An Examination of the Accuracy of Objective Analysis Schemes

Alexander Ingalls[1,2] and Bradley G. Illston[3]
[1]National Weather Center Research Experiences for Undergraduates Program, Norman, Oklahoma USA
[2]University of Maine, Farmington, Maine, USA
[3]University of Oklahoma, Norman, Oklahoma USA

ABSTRACT

The properties of errors associated with objective analysis models are examined in terms of atmospheric variables to obtain a better understanding of the accuracy of such methods. Several schemes have been developed since the late 1940s which are used for weather forecasting and diagnosis. This information is ultimately conveyed to the public; therefore, an examination of their accuracy is prudent. A removal and replacement technique is utilized to compare interpolated values to their corresponding observational values. The observation data are collected from a network of mesoscale weather stations from which three variables, air temperature, relative humidity, and wind speed, are analyzed. In order to determine some metric of accuracy, errors are considered to be interpolated values that exceed the observation instrument's calibration range. These errors are then investigated further through a statistical analysis. The resulting analysis shows three levels of errors. The first and most fundamental level shows a comparison of the magnitude of errors. For a particular day and time, the results showed a root mean squared error for air temperature to be 1.37 C. The second shows the relative amount of error per variable per observation station and geographically where these errors occur most often. The results show that the majority of errors were air temperature (55.62%) and relative humidity errors (40.13%), while wind speed errors account for just 4.25% of errors. The third is an accumulative view of how many errors stations had over 12 candidate days (one per month) over a year with an average root mean squared error of 13 errors for any given Mesonet station. Further analysis of the results suggests that there may be a relationship between so called 'edge-cases' and frequency of errors, where an observation near the boundary of some finite area may not have sufficient input data to perform the interpolation effectively. It was found that 60% of the top ten stations with errors were indeed boundary cases. The results could also suggest that seasonal variability influences the scheme's accuracy, particularly during winter months, however a more robust investigation is likely required.

1. INTRODUCTION

Objective analysis (OBAN) can be defined as a method that "follows a prescribed set of rules, or 'model,' and will produce the same analysis given the same set of data." (McPherson, 1986). Discovering truth from empirical evidence is the overall goal of science, therefore, objective analysis is a natural extension for such efforts. In a generalized view, these schemes are a tool for researchers to automate much of the computation involved with the interpolation process. In the context of this study, an OBAN scheme is examined in terms of atmospheric variables collected from the Oklahoma Mesonet, a statewide mesoscale network of meteorological observation stations. Consequently, a foundational understanding for how such schemes are deployed in the meteorological field is recommended.

One of the primary reasons OBAN schemes are used for weather forecasting or diagnosis, "is to interpolate irregularly spaced observational data to a uniform grid." (Lu and Browning, 1998). It was found that "the type of scheme which has been most successful at doing this is a surface fitting scheme, that is, the method of fitting a geometrical surface to the reported data and calculating the values determined by that surface at any other points of interest, specifically, the grid points. " (Barnes, 1964). Several varying schemes have been developed since "the late 1940s, as numerical weather prediction was getting started" (Richard P. McNulty, 2011). One of the most often used, and the primary study of this paper, is the Barnes scheme.

The Barnes scheme utilizes a mathematical structure founded on the assertion that "the two-dimensional distribution of an atmospheric variable can be represented by the summation of an infinite number of independent harmonic waves, that is, by a Fourier integral

[1] *Correspondence to:* Alexander Ingalls, University of Maine - Farmington, 224 Main St, Farmington, ME 04938, USA, alexander.ingalls@maine.edu

representation." (Barnes, 1964). However, Barnes acknowledges the limitation of the scheme's dependency that the spatial distribution of data be reasonably uniform. One of the main reasons for this is because "sparse data distribution can lead to rapid convergence on observational errors, which can distort the analysis field." (Lu and Browning, 1998). The Barnes scheme utilizes inverse distant weighting to estimate unknown grid point values, which is a process where the sum of nearby data points determines the unknown value by placing weights on the data point values according to their distance from the unknown grid point.

Since this scheme is often used in meteorological analysis, this study looks to build upon previous work (Barnes, 1964; Achtemeier, 1987; Lu, Browning, 1998) by assessing the scheme's behavior. The Oklahoma Mesonet's (Brock et al., 1995; McPherson et al., 2007) 120 stations are relatively uniform in their spatial distribution as well as elevation and are therefore a compelling candidate to use with the Barnes scheme in examination of its accuracy. In accordance with Barnes own assessment of the scheme, it is noted that grid points along the grid boundary may have erroneous estimations due to a lack of observed data in spatial proximity to such boundary grid points (Barnes, 1964). This study explores the properties of errors associated with interpolation, and hypothesizes that if interpolation errors do occur, then they most often occur at boundary grid points.

2. METHODOLOGY

The Barnes scheme is a combination of the successive corrections and inverse distance weight interpolation concepts. The premise of a successive corrections model "involves an iterative procedure. For each iteration, increments between the previous analysis field and the observations are computed at each observation location. These increments are then multiplied by a priori specified weights, summed over all stations, and the result added to the previous analysis field on the analysis grid. The iteration, coupled with a properly chosen weighting function, will converge to the observation values at the station locations." (Lu and Browning, 1998). It was found in later revisions to the Barnes scheme that a three-iteration (pass) or four-pass method rendered better results than the conventional two-pass method seen in Barnes's 1973 revision of the scheme (Barnes, 1994). This study examines the behavior of the Barnes scheme with three passes as to avoid the inherent problem of over-smoothing the analysis field seen with a larger number of iterations (Nuss and Titley, 1994).

In conjunction with successive corrections, inverse distance weighting (IDW) involves a central process within each iteration where unknown grid point values are derived from known data points through "a least-squares fit of the surface to the data with the influence of each datum weighted according to its distance from the grid point." (Barnes, 1964). IDW is limited to a finite area around the examined grid point known as the radius of influence (ROI). Further, the ROI "determines the values of the variable at grid points as the sum of weighted values of the individual data. The closer a data point to the grid point in question, the greater influence the datum at that point exerts." (Barnes, 1964). It was found that using a fixed ROI within each successive correction pass rendered better results than a previous convention based on the premise of "first analyzing for the long wave lengths and then building in the short waves through decreasing the radius of influence on each correction pass through the data." (Achtemeier, 1987). In this study, a fixed ROI value of 5 grid cells is utilized in accordance with previous work that showed superior results using a radius equal to two times the average distance between data points (Achtemeier, 1987).

Concurrent with the radius of influence is an exponential decay function known as the convergence parameter. This function applies a non-linear weight factor to known grid point values within the ROI by multiplying the data points distance from the unknown grid point by the convergence parameter's decay constant (lambda). It was found in Barnes' revision of the original scheme that a lambda value of 0.25 rendered more accurate interpolation values (Barnes, 1993). Therefore, this study uses 0.25 as the convergence parameter value.

The computation to determine an unknown value at a particular grid point can be expressed where the weight factor (1) is applied to each datum value within the ROI and the sum of the weighted values, divided by the sum of their weights is equal to the interpolated value (2).

$$w_j = e^{\left(-\frac{r^2}{4k}\right)} \tag{1}$$

$$X_{i,j} = \frac{\sum_{n=1}^{n} w(r,d) \cdot X_{on}}{\sum w(r,d)} \quad (2)$$

The dimensions of each grid cell are chosen as 10km^2 with a buffer of 10km surrounding the outside edges of the state boundary to ensure grid points are satisfied such that their ROI contains a sufficient observation values to perform the interpolation (Barnes, 1964). Barnes does acknowledge that in some cases, specifically grid boundary cases, there may not be data points within these grid point's ROI. (Barnes, 1964) It should be noted that an appropriate grid cell size is to be used depending on the total area being examined. For instance, a 100km cell size could result in poor resolution for an area the size of a single state but may be appropriate for the continental United States whereas a 10km cell size could be considered more appropriate for the state level. Computational time becomes a factor in accordance with a higher resolution grid. Generally, the higher the resolution, the longer it takes for the interpolation process to complete. An effective ROI is determined in part by grid spacing, it was found that two times the average spacing between data points resulted in superior results (Achtemeier, 1987).

The analysis grid is initiated with 'missing' values, and subsequently populated with station data according to the stations' coordinates in relation to the grid structure. To assess how well the Barnes scheme estimates grid values, a removal and replacement process is implemented where for each grid cell containing station data, the station's value is removed, and the estimation of the cell's value is derived from the IDW values of nearby stations within the cell's ROI.

Upon completion of the three-pass correction, a statistical analysis compares the objective analysis values to their corresponding observations. A root mean squared error is calculated to show the averaged error rate, which identifies outlying errors not only by value, but also by station name. Such stations are then compared geographically to determine whether the station in question is a boundary grid point. In addition, the interpolated data grid is filtered by excluding values that fall within the analysis variable's instrument calibration range from further examination. In this study, interpolated values that positively or negatively exceed the calibration range are considered to be true errors. Interpolated values within the calibration range are subject to as much scrutiny as the actual observation values and are therefore not pertinent to studying the Barnes scheme's accuracy.

3. DATA

To assess the Barnes scheme's accuracy, several key calculations are preformed using data collected from the Oklahoma Mesonet, a statewide mesoscale network of meteorological observation stations. Each Mesonet instrument is calibrated prior to deployment and monitored regularly to ensure collection of the highest quality data. (mesonet.org; Brock et al., 1995; McPherson et al., 2007). Air temperature is measured at 1.5m above ground level and has an instrumentation calibration range of +/- 0.5^o C. The sensor is housed in an aspirated radiation shield which continuously draws ambient air over the sensor, while protecting it from solar radiation. Humidity is measured at 1.5m above ground level as a percentage and has an instrumentation calibration range of +/- 3% in the relative humidity range of 10% to 98%. Wind speed is measured at 10m above ground level and has an instrumentation calibration range of +/- 0.3 m/s

The analysis variables were chosen because of their resistance to high variance and localization. Other variables such as rainfall amount or atmospheric pressure can be highly localized, which has the potential to add unnecessary complexity to the examination of the scheme's accuracy. One complexity being the problem of the 'bullseye effect' in IDW, where among a relatively normalized field there exists unusually high variance at a particular grid point. The interpolation tends to create concentric and most likely erroneous values around the grid point.

Candidate days (one per month) were selected which had relatively calm atmospheric conditions; times when there were no major frontal boundaries, no rainfall events, and a relatively smooth gradient of atmospheric pressure across the state to further reduce the number of factors influencing the analysis variables' values. Further, the candidate days were chosen in this range to possibly see if seasonal variability influences the scheme's accuracy.

4. RESULTS

Air temperature errors are visually represented by the difference of a station's objective analysis (OBAN) value, and its real value (fig 1). Subsequent results are derived from this definition of the error. The error magnitude includes the analysis variable's instrument calibration range,

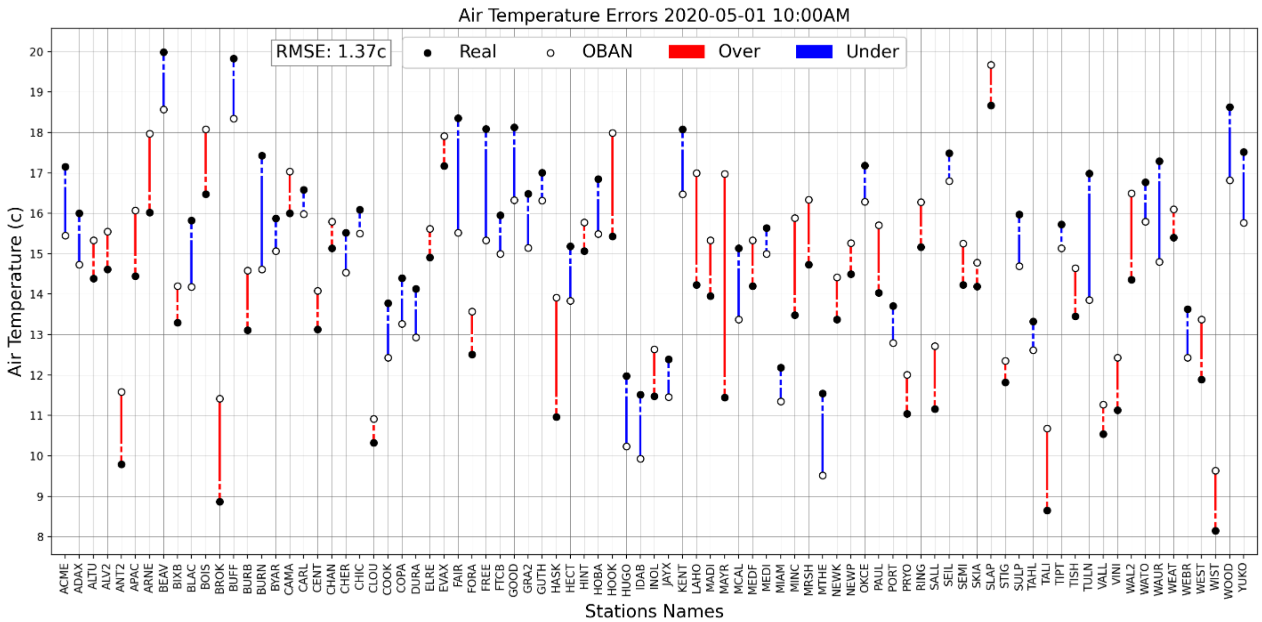

Figure 1. The magnitude of error between interpolated and real values

represented by a dashed line. While there is no relationship between air temperature and station names, particular stations are identified for further analysis by examining their geographic location.

Figure 2 shows the same information of figure 1, however station OBAN values and real values are compared directly. Stations are denoted by the blue scatter dots, and the black 1:1 line represents a perfect interpolation where the OBAN value is the real value. Roughly one third of the stations had OBAN values within the upper and lower instrument calibration limit. For errors exceeding this range, the root mean squared error

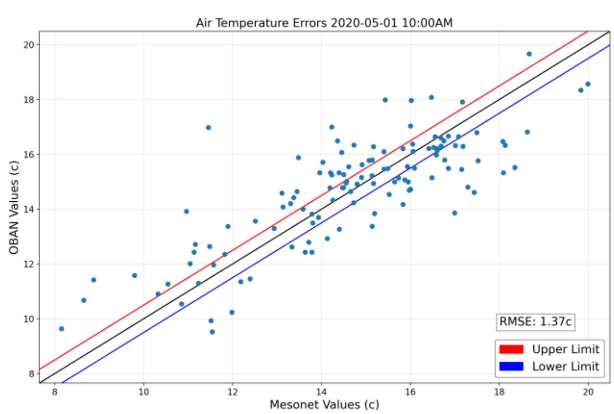

Figure 2. A comparison between interpolated and real values

magnitude is 1.37° C. The most notable outlier for this particular day, time, and analysis variable is seen in both figure 1 and figure 2, with an error magnitude of roughly six degrees at the MAYR station.

The previous figures have shown a visual representation of errors on a particular candidate day, at a particular time and only one of the three analysis variables. To gain a deeper understanding of the Barnes scheme's accuracy it is necessary to examine errors over a period of time. Accumulated total amount of errors for each analysis variable are seen in the sorted arrangement of Mesonet stations in figure 3. Air temperature and relative humidity errors comprised the majority of errors at any given station, however wind speed errors mostly occurred at stations that had the most errors of all stations. The average amount of errors for all stations over the twelve candidate days is 13.19, however it should be noted this value is simply 13 as errors are discrete rather than continuous.

A comparison between the relative number of errors for each analysis variable is seen in figure 4 where among the twelve candidate days, 55.62% are air temperature errors, 40.13% are relative humidity errors, and 4.25% are wind speed errors. Since changes in the temperature of air can change the relative humidity, even when the absolute humidity remains constant, an error associated with air temperature may influence an error associated with relative humidity. Further analysis with more candidate days and different times of the day is required to explore this relationship in a meaningful way.

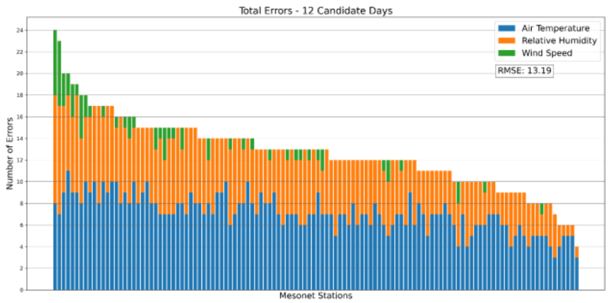

Figure 3. Accumulated number of errors for each analysis variable over 12 candidate days. The x axis contains a sorted arrangement of Mesonet stations

A possible explanation for the left most stations in figure 3 could be a lack of initial data points within the respective station's radius of influence. This leads to a situation where only after the first pass of the scheme, where a majority of the interpolation occurs, do input values become available for these stations. However, since these newly available input values themselves are the interpolated values from surrounding stations, their

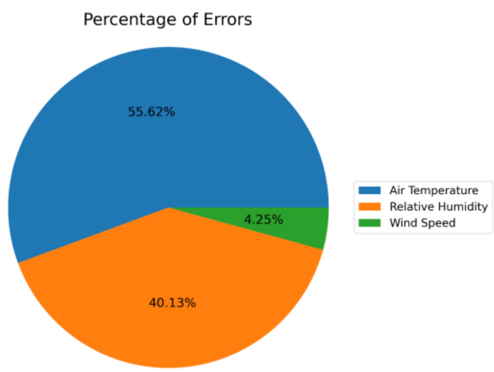

Figure 4. A comparison of the relative percentage of errors for each analysis variable

interpolated values are derived from calculated data instead of real data. This study hypothesizes that such situations most often occur at so called 'boundary' cases along the edge of the examined area's border. This is because a section of these station's radius of influence exceeds the edge and therefore effectively has less overall area to include known data values. In some cases, no known data values are present.

This idea is explored in figure 5 where for each Mesonet station, the number of total errors is compared to the number of known data points within the respective station's radius of influence (ROI). Generally, as seen with the regression line, as the number of ROI data points increases, the total number of errors decreases. However, the greatest and least number of errors both occur when a station has a single known data point within the ROI. It should be noted that while the figure appears to only show a subset of Mesonet stations, each blue station point may have more than one station 'stacked' on top of one another at a particular point. All 120 Mesonet stations are shown here even if not directly.

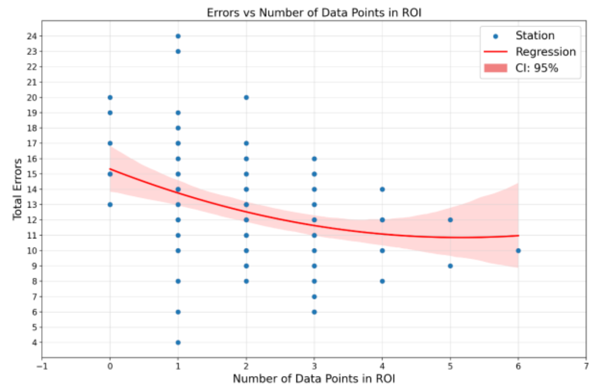

Figure 5. A comparison between the number of initial data points within each station's radius of influence, and the number of errors at that station.

To further illustrate the comparison of errors to data points within the ROI for a particular station, the top ten most erroneous stations and stations with no data points in their ROI are compared on a map of Oklahoma (fig. 6). The circles on the map are approximately equal to the area of the ROI used in this study. Red circles are the top ten most erroneous stations, blue circles are stations with no data points. The two stations with the most errors were KENT and BOIS, which are both furthest west in the panhandle. It should be noted their case is somewhat unique where they include only each other as the initial data points in their ROI. It follows that if one of the stations is erroneous, the other will also be erroneous.

Other notable stations include BUFF and ARNE, which happen to be in both categories and are also both 'boundary' cases. However, not all of the top ten stations with errors are boundary cases. Stations HOBA, FAIR, MCAL and FREE are all considered interior cases, and therefore 60% of the top ten stations with errors are boundary cases. This may suggest a weak but noticeable relationship between boundary cases and number of errors at a station.

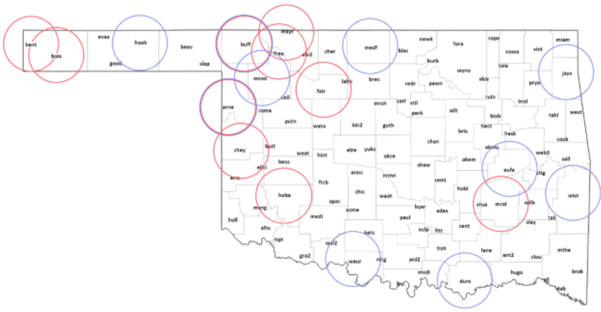

Figure 6. Circle - Approximate area of the radius of influence (ROI), red - top ten most erroneous stations, blue - stations with no data points in their ROI

The total amount of stations who had at least one error is seen in figure 7 in a comparison with stations who had no errors. While the x axis does list months sequentially starting in May (2020) and ending in April (2021), the plot is not meant to describe an averaged value for the given month. Since the twelve candidate days are one day at one time per month, it is inappropriate to assume this accounts for the behavior of the scheme over an entire month. However, the plot does show that the highest number of errors occurred in April 21', and the lowest in July 20' with a gradual increase in errors from the July candidate day to the April candidate day.

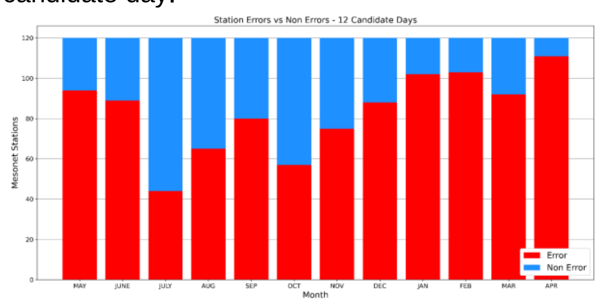

Figure 7. A comparison of the total number of errors for each candidate day

5. CONCLUSION

The accuracy of objective analysis schemes was examined using data from the Oklahoma Mesonet. A removal and replacement process was utilized where for each Mesonet observation value, a corresponding objective analysis value was calculated using the Barnes scheme. The calculated values were derived from three analysis variables chosen to be examined, air temperature, relative humidity, and wind speed. These variables are examined during candidate days which were chosen to be studied when atmospheric conditions were relatively calm, meaning no major frontal boundaries, rainfall events, or high variance in atmospheric pressure was present across the state.

Accuracy is best defined in this study as a discrete classification of values being either an error or non-error where the magnitude of the difference between an OBAN value and the corresponding observation value which exceeds an instrument's calibration range is considered erroneous. Using this metric, accuracy is examined at three levels, beginning with a particular analysis variable on a particular candidate day.

It was found that the root mean squared error for air temperature on May 01, 2020 was 1.37^{o} C, and the largest magnitude of all errors was roughly six degrees over the real value. The second level of analysis examines accumulative errors across all Mesonet stations, analysis variables, and candidate days. It was found that 55.62% of errors were air temperature errors, 40.13% were relative humidity errors, and 4.25% were wind speed errors. The third level of analysis examines geographically where errors occur most. It was found that a majority of the top ten most erroneous stations were considered to be 'edge', or 'boundary' cases, and of these ten stations, two of which did not have initial input data. Further analysis was conducted on a comparison between the number of errors at a station and the number of initial data points within a stations radius of influence (ROI). It was found that there is a general relationship between the two, with a gradual lowering of errors as the number of data points within the ROI increases.

While this study is a high-level overview of the accuracy of the Barnes scheme, it is not comprehensive. Future research is necessary for a more deterministic approach in the examination of objective analysis accuracy. This should include a greater data set comprised of many more candidate days, different times within each candidate day, and perhaps more analysis variables. Since this study was focused particularly on the Barnes scheme, it may be prudent to use similar metrics on other objective analysis schemes such as the Cressman scheme. Further, the parameters of the Barnes scheme may be adjusted to minimize error emergence for each analysis variable, and also for seasonal variability if it is found to affect the accuracy.

6. ACKNOWLEDGMENTS

The authors would like to thank the National Weather Center Research Experiences for Undergraduates Program, as well as Dr. Daphne LaDue and Alex Marmo for their support in said program, and the taxpayers of Oklahoma for their continued support of the Oklahoma Mesonet. This research is made possible by the National Science Foundation under Grant No. AGS-1560419.

7. REFERENCES

Achtemeier, G.L., 1987: On the concept of varying influence radii for a successive corrections objective analysis. *Monthly Weather Review*, *115*(8), pp.1760-1772.

Barnes, S. L.,1964: A Technique for Maximizing Details in Numerical Weather Map Analysis, *Journal of Applied Meteorology and Climatology*, *3*(4), 396-409.

Barnes, S.L., 1994: Applications of the Barnes objective analysis scheme. Part I: Effects of undersampling, wave position, and station randomness. *Journal of Atmospheric and Oceanic Technology*, *11*(6), pp.1433-1448.

Barnes, S.L., 1994: Applications of the Barnes objective analysis scheme. Part II: Improving derivative estimates. *Journal of Atmospheric and Oceanic Technology*, *11*(6), pp.1449-1458.

Barnes, S.L., 1994: Applications of the Barnes objective analysis scheme. Part III: Tuning for minimum error. *Journal of Atmospheric and Oceanic Technology*, *11*(6), pp.1459-1479.

Brock, F. V., K. C. Crawford, R. L. Elliott, G. W. Cuperus, S. J. Stadler, H. L. Johnson, and M. D. Eilts, 1995: The Oklahoma Mesonet: A technical overview. J. Atmos. Oceanic Technol., 12, 5-19.

Lu, C. and G.L. Browning, 1998. The impact of observational errors on objective analyses. *Journal of the atmospheric sciences*, *55*(10), pp.1791-1807.

McPherson, R.D., 1986: Operational Objective Analysis Techniques and Potential Applications for Mesoscale Meteorology. In: Ray P.S. (eds) *Mesoscale Meteorology and Forecasting*. American Meteorological Society, Boston, MA.

McPherson, R.A., C.A. Fiebrich, K.C. Crawford, J.R. Kilby, D.L. Grimsley, J.E. Martinez, J.B. Basara, B.G. Illston, D.A. Morris, K.A. Kloesel, and A.D. Melvin, 2007: Statewide monitoring of the mesoscale environment: A technical update on the Oklahoma Mesonet. *Journal of Atmospheric and Oceanic Technology*, *24*(3), pp.301-321.

Nuss, W.A. and D.W. Titley, 1994: Use of multiquadric interpolation for meteorological objective analysis. *Monthly Weather Review*, *122*(7), pp.1611-1631.

Richard P. McNulty, 2011: Operational Weather Analysis, http://www.wxonline.info/ebook/Chapter4_final.pdf