

An Objective Method for Clustering Observed Vertical Thermodynamic Profiles by Their Boundary Layer Structure

DILLON V. BLOUNT,^a CLARK EVANS,^a ISRAEL L. JIRAK,^b ANDREW R. DEAN,^b AND SERGEY KRAVTSOV^a

^a *University of Wisconsin–Milwaukee, Milwaukee, Wisconsin*

^b *NOAA/Storm Prediction Center, Norman, Oklahoma*

(Manuscript received 17 November 2022, in final form 27 April 2023, accepted 1 May 2023)

ABSTRACT: This study introduces a novel method for comparing vertical thermodynamic profiles, focusing on the atmospheric boundary layer, across a wide range of meteorological conditions. This method is developed using observed temperature and dewpoint temperature data from 31 153 soundings taken at 0000 UTC and 32 308 soundings taken at 1200 UTC between May 2019 and March 2020. Temperature and dewpoint temperature vertical profiles are first interpolated onto a height above ground level (AGL) coordinate, after which the temperature of the dry adiabat defined by the surface-based parcel's temperature is subtracted from each quantity at all altitudes. This allows for common sounding features, such as turbulent mixed layers and inversions, to be similarly depicted regardless of temperature and dewpoint temperature differences resulting from altitude, latitude, or seasonality. The soundings that result from applying this method to the observed sounding collection described above are then clustered to identify distinct boundary layer structures in the data. Specifically, separately at 0000 and 1200 UTC, a k -means clustering analysis is conducted in the phase space of the leading two empirical orthogonal functions of the sounding data. As compared to clustering based on the original vertical profiles, which results in clusters that are dominated by seasonal and latitudinal differences, clusters derived from transformed data are less latitudinally and seasonally stratified and better represent boundary layer features such as turbulent mixed layers and pseudoadiabatic profiles. The sounding-comparison method thus provides an objective means of categorizing vertical thermodynamic profiles with wide-ranging applications, as demonstrated by using the method to verify short-range Global Forecast System model forecasts.

KEYWORDS: Soundings; Empirical orthogonal functions; Clustering

1. Introduction

The history of vertical soundings, which provide meteorological observations through parts of Earth's atmosphere, begins in the nineteenth century. The first recorded vertical sounding was taken in 1894 by Abbott Lawrence Rotch with a kite that carried a lightweight thermograph (Blue Hill Observatory and Science Center 2021). This initiated regular atmospheric soundings of air temperature, dewpoint temperature (hereafter simply temperature and dewpoint, respectively), and pressure, as well as wind speed and direction. Organizations such as the U.S. Weather Bureau and Germany's Aeronautical Observatory continued using kite soundings into the 1930s, while vertical soundings using aircraft and free-flying balloons became more common in the 1930s and 1940s. However, there were downsides to using balloons and kites, as kites could only reach altitudes of around 4 km above ground level and balloons had to be recovered to obtain the recorded data (Stith et al. 2018).

These pitfalls led to the development of radio-transmitting instrument packages attached to balloons called radiosondes. Radiosondes observe temperature, dewpoint, and pressure in

the lower atmosphere and radio these data back to a remote receiving station. The instrument packages that can also record and transmit horizontal wind data are known as rawinsondes (Stith et al. 2018). Routine upper-air observations, which collect data using rawinsondes in the troposphere and lower to middle stratosphere, are taken at fixed locations across the globe up to two times per day, typically at 0000 and 1200 UTC.

The wide range of locations from which the atmosphere is sampled by rawinsondes ensures that rawinsonde observations can sample many different atmospheric phenomena. For example, precipitation is often accompanied by near-saturation and a pseudoadiabatic vertical temperature profile (Fig. 1a). Clear skies, cool temperatures, and calm winds at night can lead to the formation of a near-surface radiation inversion (Fig. 1b). Strong surface sensible heating can result in the formation of deep turbulent eddies and associated turbulent mixed layer (Fig. 1c). Finally, frontal inversions separate cooler, drier air masses near the surface from warmer, moister air masses above (Fig. 1d). These features are associated with distinct temperature and dewpoint profile shapes on skew T - $\ln p$ diagrams that are functions of the features themselves and of the synoptic meteorological conditions within which they occur.

Previous studies have introduced subjective and objective methods for clustering vertical soundings. Subjective clustering methods include those by geographic location (Evans et al. 2018), surface-based instability magnitude (Coniglio et al. 2013; Evans et al. 2018), and the presence of a capping inversion (Coniglio et al. 2013; Nevius and Evans 2018), with these methods subsequently used to better understand numerical model

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/WAF-D-22-0195.s1>.

Corresponding author: Clark Evans, evans36@uwm.edu

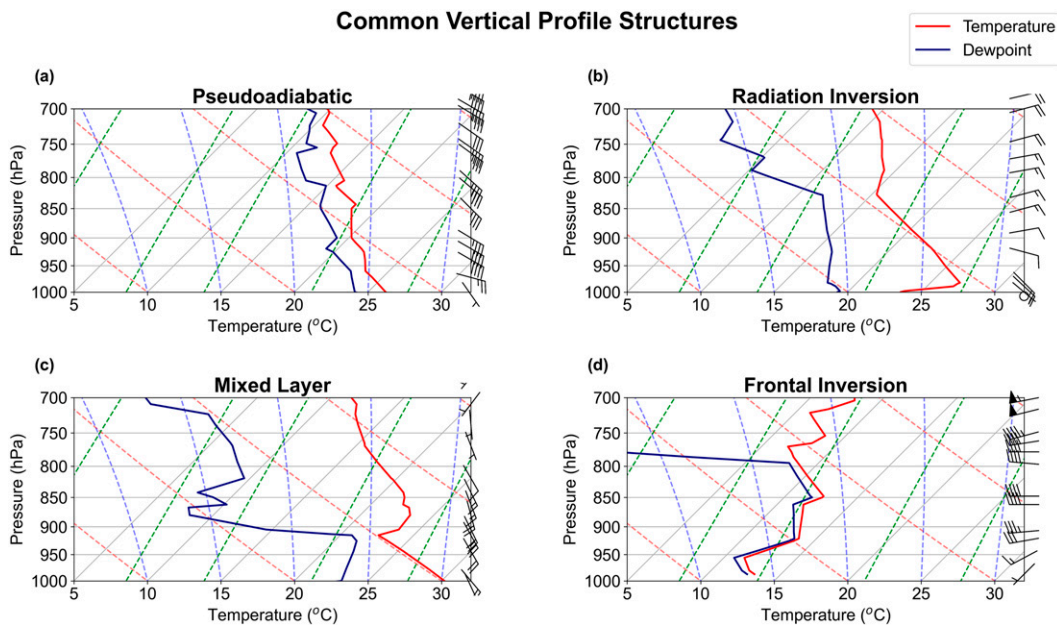


FIG. 1. Observed skew T - tp diagrams (temperature in $^{\circ}\text{C}$ in solid red lines; dewpoint temperature in $^{\circ}\text{C}$ in solid blue lines; horizontal wind speed and direction in barbs, with half barb = 5 kt, barb = 10 kt, and pennant = 50 kt, where 1 kt $\approx 0.51 \text{ m s}^{-1}$) constrained to below 700 hPa to depict examples of boundary layer structures. (a) A moist sounding profile at Slidell, LA, at 0000 UTC 13 Jul 2019 with a nearly pseudoadiabatic layer above 850 hPa; (b) a radiation inversion extending from the surface to 980 hPa at Jackson, MS, at 1200 UTC 15 Sep 2019; (c) a vertically mixed layer extending from the surface to 915 hPa at Brownsville, TX, at 0000 UTC 22 Jul 2019; and (d) a frontal inversion over the 950–925-hPa layer at Buffalo, NY, at 0000 UTC 4 Oct 2019.

biases in such environments. Among objective clustering methods, methods previously applied include self-organizing maps (SOMs; Kohonen 1995) and k -means clustering (Forgy 1965; Lloyd 1982). For example, SOMs have been used to cluster vertical soundings in proximity to distinct thunderstorm modes (e.g., tornadic versus nontornadic supercells; Nowotarski and Jensen 2013), and ozone mixing ratio profiles (Jensen et al. 2012), whereas k -means clustering has been used to cluster vertical soundings to identify distinct Amazonian meteorological regimes (Giangrande et al. 2020). These methods have demonstrated the ability to identify distinct thermodynamic-profile structures when there is little variation in the sounding climatology. However, it is still unclear whether these methods can efficiently identify distinct thermodynamic profile structures in the presence of more substantial variability in the sounding climatology.

This study introduces a novel objective method to account for climatological variability in large samples of atmospheric boundary layer vertical thermodynamic profiles by subtracting the values of the dry adiabat that extends upward from the surface parcel's air temperature from the temperature and dewpoint profiles. The resulting vertical profiles are hereafter referred to as transformed profiles. This method largely preserves the vertical profiles' shapes, which represent unique atmospheric processes that can occur throughout the year (and thus be associated with a wide range of temperatures and dewpoints) at any geographic location. The transformed profiles are then categorized into groups via k -means clustering to identify similar profiles. The hypothesis guiding this study is

that this sounding transformation method removes sufficient background environmental variability to enable the distinct structures depicted by vertical thermodynamic profiles—here focusing on the atmospheric boundary layer—to be objectively identified and clustered across a large, highly variable sounding climatology.

The rest of this study is structured as follows. Section 2 outlines the data analyzed before discussing the transformation, data compression, and clustering methodology used to classify soundings. Section 3 demonstrates the efficacy of this methodology for a large set of observed soundings from across the United States during May 2019–March 2020. Section 4 illustrates how the classification and clustering method can be used to document environment-specific biases in numerical model-forecast vertical thermodynamic profiles. Finally, section 5 summarizes the study's key findings and outlines and further potential uses of the objective sounding classification method.

2. Data and methods

The data used in this study are observed temperature and dewpoint vertical profiles from 112 routine upper-air observation stations across the United States, Canada, and Mexico launched by the National Oceanic and Atmospheric Administration (NOAA), Environment and Climate Change Canada, and Servicio Meteorológico Nacional, respectively. The data cover the period from 1200 UTC 7 May to 1200 UTC 31 March 2020, excluding 0000 UTC 8 June–1200 UTC 20 June 2019 and

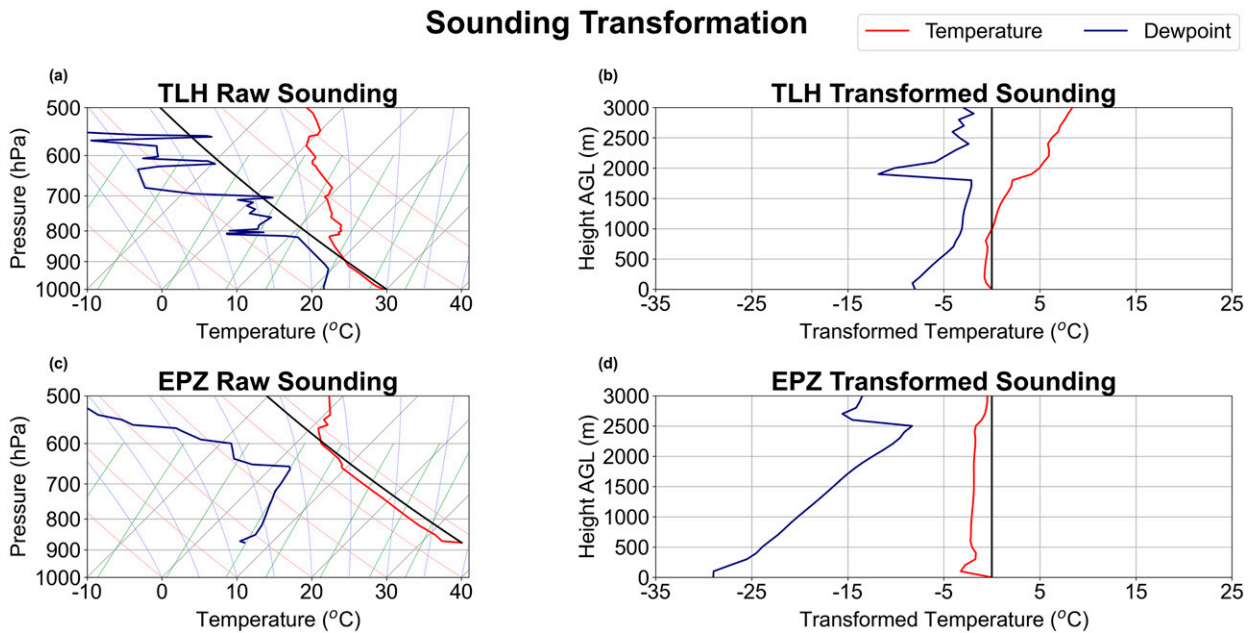


FIG. 2. Observed skew T - $\ln p$ diagrams from (a) Tallahassee, FL, at 0000 UTC 31 May 2019 and (c) El Paso, TX, at 0000 UTC 26 Jul 2019. The temperature and dewpoint are depicted in red and blue lines, respectively. (b),(d) As in (a) and (c), but transformed to a height AGL vertical coordinate with the surface temperature’s dry adiabat [black line in (a) and (c)] subtracted from the temperature and dewpoint at all altitudes.

0000 UTC 12 October–0000 UTC 1 November 2019 due to gaps in the Storm Prediction Center (SPC) sounding archive used in this study, and include $N_{\text{soundings}}^{00} = 31\,153$ soundings at 0000 UTC and $N_{\text{soundings}}^{12} = 32\,308$ soundings at 1200 UTC. These data are obtained from SPC’s internal archive in JSON format, a format which is easy to read using Python’s pandas (Pandas Development Team 2023) package and for which we previously developed a data processing and visualization workflow. The large data volume limited the amount of data which could be transferred from SPC’s internal systems to those at UWM used to process and analyze the data, such that only 11 months of data are used in this study. However, since these data capture a nearly complete annual cycle, we believe the climatological sounding structures would not change if additional years are included.

Observed soundings are first interpolated onto a common height above ground level (AGL) vertical grid with uniform vertical grid spacing of 100 m. This accounts for differences in altitudes between stations (e.g., Fig. 1 of Fovell and Gallagher 2020). The new grid’s vertical extent is restricted to the lower troposphere (below 3 km AGL) given its importance to surface sensible weather, atmospheric stability, and particulate transport. This results in a grid with $N_z = 31$ vertical levels. Next, the values of the dry adiabat extending upward from the surface parcel’s air temperature are subtracted from each sounding’s temperature and dewpoint profiles at all vertical levels (Fig. 2). The dry adiabat is chosen for this transformation because of its constant lapse rate (9.8 K km^{-1}) no matter the altitude or moisture content. The transformed temperature profile is directly related to parcel stability; e.g., a negatively

sloped transformed temperature profile inherently indicates that an air parcel lifted from the altitude at which the negative slope begins is absolutely unstable. As in Fig. 2, wherein Tallahassee, Florida, and El Paso, Texas, soundings are characterized by turbulent mixed layers despite substantially different station altitudes and surface meteorological conditions, these transformations retain fundamental sounding shapes while reducing the variability that results from altitude, latitudinal, and seasonal differences between soundings. The raw soundings in this plot are limited to 500 hPa to focus on lower-tropospheric features. Allowing these examples to extend to 500 hPa ensures that the lowest 3 km AGL are included within each sounding (e.g., 3 km AGL is approximately 600 hPa at El Paso in Fig. 2c) and provides context for the grid transformation, i.e., allows for readers to see which data are included and not included after this transformation. Hereafter, the untransformed data are referred to as the raw data, whereas the transformed data are referenced as such.

The raw and transformed temperature and dewpoint data are arranged into an input data matrix \mathbf{X} , which has dimensions $N_{\text{soundings}}^{00} \times (2N_z)$ for 0000 UTC data and $N_{\text{soundings}}^{12} \times (2N_z)$ for 1200 UTC data; note that the factor of 2 on N_z represents the number of variables in the input data (temperature and dewpoint). 0000 and 1200 UTC are considered separately given the influence of the diurnal cycle on near-surface conditions. The mean thermodynamic profile (computed by averaging over all samples in \mathbf{X}) is subtracted from each sample profile prior to further analysis so that the subsequent clustering is performed on anomalous profiles. Next, \mathbf{X} is subjected to an empirical orthogonal function (EOF; Obukhov 1947; Lorenz 1956; Davis 1976)

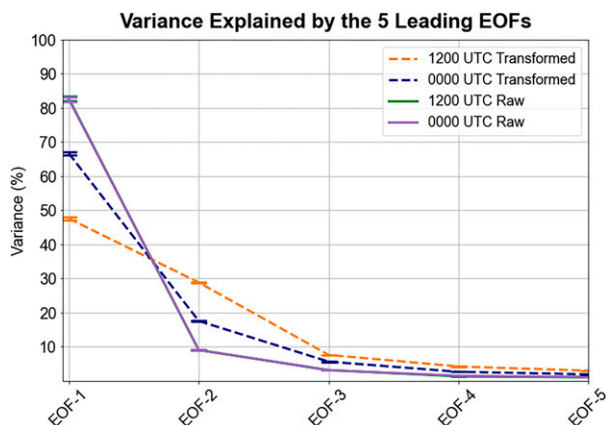


FIG. 3. Percentage of variance accounted for by the five leading EOFs for raw and transformed sounding data at 0000 and 1200 UTC. The heuristic error bars (North et al. 1982) represent the PC-variance uncertainty associated with each EOF. Note that the lines for the 0000 and 1200 UTC raw data largely overlap each other.

analysis to reduce the input data's dimensionality (e.g., Monahan et al. 2009). Specifically, \mathbf{X} is decomposed as $\mathbf{X} = \mathbf{UV}^T$, where \mathbf{V} is a $(2N_z) \times (2N_z)$ orthogonal matrix of EOFs, $\mathbf{U} = \mathbf{XV}$ is the matrix of the principal components (PCs) which has the same size as \mathbf{X} and contains the uncorrelated series of the magnitudes associated with each EOF pattern across the input soundings, and the superscript T indicates the matrix's transpose. The loadings \mathbf{V} are found so that the variance accounted for by the leading K EOFs (ordered by decreasing explained variance; Wilks 2019) is maximized and the residual variance is minimized.

For the raw data, the leading EOF mode is well separated from the remaining EOF modes in the 0000 and 1200 UTC data, with the leading EOF mode accounting for over 80% of the variance in the data (solid lines in Fig. 3). Thus, only the leading EOF is hereafter retained for the raw data. Interestingly, despite representing distinct times within the diurnal cycle, the variance explained by each of the leading five EOF modes is nearly identical between the raw data at 0000 and 1200 UTC (Fig. 3). This indicates that the dominant mode of variability in the data is identical between the two times, which we believe to be associated with airmass properties that are highly variable across the sounding climatology (which covers nearly a full year and over 50° of latitude) given the cluster-mean profiles presented in section 3a.

Conversely, for the transformed data, the leading two EOF modes are well-separated from the remaining EOF modes in the 0000 and 1200 UTC data (dashed lines in Fig. 3). Thus, the leading two EOFs are hereafter retained for the transformed data. For the data considered in this study, the sounding transformation leads to a lower variance in the transformed data's first EOF at each time but a higher variance in the transformed data's second EOF at each time as compared to the corresponding raw soundings' EOFs (solid versus dashed lines in Fig. 3), which must result from the sounding transformation process removing variance associated with the variable airmass properties represented in the input sounding data.

Next, the data from each dataset (raw and transformed data) at each observation time (0000 and 1200 UTC) are subjected to k -means clustering in the corresponding EOF-1 phase space for the raw data and the EOF-1–EOF-2 phase space for the transformed data. The k -means clustering is a nonhierarchical method for grouping data, wherein data can be reassigned between clusters as the analysis is performed, into a user-specified number of clusters (Forgy 1965; Lloyd 1982; Wilks 2019). In this method, an initial clustering is formed based on input points' distances from randomly assigned initial points. The algorithm then computes cluster centroids, calculates the Euclidean distance of each data point from the different cluster centroids, and assigns the data to the cluster with the smallest Euclidean distance between the data point and its centroid. This process is iterated until the distance from each data point to its respective cluster centroid is minimized (Wilks 2019). The efficiency of this clustering technique is represented by silhouette scores, which measure the mean intracluster Euclidean distances compared to the mean intercluster Euclidean distances. The silhouette score can be as large as 1, which represents a perfect efficiency of the clustering method. For both the 0000 and 1200 UTC analyses using the raw and transformed sounding data, $k = 2$ represents the number of clusters (in the range of 2–10) that produces the highest cluster-average silhouette score (Rousseeuw 1987) and lowest number of negative silhouette scores; i.e., two clusters are optimal for maximizing intercluster variance and minimizing intracluster variance (Fig. 4).

For the raw data, the resulting cluster membership is strongly influenced by the time of the year, with Cluster-1 soundings most prevalent during the warm season and Cluster-2 profiles most prevalent during the cold season (Figs. 5a,b). This is further explored in section 3a.

Because of the large variance explained by EOF-1 (Fig. 3), how well the transformed data correspond to EOF-1 is the primary distinguishing characteristic between Clusters 1 and 2 (Figs. 5c,d). This is most notable for the 0000 UTC data (Fig. 5c), for which EOF-1 explains almost 70% of the input data's variance (Fig. 3), and less notable for the 1200 UTC data (Fig. 5d), for which EOF-2 explains slightly less than 50% of the input data's variance (Fig. 3). To further examine a possible role of variability associated with the EOF-2 mode in our classifications for the transformed dataset, we repeat the paper's analyses in a phase space in which the EOF axes are normalized by the standard deviations of their associated PCs. The normalization changes the Euclidean distance in the EOF-1–EOF-2 space used in assigning datapoints to clusters using k -means clustering and thus changes the resulting cluster populations. However, as described in the supplemental material, these changes are minor in nature, such that the standard process of not normalizing the EOF axes is retained herein.

3. Classification

a. Raw versus transformed data

Clusters obtained using the raw 0000 UTC temperature and dewpoint observations are characterized by similarly

Silhouette Scores for 0000 and 1200 UTC

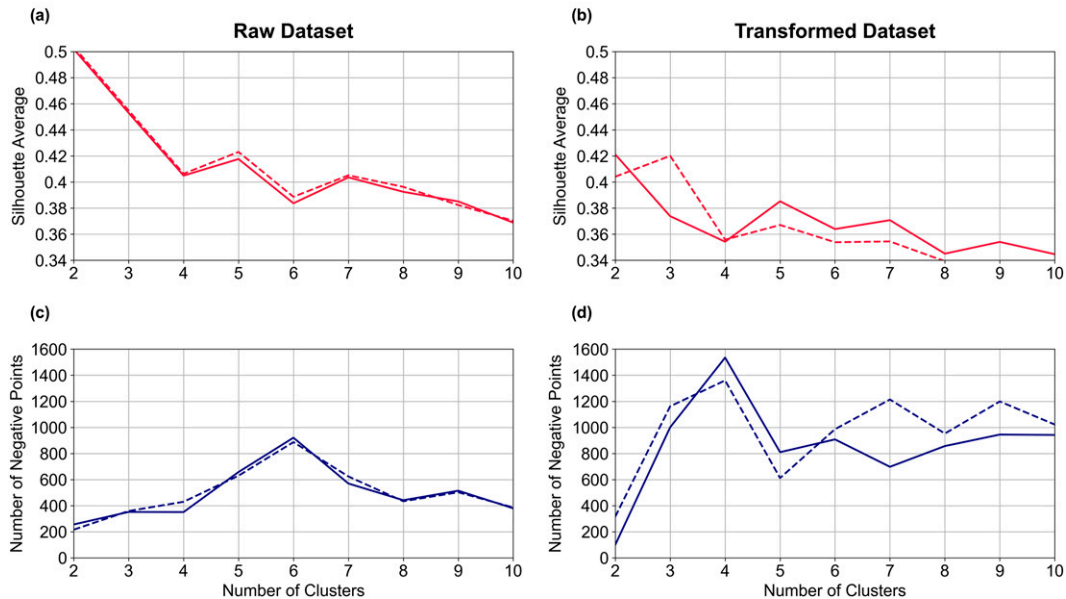


FIG. 4. (a),(b) The cluster-average silhouette score (nondimensional) and (c),(d) total number of negative points for $k = 2$ through $k = 10$ for the 0000 UTC (solid) and 1200 UTC (dashed) (left) raw and (right) transformed datasets.

shaped temperature and dewpoint profiles between the surface and 3 km AGL (Figs. 6a,c). The primary difference between these two clusters lies with their surface airmass characteristics: the Cluster-1 (Fig. 6a) mean surface temperature and dewpoint are approximately 22° and 13°C, respectively, whereas the corresponding Cluster-2 (Fig. 6c) mean

values are approximately -1° and -8°C. These differences stem from latitudinal (Fig. 7) and temporal (Fig. 8) variability between the cluster populations: the colder Cluster-2 profiles predominantly occur in climatologically colder locations such as Canada and the United States Intermountain West (Fig. 7a) on the shoulders of the warm season and

Full Dataset EOF Phase Space

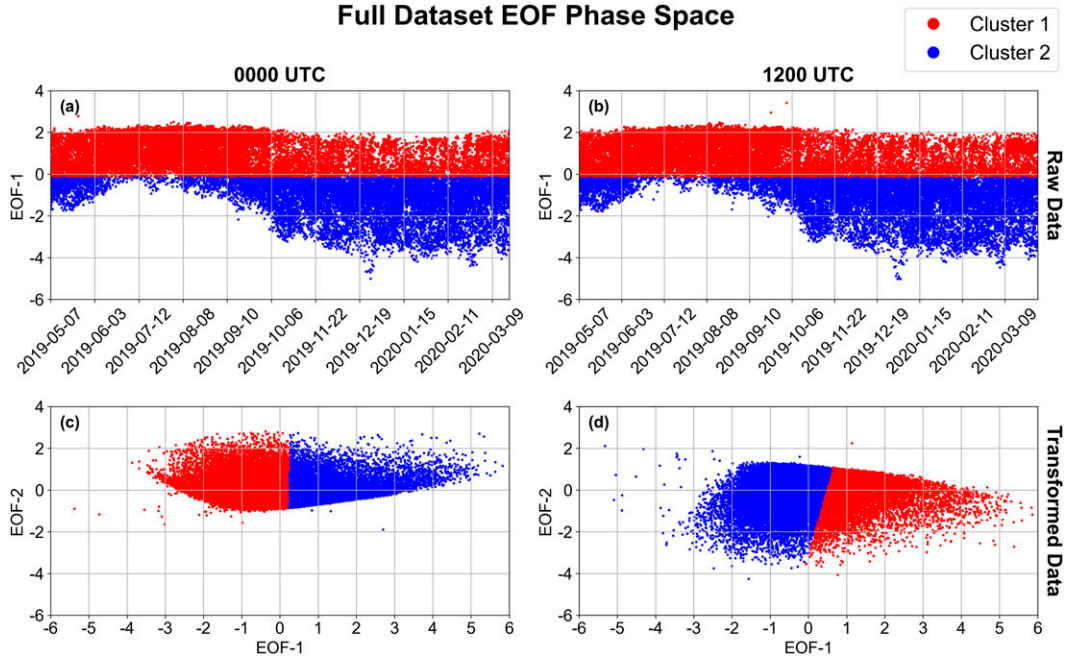


FIG. 5. Scatterplots of k -means cluster identification in the EOF-1 phase space for (a) 0000 and (b) 1200 UTC raw data and in the EOF-1–EOF-2 phase space for (c) 0000 UTC and (d) 1200 UTC transformed data.

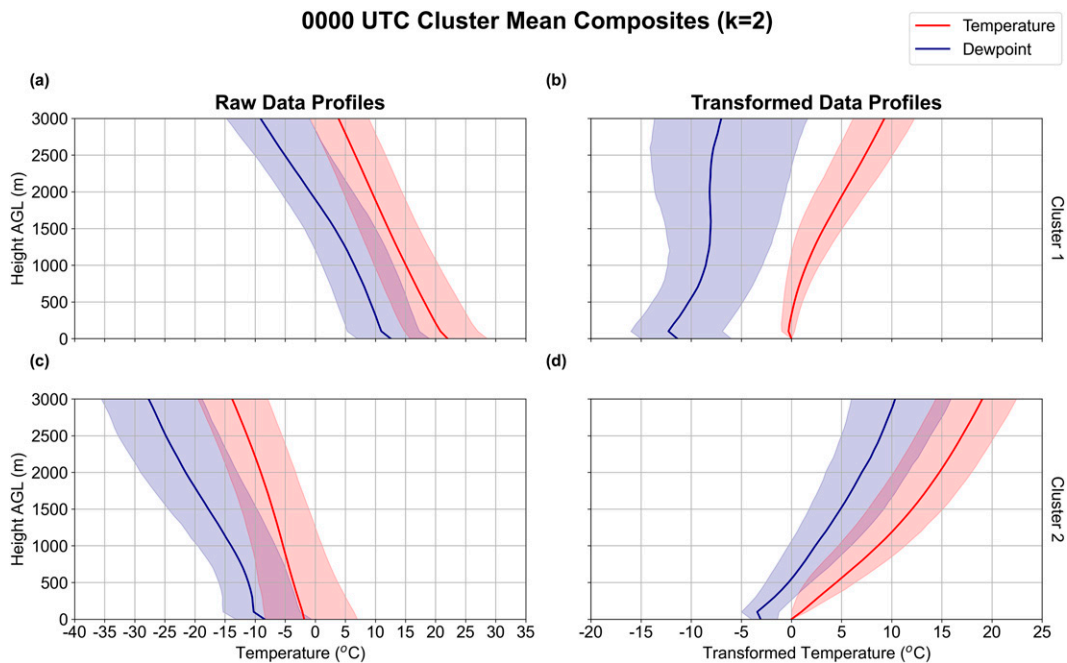


FIG. 6. Cluster-mean temperature (red lines) and dewpoint (blue lines) for (a),(b) Cluster 1 and (c),(d) Cluster 2 for the (left) raw and (right) transformed observed profiles for the 0000 UTC dataset. The semitransparent shading centered on each cluster-mean profile represents the interquartile range (25th–75th percentile) of the data.

during the cold season (Fig. 8) whereas the warmer Cluster-1 profiles predominantly occur in climatologically warmer and moister locations such as the southeastern United States (Fig. 7a) during the warm season (Fig. 8).

By contrast, the cluster-mean profiles derived from the transformed 0000 UTC data (Figs. 6b,d) indicate common boundary layer structures. Cluster 1 (Fig. 6b) depicts a vertically mixed layer, characterized by the cluster-mean transformed temperature

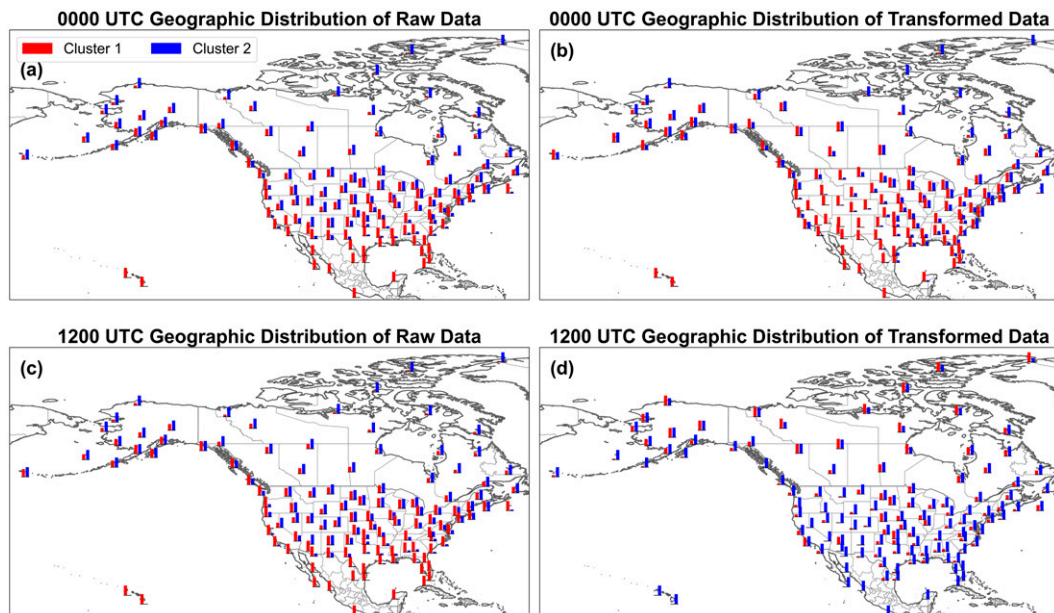


FIG. 7. The geographic distributions of soundings by cluster in the (a),(c) raw and (b),(d) transformed data at (top) 0000 UTC and (bottom) 1200 UTC. Each sounding location is denoted with a bar graph indicating the number of soundings per cluster at that location, with Cluster-1 soundings denoted in red and Cluster-2 soundings denoted in blue.

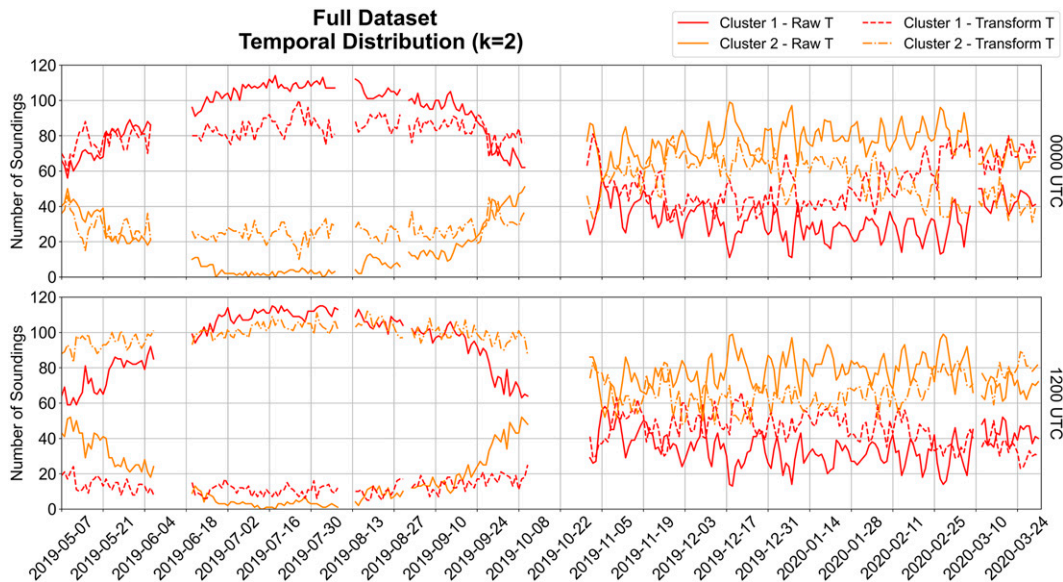


FIG. 8. Sounding counts for the (top) 0000 UTC and (bottom) 1200 UTC observed soundings. In both panels, red lines denote Cluster 1 whereas orange lines denote Cluster 2; solid lines indicate raw data whereas dot-dashed lines indicate transformed data.

approximately equal to zero (intrinsically representing a dry adiabatic lapse rate) to ~ 1.25 km AGL. Although these soundings predominantly occur in the U.S. Intermountain West (Fig. 7b) where strong surface sensible heating of the climatologically arid surface readily facilitates turbulent vertical mixing during the warm season, Cluster 1 also contains soundings from across North America. Conversely, Cluster 2 (Fig. 6d) depicts a nearly pseudoadiabatic profile, as characterized by the cluster-mean transformed temperature increasing by approximately $4^{\circ}\text{C km}^{-1}$ (roughly the difference between the dry and pseudoadiabatic lapse rates below 3 km AGL during the warm season) and a dewpoint depression of approximately 4°C near the surface. These soundings are predominantly located near major coastlines and in eastern North America, locations at which moisture availability is greater during the warm season. Note, however, that these clusters are less temporally stratified than are their counterparts derived from raw sounding data (Fig. 8). In all, the sounding-transformation process appears to reduce the extent to which soundings are clustered based by latitude and the annual cycle.

Similar results are obtained for the observed 1200 UTC soundings. For the raw data, approximately 92% of the soundings have the same cluster assignment as the following (i.e., 12 h later) 0000 UTC data. In fact, the cluster-mean composite profiles (cf. Figs. 6a,c and 9a,c) and the geographical (Figs. 7a,c) and temporal (Fig. 8) distributions associated with the 0000 and 1200 UTC clusters based on raw data are very similar. Slight differences between the cluster composites between the 1200 and 0000 UTC data (cf. Figs. 6a,c and 9a,c) (most notably, lower composite-mean surface temperatures with the 1200 UTC data) are likely a function of the diurnal cycle. Altogether, the clusters derived from 1200 UTC raw soundings are primarily stratified by latitude and time of year, as was true for the clusters derived from 0000 UTC raw soundings.

The sounding transformation process applied to 1200 UTC soundings is not as effective in reducing the latitudinal and seasonal stratification evident in the raw data as it is for 0000 UTC soundings. Here, Cluster-1 soundings are preferentially found at higher latitudes (Fig. 7d) in the cold season (Fig. 8, bottom panel), whereas Cluster-2 soundings are preferentially found at lower latitudes (Fig. 7d) and dominate the warm-season sounding population (Fig. 8, bottom panel). The cluster-mean profiles at this time are primarily distinguished by the parcel stability below 1 km AGL, with the rapid increase in cluster-mean transformed temperature with height over the 0–1 km AGL layer in Cluster 1 implying that parcels lifted from within this layer are more stable than their Cluster-2 counterparts (Figs. 9b,d). This is consistent with stronger near-surface radiative cooling at higher latitudes in the Cluster 1 population. Because of the reduced effectiveness in reducing the latitudinal and seasonal stratification as compared to the 0000 UTC data, only 26% of 1200 UTC soundings have the same cluster assignment as do those at the same observing location taken 12 h later at 0000 UTC (not shown).

b. Thunderstorm-supporting environments

Thunderstorm-supporting environments are examined to demonstrate the efficacy of using the sounding transformation and clustering method on soundings collected in a narrower latitudinal range (the conterminous United States) and portion of the year (primarily the warm season) than in the full sounding dataset. This is done by using NOAA’s Storm Prediction Center (SPC) 1200 UTC Day-1 convective outlooks. Observed soundings were filtered to only retain those located in an outlook category of general thunderstorm, representing a 10% of greater probability of thunderstorms between 1200 UTC

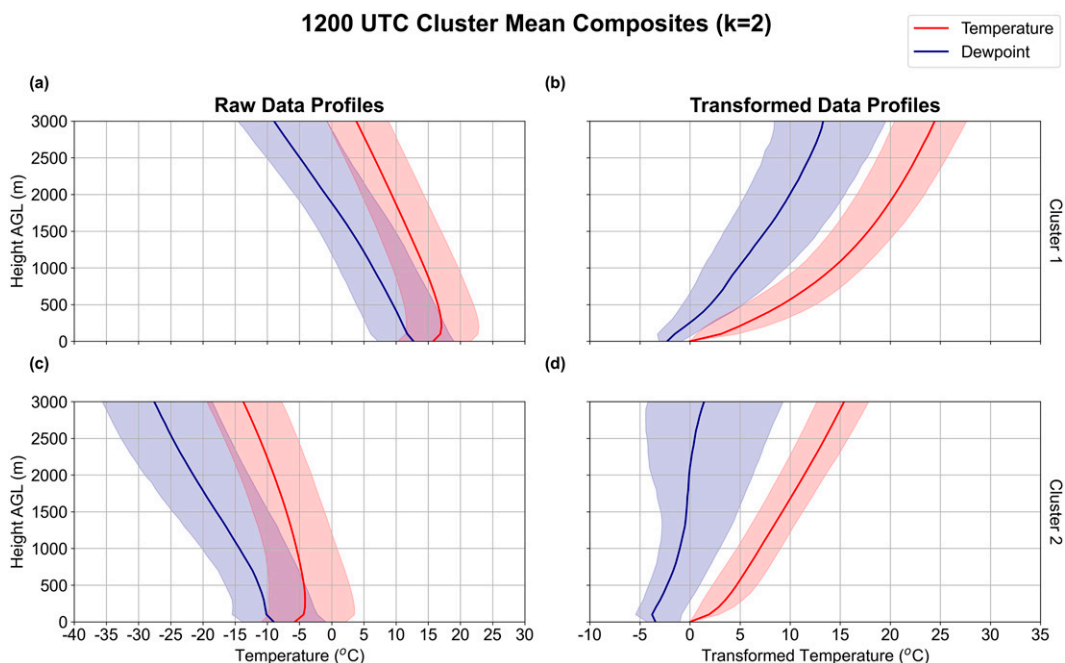


FIG. 9. As in Fig. 6, but for 1200 UTC data.

on that day and 1159 UTC on the next day (Storm Prediction Center 2022), and higher. Thus, the 0000 UTC soundings correspond to the Day-1 outlook issued on the preceding day, whereas the 1200 UTC soundings correspond to the Day-1 outlook issued at 1200 UTC on the same day. This filtering retains 6082 (19.5%) of the 0000 UTC soundings and 5815 (18%) of the 1200 UTC soundings. After subsetting these data, the data transformation, compression, and clustering methods outlined in section 2 for the full dataset are replicated for this subset, with a silhouette-score analysis again supporting retaining two clusters for both the raw and transformed data at both analysis times (not shown).

As with the full sounding dataset, the cluster-mean vertical profiles for the clusters derived from raw 0000 UTC sounding data have similar shapes, but with substantially different cluster-mean surface temperatures and dewpoints (Figs. 10a,c). The warmer, moister Cluster 1 has a mean surface temperature of 28°C and a mean surface dewpoint of 18°C, whereas the colder, drier Cluster 2 has a mean surface temperature of 17°C and a mean surface dewpoint of 4°C. Further, Cluster-1 soundings are preferentially located in the southeastern United States (Fig. 11a) and preferentially occur during the warm season (Fig. 12, top panel), whereas Cluster-2 soundings are preferentially located in the northwestern United States (Fig. 11a) and preferentially occur during the cold season (Fig. 12, top panel). Altogether, this suggests that clusters derived from raw sounding data are again primarily latitudinally and seasonally stratified despite the reduced variability within the input data.

Conversely, the transformed 0000 UTC data indicate the same distinct boundary layer structures as for the full 0000 UTC transformed dataset (cf. Figs. 6b,d and 10b,d), with Clusters 1

and 2 depicting a vertically mixed layer and nearly pseudo-adiabatic profile, respectively. The geographic and temporal distributions of soundings within these clusters are largely unchanged from the full 0000 UTC transformed dataset, with Cluster-1 soundings most prevalent in the U.S. Intermountain West (Fig. 11b) and the warm season (Fig. 12, bottom panel) and Cluster-2 soundings most prevalent near coastlines and in the eastern United States (Fig. 11b) and less preferentially occurring during the warm season (Fig. 12, bottom panel).

The correspondence of the cluster-mean profiles for the raw and transformed 0000 UTC soundings in thunderstorm-supporting regions to those for the full data (section 3a) result from a large overlap in the cluster populations. For soundings in the full and thunderstorm-supporting environment sets, 73% of the raw data and 61% of the transformed data have the same cluster assignments. Altogether, even when soundings are manually subset over a narrower latitudinal and seasonal range, the sounding-transformation process appears to reduce the latitudinal and seasonal stratification within clusters derived from these data, allowing the clustering process to better identify distinct boundary layer structures within the data.

As is true for clusters derived from the raw 0000 UTC sounding data in thunderstorm-supporting environments, clusters derived from raw 1200 UTC sounding data in thunderstorm-supporting environments are largely distinguished by their cluster-mean surface temperatures and dewpoints (Figs. 13a,c). Specifically, the cluster-mean surface temperature and dewpoint for Cluster 1 are 22° and 18°C, respectively, whereas the cluster-mean surface temperature and dewpoint for Cluster 2 are 10° and 7°C, respectively. In fact, 91% of the soundings have the

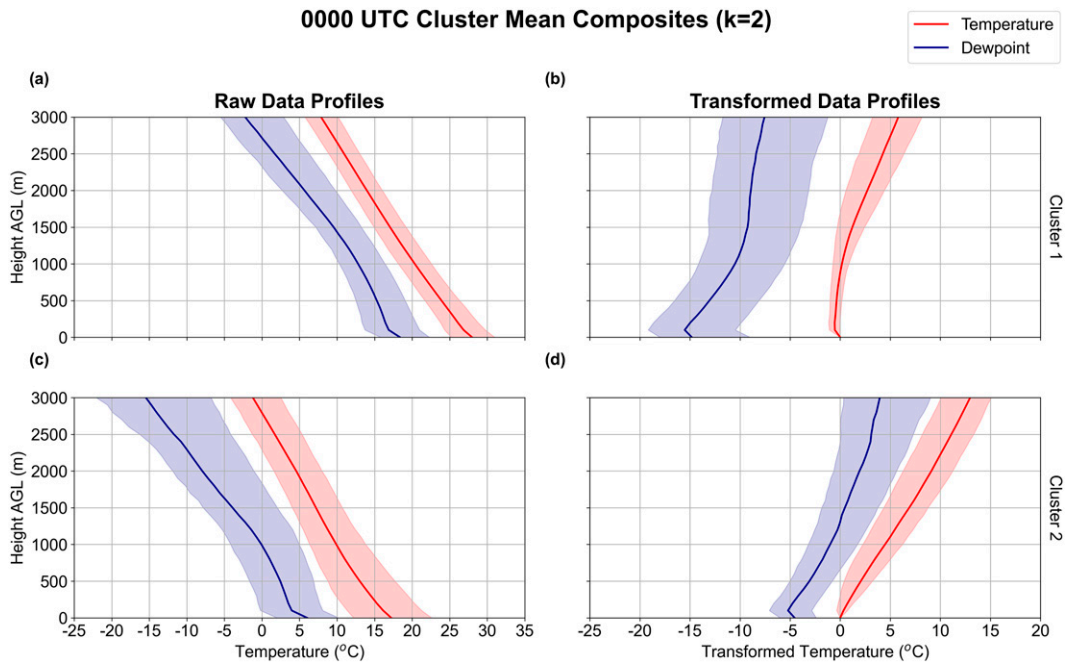


FIG. 10. As in Fig. 6, but only for vertical thermodynamic profiles contained within thunderstorm-supporting environments, as assessed by whether they are located within the Storm Prediction Center’s 1200 UTC Day-1 Convective Outlook valid for the time at which each sounding is observed.

same cluster assignment as the following (i.e., 12 h later) 0000 UTC data, emphasizing the degree to which the two cluster populations overlap. Thus, as at 0000 UTC, these differences primarily result from latitudinal and temporal variability in the cluster populations: Cluster-1 soundings are preferentially located in the southeastern United States (Fig. 11d) and primarily occur in summer (Fig. 12, bottom panel) whereas Cluster-2 soundings are preferentially located in the northwestern

United States (Fig. 11d) and primarily occur during the cold season (Fig. 12, bottom panel).

Likewise, as is true for clusters derived from the full 1200 UTC transformed soundings, the sounding-transformation process is less effective in reducing the latitudinal and seasonal stratification evident in the raw data in thunderstorm-supporting environments as it is for 0000 UTC soundings. Composite-mean transformed temperature and dewpoint profiles are similar between Clusters 1

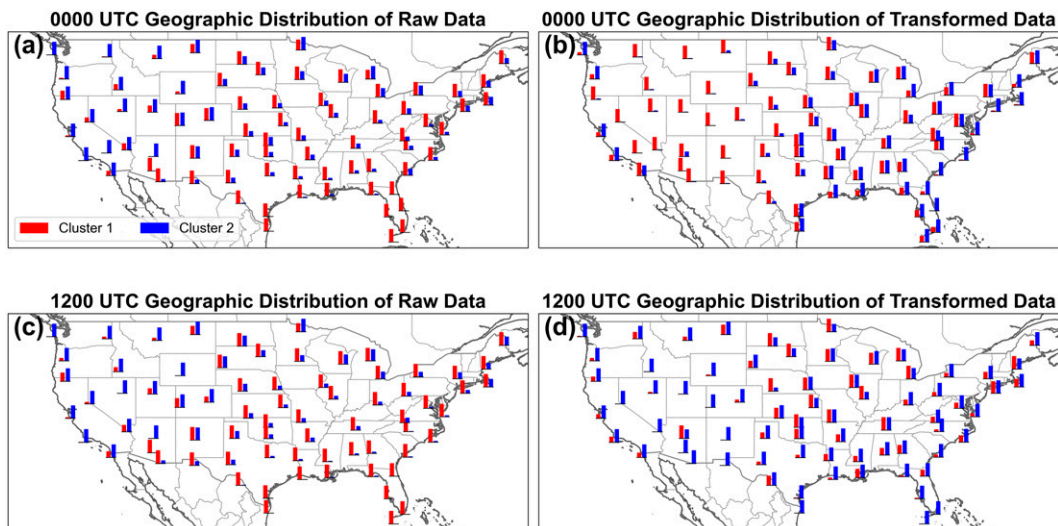


FIG. 11. As in Fig. 7, but only for vertical thermodynamic profiles contained within thunderstorm-supporting environments determined from SPC general thunderstorm areas (forecasts of which only cover the conterminous United States).

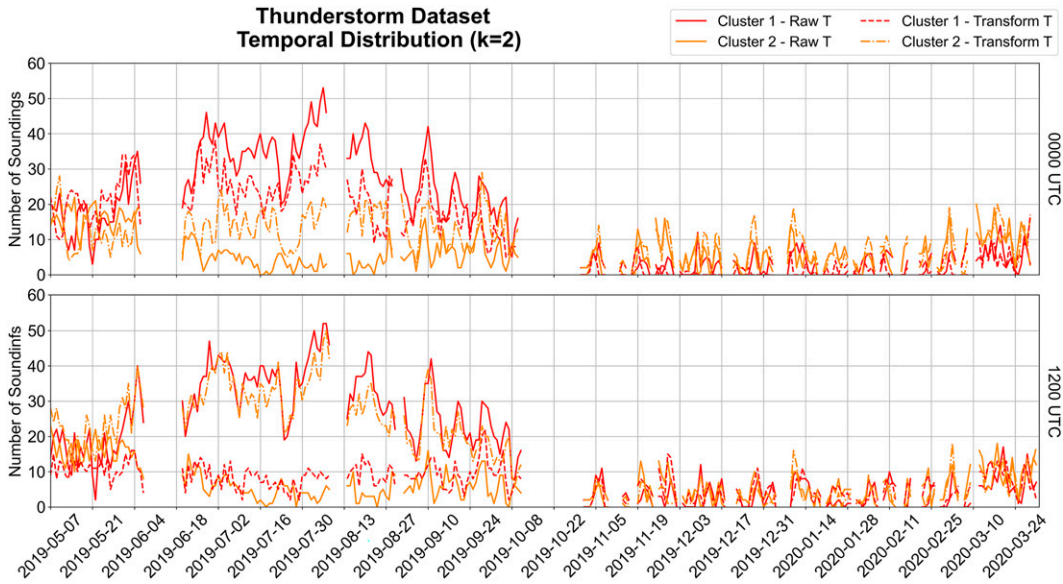


FIG. 12. As in Fig. 8, but only for vertical thermodynamic profiles contained within thunderstorm-supporting environments.

and 2 (Figs. 13b,d), with Cluster 1 having slightly greater stability for parcels lifted from the lowest 1 km AGL as compared to Cluster 2. Cluster-1 soundings preferentially are found in the north-central United States (Fig. 11d) during the cold season (Fig. 12, bottom panel), whereas Cluster-2 soundings preferentially are found at lower latitudes (Fig. 11) during

the warm season (Fig. 12, bottom panel). Because of the reduced effectiveness in reducing the latitudinal and seasonal stratification as compared to the 0000 UTC data, only 45% of 1200 UTC soundings in thunderstorm-supporting environments have the same cluster assignment as those at the same observing location taken 12 h later at 0000 UTC (not shown).

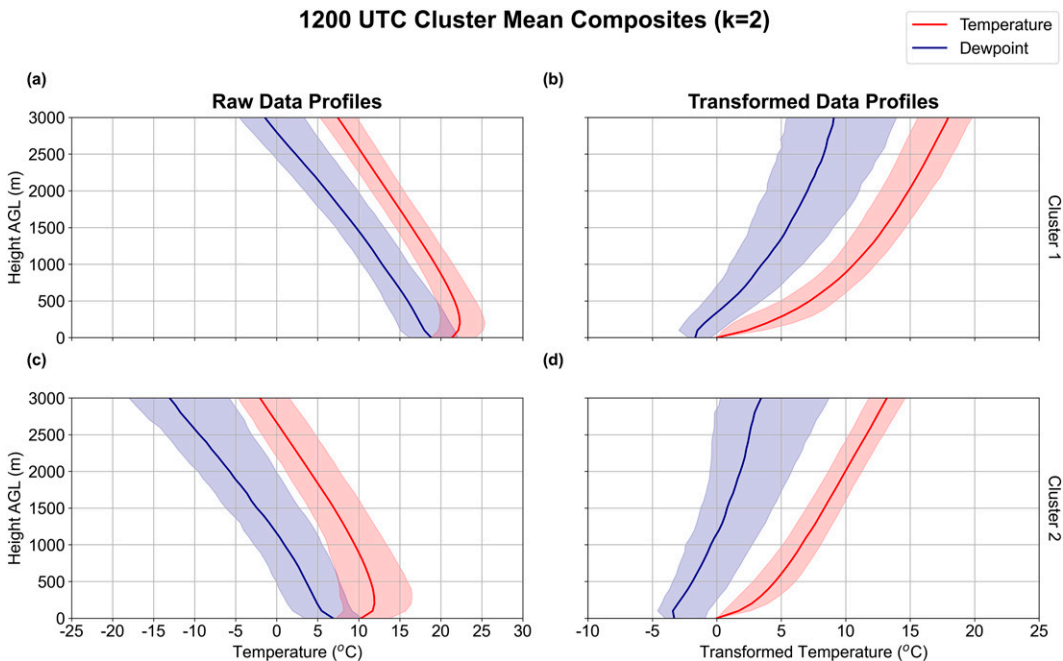


FIG. 13. As in Fig. 10, but only for vertical thermodynamic profiles contained within thunderstorm-supporting environments determined from SPC general thunderstorm areas (forecasts of which only cover the conterminous United States).

As with the 0000 UTC data, the correspondence of the cluster-mean profiles for the raw and transformed 1200 UTC soundings in thunderstorm-supporting regions to those for the full data (section 3a) result from a large overlap in cluster populations. For soundings in the full and thunderstorm-supporting environment sets, 73% of the raw data and 72% of the transformed data have the same cluster assignments.

4. Application to model verification

Model-analyzed and forecast vertical thermodynamic profiles in the boundary layer are typically biased due to the imperfect approximations used in their turbulence parameterizations (e.g., Bright and Mullen 2002; Burlingame et al. 2017; Cohen et al. 2015, 2017; Coniglio et al. 2013; Evans et al. 2018; Hu et al. 2010; Stensrud and Weiss 2002). Such biases are not constant across environments, however. For example, internal SPC evaluations of pre-implementation GFS releases have long indicated that the model *overparameterizes* turbulent vertical mixing in unstable warm-season, thunderstorm-supporting environments, particularly near drylines in the central United States (not shown). However, most studies use subjective data stratifications—such as by geography, surface-based instability magnitude, and/or the presence of a capping inversion (e.g., Coniglio et al. 2013; Evans et al. 2018; Nevius and Evans 2018)—to document biases in model-analyzed and forecast vertical thermodynamic profiles in the boundary layer.

Here, we use the full-dataset clusters from section 3a for the raw and transformed data to verify short-range Global Forecast System (GFS) version 15.1 model forecast soundings for the May–November 2019 period. In GFS version 15.1, released in 2019 (Maxson 2019) and superseded by GFS version 16 in March 2021 (Farrar 2021), turbulent vertical mixing is parameterized using a hybrid eddy-diffusivity (ED), countergradient (CG), and mass-flux (MF) approach (Han et al. 2016). The ED method, which applies to stable conditions, parameterizes turbulent mixing locally (i.e., only between adjacent vertical levels). The CG method, which applies to weakly unstable conditions, mimics nonlocal vertical transport by large eddies through a parameterized countergradient transport from low to high values. The MF method, which applies to strongly unstable conditions, mixes nonlocally by mathematically relating turbulent mixing to the vertical transport accomplished by entraining surface thermals. Given this stability-dependent formulation for parameterizing turbulent vertical mixing, we hypothesize that cluster-mean verification statistics (e.g., bias for temperature and dewpoint forecasts) for the transformed sounding data—which largely stratify by meteorological phenomena—will better elucidate environment-specific model biases than will verification statistics for the raw sounding data—which largely stratify by latitude and the time of year.

The verification statistic considered herein is bias, defined as model minus observation (taken to approximate truth) and averaged over each cluster (here representing those derived from the full rather than thunderstorm-supporting environment dataset). Bias is computed for four forecast hours, 0, 12, 24, and 36 h, separately for 0000 and 1200 UTC observations.

Cluster-mean results are presented for both the raw and transformed data to further demonstrate the utility of the sounding-transformation, dimension-reduction, and clustering process.

The cluster-mean bias curves for the transformed sounding data have more pronounced shape differences than do those for the raw sounding data. The 0000 and 1200 UTC–verifying Cluster-1 and Cluster-2 mean temperature and dewpoint biases have similar shapes, albeit with different magnitudes, when considering the raw sounding data except for the 1200 UTC cluster-mean dewpoint biases (Figs. 14 and 15a–d). For example, the 0000 UTC–verifying Cluster-1 and Cluster-2 mean dewpoint bias profiles exhibit an increasingly large moist bias with altitude, with the Cluster-1 mean dewpoint bias being 0.5°–1°C larger than that with Cluster 2 (Figs. 14a–d). Likewise, the 1200 UTC–verifying Cluster-1 and Cluster-2 mean temperature bias profiles are both cold-biased throughout the column, with the Cluster-1 mean temperature bias being 0.5°C warmer than with Cluster 2 (Figs. 15a–d).

Conversely, the 0000 and 1200 UTC–verifying Cluster-1 and Cluster-2 mean temperature and dewpoint bias profiles have different shapes when considering the transformed sounding data, with reduced overlap between 25th and 75th percentile temperature and dewpoint values for each cluster (shading in Figs. 14 and 15) as compared to the raw sounding data. This is most notable for cluster-mean dewpoint, with the 0000 UTC–verifying Cluster 1 and 1200 UTC–verifying Cluster 1 exhibiting increasingly large moist biases with altitude and forecast lead time that are not shared by the 0000 UTC–verifying Cluster 2 and 1200 UTC–verifying Cluster 2 (Figs. 14 and 15e–h). The increasingly large moist bias with altitude likely results from the model’s inability to accurately represent the altitude and sharpness of the strong inversion that often occurs atop the boundary layer in the semiarid environments that dominate these clusters’ populations (e.g., Evans et al. 2018). Different bias-profile shapes are also seen with the cluster-mean temperature bias profiles below 1 km AGL for 0000 UTC–verifying forecasts, with Cluster-1 cold biased in the mean by up to 1°C and Cluster-2 near-zero biased in the mean over this layer (solid curves in Figs. 14e–h).

Altogether, these limited results suggest that the sounding transformation and clustering method introduced in this study has promise for facilitating environment-specific forecast verification and subsequent model development, testing, and evaluation activities.

5. Conclusions

This study introduces a novel method to transform, dimensionally reduce, and cluster observed soundings to aid in objectively identifying boundary layer sounding structures. This method involves first interpolating soundings at a common analysis time (0000 or 1200 UTC) to a uniform height AGL grid, then subtracting the dry adiabat that extends upward from the surface parcel’s air temperature from both the temperature and the dewpoint at all altitudes. Both raw (or untransformed) and transformed soundings are clustered using *k*-means clustering in the phase space of their respective

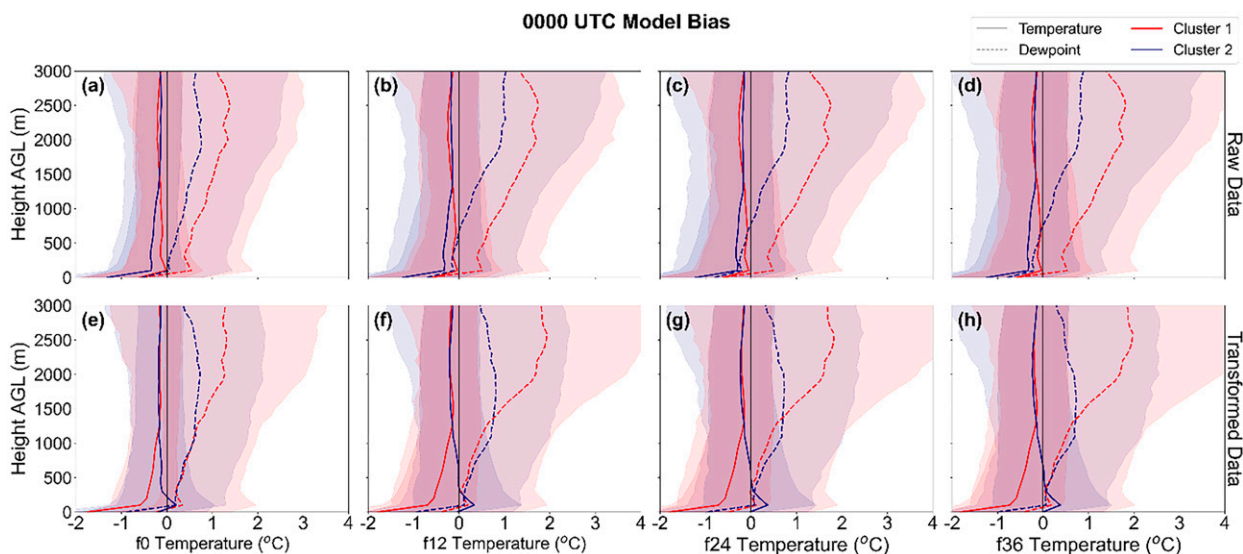


FIG. 14. Cluster-mean bias, where bias is defined as model minus observations, for GFS (a),(e) 0-; (b),(f) 12-; (c),(g) 24-; and (d),(h) 36-h forecasts valid at 0000 UTC using the full sounding dataset. (top) The results for the raw sounding data and (bottom) the results for the transformed sounding data. Cluster 1 is depicted in red and Cluster 2 is depicted in blue, with temperature depicted in solid lines and dewpoint depicted in dashed lines, for all panels. Semitransparent shading indicates the values between the 25th and 75th percentile of each cluster's respective distributions.

leading EOFs, with two clusters retained for each data at both analysis times based on a silhouette-score analysis.

Transforming the soundings prior to clustering them allows for the resulting clusters to represent distinct boundary layer structures instead of climatological airmass characteristics as is seen in the clusters derived from the non-transformed data. Specifically, 0000 and 1200 UTC cluster-mean profiles derived from the raw data are distinguished primarily by differences in the temperature and dewpoint profiles, with Cluster-1 profiles preferentially occurring during the warm season (Figs. 5a,b and

8) and being substantially warmer and moister than Cluster 2 (Figs. 6 and 9). Conversely, 0000 UTC cluster-mean profiles derived from transformed sounding data better represent common boundary layer structures, such as vertically mixed layers, and exhibit greater geographical and temporal variability (Figs. 6–8 and 10–12). It is less effective at reducing the latitudinal and seasonal stratifications within the raw data at 1200 UTC however, although the reasons as to why are unclear and warrant further research. The transformation method's ability to distinguish boundary layer structures in a sounding climatology is

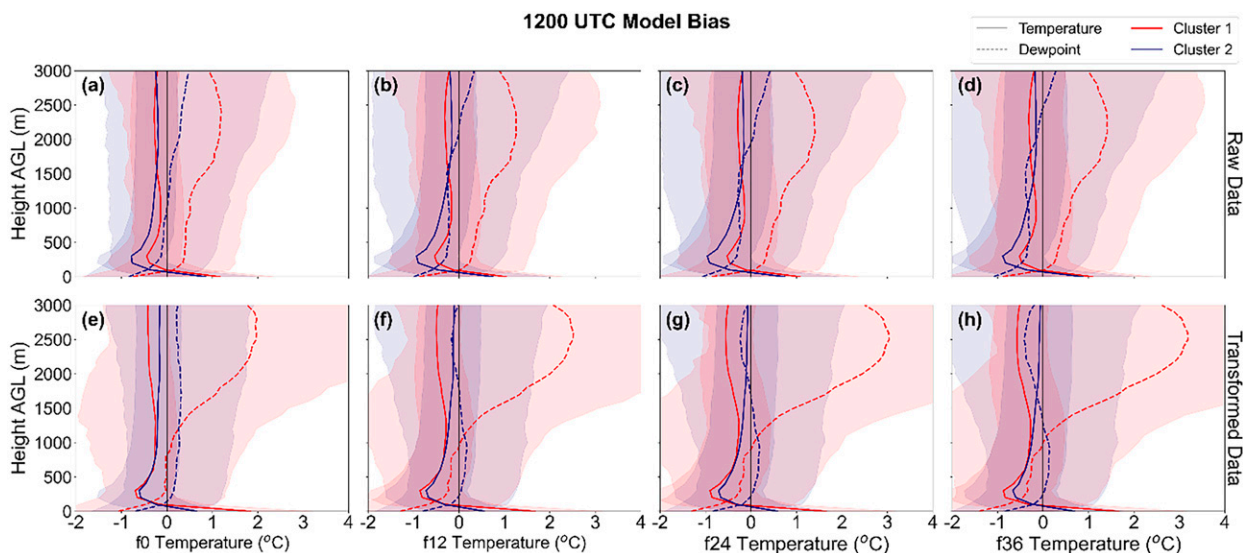


FIG. 15. As in Fig. 14, but using 1200 UTC observed soundings.

only slightly reduced when the variability in the input data is reduced, such as is done herein to isolate soundings in thunderstorm-supporting environments before clustering. The transformation method's efficacy allows for the derived clusters to be used in applications ranging from model verification, wherein model biases are often variable across meteorological environments, to validating remote sensing instrument retrieval algorithms (e.g., temperature retrievals from microwave radiometers).

Although the method introduced here shows promise for isolating environmental variability within large sounding datasets, there are nevertheless several limitations that must be kept in mind. First, the sounding data considered in this study only cover the 11 months from early May 2019 to late March 2020 over North America. It is possible that the method's efficacy or its outputs (e.g., the optimal number of EOFs or clusters to retain) would be different if a larger, more variable input dataset is used. Second, the method is limited by assumptions inherent to its compression and clustering algorithms, most importantly the Euclidean distance formulation of the silhouette score. Other compression or clustering approaches may produce different results. Third, choices such as limiting our analysis to the lowest 3 km AGL (instead of considering the entire troposphere) or defining thunderstorm-supporting environments based on SPC convective outlooks (which cover a 24-h period, such that a sounding may not truly be in a thunderstorm-supporting environment at both of the two times considered) are somewhat arbitrary. Here, too, other approaches may produce different results. Additional research is necessary to evaluate these limitations and evaluate the method's efficacy for applications beyond sounding classification and forecast verification.

Acknowledgments. This research was sponsored by the NOAA Testbeds program under Award NA18NWS4680062. Fruitful discussions with Gretchen Mullendore and Caitlyn Mensch are greatly appreciated. Constructive feedback from three anonymous reviewers and *Weather and Forecasting* Chief Editor Gary Lackmann helped to improve the manuscript.

Data availability statement. The observed and model soundings used in this study, as well as Python code to transform the height coordinate and temperature formulation in both datasets, can be obtained from a Zenodo repository at <https://zenodo.org/record/7097496>.

REFERENCES

- Blue Hill Observatory and Science Center, 2021: A brief history of the Blue Hill Meteorological Observatory. Accessed 10 July 2023, <https://bluehill.org/about/>.
- Bright, D. R., and S. L. Mullen, 2002: The sensitivity of the numerical simulation of the southwest monsoon boundary layer to the choice of PBL turbulence parameterization in MM5. *Wea. Forecasting*, **17**, 99–114, [https://doi.org/10.1175/1520-0434\(2002\)017<0099:TSOTNS>2.0.CO;2](https://doi.org/10.1175/1520-0434(2002)017<0099:TSOTNS>2.0.CO;2).
- Burlingame, B. M., C. Evans, and P. J. Roebber, 2017: The influence of PBL parameterization on the practical predictability of convection initiation during the Mesoscale Predictability Experiment (MPEX). *Wea. Forecasting*, **32**, 1161–1183, <https://doi.org/10.1175/WAF-D-16-0174.1>.
- Cohen, A. E., S. M. Cavallo, M. C. Coniglio, and H. E. Brooks, 2015: A review of planetary boundary layer parameterization schemes and their sensitivity in simulating southeastern U.S. cold season severe weather events. *Wea. Forecasting*, **30**, 591–612, <https://doi.org/10.1175/WAF-D-14-00105.1>.
- , —, —, —, and I. L. Jirak, 2017: Evaluation of multiple planetary boundary layer parameterization schemes in southeast U.S. cold season severe thunderstorm environments. *Wea. Forecasting*, **32**, 1857–1884, <https://doi.org/10.1175/WAF-D-16-0193.1>.
- Coniglio, M. C., J. Correia, P. T. Marsh, and F. Kong, 2013: Verification of convection-allowing WRF Model forecasts of the planetary boundary layer using sounding observations. *Wea. Forecasting*, **28**, 842–862, <https://doi.org/10.1175/WAF-D-12-00103.1>.
- Davis, R. E., 1976: Predictability of sea surface temperature and sea level pressure anomalies over the North Pacific Ocean. *J. Phys. Oceanogr.*, **6**, 249–266, [https://doi.org/10.1175/1520-0485\(1976\)006<0249:POSSTA>2.0.CO;2](https://doi.org/10.1175/1520-0485(1976)006<0249:POSSTA>2.0.CO;2).
- Evans, C., S. J. Weiss, I. L. Jirak, A. R. Dean, and D. S. Nevius, 2018: An evaluation of paired regional/convection-allowing forecast vertical thermodynamic profiles in warm-season, thunderstorm-supporting environments. *Wea. Forecasting*, **33**, 1547–1566, <https://doi.org/10.1175/WAF-D-18-0124.1>.
- Farrar, M., 2021: Service change notice 21-20 (updated). National Weather Service Headquarters, Silver Spring, MD, 13 pp., https://www.weather.gov/media/notification/pdf2/scn21-20_gfsv16.0_aac.pdf.
- Forgy, E. W., 1965: Cluster analysis of multivariate data: Efficiency vs. interpretability of classifications. *Biometrics*, **21**, 768–769.
- Fovell, R. G., and A. Gallagher, 2020: Boundary layer and surface verification of the High-Resolution Rapid Refresh, version 3. *Wea. Forecasting*, **35**, 2255–2278, <https://doi.org/10.1175/WAF-D-20-0101.1>.
- Giangrande, S. E., D. Wang, and D. B. Mechem, 2020: Cloud regimes over the Amazon basin: Perspectives from the GoAmazon2014/15 campaign. *Atmos. Chem. Phys.*, **20**, 7489–7507, <https://doi.org/10.5194/acp-20-7489-2020>.
- Han, J., M. Witek, J. Teixeira, R. Sun, H.-L. Pan, J. K. Fletcher, and C. S. Bretherton, 2016: Implementation in the NCEP GFS of a Hybrid Eddy-Diffusivity Mass-Flux (EDMF) boundary layer parameterization with dissipative heating and modified stable boundary layer mixing. *Wea. Forecasting*, **31**, 341–352, <https://doi.org/10.1175/WAF-D-15-0053.1>.
- Hu, X.-M., J. W. Nielsen-Gammon, and F. Zhang, 2010: Evaluation of three planetary boundary layer schemes in the WRF Model. *J. Appl. Meteor. Climatol.*, **49**, 1831–1843, <https://doi.org/10.1175/2010JAMC2432.1>.
- Jensen, A. A., A. M. Thompson, and F. J. Schmidlin, 2012: Classification of Ascension Island and natal ozonesondes using self-organizing maps. *J. Geophys. Res.*, **117**, D04302, <https://doi.org/10.1029/2011JD016573>.
- Kohonen, T., 1995: *Self-Organizing Maps*. Springer Series in Information Sciences, Vol. 30, Springer-Verlag, 362 pp.
- Lloyd, S. P., 1982: Least squares quantization in PCM. *IEEE Trans. Info. Theory*, **28**, 129–136, <https://doi.org/10.1109/TIT.1982.1056489>.
- Lorenz, E. N., 1956: Empirical orthogonal functions and statistical weather prediction. Statistical Forecasting Project Scientific Rep. 1, MIT Department of Meteorology, 52 pp.,

- https://eapsweb.mit.edu/sites/default/files/Empirical_Orthogonal_Functions_1956.pdf.
- Maxson, B., 2019: Service change notice 19-40. National Weather Service Headquarters, Silver Spring, MD, 8 pp., https://www.weather.gov/media/notification/pdf2/scn19-40gfs_v15_1.pdf.
- Monahan, A. H., J. C. Fyfe, M. H. P. Ambaum, D. B. Stephenson, and G. R. North, 2009: Empirical orthogonal functions: The medium is the message. *J. Climate*, **22**, 6501–6514, <https://doi.org/10.1175/2009JCLI3062.1>.
- Nevius, D. S., and C. Evans, 2018: The influence of vertical advection discretization in the WRF-ARW Model on capping inversion representation in warm-season, thunderstorm-supporting environments. *Wea. Forecasting*, **33**, 1639–1660, <https://doi.org/10.1175/WAF-D-18-0103.1>.
- North, G. R., T. L. Bell, R. F. Cahalan, and F. J. Moeng, 1982: Sampling errors in the estimation of empirical orthogonal functions. *Mon. Wea. Rev.*, **110**, 699–706, [https://doi.org/10.1175/1520-0493\(1982\)110<0699:SEITEO>2.0.CO;2](https://doi.org/10.1175/1520-0493(1982)110<0699:SEITEO>2.0.CO;2).
- Nowotarski, C., and A. Jensen, 2013: Classifying proximity soundings with self-organizing maps toward improving supercell and tornado forecasting. *Wea. Forecasting*, **28**, 783–801, <https://doi.org/10.1175/WAF-D-12-00125.1>.
- Obukhov, A. M., 1947: Statistically homogeneous fields on a sphere. *Usp. Mat. Nauk*, **2**, 196–198.
- Pandas Development Team, 2023: pandas-dev/pandas: Pandas (v2.0.1). Zenodo, accessed 10 July 2023, <https://doi.org/10.5281/zenodo.7857418>.
- Rousseeuw, P. J., 1987: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.*, **20**, 53–65, [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
- Stensrud, D. J., and S. J. Weiss, 2002: Mesoscale model ensemble forecasts of the 3 May 1999 tornado outbreak. *Wea. Forecasting*, **17**, 526–543, [https://doi.org/10.1175/1520-0434\(2002\)017<0526:MMEFOT>2.0.CO;2](https://doi.org/10.1175/1520-0434(2002)017<0526:MMEFOT>2.0.CO;2).
- Stith, J. L., and Coauthors, 2018: 100 years of progress in atmospheric observing systems. *A Century of Progress in Atmospheric and Related Sciences: Celebrating the American Meteorological Society Centennial*, Meteor. Monogr., No. 59, Amer. Meteor. Soc., <https://doi.org/10.1175/AMSMONOGRAPHS-D-18-0006.1>.
- Storm Prediction Center, 2022: SPC products. Accessed 10 July 2023, <https://www.spc.noaa.gov/misc/about.html>.
- Wilks, D. S., 2019: *Statistical Methods in the Atmospheric Science*. 4th ed. Elsevier, 840 pp.