VERIFICATION OF ENVIRONMENTAL-REGIME–STRATIFIED GFS SHORT-RANGE VERTICAL SOUNDING FORECASTS

by

Dillon Blount

A Thesis Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Master of Science

in Atmospheric Science

at

The University of Wisconsin-Milwaukee

August 2021

ABSTRACT

VERIFICATION OF ENVIRONMENTAL-REGIME–STRATIFIED GFS SHORT-RANGE VERTICAL SOUNDING FORECASTS

by

Dillon Blount

The University of Wisconsin-Milwaukee, 2021
Under the Supervision of Professor Clark Evans

In recent years, the United States' operational global numerical weather prediction model, the Global Forecast System (GFS), has been upgraded to include a new dynamical core and an updated turbulence parameterization. This updated turbulence parameterization uses a hybrid eddy-diffusivity, countergradient, and mass-flux formulation to approximate near-surface turbulent vertical mixing. The precise formulation used is based on the local stability, with the eddy-diffusivity, countergradient, and mass flux formulations used under stable, weakly unstable, and strongly unstable conditions, respectively.

In this study, an objective classification of environmental regimes is used to verify GFS short-range vertical soundings, focusing on the planetary boundary layer where the turbulence parameterization plays an important role in determining vertical mixing, thus sounding characteristics. Observed temperature and dewpoint temperature from 15,488 soundings taken at 0000 UTC and 16,118 soundings taken at 1200 UTC between May – November 2019 are first interpolated into a height above-ground-level (AGL) coordinate and normalized to the pseudoadiabat defined by the surface-based parcel's wet-bulb temperature. This allows for sounding shapes to be classified together regardless of their temperature and dewpoint temperature

differences due to altitude, geography, or seasonality. A multivariate empirical orthogonal function (EOF) analysis is then performed on the normalized sounding data, after which a *k*-means clustering analysis is conducted on the leading two principal components retained from the EOF analysis. The output of this analysis classifies soundings into three different environmental regimes at each time, leading into a regime-specific model sounding verification. The classification method identifies distinct environments at each time, including deeply mixed layers (strongly unstable), shallow mixed layers (weakly/moderately unstable), and radiation inversions (stable), and each profile has varying biases due in part to turbulent mixing issues within the boundary layer.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank Dr. Clark Evans for his guidance throughout this project. Under his teaching, I was able to grow as a student and researcher while using the concepts and skills learned during this study.

Thank you to Dr. Kahl and Dr. Kravtsov for serving on my thesis committee.

Thank you to my family and friends for their support. I would not have been able to achieve this accomplishment without their unwavering support.

1. Introduction

The history of vertical soundings, which provide meteorological observations through parts of the 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 2021). This began regular atmospheric soundings of temperature, dewpoint temperature (hereafter simply dewpoint), pressure, and 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 and balloons had to be recovered to obtain the recorded data. These pitfalls led to the development of radio-transmitted instrument packages called radiosondes. Radiosondes are attached to balloons, observe temperature, dewpoint, and pressure data in the lower atmosphere, and radio these data back to a remote receiving station. When these packages can also record and transmit horizontal wind data, they 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 at two times per day, 0000 and 1200 UTC.

The United States is one of the best coverage areas for upper-air soundings in the world (Fig. 1), facilitating routine sampling of many different environmental regimes across the country. These environmental regimes are defined by the atmospheric conditions that lead to distinct meteorological phenomena. For example, clear skies, cool temperatures, and calm winds at night can lead to the formation of a near-surface radiation inversion (Fig. 2). Ongoing precipitation and moist, tropical air masses are typically associated with a nearly saturated vertical profile. Hot and

dry environmental conditions, due to strong surface sensible heating, lead to the formation of deep turbulent eddies and associated vertically mixed layer within the planetary boundary layer (PBL; Fig. 3). These and other environmental regimes are associated with distinct temperature and dewpoint profile sounding shapes. These profiles can be depicted by rawinsonde observations as well as by numerical weather prediction model analyses and forecasts. However, coarse model vertical resolution, imperfect model parameterizations of sub-grid-scale processes such as radiation and turbulence, and numerical approximations used in discretizing the equations governing atmospheric motion and heat transfer, model-analyzed and model-forecast vertical soundings can be different from their observational counterparts.

Model verification across multiple environmental regimes is needed to better quantify the effects associated with model parameterizations, including those of the PBL, which often exhibit biases that are reflective of inaccuracies in representing turbulent vertical mixing. There are two main approaches to parameterizing turbulent vertical mixing, local and non-local mixing. Local mixing is the mixing of meteorological variables between adjacent vertical grid boxes, whereas non-local mixing is the mixing of variables of non-adjacent grid boxes (Stensrud 2007). There is substantial variability in PBL parameterization biases between non-local mixing approaches, however (e.g., Han and Bretherton 2019), with modern parameterizations emphasizing mass-flux over countergradient methods (each described in more detail below) due to the former's more physically consistent formulation (Coniglio et al. 2013). A recent study analyzed the upgraded Global Forecast System version 15.1 (GFS), the National Oceanic and Atmospheric Administration’s operational global numerical weather prediction model (NOAA 2021), and showed that the inclusion of mass-flux formulation in the PBL parameterization improves the growth of the convective boundary layer (Han et al. 2016). Other studies have further identified

distinct model errors within different environmental conditions with both local and non-local turbulent vertical mixing formulations (e.g., Krogsaeter and Reuder 2014, Clark et al. 2015, Cohen et al. 2017, Burlingame et al. 2017, Evans et al. 2018, Nevius and Evans 2018).

In GFS version 15.1, released in 2019 (Maxson 2019), the PBL parameterization uses a hybrid eddy-diffusivity (ED), countergradient (CG), and mass-flux (MF) formulation to represent turbulent vertical mixing (Han et al. 2016). The ED formulation uses a local mixing approach in which turbulent vertical mixing only occurs between adjacent vertical levels. Conversely, the CG formulation mimics non-local vertical transport by large eddies between nonadjacent grid boxes through a parameterized transport from low values to high values, such as the downward transport of high potential-temperature air. Likewise, the MF formulation mixes non-locally, except by mathematically relating this mixing to the vertical transport accomplished by entraining surface thermals. The approach used at a given location and forecast time is based on the PBL stability, with the ED approach used for stable conditions, the CG approach used in weakly unstable environments, and the MF approach used in strongly unstable environments (Han et al. 2016). This PBL formulation replaced the eddy-diffusivity/countergradient (EDCG) PBL parameterization in the GFS version 14, which routinely underestimated PBL growth in strongly unstable environments due to the CG approach insufficiently representing turbulent vertical mixing under these conditions (Han et al. 2016). However, NOAA's Storm Prediction Center (SPC) notes that the EDCG parameterization *overparameterizes* turbulent vertical mixing in warm-season, thunderstorm-supporting environments (which represent a specific type of strongly unstable environments), particularly near drylines in the central United States (not shown). This discrepancy, and more broadly the distinct forecast errors identified by SPC and Han et al (2016), largely motivate this study.

Model verification approaches typically rely on subjective classifications to determine model skill and performance characteristics. Examples include classifications by geographic location (Evans et al. 2018), surface-based instability magnitude (Coniglio et al. 2013, Evans et al. 2018), and presence of a capping inversion (Coniglio et al. 2013, Nevius and Evans 2018). However, previous studies have shown that quasi-objective statistical clustering techniques are able to reliably identify environmental regimes. One method of clustering vertical meteorological profiles is self-organizing maps (SOM; Kohonen 1995). For example, SOMs have been used to cluster vertical soundings taken in proximity to warm-season thunderstorms, with each cluster uniquely associated with a distinct storm type (e.g., supercellular vs. non-supercellular; Nowotarski and Jensen 2013). SOM clustering have also been used to classify ozone mixing ratio profiles (Jensen et al. 2012). A variety of other meteorological phenomena have utilized SOMs to classify data, including near-storm tornadic environments (Anderson-Frey et al. 2017) and a synoptic climatology of geopotential height, sea-level pressure, and extreme climate events (Liu and Weisberg 2011). A weakness of the SOM method is that it requires the user to specify the number of nodes for the dataset; however, as with other objective clustering techniques, it is possible to objectively determine this number by iterating over multiple possible node counts and choosing that which maximizes intercluster variance and minimizes intracluster variance (e.g., Nowotarski and Jensen 2013).

Another method that can be used to cluster meteorological vertical profiles is *k*-means clustering (Forgy 1965, Lloyd 1982). For example, *k*-means clustering has been used to identify distinct Amazonian meteorological regimes, including three dry regimes and two moist regimes representing the transition from shallow cumulus to deep cumulus clouds, from routine radiosonde observations (Giangrande et al. 2020). A variety of other meteorological phenomena have also

been classified using this technique, including the Madden-Julian (e.g., Wheeler and Hendon 2004) and Arctic (e.g., Thomson and Wallace 1998) Oscillations, tropical mesoscale convective systems (Pope et al. 2009), solar irradiation (Thanh Nga et al. 2021), and the cyclone-phase-space (Hart 2003) pathways of tropical cyclones during extratropical transition (Arnott et al. 2004). One possible weakness with *k*-means clustering is the need to predetermine the number of clusters into which to parse the data, but as with SOMs there are objective methods to iteratively determine the appropriate number of clusters through maximizing the intercluster variance and minimizing the intracluster variance (Rousseeuw 1987).

This study presents an environmental-regime-specific verification of GFS v15.1 short-range forecast near-surface vertical thermodynamic profiles, wherein environmental regimes are identified using *k*-means clustering applied to a dimensionally reduced dataset obtained from a multivariate empirical orthogonal function (EOF) analysis. A unique aspect of this study is its novel method for representing vertical thermodynamic profiles by their shapes, which allows for unique sounding features (e.g., radiation and frontal inversions, deep mixed layers, etc.) to be identified independent of the time of year and geographic location. The hypothesis guiding this study is that the GFS version 15.1 does not have consistent forecast skill in predicting near-surface vertical thermodynamic profiles across different environmental regimes because the methods used to parameterize turbulence under varying stability conditions perform differently because of the fundamental differences in the physical processes being parameterized by each.

The rest of this study is structured as follows. Section 2 outlines the data presented before discussing the sounding transformation, data compression, and clustering method used to classify soundings. Section 3 presents the results from the environmental-regime–specific model

verification. A discussion of these results is presented in section 4. Finally, concluding remarks and potential uses of the objective sounding classification method are presented in section 5.

2. Data and Methods

The data used in this study are observed and GFS v15.1 short-range model-forecast vertical temperature and dewpoint profiles from routine upper-air observation stations across the United States, Canada, and Mexico. This dataset covers the period from 1200 UTC 7 May – 0000 UTC 3 November 2019, with two notable gaps between 0000 UTC 8 June – 1200 UTC 20 June and 0000 UTC 12 October – 0000 UTC 1 November, and includes both observed soundings taken and output from model forecasts conducted at 0000 and 1200 UTC. There are 15,489 observed soundings at 0000 UTC and 16,118 observed soundings at 1200 UTC. Model verification is conducted at four forecast lead times: 0, 12, 24, and 36 h. Although observational soundings are typically released approximately one hour before 0000 and 1200 UTC, such that near-surface conditions depicted by these data are more reliably attributed to 2300 and 1100 UTC, respectively, the archived data from the SPC used in this study do not include the corresponding forecast times of 11 h, 23 h, and 35 h.

Both observed and model soundings are preprocessed to a new height grid and temperature formulation. First, soundings are interpolated to a common above ground level (AGL) vertical grid with uniform vertical grid spacing of 100 m. This allows for each sounding to be analyzed on a common vertical grid. This addresses the differences in altitudes between stations (e.g., Fig. 1 of Fovell and Gallagher 2020) and inconsistent vertical grids between the observed and modeled soundings. The vertical extent of the new grid is restricted to 3 km AGL to focus on the PBL, although we acknowledge that some PBLs may extend above 3 km AGL in arid regions during the warm season. Next, each sounding's temperature and dewpoint vertical profiles are subtracted at

all vertical levels by the temperature value of the pseudoadiabat ascending upward from the surface-based parcel's wet-bulb temperature. These new temperatures and dewpoints represent what are referred to as the normalized temperature and dewpoint, and this transformation is depicted in Fig. 4. This retains fundamental sounding shapes while reducing the large profile variability between soundings that result from altitude, geographic, and seasonal differences between soundings.

The normalized temperature and dewpoint data are then used to identify environmental regimes within the data. First, the normalized *observed* temperature and dewpoint data (separately at 0000 UTC and 1200 UTC given the influence of the diurnal cycle on near-surface conditions) are input into a multivariate empirical orthogonal function (EOF) analysis (Obukhov 1947, Lorenz 1956, Davis 1976). The input vector, $\boldsymbol{x}$, is used to calculate $\boldsymbol{u}$, the principal components (PCs) of the data in which the variance is contained. The PCs are calculated using the anomalies to the temporal mean within the data, $\boldsymbol{x'}$, because these will allow for a variance consideration due to the linear combinations representing the maximum variability. Of the resulting PCs, u_1 is the first PC and contains the largest amount of variance within the data, u_2 is the second PC and contains the second-largest amount of variance, and so on (Wilks 2019). For the data considered in this study, the two leading PCs are well-separated from the third and subsequent PCs at each analysis time, as given by the substantial reduction in variance explained between the second and third PCs (Table 1), and thus are retained as input to a *k*-means clustering analysis.

Finally, the dimensionally reduced data are subjected to *k*-means clustering. *K*-means clustering is a nonhierarchical method for grouping data, wherein data can be reassigned between clusters as the analysis is performed. The *k*-means method clusters data in a user-specified number of clusters. In this method, an initial clustering is formed based on input points' distances from

randomly assigned initial points. The algorithm then computes each cluster's centroids, calculates the Euclidean distance of each data point from the different cluster centroids, and assigns the data to the cluster with the smallest distance between the data point and its centroid. This process is iterated until the distance from each data point is minimized to its respective cluster centroid (Wilks 2019). The number of clusters is set to three (k = 3) at both 0000 UTC and 1200 UTC, as this is the number of clusters (in the range of 2-10) that produces the highest cluster-average silhouette scores (Rousseeuw 1987), indicating the efficiency of the clusters in maximizing the intracluster variance and minimizing the intercluster variance (Figs. 5-6, top panels). Although the number of data points with negative silhouette scores, indicating poor matches to their assigned cluster, is relatively high for k = 3 at both times (Figs. 5-6, bottom panels), it is not inordinately high compared to the other cluster counts considered. The clustered sounding data represent distinct sounding profiles between clusters, as is demonstrated in Section 3.

Model performance is characterized using bias, with the observed soundings taken to approximate truth. Biases are computed separately at four forecast hours (0 h, 12 h, 24 h, 36 h) for each sounding, then the mean bias across all soundings in a cluster at the specified observation time and forecast hour is computed.

3. Results

3.1 0000 UTC Results

The *k*-means clustering analysis results in three distinct sounding profiles for the 0000 UTC data. Cluster 1 is the largest of the three clusters and contains 8,780 soundings (or 56.7% of the total). The soundings are preferentially located in the southeastern and eastern United States and southern Alaska, with the fewest soundings located in the desert southwest (Fig. 7). The cluster-

mean profile is characterized by a shallow mixed layer extending to approximately 1.5 km AGL (Fig. 8a), as shown by the temperature approaching the zero line (indicating a lapse rate greater than the pseudoadiabat). In the mean normalized profile above 1.5 km AGL, the temperature profile becomes nearly vertical whereas the dewpoint profile becomes more negative. The variation of the cluster-mean normalized dewpoint profile grows larger with altitude whereas the cluster-mean temperature variation does not grow substantially. The forecast-hour zero (f0; Fig. 8c) cluster-mean temperature bias is negative (cold) at all levels with largest values in the PBL (up to 1.5 km AGL) of approximately -0.5°C. Above 1.5 km AGL, the temperature has a near-zero but still slightly negative mean bias. The variation in the temperature biases is approximately 2°C at all vertical levels. The f0 dewpoint bias is negative (dry) below 100-200 m AGL with a positive (moist) bias above this layer of approximately 0.75°C. The cluster-mean dewpoint bias is approximately 0.5°C above the near-surface layer, but with a large variation of more than 3°C on both sides of the mean that grows larger with altitude. The bias profiles at later forecast hours (f12-36; Fig. 8d-f) remain largely the same to those at f0. However, the mean temperature bias becomes less cold above 1.5 km AGL throughout time.

Cluster 2 contains 2,855 soundings (or 18.4% of the total). These soundings are largely located in Canada, particularly extreme northern Canada (Fig. 9). The cluster-mean profile indicates stable conditions throughout the lowest 3 km AGL, as given by the increasingly positive normalized temperature values (denoting a parcel ascent curve that lies progressively further to the left of the environmental temperature curve on a skew-*T*, ln-*p* diagram) with height, with a dewpoint depression of approximately 5°C (Fig. 10a). The cluster-mean temperature bias (Fig. 10c) is warm (around 0.25°C) below 400 m AGL but is a cold bias above 400 m AGL. The cold bias increases with altitude until 2.5 km AGL, at which altitude the temperature bias becomes

closer to zero. The largest cluster-mean cold bias is between 2–2.5 km AGL, at which altitude it has a magnitude of approximately -0.25°C. The cluster-mean dewpoint bias is moist (positive) throughout the profiles. The cluster-mean moist bias from just above the surface to 2 km AGL is nearly 1°C but decreases to approximately zero between 2–3 km AGL. The temperature bias has a large variation of approximately 4°C in the lowest levels with a decreasing variation with altitude. The dewpoint bias has a variation of several degrees in the low levels with an increasing variation with altitude. As for Cluster 1, the cluster-mean temperature and dewpoint biases at later forecast hours (Fig. 10d-f) are similar to those at f0.

Cluster 3 contains 3,854 soundings (24.9% of the total). These soundings are primarily located in the southwestern US, with over 100 soundings throughout the period in several locations across this region, and substantially lower counts at other locations (Fig. 11). The cluster-mean normalized profile (Fig. 12a) indicates a warm and dry, deep mixed layer, as indicated by the large dewpoint depression (20–30°C) at all altitudes and normalized temperature and dewpoint profiles with similar shapes to those in Cluster 1 except extending to higher altitudes. The variation of the cluster-mean temperature profiles grows smaller with increasing altitude whereas the variation of the cluster-mean dewpoint profiles grows larger with altitude. This cluster contains the largest biases of temperature and dewpoint of the three clusters. At f0, the cluster-mean temperature bias is cold throughout the vertical levels with the largest mean cold biases of approximately -0.5°C near 500 m and 2.5 km AGL (Fig. 12c). At f0, there is a cluster-mean dry bias of approximately -0.5°C until 700 m AGL, with a mean moist bias of up to 2°C at higher altitudes (Fig. 12c). These cluster-mean temperature and dewpoint bias profiles remain consistent with time, albeit with the bias magnitudes becoming somewhat larger at higher elevations (Fig. 12d-f). The temperature bias variation is largest near the surface and decreases with increasing altitude. The dewpoint bias

variation is smallest bellow 1.5 km AGL and largest at 2.5 km AGL where it is greater than 10°C in magnitude.

3.2 1200 UTC Results

The *k*-means clustering analysis results in three distinct sounding profiles for the 1200 UTC data. Cluster 1 contains 4,444 soundings (27.6% of the total). These soundings are in colder climates with a large number coming from stations in northern Canada and along the Rocky Mountains (Fig. 13). The cluster-mean normalized and actual temperature profiles (Fig. 14a,b) are characterized by a surface-based inversion with an approximate depth of 200-300 m that is likely to be a radiative inversion given the time of day and inversion base at the surface. Above the inversion, the normalized temperature remains approximately constant until 3 km AGL (Fig. 14a). The profile has a temperature variation of approximately 10°C within the inversion layer and 5°C to 7°C above the inversion layer. At f0, notable cluster-mean warm (up to 2°C) and dry (up to -0.5°C) biases exist within the inversion layer, and these biases switch signs to cold and moist, respectively, above the inversion layer meaning the model is underrepresenting the inversion strength (Fig. 14c). The mean cold bias is largest at inversion top, where it is approximately -0.5°C, and becomes smaller over the next 200-300 m AGL above this altitude. The remaining vertical levels have a small cluster-mean temperature bias of less than 0.5°C. The cluster-mean dewpoint bias remains moist above inversion top with the largest moist biases of approximately 0.75–1°C between 1.5–2.5 km AGL These profile shapes remain largely the same through the forecast hours, except with a reduction in mean moist-bias magnitudes with time (Fig. 14d-f). The temperature bias variation is largest within the inversion layer at approximately 4°C and decreases in magnitude with increasing altitude. The dewpoint bias variation is greater than 5°C at all vertical levels, with the smallest variation located near the surface in the inversion layer.

Cluster 2 contains 8,679 soundings (53.8% of the total). These soundings are primarily located in the southeastern US and southern Alaska with a smaller density at other stations across the US and Canada (Fig. 15). As in Cluster 1, the cluster-mean normalized and raw profiles indicate a surface-based inversion with an approximate depth of 200-300 m (Fig. 16a,b); however, the cluster-mean inversion is substantially weaker in Cluster 2 than in Cluster 1. Above the inversion, the normalized temperature profile approximately follows the pseudoadiabat along which the cluster-mean surface-based parcel ascends (Fig. 16a). The normalized temperature and dewpoint profiles have a small variation of approximately 5°C below 1.5 km AGL, increasing above this level (Fig. 16a). At f0, the cluster-mean temperature and dewpoint biases are similar to those in Cluster 1 biases 500 m AGL (Fig. 16c), with is a warm and dry bias near the surface. As in Cluster 1, the cluster-mean temperature bias becomes cold bias by inversion top, with largest values between -0.5 and -1°C at approximately 400 m AGL (Fig. 16c). The temperature and dewpoint profiles remain cold- and moist-biased in their respective means throughout the rest of the profile, albeit with relatively small magnitudes above 1 km AGL. By f12, the cluster-mean temperature bias grows more negative below 1.5 km AGL, with a peak value of approximately -1°C at 400 m AGL (Fig. 16d). The cluster-mean dewpoint bias profile develops an S-shape beginning at this forecast hour with the largest negative values near 500 m AGL and near 2.5 km AGL. These characteristics become increasingly evident at later forecast times (Fig 16e,f). The temperature bias variation is several degrees throughout all vertical levels and forecast times, whereas the dewpoint variation is largest variation at higher altitudes.

Finally, Cluster 3 is the smallest cluster, containing only 2,995 soundings (18.6% of the total). These soundings are located throughout the continental US with two West Coast locations having a particularly high number of soundings (Fig. 17). The cluster-mean normalized sounding

profile for this cluster represents a shallow radiation inversion with residual mixed layer above this inversion, as indicated by drier upper-levels and residual nearly pseudoadiabatic layer (Fig. 18a). This cluster contains the largest temperature and dewpoint biases of the 1200 UTC data (Fig. 18c-f). At f0, the cluster-mean temperature bias is cold throughout many of the vertical levels with the largest bias below 500 m AGL, where it approaches -1°C (Fig. 18c). The bias then becomes smaller between 1–2 km AGL where it then gets slightly larger above 2 km AGL. The cluster-mean dewpoint bias is dry near the surface, then becomes zero at inversion top and moist above 1 km AGL. The cluster-mean moist bias is the largest between 2.5–3 km AGL, where it is approximately 2.5°C. The cluster-mean biases are similar at later forecast hours (Fig. 18d-f), albeit with the mean lower-level biases becoming more negative and the upper-level mean biases becoming more positive with time. The temperature bias variation, throughout all vertical levels and forecast hours, remains several degrees Celsius. The dewpoint bias is smallest at low levels and larger than 10°C above approximately 1.7 km AGL.

4. Discussion

Both GFS v15.1 and its successor GFS v16 (released in 2020) exhibit notable cold biases, with these biases extending through the troposphere in GFS v15.1 (NOAA 2019) but limited to near the surface in GFS v16 (Manikin et al. 2021). These biases exist because of improper handling of radiation and snow cover. These low-level cold biases are noted in the data considered in this study. The results can impact the magnitude of the biases and mixing issues like the initial model biases impacting the turbulence parameterizations in a previous study (Coniglio et al. 2013).

The *k*-means clustering analysis results in three distinct environmental regimes in the data at 0000 UTC and 1200 UTC. The three environmental regimes in the 0000 UTC data are a shallow mixed layer (Cluster 1), a stable PBL (Cluster 2), and deep and dry mixed layer (Cluster 3). Each

regime consists of forecast errors related to problems with the vertical mixing in the PBL. In the shallow mixed layer regime, the lower-level temperature and dewpoint biases indicate the GFS is underrepresenting the warmth and dryness of the near surface layer with a cold and moist bias present below 1.5 km AGL. The temperature bias remains near zero (likely due to the cold bias presented above as it does not become warm) at higher altitudes, but the moist dewpoint bias is still present. This indicated that the GFS remains too moist through the boundary layer. In the stable layer, a notable cold and moist bias is found in the upper levels indicating that the GFS is overmixing the boundary layer below and bringing colder and more moist air from below the boundary layer. Finally, the deep and dry mixed layer also indicates the GFS is overmixing the PBL to a larger degree than the shallow mixed layer leading to larger errors in the model.

In the 1200 UTC data, the three environmental regimes are a deep radiation inversion, a shallow radiation inversion, and a shallow radiation inversion with a residual mixed layer. The deep mixed layer indicates that the GFS is overmixing the boundary layer due to a warm and moist bias near the surface and a cold and dry bias aloft. The shallow radiation inversion shows a cold, dry bias in the inversion layer and in later times shows a warm (relative to the known cold bias), dry bias aloft. This could show that the GFS underrepresenting the strength of the radiation inversion. Finally, the shallow radiation inversion with the residual mixed layer is overmixing the residual mixed layer due to the smaller cold bias and dry bias between 500 m AGL and 1.5 km AGL with a colder, moist bias aloft. In these data, the GFS seems to be undermixing more stable environments while overmixing unstable environments.

After the f0 analysis, subsequent forecast hour biases are generally similar temperature and dewpoint biases as the f0 forecast. It is noted that the bias structures could be expected to change throughout forecast times. Model forecasts initialize using analyses that have the previous model

cycle's 6-h forecast as their first guess, and observation assimilation does not entirely ameliorate (nor is it intended to entirely ameliorate) the forecast biases in the resulting analysis. This allows the bias structures to propagate similarly through the forecast hours considered here, similar to that seen in Coniglio et al. (2013).

5. Conclusions and Future Work

In this study, soundings are reformulated to a height above ground level coordinate as well as a normalized temperature and dewpoint (using a novel method). This is calculated by using a soundings parcel path to the lifted condensation level and down a pseudoadiabat to the surface. This indicated the surface wet-bulb temperature which is then used to calculate the pseudoadiabat used to subtract temperature and dewpoint. This allows for soundings to be considered regardless of the station altitude, geographic region, and seasonality. The 0000 UTC and 1200 UTC soundings are dimensionally reduced using a multivariate EOF analysis, separately. The leading two PCs explain approximately 70% of the variance within the dataset and are well-separated from the third and subsequent PCs. These leading PCs are then used as input into a *k*-means clustering analysis, from which three clusters are retained at each analysis time. Three distinct environmental regimes are identified at each time. In the 0000 UTC data, the identified regimes are a shallow mixed layer, a stable profile, and a deep and dry mixed layer. The biases from this verification showed that perhaps more unstable soundings have an issue overmixing the boundary layer. In the 1200 UTC data, the three distinct regimes were a deep radiation inversion, a shallow radiation inversion with pseudoadiabatic layer above the inversion, and a shallow radiation inversion with a residual mixed layer above the inversion.

Along with this study, the same method described here was separately applied to soundings only within thunderstorm-supporting environments in the United States, obtained from the Storm

Prediction Center's 1200 UTC convective outlook general thunderstorm areas. While this approach led to less than half of the current dataset, it showed promise as a future application of applying this novel sounding normalization and classification method to quantify environment-specific model sounding biases. Another approach to determine the cause of large errors was filtering the bias data for each cluster and lead time into the 10% most-positive and 10% most-negative biases to determine the environmental sounding structures associated with the largest biases. However, this also led to a small dataset (3,100 soundings for 0000 UTC and 3,226 soundings for 1200 UTC) that did not allow for any reliable insights to be drawn. Additionally, the loadings from this study can be used for an application (e.g., classify a different sounding dataset) without having to repeat this process for a large dataset.

Considerations could be made to improve the ability to capture soundings structures. The soundings could be filtered based on how many vertical levels are retained in the lowest 3 km AGL which would allow for sharper temperature and dewpoint variations to be captured. Another consideration is using a local standard time instead of only 0000 and 1200 UTC for variations in the time of day to be better represented. Finally, the vertical interpolation to an AGL coordinate can be improved by using other methods besides linear interpolation (e.g., cubic spline interpolation).

Figures

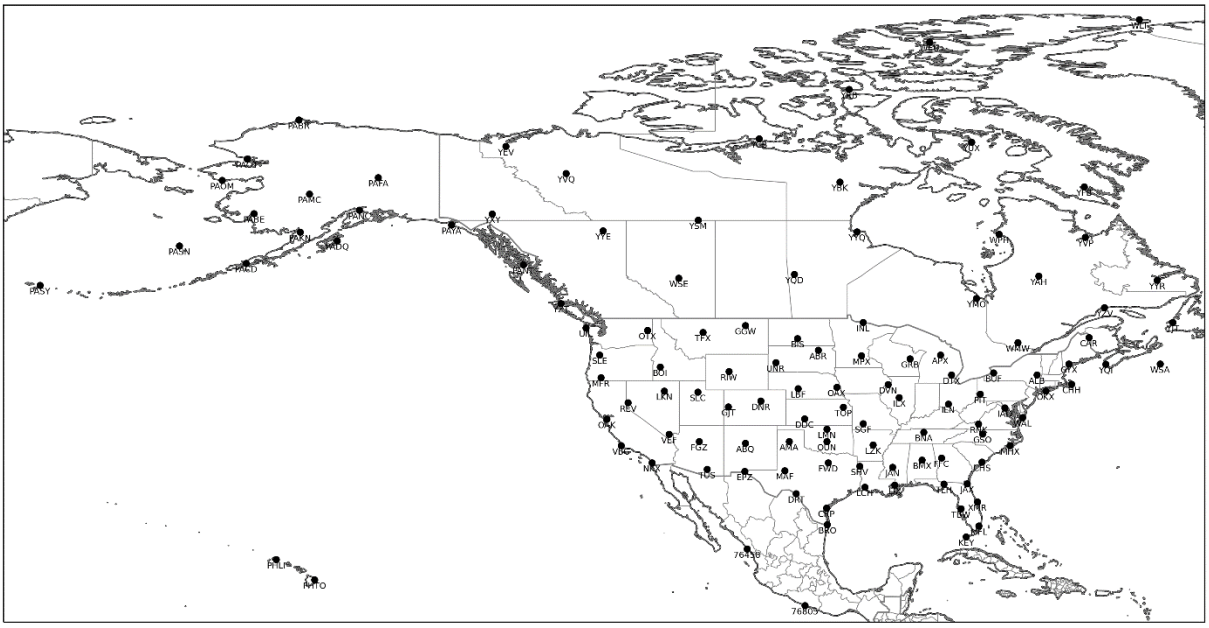

Figure 1: A depiction of the sounding locations across much of North America.

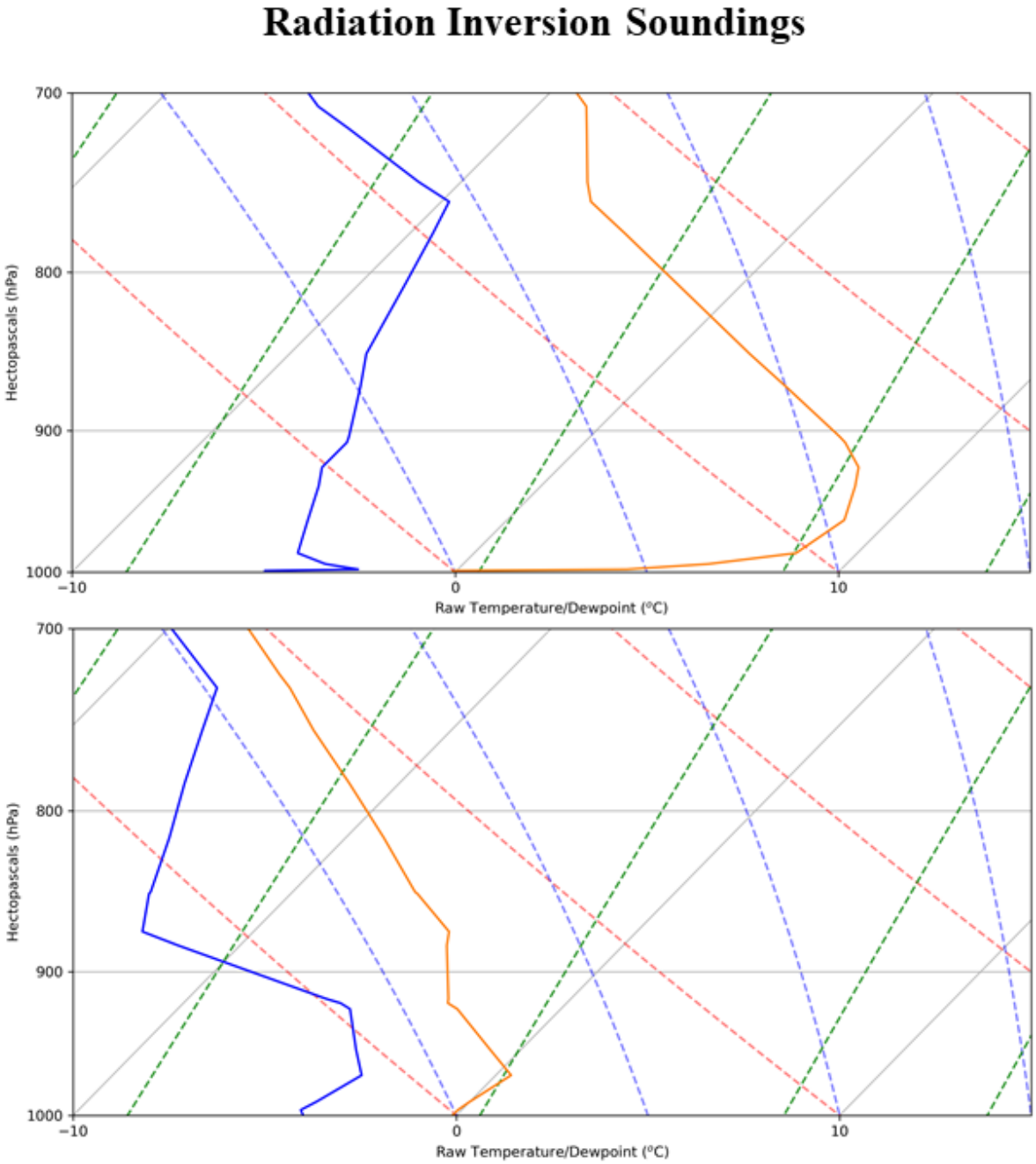

Figure 2: Observed skew-*T*, ln-*p* diagrams depicting surface-based radiation inversions at Norman Wells, Northwest Territories, Canada at (top) 1200 UTC 7 May 2019 and (bottom) 1200 UTC 27 September 2019.

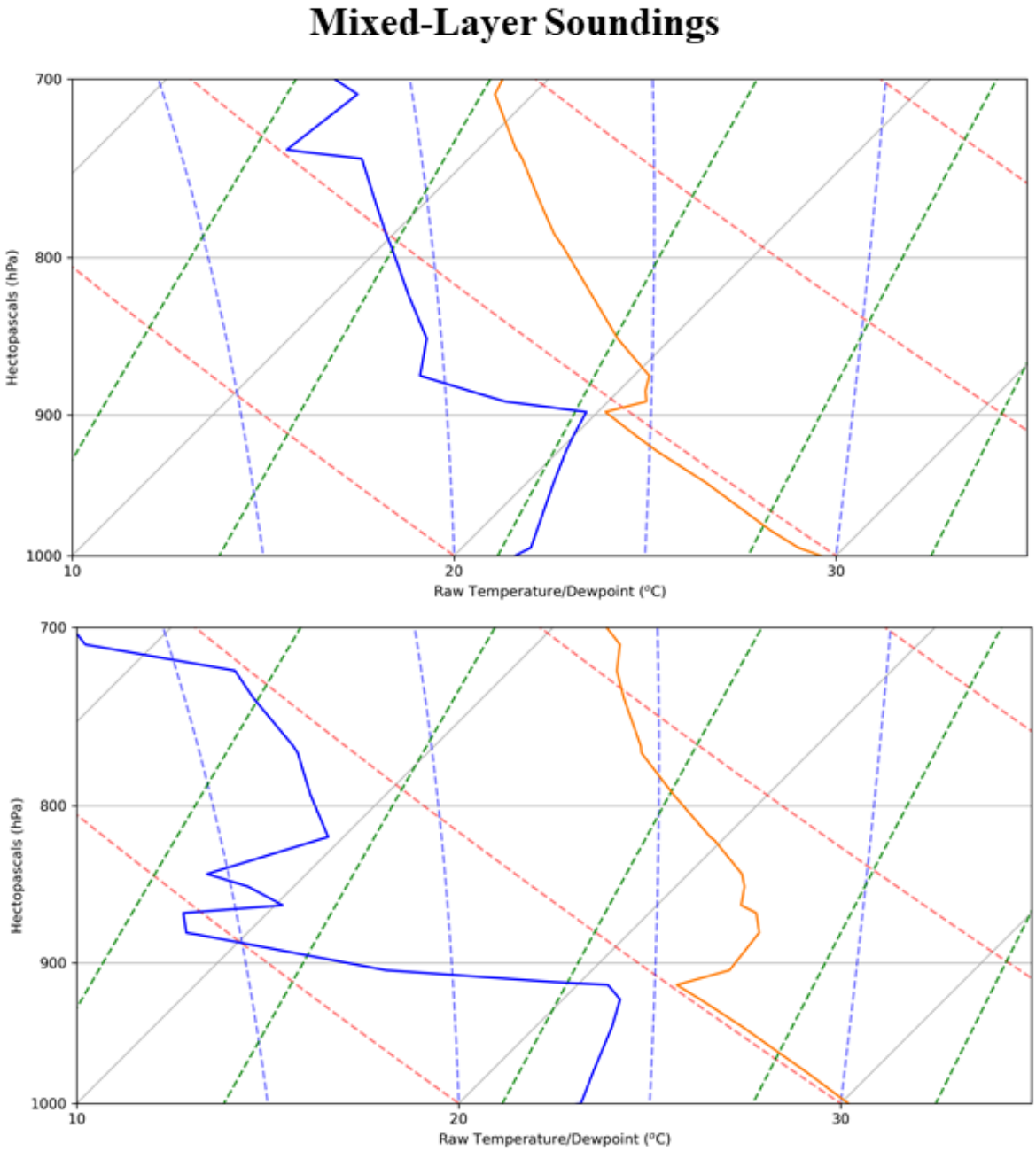

Figure 3: Observed skew-*T*, ln-*p* diagrams depicting surface-based mixed layers at (top) Key West, FL (KEY) at 0000 UTC on 23 July 2019 and (bottom) Brownsville, TX (BRO) at 0000 UTC on 22 July 2019.

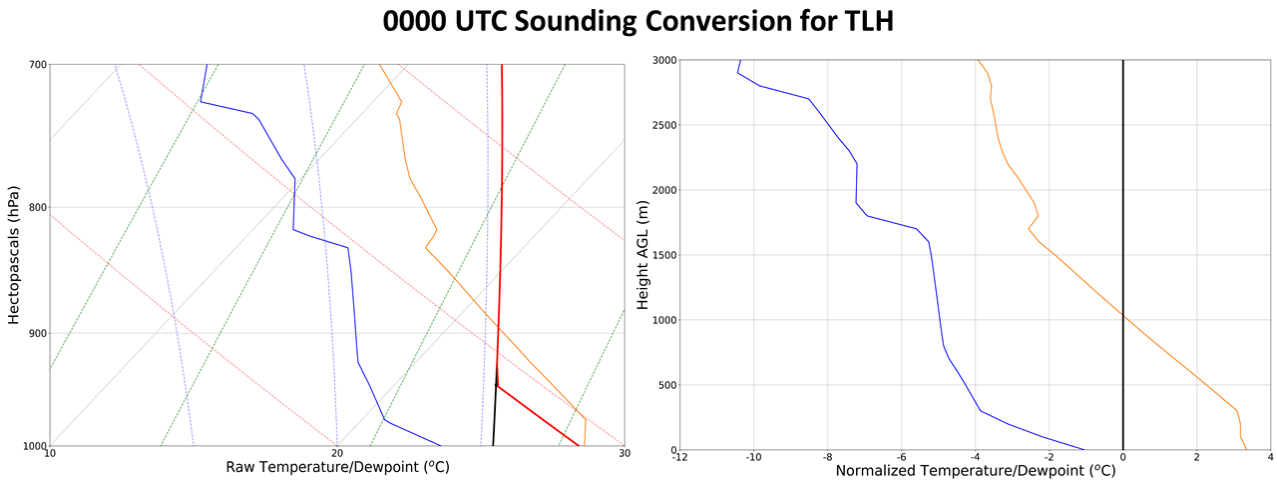

Figure 4: The transformation of vertical temperature (orange) and dewpoint (blue) profiles, here from the observed 0000 UTC 7 August 2019 Tallahassee, FL (TLH) sounding, into a normalized framework. The left panel shows the temperature and dewpoint on a skew-*T*, ln-*p* diagram with the calculated pseudoadiabat (red) corresponding to the surface-based parcel's wet-bulb temperature in the solid black line. This panel also shows the parcel path to the LCL (black) and following the pseudoadiabat. The right panel shows the normalized temperature and dewpoint, with the pseudoadiabat given by the solid black line at 0°C.

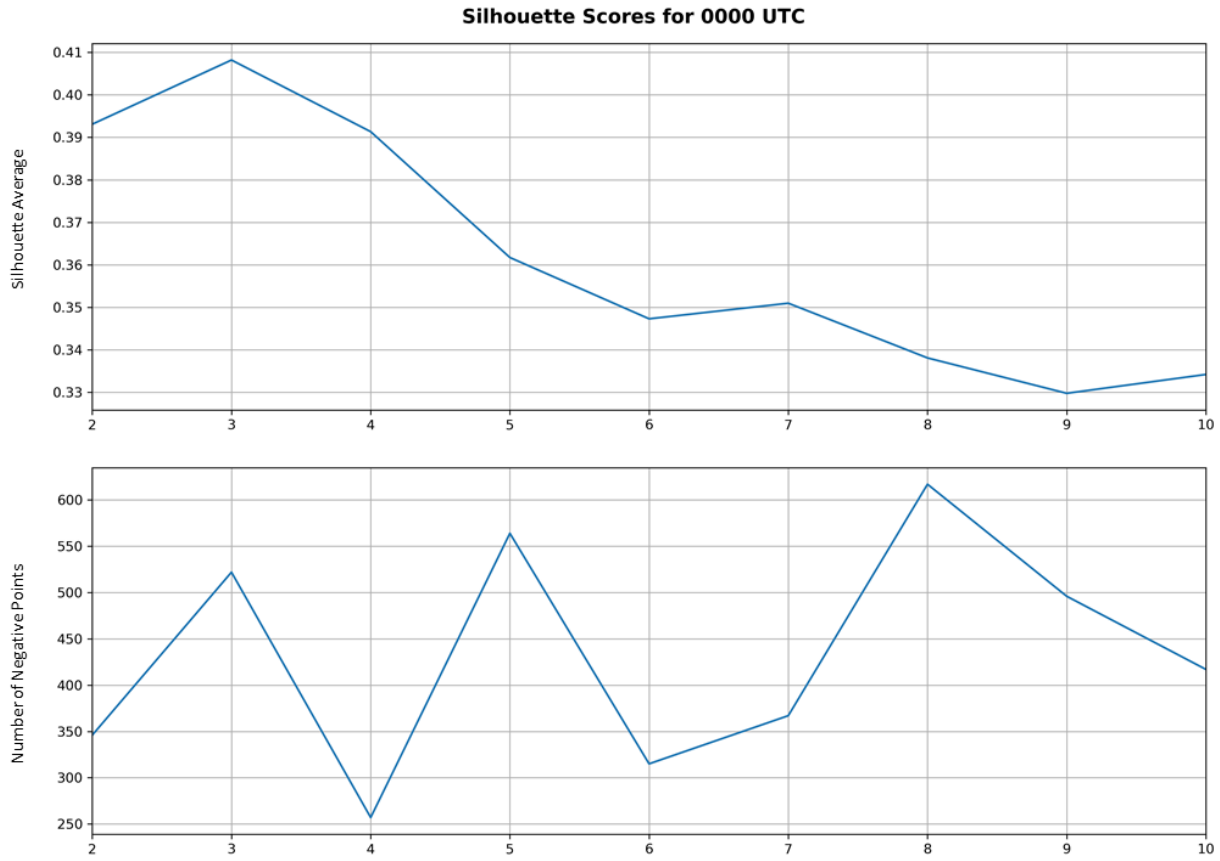

Figure 5: The cluster-average silhouette score (top) and total number of negative points (bottom) for k = 2 through k = 10 for the 0000 UTC observations.

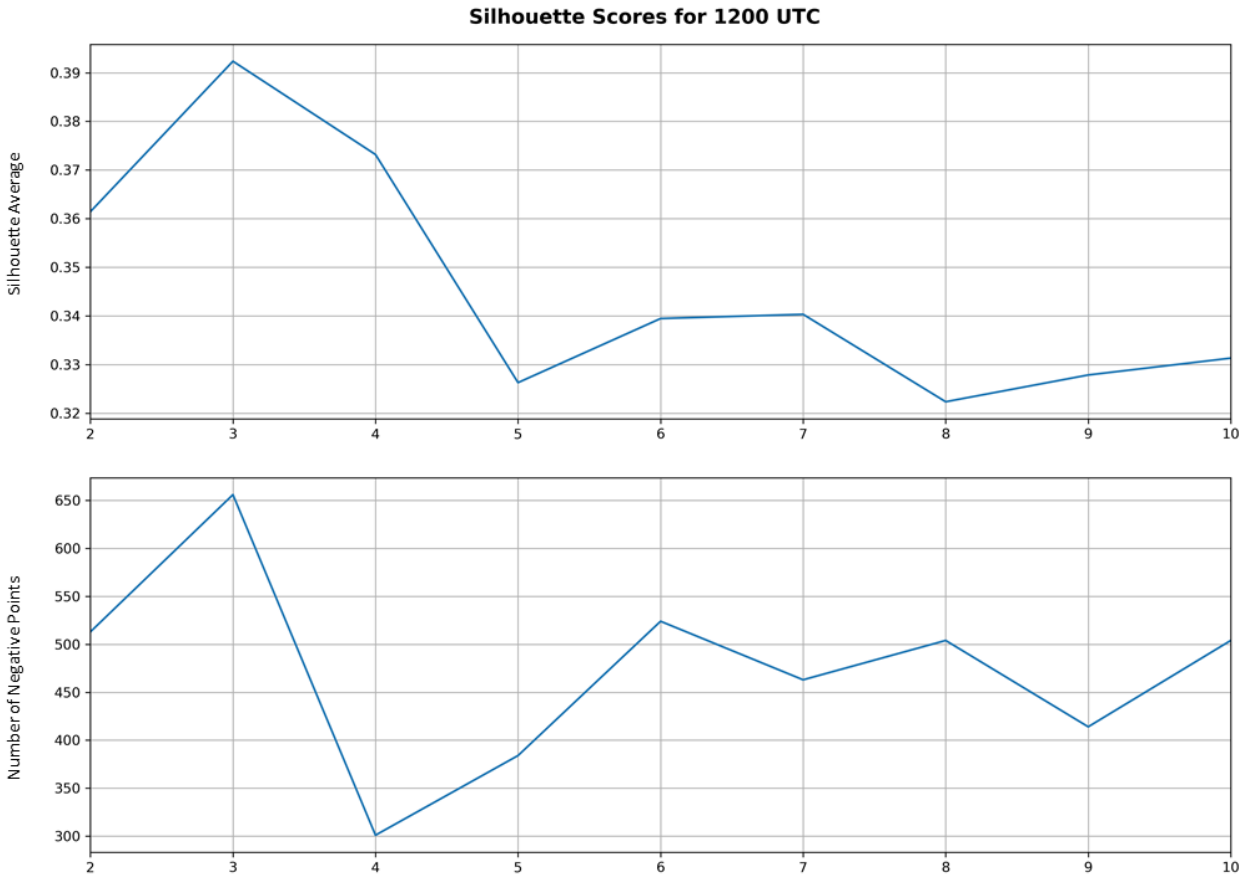

Figure 6: As in Fig. 5, except for the 1200 UTC observations.

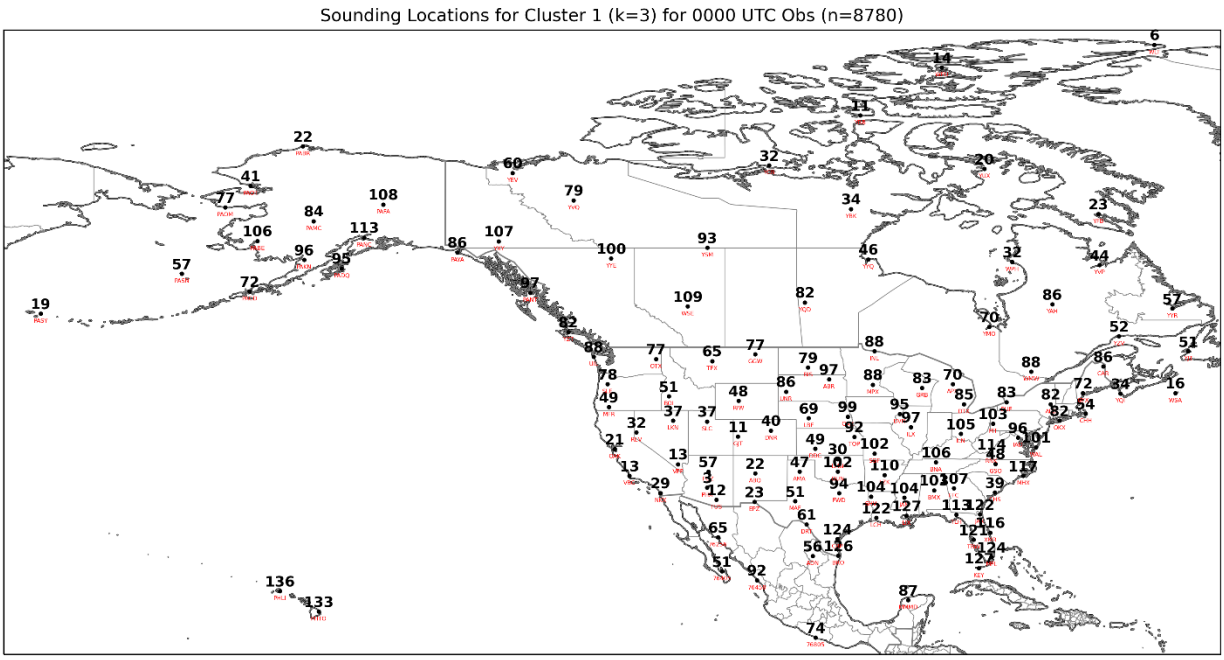

Figure 7: Observed 0000 UTC sounding counts by location for Cluster 1. There is a total of 8,780 soundings.

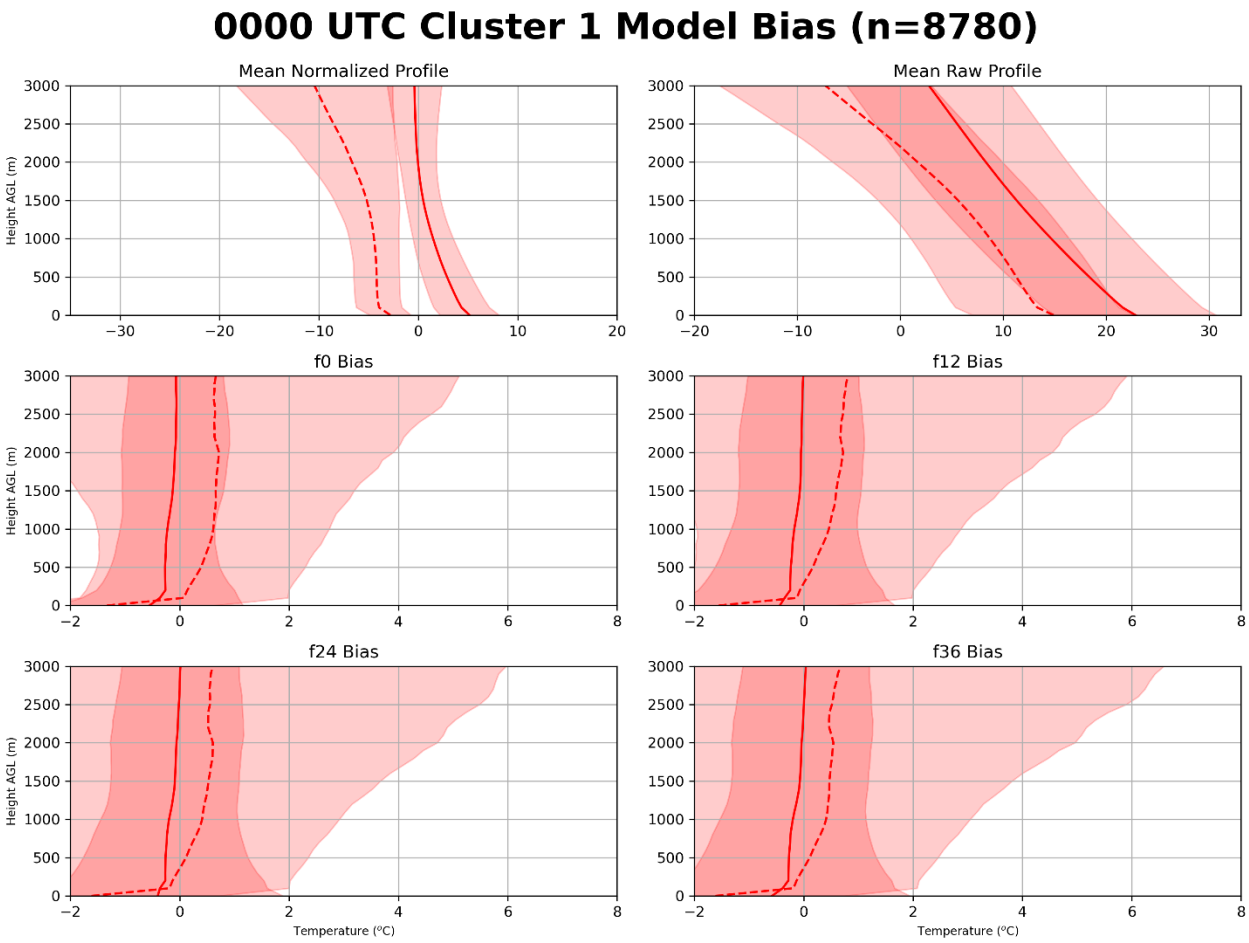

Figure 8: Cluster-mean (a) normalized and (b) raw temperature (solid; °C) and dewpoint (dashed; °C) profiles for Cluster 1 at 0000 UTC. The bottom two rows depict the cluster-mean temperature (solid; °C) and dewpoint (dashed; °C) biases at (c) f0, (d) f12, (e) f24, and (f) f36. The shading in each panel depicts +/- one standard deviation around the respective cluster-mean quantity. All panels depict quantities between 0-3 km AGL.

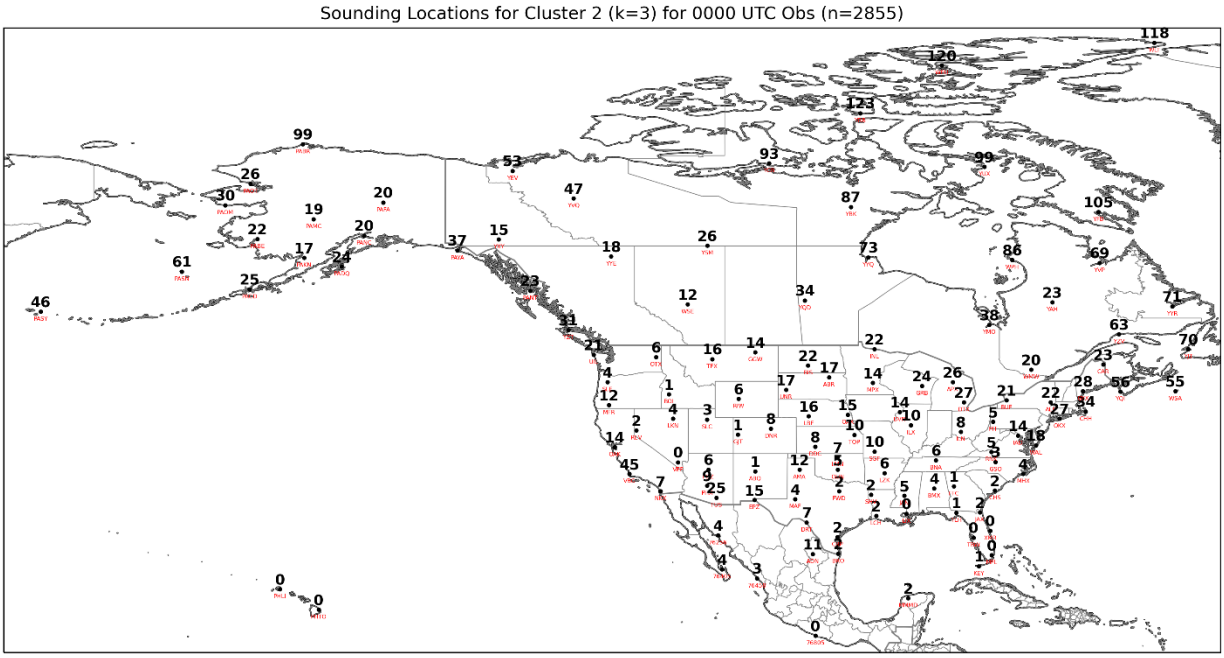

Figure 9: Observed 0000 UTC sounding counts by location for Cluster 2. There is a total of 2,855 soundings.

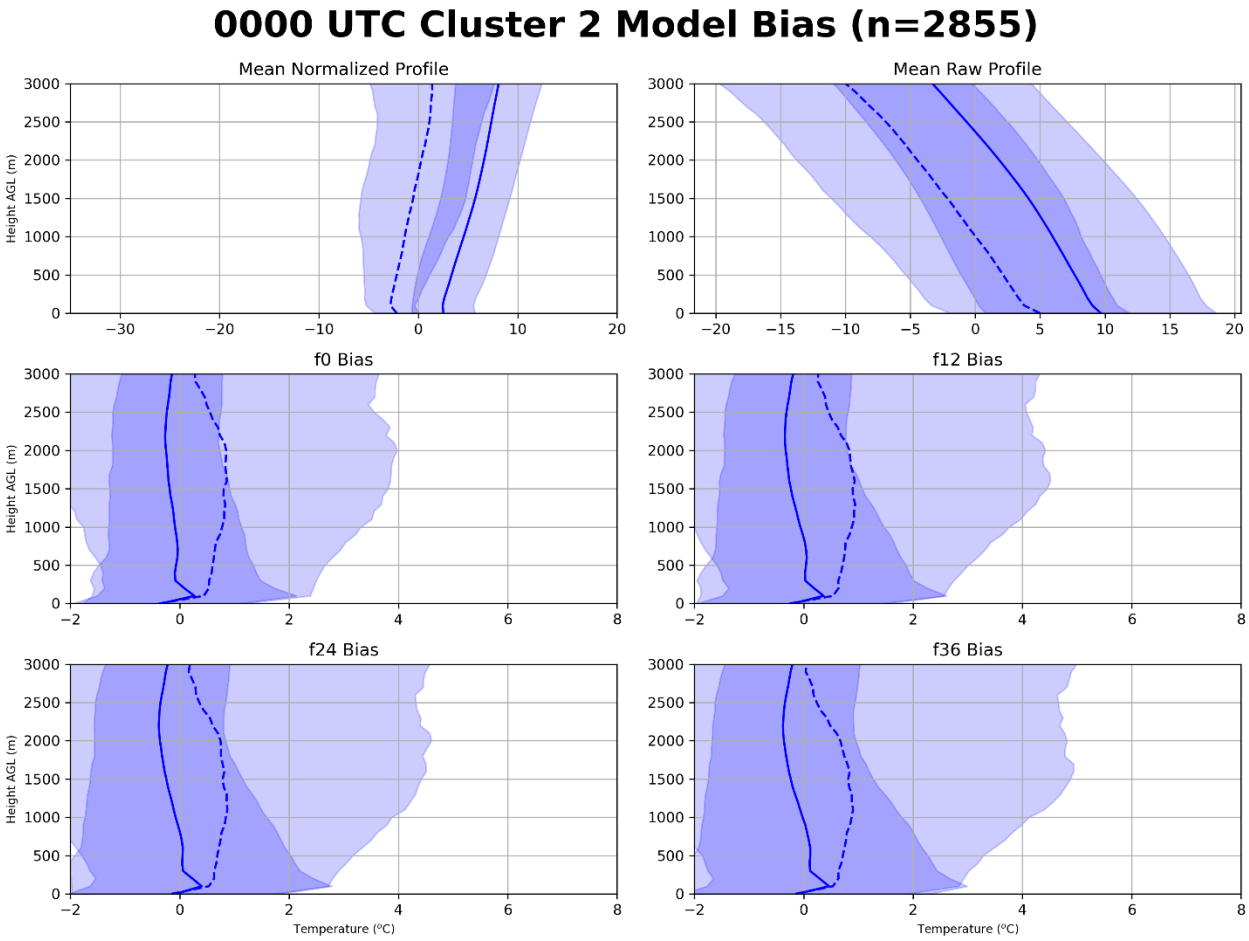

Figure 10: As in Fig. 8, except for Cluster 2 at 0000 UTC.

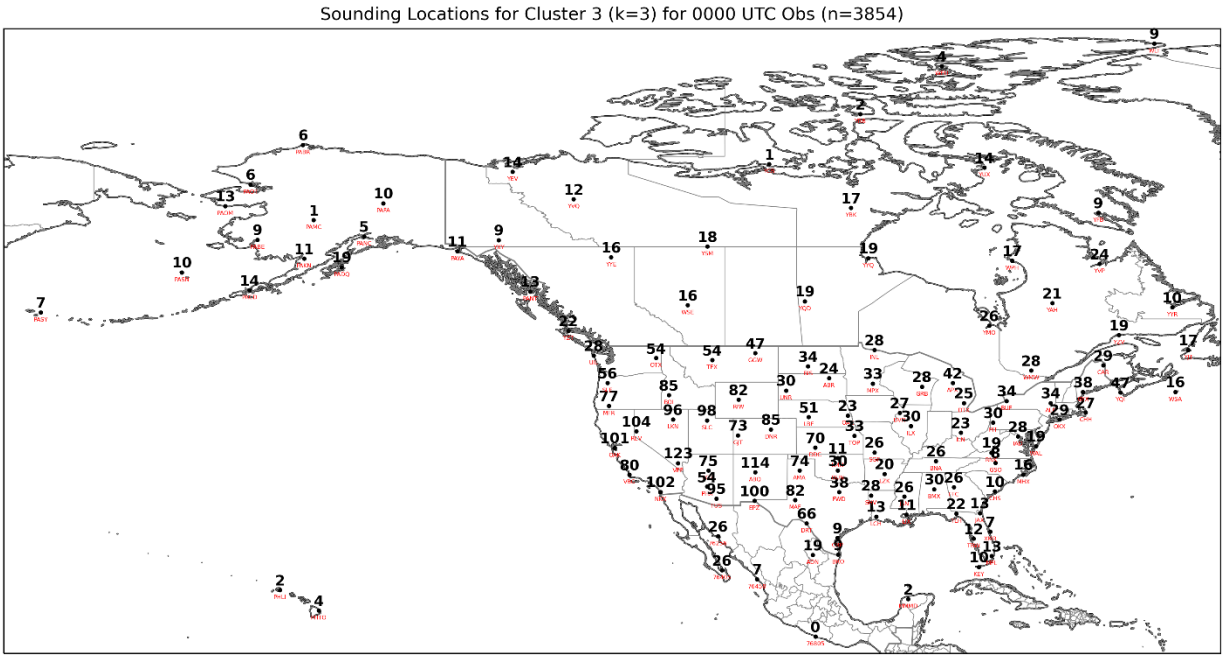

Figure 11: Observed 0000 UTC sounding counts by location for Cluster 3. There is a total of 3,854 soundings.

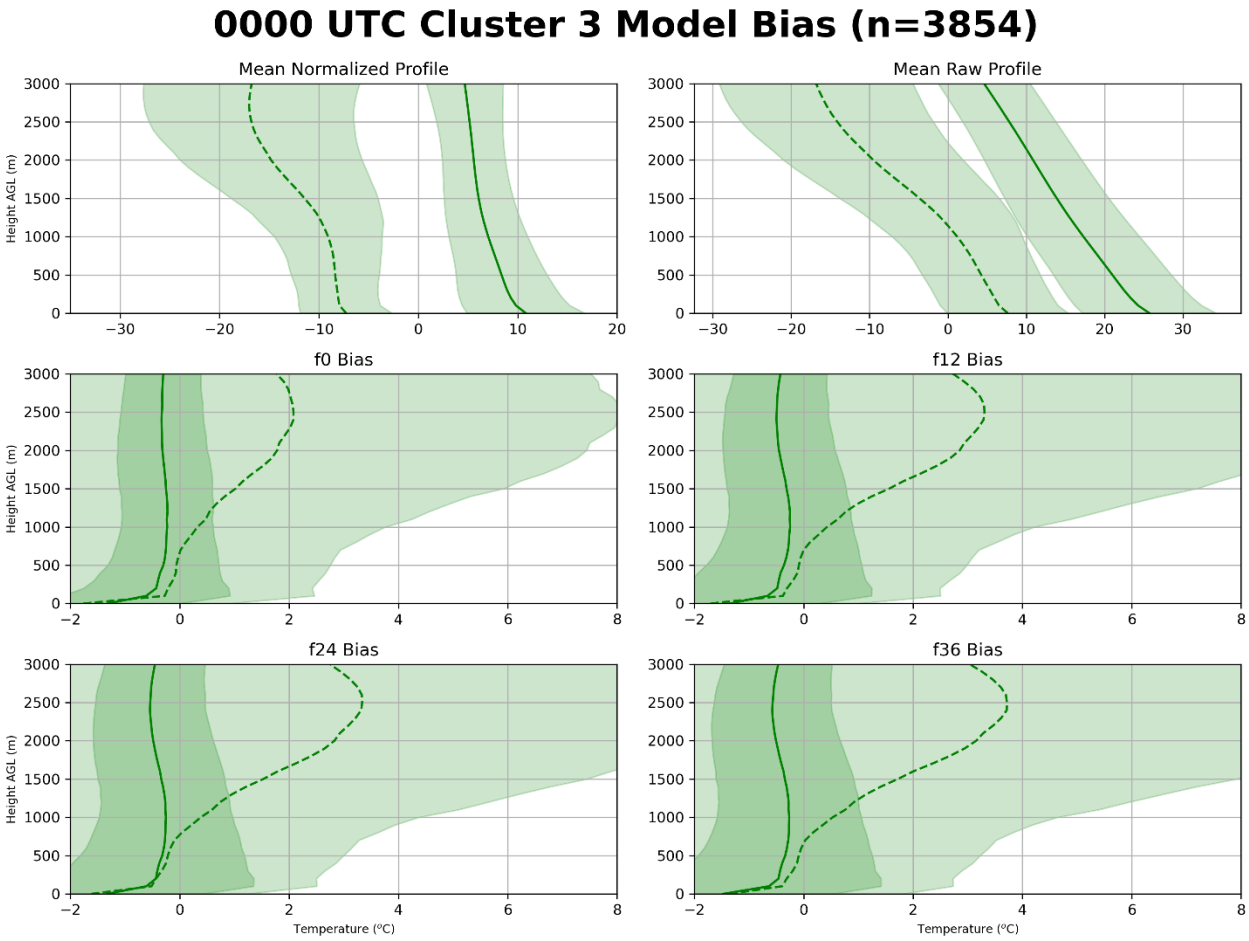

Figure 12: As in Fig. 8, except for Cluster 3 at 0000 UTC.

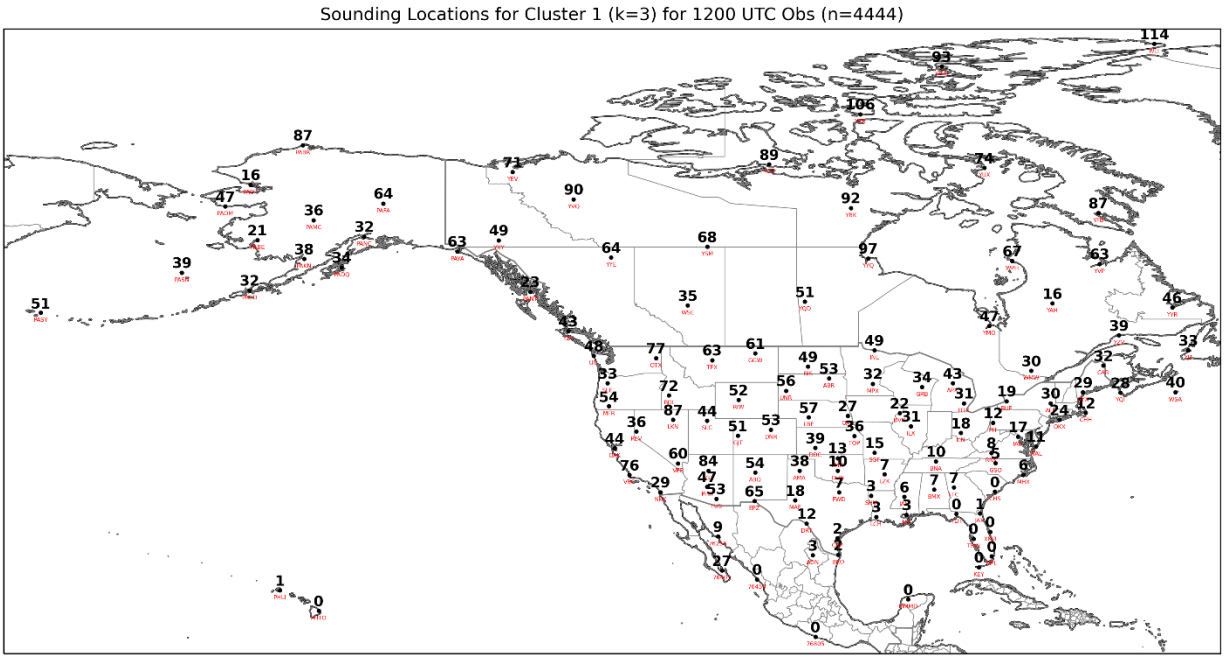

Figure 13: Observed 1200 UTC sounding counts by location for Cluster 1. There is a total of 4,444 soundings.

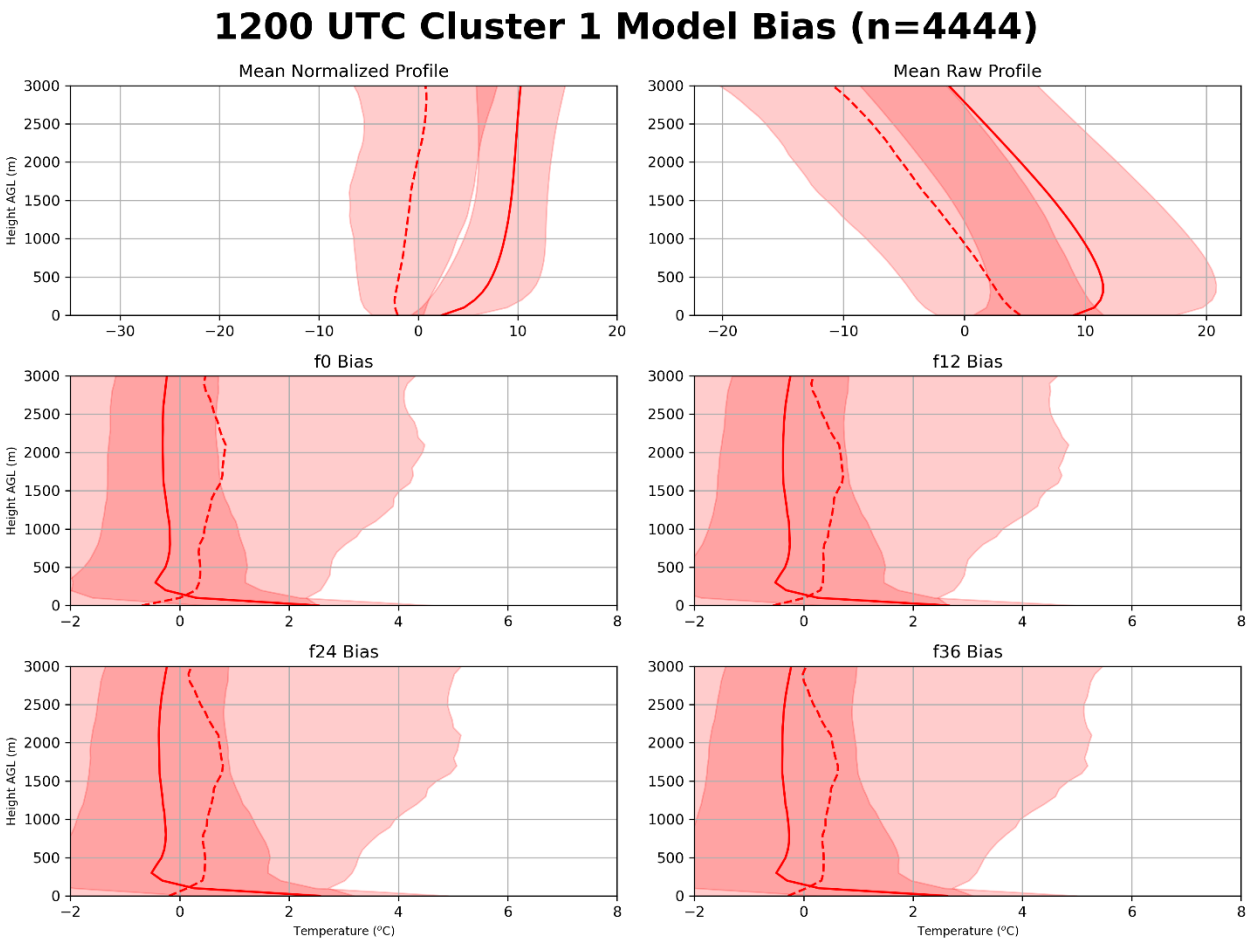

Figure 14: As in Fig. 8, except for Cluster 1 at 1200 UTC.

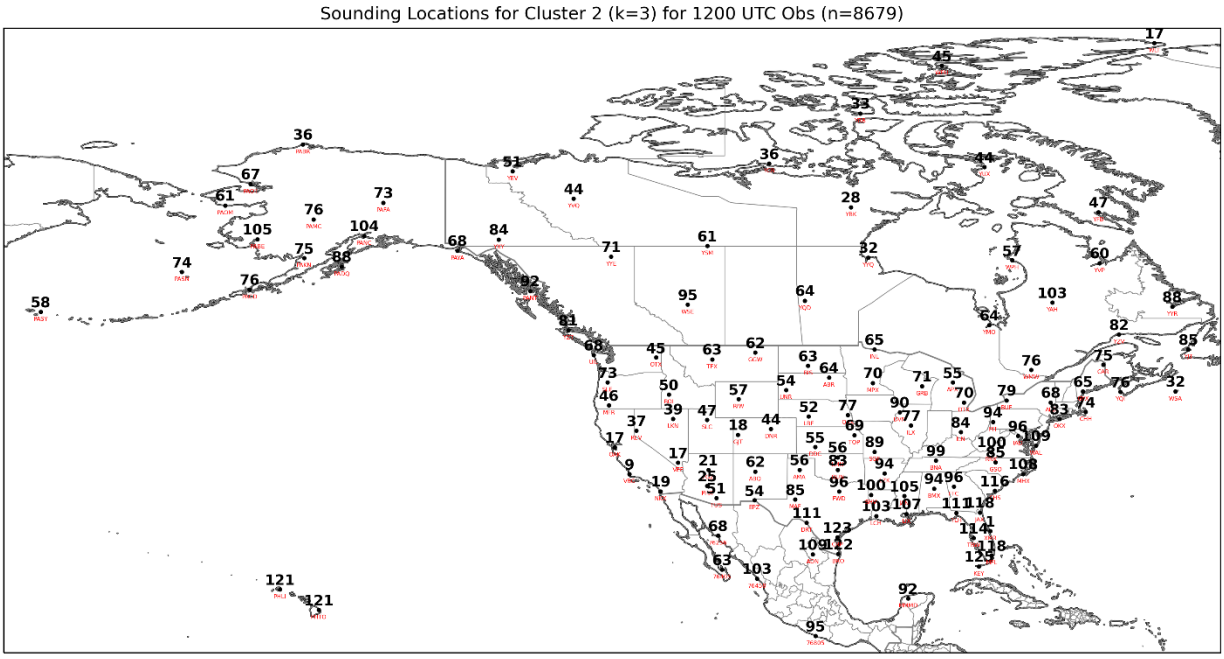

Figure 15: Observed 1200 UTC sounding counts by location for Cluster 2. There is a total of 8,679 soundings.

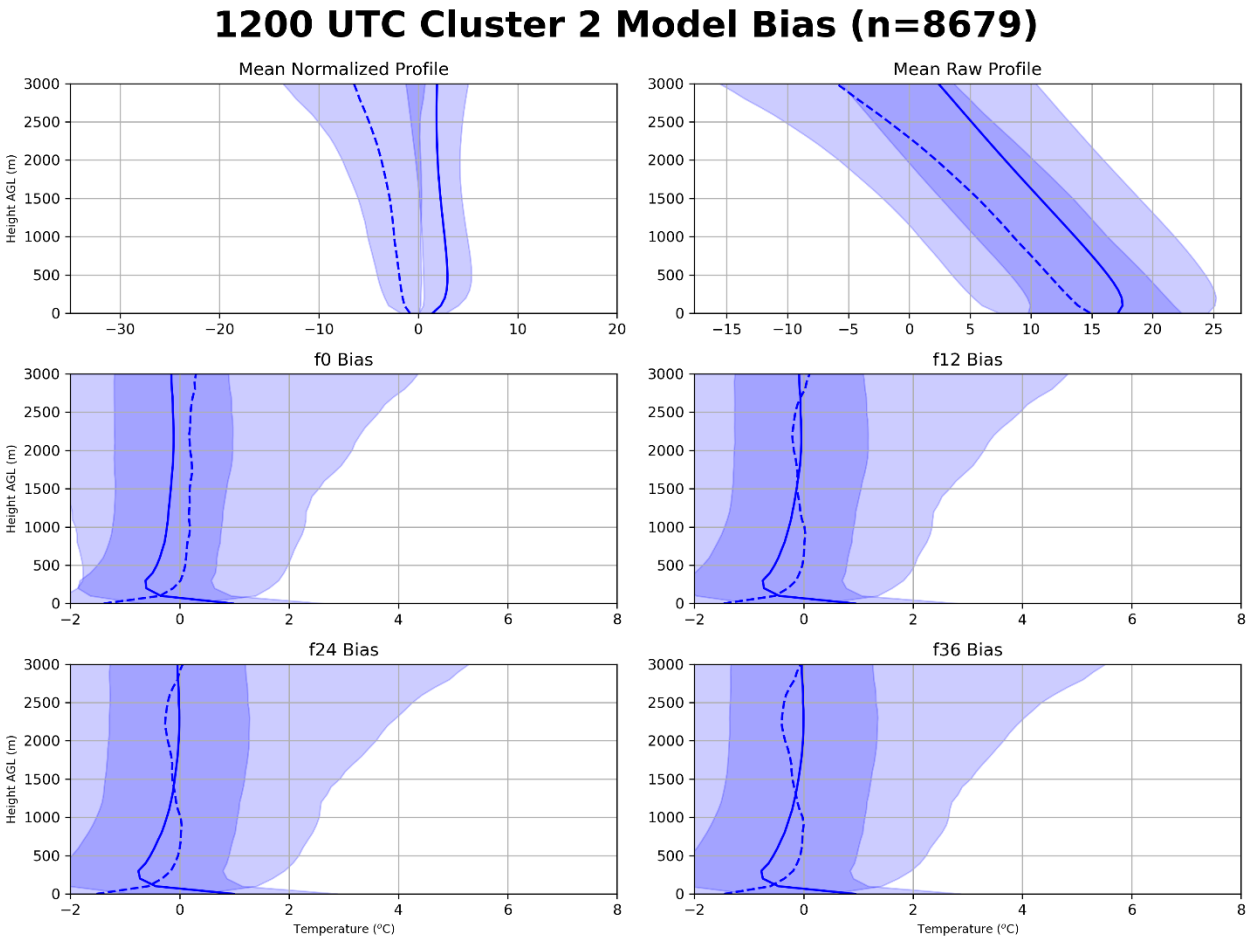

Figure 16: As in Fig. 8, except for Cluster 2 at 1200 UTC.

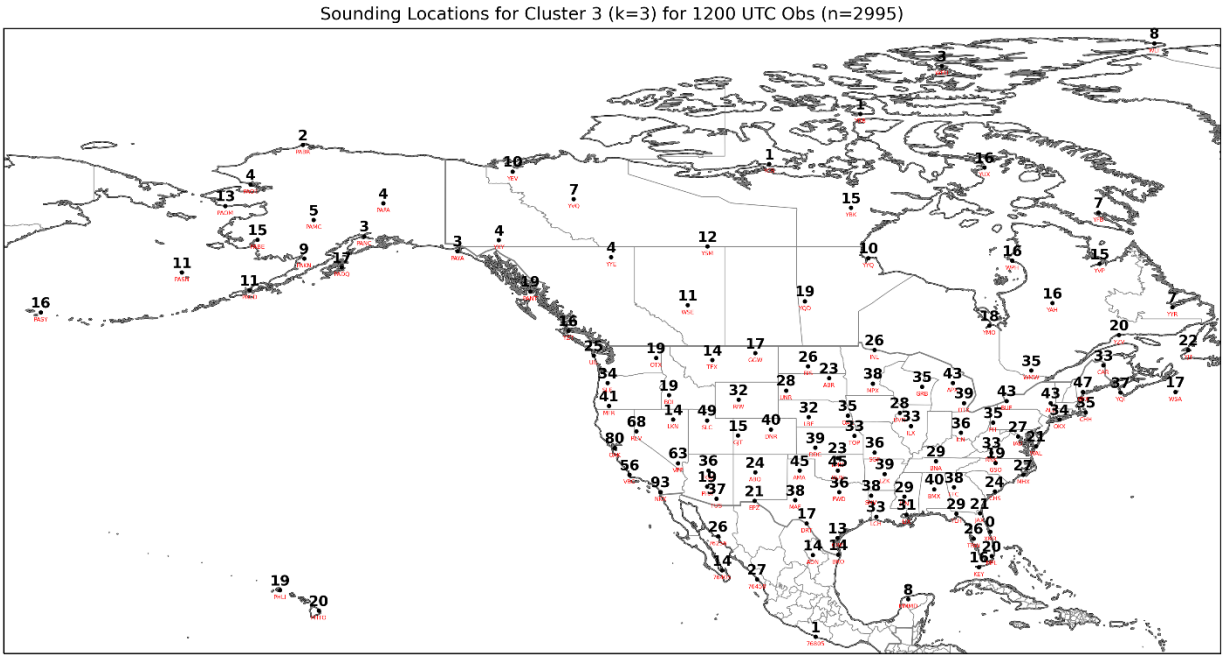

Figure 17: Observed 1200 UTC sounding counts by location for Cluster 3. There is a total of 2,995 soundings.

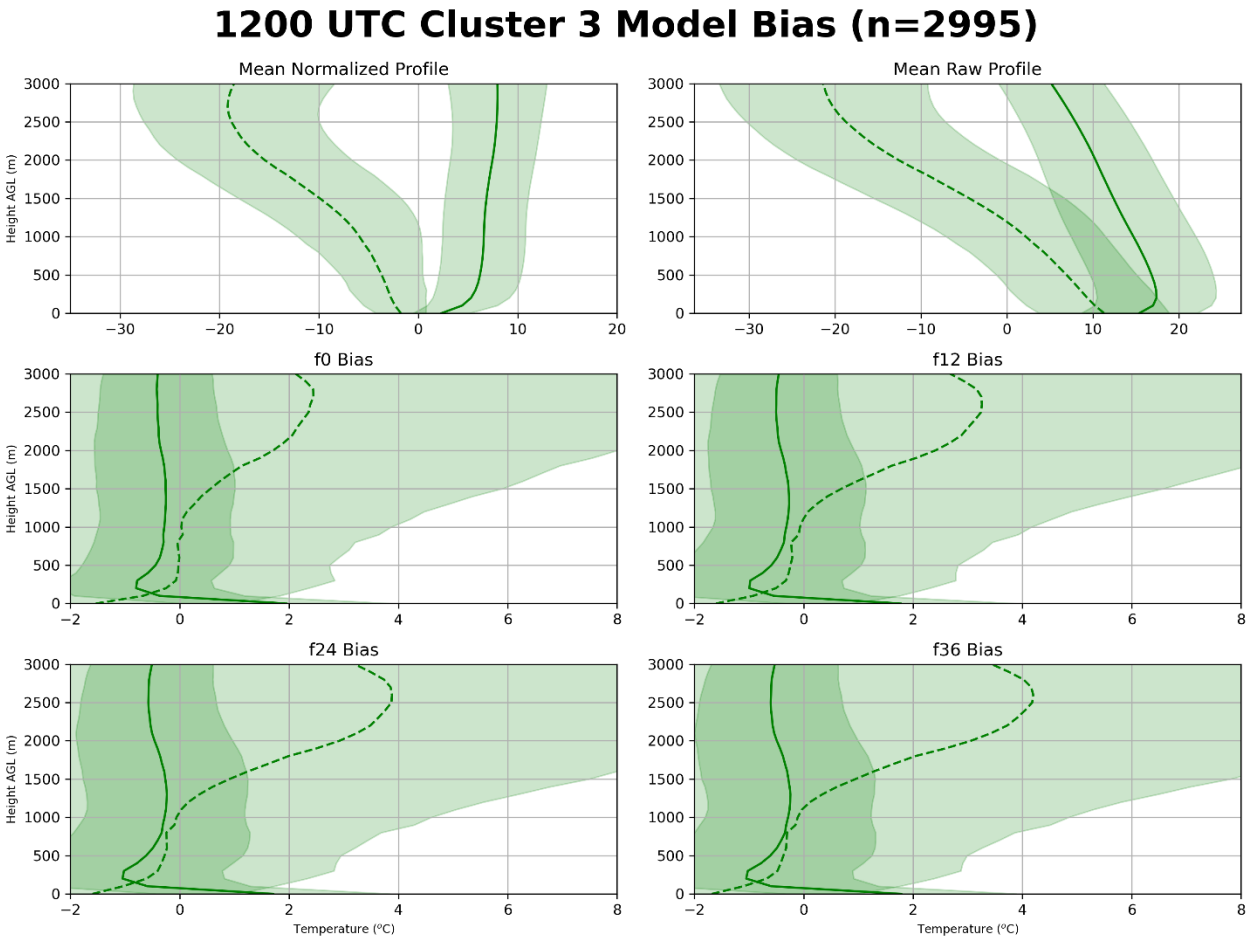

Figure 18: As in Fig. 8, except for Cluster 3 at 1200 UTC.

Tables

	PC 1	PC 2	PC 3	PC 4	PC 5
0000 UTC	36.08%	34.32%	14.00%	6.05%	2.52%
1200 UTC	39.19%	35.36%	8.18%	6.53%	2.87%

Table 1: Percentage of variance contained within the first 5 PCs of the normalized temperature and dewpoint data.

References

Anderson-Frey, A., Y. Richardson, A. Dean, R. Thompson, and B. Smith, 2017: Self-organizing maps for the investigation of tornadic near-storm environments. *Wea. Forecasting*, **32**, 1467–1475, https://doi.org/10.1175/WAF-D-17-0034.1.

Arnott, J. M., J. L. Evans, and F. Chiaromonte, 2004: Characterization of extratropical transition using cluster analysis. *Mon. Wea. Rev.*, **132,** 2916–2937.

Blue Hill Observatory and Science Center, 2021: A Brief History of the Blue Hill Meteorological Observatory. [Available online at http://bluehill.org/observatory/about-us/history/].

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.

Clark, A. J., M. C. Coniglio, B. E. Coffer, G. Thompson, M. Xue, and F. Kong, 2015: Sensitivity of 24-h forecast dryline position and structure to boundary layer parameterizations in convection-allowing WRF Model simulations. *Wea. Forecasting*, **30**, 613–638, https://doi.org/10.1175/WAF-D-14-00078.1.

Cohen, A. E., S. M. Cavallo, M. C. Coniglio, H. E. Brooks, 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, doi: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.

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.

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.

Giangrande, S.E., D. Wang, and D. B. Mechem, 2020, Cloud regimes over the Amazon Basin: perspectives from the GoAmazon2014/15 campaign, *Atmos. Chem. and 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.

Han, J., and C. S. Bretherton, 2019: TKE-based moist Eddy-Diffusivity Mass-Flux (EDMF) parameterization for vertical turbulent mixing. *Wea. Forecasting*, **34**, 869–886, https: //doi.org/10.1175/WAF-D-18-0146.1.

Hart, R. E., 2003: A cyclone phase space derived from thermal wind and thermal asymmetry. *Mon. Wea. Rev.*, **131,** 585–616.

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. Inf. Theory*, **28,** 129–136.

Liu, Y., and R. Weisberg, 2011: A review of self-organizing map applications in meteorology and oceanography. *Self-Organizing Maps—Applications and Novel Algorithm Design*, J. Mwasiagi, Ed., InTech, 253–272.

Lorenz, E. N., 1956: Empirical orthogonal functions and statistical weather prediction. Department of Meteorology, MIT Statistical Forecast Project Rep. 1, 49 pp.

Manikin, G., A. Bentley, S. Shields, C. MacIntosh, L. Dawson, M. Caron, and P. Papin, 2021: Review of the GFSv16 Upgrade. [Available online at https://www.emc.ncep.noaa.gov/ users/meg/gfsv16/pptx/MEG_3-18-21_GFSv16_Review.pptx].

Maxson, B., 2019: Public Information Statement 19-09, 4 pp. [Available online at https://www.weather.gov/ media/notification/pns19-09gfs_v15_1.pdf].

National Oceanic and Atmospheric Administration, 2021: The Global Forecast System (GFS). [Available online at https://www.emc.ncep.noaa.gov/emc/pages/ numerical_forecast_systems/gfs/ documentation.php].

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.

Nowotarski, C., and A. Jensen, 2013: Classifying proximity soundings with self-organizing maps toward improving supercell and tornado forecasting. *Wea. Forecasting*, **28**, 783–801, doi:10.1175/WAF-D-12-00125.1.

Obukhov, A. M., 1947: Statistically homogeneous fields on a sphere. *Usp. Mat. Nauk*, **2**, 196–198.

Pope, M., C. Jakob, and M. Reeder, 2009: Objective classification of tropical mesoscale convective systems. *J. Climate*, **22,** 5797–5808.

Rousseeuw, P. J., 1987: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comp. and Appl. Mathematics*, 53–65.

Stensrud, D. J., 2007, *Parameterization Schemes: Keys to Understanding Numerical Weather Prediction Models*, Cambridge University Press, 480pp.

Stith, J. L. and Coauthors, 2018: 100 Years of Progress in Atmospheric Observing Systems, *Meteorological Monographs*, **59**, 2.1–2.55, https://doi.org/10.1175/AMSMONOGRAPHS- D-18-0006.1.

Thanh Nga, P. T., P. Thanh Ha, P., and V. Thanh Hang, 2021, Satellite-BSED regionalization of solar irradiance by *k*-means clustering, *J. Appl. Meteor. Climatol.*, **60**¸ 391–402, https://doi.org/10.1175/JAMC-D-20-0070.1.

Thompson, D. W. J, and J. M. Wallace, 1998: The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophys. Res. Lett.*, **25**, 1297–1300.

Wheeler, M. C., and H. H. Hendon, 2004: An all-season real-time multivariate MJO index: development of an index for monitoring and prediction. *Mon. Wea. Rev.*, **132**, 1917–1932.

Wilks, D.S, 2019: *Statistical Methods in the Atmospheric Science*, Elsevier, 840pp.