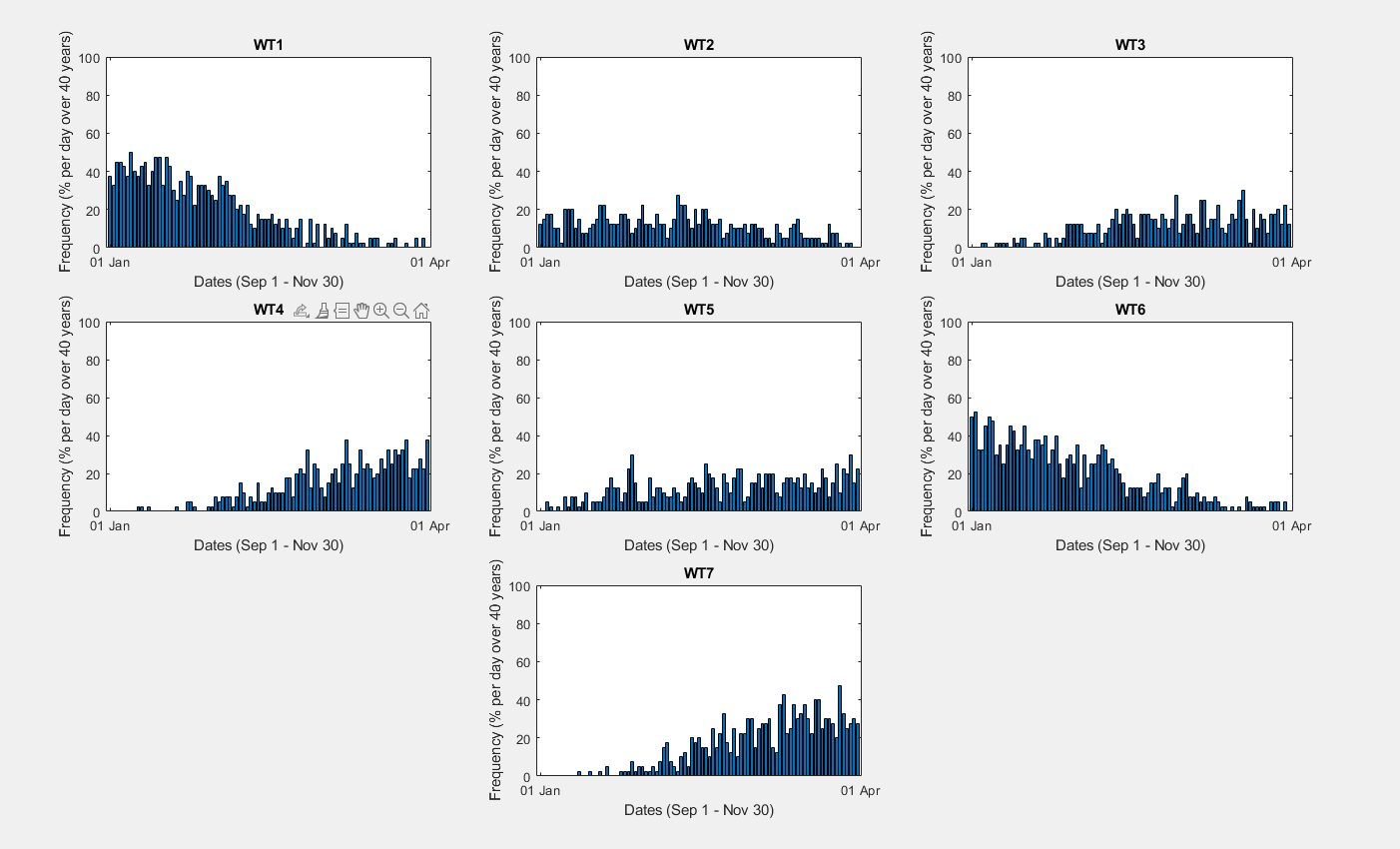
Thank you for your comments. While we agree that showing the clusters separately will help to show the trend in the individual types, we feel that how the figure as it is now best represents our data. Figure R4 shows the data as you suggested, with the individual WT trends each plotted separately. While we can infer the same information here, we are concerned in our study with the overall trends of occurrence of the Early Season and Late Season WTs together and use that to show timing in October around when the transition between Early and Late Season tends to occur.

Figure R4: WT Frequency of occurrence per day in SON over all 40 years.