Presented at the 20th Conference on Integrated Observing and Assimilation Systems
for the Atmosphere, Oceans and Land Surface (IAOS-AOLS),
New Orleans, LA, January 11-14, 2016,
Amer. Meteor. Soc., Poster 262

262 OBSERVING SYSTEM EXPERIMENTS IN THE DALLAS-FORT WORTH TESTBED

Frederick H. Carr[1,2], Matthew T. Morris*[1,2], Andrew P. Osborne[1,2], and Keith A. Brewster[1]

[1]Center for Analysis and Prediction of Storms
[2]School of Meteorology
University of Oklahoma
Norman, OK 73072

A. INTRODUCTION

In 2009, the National Research Council (NRC) published a report titled "Observing Weather and Climate from the Ground Up: A Nationwide Network of Networks" (NRC 2009). The report noted insufficiencies in the spatiotemporal resolution of mesoscale observations in the U.S., particularly those observing the three-dimensional mesoscale structure of the atmosphere. A key recommendation of the report was for a nationwide network of networks to be formed by integrating existing mesoscale observing systems with new systems.

The National Mesonet Pilot Program was founded by the National Oceanic and Atmospheric Administration (NOAA) in 2013, with the overarching goal of the program being to extend surface-based observing systems to land-based platforms (Dahlia 2013). Global Science & Technology (GST) was selected as the private-sector manager of National Mesonet and contributes a new observing system known as Mobile Platform Environmental Data (MoPED), observations from instruments mounted on trucks and other vehicles. These vehicles have been fitted with sensors that measure such variables as temperature, humidity, pressure, and precipitation. Turbulence due to vehicle motion prevents the sensors from measuring wind speed and direction, while pressure measurements must be corrected to account for motion. Active vehicles collect data approximately every ten seconds.

Previous studies have assessed the impact of increasing the spatiotemporal resolution of mesoscale observations on high-resolution model analyses and forecasts. High-impact events such as the Cleburne, Texas tornado which occurred on 15 May 2013 are important to predict accurately in order to save lives. One way to improve the forecasts of these types of events may be to integrate new observation data sets from unconventional sources. Carlaw et al. (2015) found that the inclusion of Earth Networks WeatherBug and Citizen Weather Observer Program (CWOP) surface data were instrumental in both the generation and correct placement of an intense low-level mesocyclone in the Cleburne case. Carlaw et al. (2015) utilized observing system experiments (OSEs), denying each of the non-conventional data sets individually, to examine their impact on the forecasts and analyses. The model used was the Advanced Regional Prediction System (ARPS), using the ARPS three-dimensional variational (3DVAR) with cloud analysis and Incremental Analysis Updating data assimilation.

In the future, observing system experiments (OSEs) will be performed to evaluate the impact of assimilating GST MoPED observations on high-resolution analyses and forecasts of convection. However, given the high temporal resolution of MoPED observations, data thinning is necessary to avoid assimilating more observations than are necessary for accurate mesoscale analyses. Knopfmeier and Stensrud (2013) compared analyses that incorporated additional mesoscale observations to the Real-Time Mesoscale Analysis (RTMA) and found that removing 75% of the observations had little impact on the resulting analyses in that system.

This paper reports on research projects addressing Network of Networks issues in two tracks. The first track is an extension of Carlaw et al. (2015), with the difference coming in the forecast model utilized. This track uses the Weather Research and Forecasting (WRF) model to carry out the OSEs rather than the ARPS model. The purpose of the second track is to determine the optimal averaging time period for implementing MoPED observations into similar OSEs. A hypothesis test is performed to assess whether 2-min., 3-min., 4-min., and 5-min. averages are statistically different from averages using a shorter 1-min window.

*Corresponding author address: Matthew T. Morris
120 David L. Boren Blvd., Suite 2500,
Norman, OK 73072, mtmorris@ou.edu

The paper is organized as follows: part B will address the first track (Cleburne OSE), while part C will address the second track (mobile observation thinning).

B1. OSE CASE BACKGROUND

15 May 2013

An environment favorable for discrete supercell thunderstorms developed on this day in the Dallas-Fort Worth metro area as a low-amplitude mid-level trough moved out of western Texas on 14 May 2013 into the Texas panhandle on the morning of 15 May 2013. A moist, unstable environment developed in north central Texas ahead of a dry line during the afternoon hours of 15 May. Clearing skies combined with the upper level forcing to promote ascent, eventually lead to discrete thunderstorm development. With easterly surface winds turning to 20-25 ms^{-1} south-southwesterly winds at 850 hPa there was sufficient wind shear to sustain supercellular thunderstorms in this environment. One of those supercells, which is the subject of this study, developed over Erath County between 2230 and 2330 UTC on 15 May. A low-level mesocyclone developed and began to intensify from 0100 UTC to 0200 UTC. An EF0 tornado developed from this mesocyclone near Cleburne at 0212 UTC, continuing to intensify to an EF3 tornado with a maximum path width of one mile as it passed over Lake Pat Cleburne.

B2. DATA AND METHODS

WRF model

The Advanced Research version of the Weather Research and Forecasting Model (WRF-ARW) was used to generate high resolution simulations of the convective environment. WRF-generated 3 km and 400 m forecasts were used as background for 3DVAR data assimilation cycling on a 3 km outer and a 400 m inner nested grid. At the end of the 400 m grid assimilation cycle, a 30 minute WRF forecast was generated to assess the effect of the new observations on the ability of the model to accurately predict the storm evolution.

Version 3.6 of the WRF-ARW was run with no cumulus parameterization for both of the nested grids. The physics suite included the Lin et al. (1983) single-moment microphysics scheme and the Mellor-Yamada-Nakanishi-Niino (MYNN) Planetary Boundary Layer scheme (Nakanishi and Niino 2006). The longwave radiation scheme implemented was the Mlawer et al. (1997) Rapid Radiative Transfer Model (RRTM) scheme along with the Goddard shortwave radiation scheme (Chou and Suarez 1999).

The experiments were performed on a one-way nested grid. The outer 3 kilometer horizontal resolution grid had dimensions of 900 km x 900 km while the inner grid had dimensions of 193 km x 161 km. The 3 km domain is centered at 32.7°N, 97.55°W, while the 400 m center is 32.65°N, 97.2°W (Figure 1). Both domains had 53 vertical levels with the vertical grid spacing stretched via a hyperbolic tangent function (Xue et al. 1995). The U.S. Geological Survey 30 second terrain-elevation dataset was used for topographic information in both domains.

Two experiments were run in this case, named CTL and NONEWSFC. The CTL experiment assimilated all conventional (radiosonde, profiler, radar, mesonet, aircraft) and non-conventional (AWS WeatherBug, CWOP, and GST MoPED) observations. NONEWSFC assimilated only the conventional observations, denying the non-conventional observations from impacting the forecasts and analyses. The experimental setup is summarized in Table 1.

ARPS 3DVAR

The ARPS 3DVAR (Gao et al. 2004) package with cloud analysis was used to analyze the observations. Hourly forecasts were run on the outer domain from an initial 13 km Rapid Refresh (RAP; Brown and Coauthors 2012) background field at 2100 UTC 15 May 2013 to 16 May 2013 0400 UTC with 3DVAR assimilation of conventional surface and upper air observations at the top of each hour. RAP forecast fields provided lateral boundary forcing at the top of each hour on the outer 3 km grid. The inner 400 meter domain had 5 minute cycling starting from 16 May 2013 at 0100 until 0145 UTC with externally forced boundary conditions coming from the outer 3 km grid forecast. At the completion of the inner grid cycling, a 30 minute forecast was launched from 0145 UTC on the 400 m grid.

The data were assimilated in several passes, with the synoptic observations (profilers and radiosondes) coming in the first pass with the largest horizontal decorrelation length scale. There were successively smaller passes for the more densely clustered observation datasets with the radar data being assimilated on the final pass with the smallest decorrelation length scale.

Observation Quality Control

During the assimilation process, observations are checked against the background values as well as with their neighboring observed values. A time consistency check is also performed on the

observations. If large discrepancies are found at any point during these checks the observation is thrown out from the assimilation process. The threshold value for maximum allowable difference is defined by the user within the ARPS 3DVAR program. Conventional observations go through quality control checks internally via the Meteorological Assimilation Data Ingest System (MADIS) system (NWS 1994), while non-conventional observations go through additional screening specifically checking for low wind speed biases related to poor siting or transmission issues.

B3. RESULTS

Examination of differences in CTL and NONEWSFC experiments will be used to indicate the potential value of these new observation types. The impact of the new observations on the 3DVAR analyses will be discussed first, followed by a qualitative comparison of the WRF forecasts of the storm.

Analysis Accuracy

Root mean squared error (RMSE) is calculated on the 400 m analysis grid using observations from ten independent Automated Surface Observing System (ASOS) sites located within the DFW metro and surrounding areas. RMS errors of temperature and dew point are computed for the two experiments every 5 minutes over the 45 minute analysis period from 0100 to 0145 UTC. Figure 2a shows the temperature RMSE, with the red and blue lines corresponding to the NONEWSFC and CTL experiments, respectively. A similar analysis for dew point temperature is shown in Figure 2b. Both variables show a pattern of greater RMS error of the analysis for the NONEWSFC experiment compared to the CTL experiment. Also of note, the difference between the ASOS and the analyses increases with increasing time into the analysis cycle, and the increase is more pronounced for temperature than for dew point. The increased discrepancy between the ASOS and the analyses omitting the non-standard data is an interesting feature that may indicate the effect of these non-conventional observations grows with time.

There is a clear signal that inclusion of the non-conventional observations is providing a positive impact on the analyses produced in the cycled experiments for this case. The statistical significance of this impact on RMSE still needs to be evaluated in order to draw more definitive conclusions for this case.

400m Forecasts

Simulated radar reflectivity from the WRF forecast output is used to assess how these differences in analyses translated to the forecast of the strongest storm within the 400 m domain for the first fifteen minutes of the free forecast, 0145 UTC to 0200 UTC. At 0145 UTC, the time of the final analysis, both CTL and NONEWSFC show a storm in northern Johnson County with a similar hook echo structure in the reflectivity field (Figures 3a and 3b). The wind field does indicate a stronger cyclonic circulation in the vicinity of the hook echo for the CTL case at this time. As the forecast length increases the differences in the structure of the storm, evident in the reflectivity field, increase.

At 0150 UTC, the CTL forecasted reflectivity field shows something that resembles a supercell thunderstorm with a hook echo, inflow notch indicated by a bounded weak echo region, and a higher reflectivity region to the north and east of the area of rotation (Figure 3c). The reflectivity field from the NONEWSFC experiment at this time shows a thunderstorm with weak, if any, signs of rotation (Figure 3d). Comparison with the observed reflectivity field in this area at this time (Figure 4a) shows better agreement with the structure simulated in CTL than that in NONEWSFC.

By 0155 UTC, the forecasted supercell in CTL continues to mature in a similar manner to the observed storm (Figures 3e and 4b), while the NONEWSFC forecast shows a weaker storm with very little signs of rotation (Figure 3f). The differences continue at 0200 UTC as the control storm continues to look supercellular in nature (Figure 3g), potentially capable of producing the tornado that was observed around 0215 UTC. The NONEWSFC storm continues to lose its organization and does not look capable of producing a tornado in the near future (Figure 3h).

The WRF forecast for the storm in this case appears to be dependent on the assimilation of the additional non-conventional observations to make the analysis and subsequent forecast more accurate. Without the new observations, the forecast fails to capture the supercellular nature of the storm that is observed. This can likely be tied back to errors in the analysis of the thermodynamic fields for the NONEWSFC experiment. The CTL experiment with the extra observations assimilated was able to more accurately depict the environment surrounding the storm which likely led to the model better capturing the strength, structure, and evolution of the storm.

B4. CONCLUSION

This case study attempts to expand on the findings of Carlaw et al. (2015), applying a similar experimental setup, but using the WRF model in place of the ARPS model. There appears to be a similar impact of the non-conventional observations within this new framework, although each of the new datasets were not tested individually and the statistical examination is not yet as thorough as that of Carlaw et al. (2015). Nevertheless, there were clear differences in the storm structure, evolution, and intensity between the forecasts with and without the non-conventional observations as shown via the reflectivity and wind fields. The experiment assimilating all the non-conventional observations also proved to produce more accurate 3DVAR analyses of the environmental thermodynamic structure, evident in the examination of RMS error values over the assimilation cycling period. In the future, we would like to expand on the analysis presented here with further statistical verification methods for this case, along with examination of the impact of each individual new data set on its own in further experiments. To obtain more robust results enabling more definitive conclusions to be drawn about the pertinence of these non-conventional data sets, data denial experiments can be run over longer time periods (i.e. 1 month) to capture a variety of weather conditions. Also, Ensemble Kalman Filter (EnKF) data assimilation will be used and compared to 3DVAR in future observing system experiments.

C1. TEST CASE

5 November 2015

A severe weather event on 5 November 2015 could serve as a potentially useful test case to examine the forecast impact of GST MoPED observations. Figure 5a shows a surface analysis valid at 21Z on 5 November 2015. A 992 hPa surface low pressure system was positioned over Canada, just northwest of Lake Superior. An associated cold front extended south through Kansas and into the Oklahoma panhandle. A dryline was analyzed ahead of this cold front, extending from eastern Kansas through central Oklahoma and into central Texas. Convection formed along this surface boundary, some of which would become severe, as evidenced by severe hail and wind reports from the Storm Prediction Center (SPC; Figure 5b). The high spatiotemporal resolution of the MoPED observations could prove useful in this case given the sharp gradient in dew point temperatures across the dryline.

C2. DATA AND METHODOLOGY

Data thinning methodology

GST MoPED data were obtained for the period from 1700 to 2200 UTC for 5 November 2015. The domain considered is shown in Figure 6, with latitude values ranging from 31.5 to 33.9 degrees N and longitude values ranging from 96.0 to 98.4 degrees W. Twelve trucks passed through the domain during the specified time period, although not all are analyzed in this study.

The data were then thinned for five different lengths of time. The data were grouped based on truck identifier and averages computed using data from individual trucks. One-minute averages were generated using data for a single minute, with the averaged observation assigned to the 30-second mark (e.g., data from 17:00:00 to 17:00:59 were averaged into a single observation for 17:00:30). A second methodology for computing 1-minute averages is to consider data centered about the beginning of a minute (e.g., data from 17:00:30 to 17:01:30 were averaged and assigned to 17:01:00). Two-minute averages were computed by dividing the time period into 2-minute periods (e.g., observations from 17:00 to 17:01:59 were assigned to 17:01:00. Three-minute averages were computed by considering 3-minute windows (e.g., observations from 17:00:00 to 17:02:59 were averaged and assigned to 17:01:30). The inherent differences in how the observation time is assigned for the 2-, 3-, 4-, and 5-minute averages explains the necessity for computing 1-minute averages in two different ways. Three-minute and 5-minute averages are compared to 1-minute averages from the first approach, whereas 2-minute and 4-minute averages are compared to 1-minute averages derived using the second methodology.

After the data thinning was complete, the thinned results were analyzed for completeness and trucks with only intermittent observations were discarded from the data set. Three distinct time periods emerged from this analysis. The period considered in this study is from truck CW14L, with a time span from approximately 1950 to 2150 UTC. There was a brief disruption to the data reporting, with data missing for a roughly 1-minute interval at 2022 UTC. Averaged results from time periods that span this missing window were excluded from this study.

Hypothesis testing

A hypothesis test is performed to compare the results of the various averaging schemes and determine if the schemes produce statistically different results. The null hypothesis (H_0) is that the

difference between the mean temperature (°C) when averaging over 2-minute windows and 1-minute windows is equal to zero. The alternative hypothesis (H_A) is that the difference between the two means is non-zero. A two-sided hypothesis test is performed, as the longer averaging periods could yield over- or under-estimates of the 1-minute averages. The level of significance is set to 90%. This procedure is then repeated to compare 3-, 4-, and 5-minute averages to the 1-minute averages, as well as assess differences in averaging schemes for dew point temperature (°C) and surface pressure (hPa).

Permutation testing

The hypothesis tests described above will be performed using permutation testing (Wilks). Although the idea for permutation testing was introduced by Pitman in 1937, it was impractical without advanced computing. The idea resurfaced in the early 1980s with papers from Mielke (1981) and Preisendorfer and Barnett (1983). A key principle of permutation testing that holds under the null hypothesis is exchangeability, or the idea that the data from both samples originate from the same distribution and the labels attributing values to a given data set are arbitrary.

To outline the process of permutation testing, consider the case comparing 2-minute averages of temperature to the respective 1-minute averages for a one-hour period with complete data. Each data set contains 30 observations (i.e., $n_1 = n_2 = 30$), as 1-minute averages without a matched observation were not included. Once all observations are pooled together, the resulting sample size is $n = n_1 + n_2 = 30 + 30 = 60$. The samples are then selected without replacement from the resulting pool and placed into one of two artificial subsets, each containing n = 30 observations. The sample mean is recorded for each subset and the difference is computed. The number of possible permutations is given by:

$$\frac{n!}{n_1! * n_2!}.$$

The permutation process is repeated 999 times for each hypothesis test that is performed, with the null hypothesis being that the sample means are statistically the same as for the 1-min averages.

After the permutations have been computed, the mean difference between the permuted subsets is computed. The standard deviation of the differences in means of the permuted subsets is reported. A p-value is reported, which indicates how likely it is to observe the mean difference assuming that the samples are drawn from the same population. This p-value is then used to determine whether or not to reject the null hypothesis.

C3. RESULTS

Table 2 reports the mean of temperature, dew point temperature, and surface pressure for the 1-minute averages and the longer averaging schemes (e.g., 2 minutes) for the time period from 1950 to 2150 UTC on 5 November 2015 for truck denoted CW14L. The difference between these means is also shown in the tables.

Time-series plots of temperature, dew point temperature, and surface pressure for the various averaging schemes are shown in Figure 7, while Figure 8 displays histograms for the results of the permutation testing. For this data set, the 2-minute, 3-minute, 4-minute, and 5-minute averaging schemes for temperature and surface pressure are shown to be statistically equivalent to the 1-minute averaging schemes. The 3-minute averages of dew point temperature are statistically different from the 1-minute averages and the null hypothesis is rejected (p-value is 0.056). The null hypothesis is not rejected for the 2-minute, 4-minute, and 5-minute averaging lengths. It can be seen in Figure 8 that the differences in mean for the permuted samples become more varied as the length of the averaging window increases, though it still remains very small, in the fourth or fifth significant digit, and below the expected measurement error. A corresponding increase in standard deviation can also be seen in Table 2.

C4. CONCLUSIONS

For the data set considered in this study, no statistical difference was found between the 2-minute and 1-minute averaging regimes. For this case study alone, the results indicate that statistically different averages are found once the averaging time length is increased to 3 minutes and above. However, it is important to note that this experiment only considers data over a limited domain for a specific date. The methodology used in this study should be extended to an increased number of trucks over a wider domain for varied dates to determine if additional patterns emerge. Regardless of any patterns that may emerge, the data used in any case study should be examined thoroughly for any obvious patterns that would diminish the benefit of using thinning by averaging.

While there are no major deviations in temperature and dew point temperature throughout the time periods considered, the surface analysis in Figure 5a shows a dryline approaching the domain

used in this study (Figure 6). If a vehicle were to intersect a dryline, the dew point temperature would be subject to rapid fluctuations that may impact the applicability of longer averaging time scales. This is especially true in cases where trucks are traveling at highway speeds (up to 120 kilometers per hour).

For the data set considered, as the length of the averaging window increased (e.g., from 2 minutes to 5 minutes), the standard deviation of the differences in permuted means increased. This increase in standard deviation could be attributed to the relatively small sample size overall and a decreased number of observations in the original data sets as the averaging time period is increased. For instance, the number of observations compared (Table 2) is 59 when 2-minute averages are tested against 1-minute averages and 22 when 5-minute averages are tested.

A major limiting factor of the averaging schemes used in this research is that they only rely on elapsed time and do not take the vehicle's speed and position into account when computing averages. Since truck speed can vary from stationary to full highway speed (upwards of 120 km/h), the distance a truck travels in the averaging window varies considerably from sample to sample. Moving forward, a spatial averaging regime will be devised that takes into account both the accumulated distance and time elapsed between observations by considering a vector "distance" in both time and space. Once this distance reaches a set threshold, those observations will be averaged together, and a new series started. Similar statistical tests could be performed to assess how well lower-resolution averaging schemes perform in comparison to their higher-resolution counterparts. Averaging including the effects of distance could help alleviate the potential discrepancies in longer averaging time lengths, which could be especially problematic when a vehicle passes through an air mass boundary at a high speed. Once vehicle motion is taken into account, the study will be repeated with a larger sample covering more cases.

Table 1: Cleburne case study experimental setup

Experiment	Conventional	Non-conventional
CTL	ASOS/AWOS, MDCRS, Raobs, Profilers, Mesonet, WSR-88D, TDWR, CASA XUTA	GST MoPED, AWS WeatherBug, CWOP stations
NONEWSFC	ASOS/AWOS, MDCRS, Raobs, Profilers, Mesonet, WSR-88D, TDWR, CASA XUTA	None

Table 2: Results for truck CW14L from 1950 to 2150 UTC

Comparison of 1 minute and 2 minute averaging schemes (n = 59):

	Temperature	Dew point temp.	Surface pressure
Mean of 1 minute averages	29.426	23.264	986.910
Mean of 2 minute averages	29.422	23.266	986.924
Difference in means	-0.004	0.002	0.014
Standard deviation	0.0030	0.0035	0.0197
Two-sided p-value	0.216	0.660	0.472
Decision	Do not reject	Do not reject	Do not reject

Comparison of 1 minute and 3 minute averaging schemes (n = 39):

	Temperature	Dew point temp.	Surface pressure
Mean of 1 minute averages	29.396	23.279	986.939
Mean of 3 minute averages	29.401	23.264	986.946
Difference in means	0.005	-0.015	0.007
Standard deviation	0.0068	0.0075	0.0250
Two-sided p-value	0.524	0.056	0.832
Decision	Do not reject	Reject	Do not reject

Comparison of 1 minute and 4 minute averaging schemes (n = 28):

	Temperature	Dew point temp.	Surface pressure
Mean of 1 minute averages	29.500	23.290	987.018
Mean of 4 minute averages	29.481	23.278	986.982
Difference in means	-0.019	-0.012	-0.036
Standard deviation	0.0149	0.0085	0.0385
Two-sided p-value	0.184	0.202	0.434
Decision	Do not reject	Do not reject	Do not reject

Comparison of 1 minute and 5 minute averaging schemes (n = 22):

	Temperature	Dew point temp.	Surface pressure
Mean of 1 minute averages	29.555	23.280	986.932
Mean of 5 minute averages	29.538	23.279	986.891
Difference in means	-0.017	-0.001	-0.041
Standard deviation	0.0192	0.0103	0.0579
Two-sided p-value	0.390	0.962	0.598
Decision	Do not reject	Do not reject	Do not reject

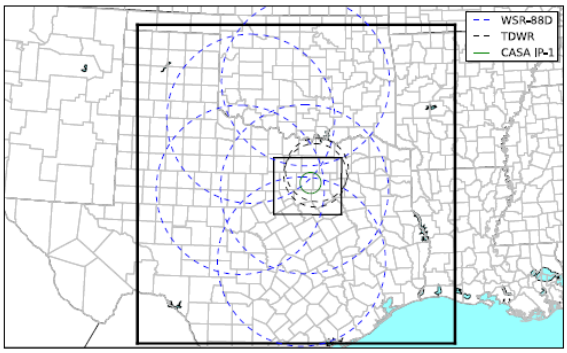
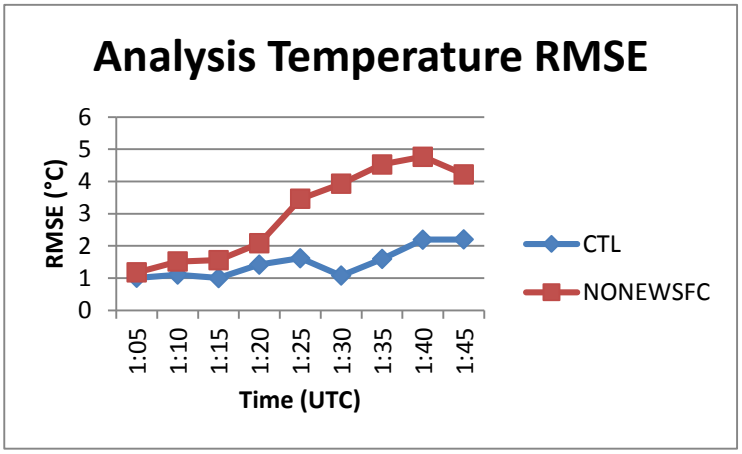


Figure 1: Depiction of experimental domains (black boxes) along with WSR-88D radar range rings (blue dotted lines). The outer box is the 3km domain while the inner box compromises the 400m domain. The CASA (Collaborative Adaptive Sensing of the Atmosphere) radar location is indicated by a green circle.

(a)



(b)

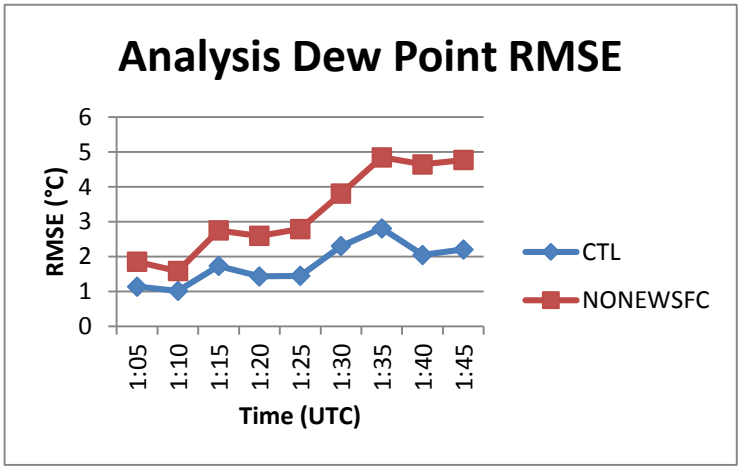


Figure 2: Plots of RMS error of the 3DVAR analysis every 5 minutes, based on 10 independent ASOS stations. Temperature calculations shown in a) and dew point calculations shown in b). CTL experiment values are shown in blue while NONEWSFC values are shown in red.

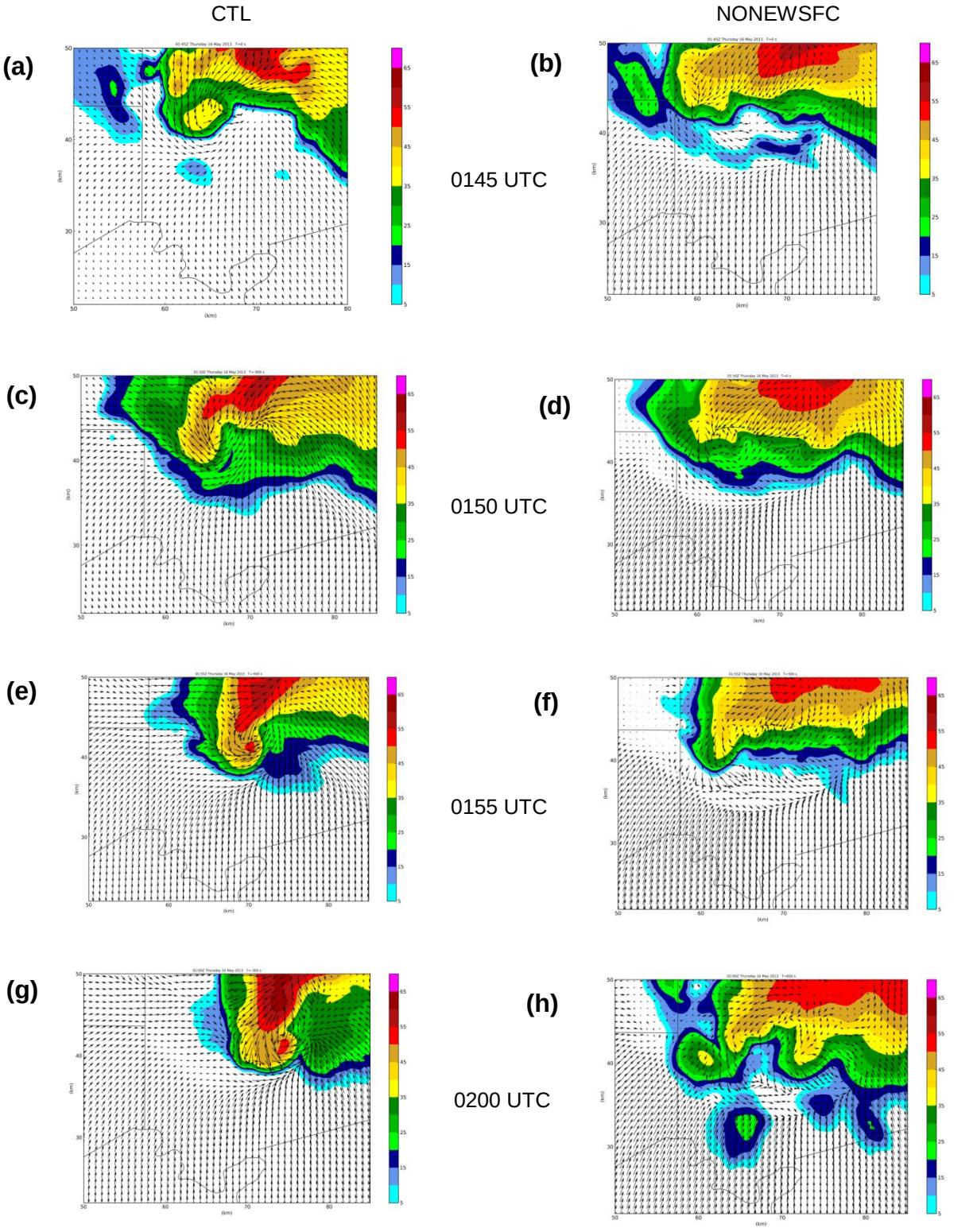
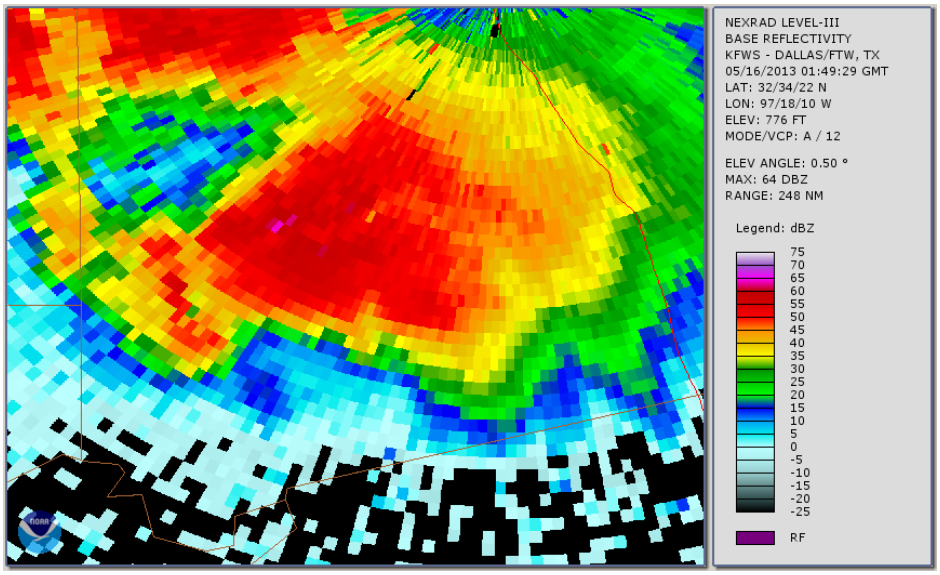


Figure 3: Plots of WRF simulated reflectivity and wind field over Johnson County for CTL (a, c, e, g) and NONEWSFC (b, d, f, h) from 0145 to 0200 UTC.

(a)



(b)

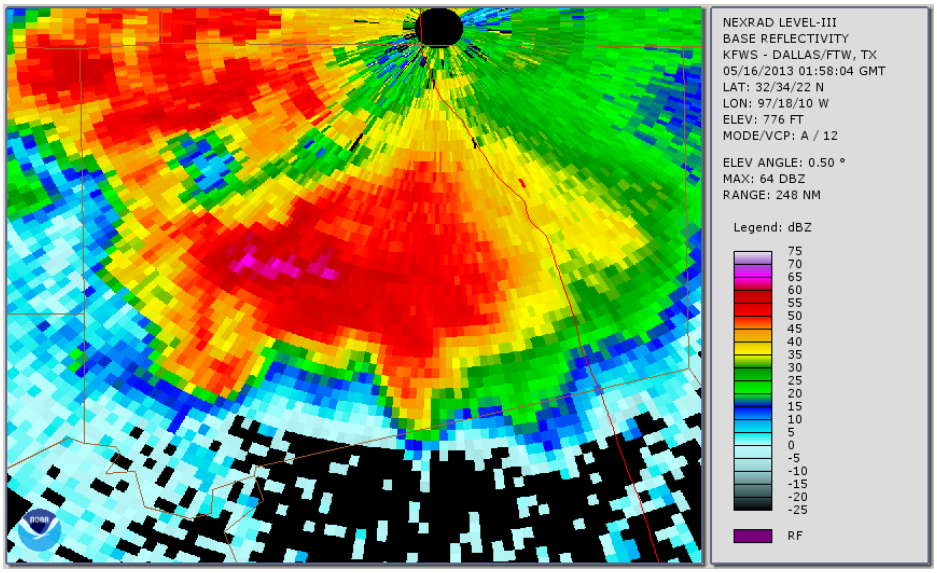
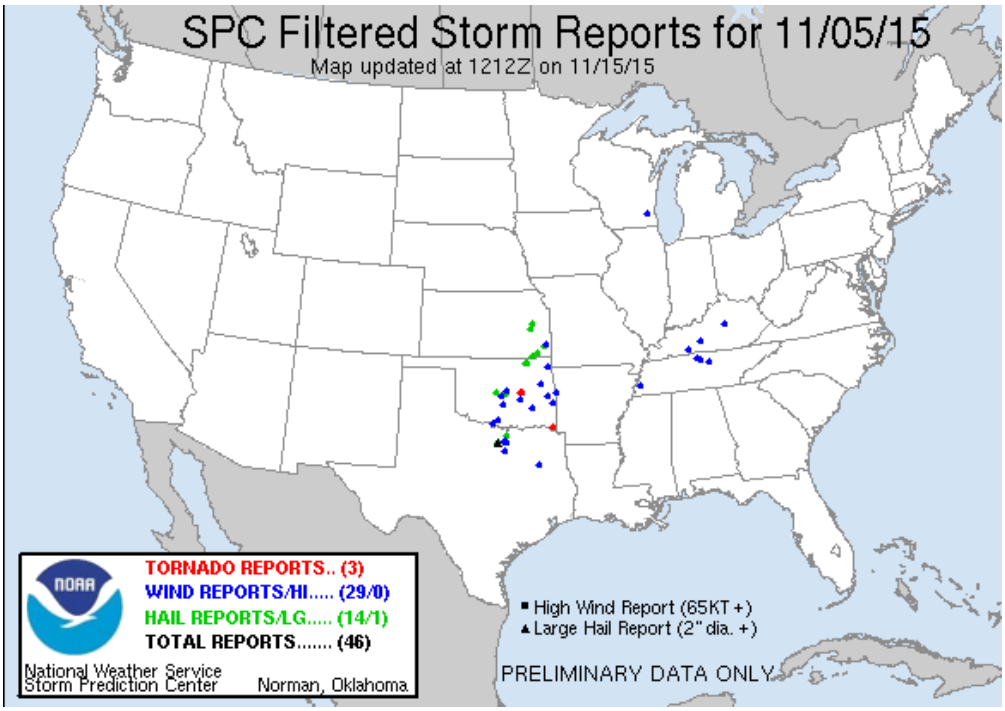


Figure 4: Observed reflectivity at 0.5° elevation angle from the KFWS WSR-88D radar at approximately 0150 UTC (a) and 0155 UTC (b).

(a)



(b)

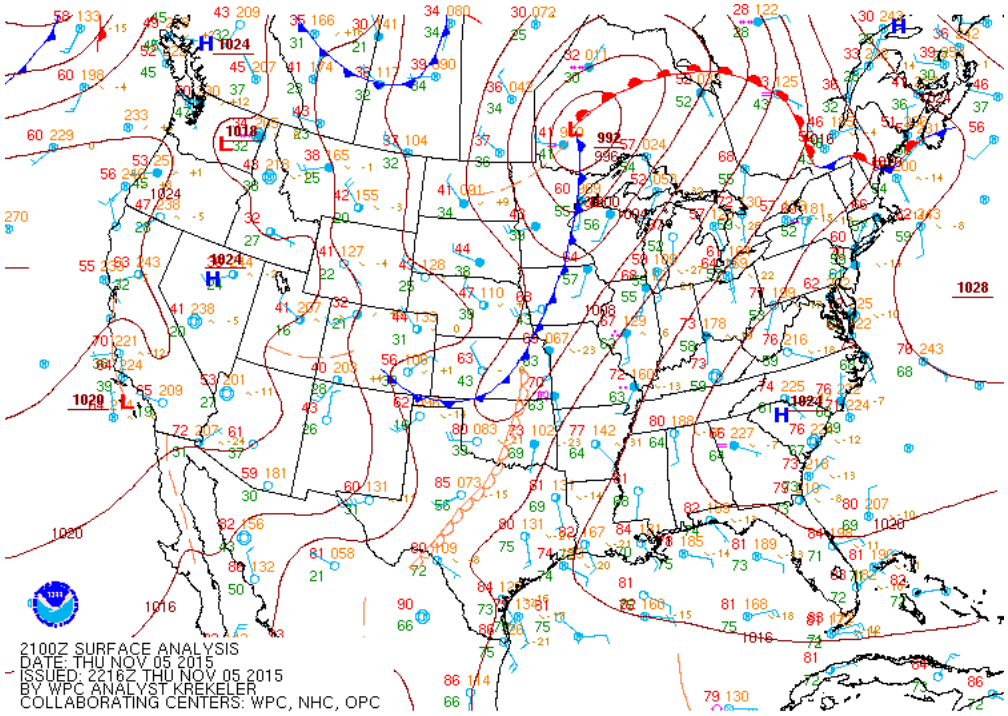


Figure 5: (a) Surface analysis from the Weather Prediction Center (WPC), valid at 2100 UTC on 5 November 2015. (b) Storm Prediction Center (SPC) storm reports from 5 November 2015. Severe hail and wind were both reported in the Dallas-Fort Worth metroplex.

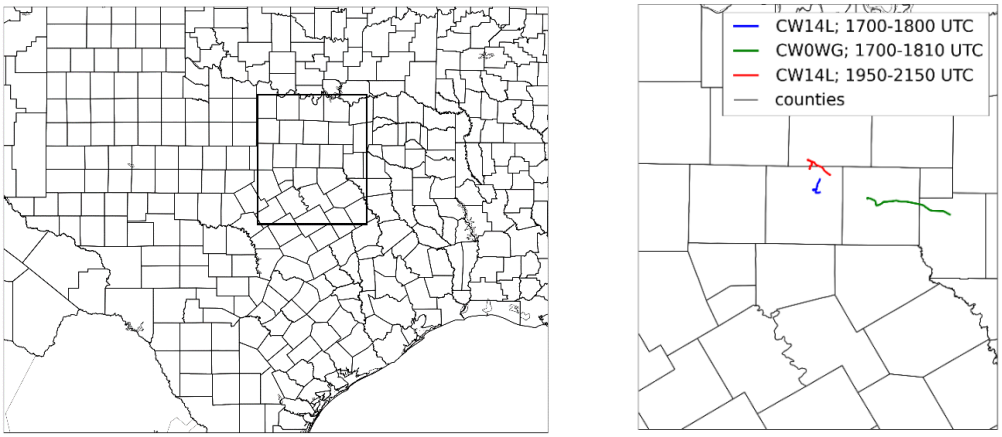


Figure 6: The left panel shows the geographic area considered in this study (outlined in black), which includes Dallas-Fort Worth. The right panel is zoomed in on the area highlighted in the left panel and shows the geographic location of the trucks considered in this study. The data considered in this study are shown by the red line (truck CW14L from 1950 to 2150 UTC).

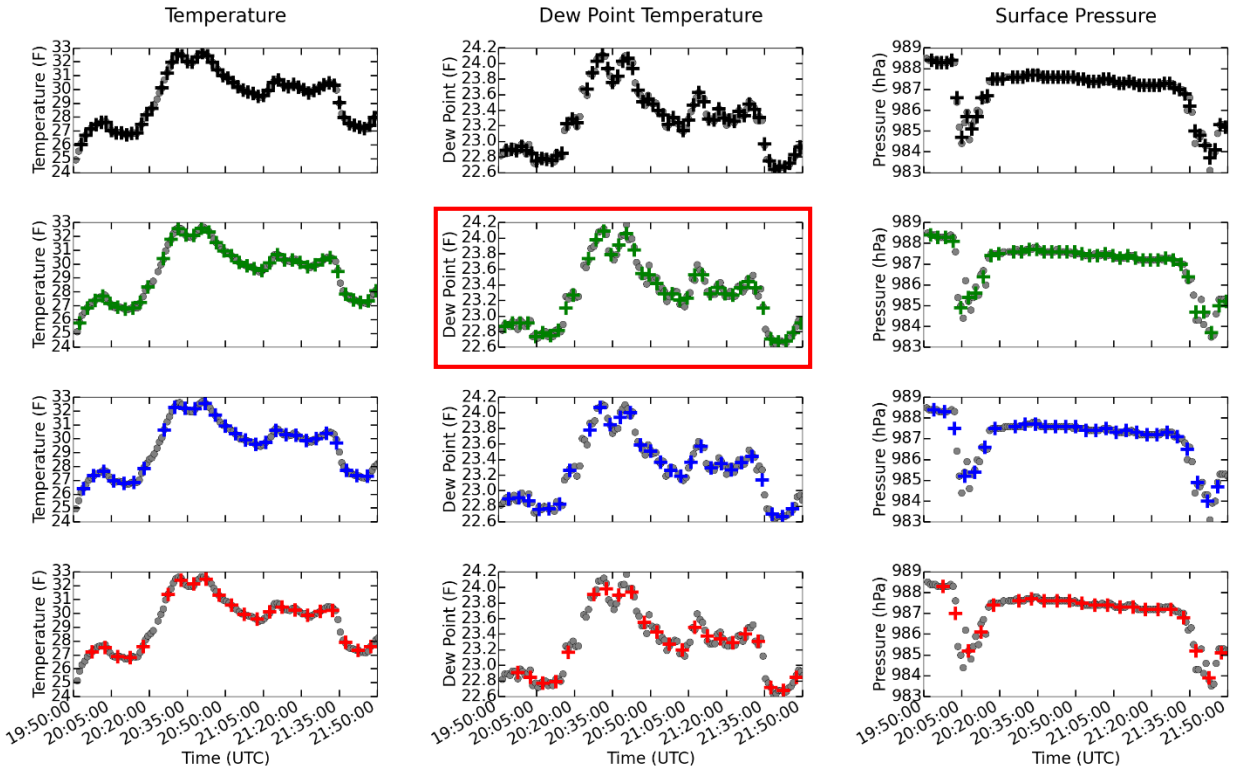


Figure 7: Results of the thinning algorithms for truck CW14L for the period from 1950 to 2150 UTC. The three columns correspond to temperature, dew point temperature, and surface pressure, respectively. The gray dots in each figure correspond to 1-minute averages of the respective variable. Black plus signs in the first row represent results of the 2-minute averaging scheme. Green plus signs in the second row are for 3-minute averages, blue plus signs in the third row represent 4-minute averages, and the red plus signs in the final row correspond to 5-minute averages. A red box indicates that an experiment resulted in statistically different results.

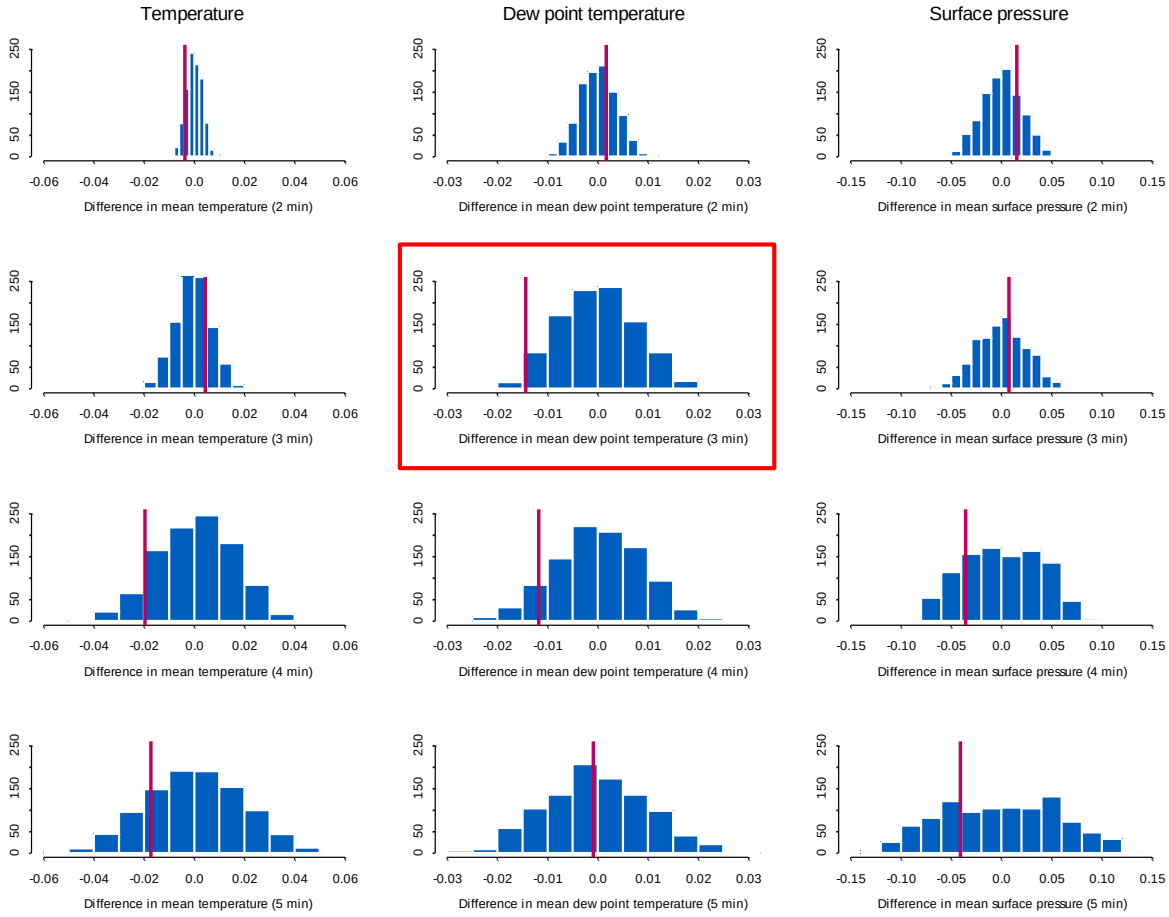


Figure 8: Results of the permutation test for truck CW14L for the period from 1950 to 2150 UTC. The rows correspond to 2-minute, 3-minute, 4-minute, and 5-minute averaging windows, respectively, where each averaging scheme is compared to the more robust 1-minute averaging scheme. The columns represent the differences in permutation means for temperature, dew point temperature, and surface pressure, respectively. The vertical red line represents the difference in mean between the two original data sets (e.g., 1-minute average temperature vs. 2-minute average temperature). A red box indicates that an experiment resulted in statistically different results.

References

Brown, J., and Coauthors, 2012: Rapid Refresh replaces the Rapid Update Cycle at NCEP. preprints. *2012 Canadian Meteorological and Oceanographic Society Congress/21st Conf. on Numerical Weather Prediction/25th Conf. on Weather and Forecasting.*

Carlaw, L. B., J. A. Brotzge, and F. H. Carr, 2015: Investigating the impacts of assimilating surface observations on high-resolution forecasts of the 15 May 2013 tornado event. *Electronic J. Severe Storms Meteor.*, 10 (2), 1–34.

Chou, M. D., and M. J. Suarez, 1999: A solar radiation parameterization for atmospheric studies. *NASA Tech. Memo.* 104606, **15**, 40 pp.

Dahlia, J., 2013: The National Mesonet Pilot Program: Filling in the Gaps, *Weatherwise*, **66.4**, 26-33.

Gao, J., M. Xue, K. Brewster, and K. K. Droegemeier, 2004: A three-dimensional variational data analysis method with recursive filter for Doppler radars. *J. Atmos. Oceanic Technol.*, 21, 457–469.

Knopfmeier, K. H. and D. J. Stensrud, 2013: Influence of Mesonet Observations on the Accuracy of Surface Analyses Generated by an Ensemble Kalman Filter. *Wea. Forecasting*, **28**, 815–841.

Lin, Y.-L. R. D. Farley, and H. D. Orville, 1983: Bulk parameterization of the snow field in a cloud model. *J. Climate Appl. Meteor.*, 22, 1065–1092.

Mielke, P. W., Jr., K. J. Berry, and G. W. Brier, 1981: Application of Multi-Response Permutation Procedures for Examining Seasonal Changes in Monthly Mean Sea-Level Pressure Patterns. *Mon. Wea. Rev.*, **109**, 120–126.

Mlawer, Eli. J., Steven. J. Taubman, Patrick. D. Brown, M. J. Iacono, and S. A. Clough (1997), Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated–k model for the longwave. *J. Geophys. Res.*, 102, 16663–16682.

Nakanishi, M., and H. Niino, 2006: An improved Mellor–Yamada level 3 model: its numerical stability and application to a regional prediction of advecting fog. *Bound. Layer Meteor.* **119**, 397–407.

National Research Council. *Observing Weather and Climate from the Ground Up: A Nationwide Network of Networks.* Washington, DC: The National Academies Press, 2009.

NWS, 1994: Technique specification package 88-21-R for AWIPS-90 RFP. Appendix G requirements numbers: Quality control incoming data, AWIPS Doc. TDP-03201992R2, NOAA/National Weather Service Office of Systems Development, 39 pp. [Available online at http://docs.lib.noaa.gov/noaa_documents/NWS/NWS_TSP_88-21-R2.pdf.]

Pitman, E. J. G., 1937: The “closest” estimates of statistical parameters, *Mathematical Proceedings of the Cambridge Philosophical Society*, **33**, 212-222.

Preisendorfer, R. W. and T. P. Barnett, 1983: Numerical Model-Reality Intercomparison Tests Using Small-Sample Statistics. *J. Atmos. Sci.*, **40**, 1884–1896.

Wilks, D. S. Statistical Methods in the Atmospheric Sciences. Elsevier, 2011.

Xue, M., K. K. Droegemeier, V. Wong, A. Shapiro, and K. Brewster 1995: ARPS Version 4.0 710 User's Guide, 380 pp. [Available online at http://www.caps.ou.edu/ARPS].