Let's Talk About the Weather

Jill Lundell[*]   Brennan Bean[†]   Jürgen Symanzik [‡]

Abstract

The accuracy of weather forecasts is dependent on many factors including the closeness of the forecast, location, and changes in weather forecasting methods. Understanding the reliability and the shortcomings of these predictions across the highly varied climate zones of the United States provides valuable context for future predictions. Graphical exploration of weather forecast data provides insight on factors associated with weather forecast accuracy. In addition, these graphical representations aid in detecting region-specific trends and anomalies associated with weather predictions in the United States.

Key Words: Climate, Clustering, Data Expo 2018, Glyph Plots, Random Forests, Visualization

1. Introduction

From the bitter cold and wet winters along Lake Superior, to the oppressively hot and dry summers in the Valley of the Sun, Arizona, the United States (U.S.) experiences a wide range of climatic extremes. These extremes, in turn, create unique challenges forecasting the weather. Characterizing forecast errors across such a diverse landscape is equally challenging, requiring multi-dimensional visualizations across space, time, and various climate measurements. In spite of these challenges, better understanding the nature and patterns in forecast errors across the U.S. helps meteorologists as they strive to improve weather forecasts. It can also help everyday Americans determine how much faith they should put in the weather forecast on the day of their picnic.

The 2018 Data Expo of the Sections on Statistical Computing and Statistical Graphics of the American Statistical Association (ASA) provided an opportunity to explore and compare weather forecast errors across the U.S. Our analysis focused on the question:

How do weather forecast errors differ across regions of the U.S.?

This motivating question prompted the subsequent questions:

- Can we cluster U.S. weather stations into regions based on weather characteristics?

- How do forecast errors change by region and by season?

- Who are the winners and losers in terms of overall forecast accuracy?

- Which variables are important in determining forecast errors?

- How do error variables correlate and do these correlations change by region?

---

[*]Department of Mathematics and Statistics, Utah State University, 3900 Old Main Hill, Logan, UT 84322, `jflundell@gmail.com`

[†]Department of Mathematics and Statistics, Utah State University, 3900 Old Main Hill, Logan, UT 84322, `brennan.bean@aggiemail.usu.edu`

[‡]Department of Mathematics and Statistics, Utah State University, 3900 Old Main Hill, Logan, UT 84322, `juergen.symanzik@usu.edu`

This article is devoted to answering these questions. In Section 2.1, we summarize the data and the associated cleaning process. We then show in Section 2.2 that the U.S. can be clustered into six well-defined weather regions using the provided climate measurements, elevation, and distance to coast. These clusters, or weather regions, form the basis of our comparison of forecast accuracy across the U.S. through a series of multi-dimensional plots and variable importance analyses described in Section 3. Next, we introduce in Section 4 the interactive application we created to enhance our data explorations. Finally, we conclude in Section 5 that the climate differences that distinguish the weather regions of the U.S. also create region-specific patterns and differences in forecast accuracy.

2. Measurement Explorations

The data contain measurements and forecasts for 113 U.S. weather stations from July 2014 to September 2017. These data can be obtained at the following URL:

```
http://community.amstat.org/stat-computing/data-expo/
data-expo-2018.
```

Daily measurements for eight different weather metrics were recorded for each location including temperature, precipitation, dew point, humidity, sea level pressure, wind speed, cloud cover, and visibility. Many notable weather events are also textually recorded. Daily measurements of the minimum, maximum, and mean were recorded for each metric.

2.1 Data Cleaning

Table 1 shows the weather variables included in our final analysis. We excluded mean daily measurements for temperature, precipitation, dew point, humidity and sea level pressure as these measurements were near perfect linear combinations of their corresponding minimum and maximum measurements. We also excluded maximum visibility from the analysis as this measurement was equal to 10 miles for more than 97% of all recorded measurements. Lastly, we combined the information provided by maximum wind speed and maximum wind gust by retaining only the lower of the two measurements after removing outliers. The decision to combine the information from these two wind variables was motivated by the fact that 13% of all maximum wind gust values were missing. In addition, it is difficult to separate unusually high, yet valid, maximum wind gust and wind speed measurements from true outliers.

We supplemented the provided location information with elevation and distance to the nearest major coast. Elevation information was obtained for each location through Google's API server (Google, 2018) via the rgbif R package (Chamberlain, 2017). Distance to coast was calculated as the closest geographical distance between each measurement location and one of the vertices in the U.S. Medium Shoreline dataset (NOAA, 2018), which includes all ocean and Great Lakes coasts for the contiguous 48 states. Because this dataset does not include the coastlines of Alaska and Hawaii, distance to coast calculations for these locations used manually extracted shorelines from NOAA's Shoreline Data Explorer (NGS, 2018). We acknowledge there are limitations to this method of distance calculation, as calculations for some locations, such as Arizona, are slightly longer than they would be had we used shoreline information for Mexico's Gulf of California. Nevertheless, these measurements effectively separate inland weather stations from coastal stations.

Some stations did not record relevant climate variables. When possible, these missing observations were replaced with corresponding measurements obtained from the nearest

Table 1: List of weather variables included in our analysis. All observations outside the indicated ranges were removed prior to our analysis.

Variable	Unit	Range
Min/Max Temperature	$^{\circ}F$	$[-37, 127]$
Precipitation	in	$[0, 12.95]$
Min/Max Dew Point	$^{\circ}F$	$[-50, 90]$
Min/Max Humidity	%	$(0, 100]$
Min/Max Sea Level Pressure	inHg	$[28.2, 31.2]$
Mean/Max Wind Speed	mph	$[0, 70]$
Min Visibility	mi	$[0, 10]$
Cloud Cover	okta	$\{0, 1, \cdots, 8\}$
Distance to Coast*	mi	$[0, 807]$
Elevation*	ft	$[3, 7422]$

* manually added data

National Weather Service (NWS) first order station as obtained through the National Climatic Data Center (NCDC) (NOAA, 2018). Missing values include wind speed in Baltimore, Maryland, precipitation in Denver, Colorado, and replacements of outlier precipitation measurements at multiple locations. When replacements were not readily obtained through the NCDC, systematic missing observations were replaced with corresponding observations from the nearest geographical neighbor within the dataset, as was the case for visibility and cloud cover in Baltimore, Maryland (replaced with Dover, Delaware, measurements) and Austin, Nevada (replaced with Reno, Nevada, measurements).

Table 1 also shows the observation ranges for each of the included variables. These measurement ranges are either definitional, such as the bounds for humidity, or simply practical, such as the bounds for temperature. All measurements falling outside the bounds shown in Table 1 were removed prior to our analysis. Several individual outliers were also removed or replaced based on location-specific inconsistencies including

- removal of one unusually low minimum temperature measurement in Honolulu, Hawaii, ($< 10°$F) and two in San Francisco, California ($< 20°$F);

- replacement of the following unusually high precipitation readings with precipitation readings at nearby weather stations (NOAA, 2018):

    - Oklahoma City, Oklahoma, on 8/10/2017 (38.33in $\rightarrow$ 0.8in)

    - Salmon, Idaho, on 4/21/2015, 5/2/2016, and 5/3-4/2017 (10.02in $\rightarrow$ 0in)

    - Flagstaff, Arizona, on 12/24/2016 (7.48in $\rightarrow$ 0.97in)

    - Indianapolis, Indiana, on 7/15/2015 (9.99in $\rightarrow$ 0in);

- removal of one unusually low minimum dew point measurement in Honolulu, Hawaii ($< 40°$F), two in Hoquiam, Washington ($< 0°$F), four in Las Vegas, Nevada ($< -15°$F), and two in Denver, Colorado ($< -20°$F).

2.2 Data Clustering

The U.S. has been divided into regions based on environmental characteristics such as watersheds and climate. We examined the set of existing environmental regions and were

unable to find one that made sense in terms of weather. We created our own weather regions by clustering the weather stations based on the metrics in Table 1. Thus, clusters are defined by weather characteristics observed at each station. We did not incorporate forecast accuracy because this can change as models are improved and we wanted to form clusters using the physical characteristics of each station.

Once cleaned, the data were aggregated across each weather station by taking the mean and standard deviation of each variable in Table 1 for each station over the period of record. This left one mean and one standard deviation for each of the variables in Table 1 for each of the 113 weather stations.

Hierarchical clustering (Friedman et al., 2001, pp. 520-526) with Euclidean distance and Ward's minimum variance clustering method (Murtagh and Legendre, 2014) were used to identify clusters. The clusters were examined spatially to determine the performance of the clustering and choose the number of clusters. Because we were looking for clusters that divide the U.S. into weather regions, we wanted to ensure they were of a sufficient size to be practical. We chose six clusters. Five clusters resulted in one cluster that included all of the stations from the Midwest to the East Coast which we felt is too large. Seven clusters produced a cluster that contained only five weather stations which seemed to be too small.

Figures 1 and 2 show the results of the cluster analysis. Figure 3 shows a parallel coordinate plot of the characteristics for each weather region. The Z-score for mean and standard deviation for each of the variables in Table 1 was computed and plotted on the parallel coordinate plot. It is difficult to distinguish the six weather regions from each other so an interactive app was created that provides a better view of the features of each cluster. The app is discussed in Section 4 in this article.

The names and characteristics of each weather cluster are as follows:

- **Cali-Florida:** Warm and humid with high dew point and pressure. Low variability in almost all measurements.

- **Southeast:** Warm and humid with lots of rain. High variability in precipitation and low variability in temperature.

- **Northeast:** Cold, humid, and low visibility. High variability in temperature, dew point, and pressure.

- **Intermountain West:** Cold and dry, with high variability in temperature, wind speed, and pressure. Low variability in precipitation and dew point.

- **Midwest:** Landlocked with high wind speed and high variability in temperature, pressure, and wind speed.

- **Southwest:** Warm, sunny, and dry with little variation. High variability in wind speed and humidity.

3. Forecast Error Explorations

Forecasts were restricted to minimum temperature, maximum temperature, and the probability of precipitation. We found no obvious outliers in the weather forecasts. Rather, the forecast data were replete with duplicate values for minimum temperature and precipitation. We retained the lowest forecast of minimum temperature and the highest forecast of precipitation probability for each forecast.

Forecast lags of six or seven days contained a large number of missing values. Forecast lag is defined as the number of days between the day of forecast and the day being forecast.
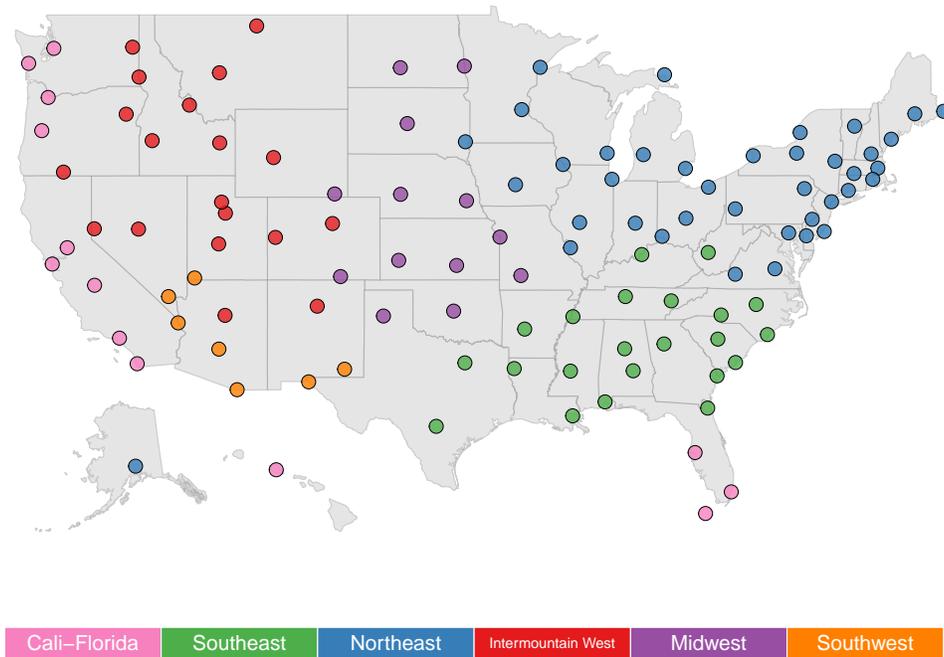
| Cali–Florida | Southeast | Northeast | Intermountain West | Midwest | Southwest |

Figure 1: Map of weather clusters.

Thus, same day forecasts would have a lag of 0, one day prior forecasts a lag of 1, and so on. We removed all forecasts past lag 5. We also removed all forecasts containing negative lags (i.e., a forecast made *after* the actual observation).

The forecast error for minimum and maximum temperature is calculated as the absolute difference between forecast and measurement. The forecast error for precipitation is measured using the Brier Skill Score (BSS), a well-known measure of probabilistic forecast accuracy (Weigel et al., 2007). It is defined for a particular weather station as

$$\text{BSS} = 1 - \frac{\sum\limits_{i=1}^{N} \sum\limits_{j=0}^{M} (Y_{ij} - O_i)^2}{\sum\limits_{i=1}^{N} \sum\limits_{j=0}^{M} (P - O_i)^2} \tag{1}$$

where

- $Y_{ij} \in [0, 1]$ is the predicted probability of rain on day i with forecast lag j;

- $O_i \in \{0, 1\}$ is a binary variable with value 1 if *any* precipitation fell during the day and 0 otherwise. We defined a precipitation event as a positive precipitation measurement or the inclusion of the words "rain" or "snow" in the event information;

- $P \in [0, 1]$ is the average daily chance of precipitation over the period of interest, defined as $P = \frac{1}{N} \sum\limits_{i=1}^{N} O_i$;

- N denotes the number of days of recorded precipitation in the period of record and $M \in \{0, \dots, 5\}$ denotes the number of forecast lags.
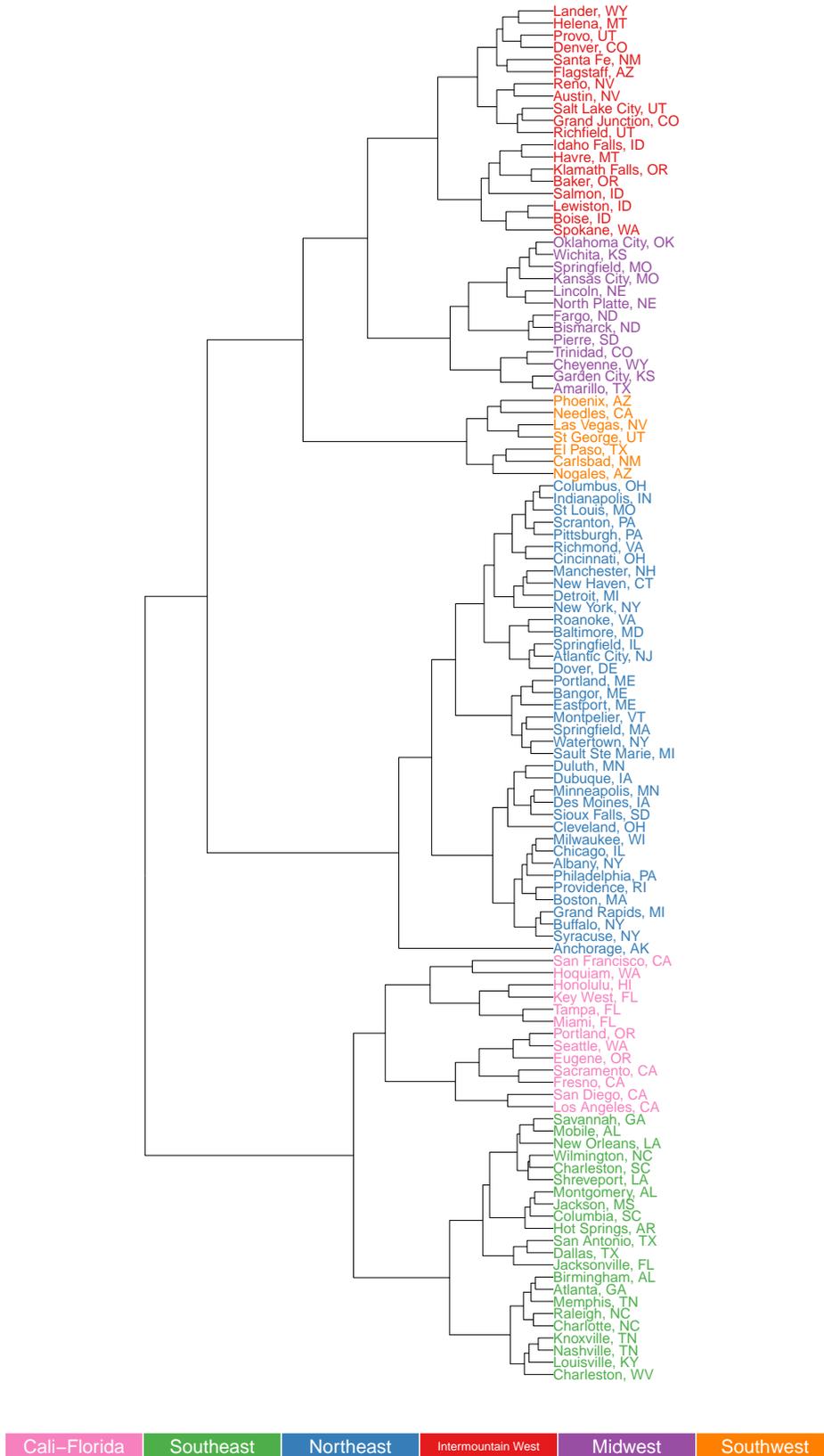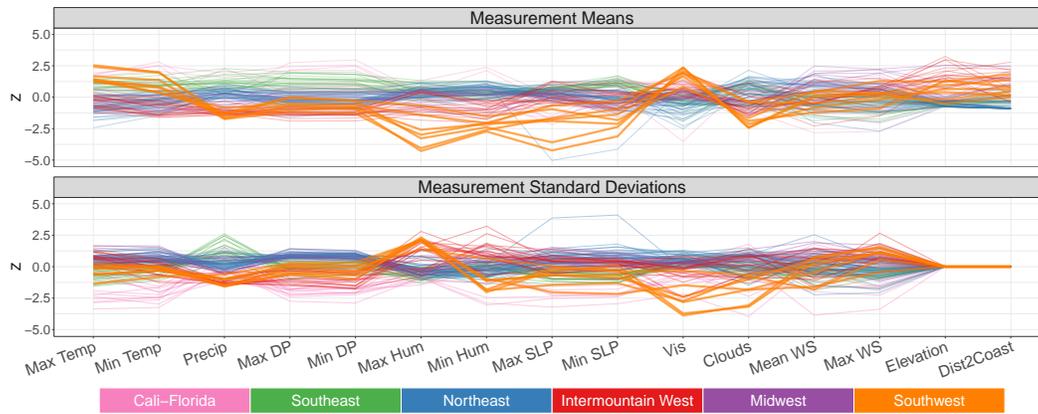
Figure 2: Dendrogram of weather clusters.

Figure 3: Parallel coordinate plot of the means and standard deviations of the weather variables listed in Table 1. Each line in the plot represents one of the 113 weather stations. The color of the lines match the weather region to which the station belongs.

Note that the BSS $\in (-\infty, 1]$, with 1 indicating a perfect forecast skill and movement towards $-\infty$ indicating worse forecasts. We chose to use $1 - $ BSS so all three error variables would be consistent in orientation. The following subsections explore differences in forecast errors both between and within the previously defined weather regions visualized in Figure 1. Forecast errors are averaged over lag and in some cases averaged over month in each graph. The visualizations in the following subsections confirm our hypothesis that different weather regions experience distinctly different weather forecast patterns.

3.1 Seasonal Trends

The position of the U.S. in the northern hemisphere makes most of the country subject to distinct weather seasons. Seasons are most pronounced in the northern U.S. We hypothesize that the forecast error behavior will be inextricably linked to this seasonality. We explore this through a series of space-time graphs. Modeling space and time simultaneously creates a three-dimensional problem usually visualized as small multiples. Small multiples are "a series of graphics, showing the same combination of variables [e.g., latitude and longitude], indexed by changes in another variable [e.g., time]" (Tufte, 2002, p. 170). The issue with this approach is that it becomes difficult to visually comprehend all but the most drastic changes from graph to graph. One alternative that allows simultaneous visualizations of both space and time is through the use of glyphs, or symbols, that allow for multi-dimensional visualizations in a spatial context (Carr et al., 1992; Wickham et al., 2012).

Figure 4 shows glyph plots of seasonal forecast errors throughout time. The forecast error is visualized as the scaled distance from a center point to the edge of a polygon with twelve observations starting with January at the 12:00 position and proceeding clockwise. The asymmetry of the glyphs about their center points illustrates how forecast errors change across time and across space. For example, locations in the Northeast are worse at forecasting precipitation in the winter than in the summer, while locations in the Southeast forecast precipitation equally well all year long.

In addition to highlighting forecasting asymmetries, Figure 4 reveals location-specific anomalies. For example, San Francisco, California, predicts minimum temperatures well all year long, but only predicts maximum temperatures well in the winter months. This is

likely due to chilling coastal fogs known to frequent the region throughout the year that can create sharp temperature differences over short distances (Nolte, 2016). The struggle to predict temperature seems reasonable in light of these facts as this measurement location is more than 11 miles away from the forecast location. The issue is likely less pronounced in the winter because the contrast between inland and coastal temperatures is reduced.

Another location-specific anomaly of note is the drastic seasonality of precipitation forecasts for locations surrounding the Great Lakes, as observed in Figure 4. This unusually bad forecasting in the winter is likely due to lake-effect snow which is prevalent in the region. Up to 100% more snow falls downwind of Lake Superior in the winter than would be expected without the lake-effect (Scott and Huff, 1996). This area has been previously identified as having the most unpredictable precipitation patterns in the nation (Silver and Fischer-Baum, 2014). The above examples demonstrate the ease with which comparisons can be made across space and time with these glyph-based plots.

3.2   Error Scatterplots

Scatterplots reveal outliers and overall trends within weather regions. We constructed an interactive scatterplot app that allows the examination of trends between the three forecast error variables for individual forecast lags or aggregated across all forecast lags. Figures 5 (a-c) show examples of plots from the interactive app. The figures show the scatterplot for the data aggregated over all forecast lags, as well as the scatterplots for lags of 5, 3, and 1, to illustrate how forecast accuracy changes over forecast lag.

Figure 5 (a) compares minimum temperature forecast accuracy with precipitation accuracy. Weather stations with the worst predictions of minimum temperature are located in New England and the Intermountain West. New England is known for extreme winter weather and the frequency of extreme weather events seems to be increasing (Cohen et al., 2018). This likely contributes to the struggle these stations have predicting minimum temperature. Cali-Florida uniformly has the best predictions of minimum temperature.

Figure 5 (b) compares maximum temperature prediction accuracy with precipitation accuracy. Four weather stations in the Great Lakes region have the worst precipitation predictions in the dataset. Poor precipitation forecast accuracy in this region illustrates the difficulty in forecasting lake-effect snow, as discussed in the previous section. Precipitation forecast accuracy for the Great Lakes region improves substantially as the forecast lag decreases.

Figure 5 (c) shows the relationship between minimum and maximum temperature forecast accuracy. Three outliers stand out in these scatterplots, namely Key West, Florida, Austin, Nevada, and San Francisco, California. Key West predicts both minimum and maximum temperature more accurately then any other weather station. Key West also ranks in the top five for lowest variability in eight of the weather variables, which likely explains the accurate forecasts. Austin is the poorest predictor of both measures. Seventy miles along the "loneliest highway in America" (GACC, 2018) separate Austin from its weather measurements in Eureka, Nevada. The poor predictions for maximum and minimum temperature can be explained by the change in climate over such a large distance, as reflected in a negative prediction bias of around 5°F for maximum temperature and a positive bias of around 7°F for minimum temperature. San Francisco has good predictions of minimum temperature and poor predictions for maximum temperature. This feature was observed and explained in the previous section.

The interactive app developed in conjunction with this project allows for further investigation of forecast accuracy trends. The app is discussed in Section 4 of this article.
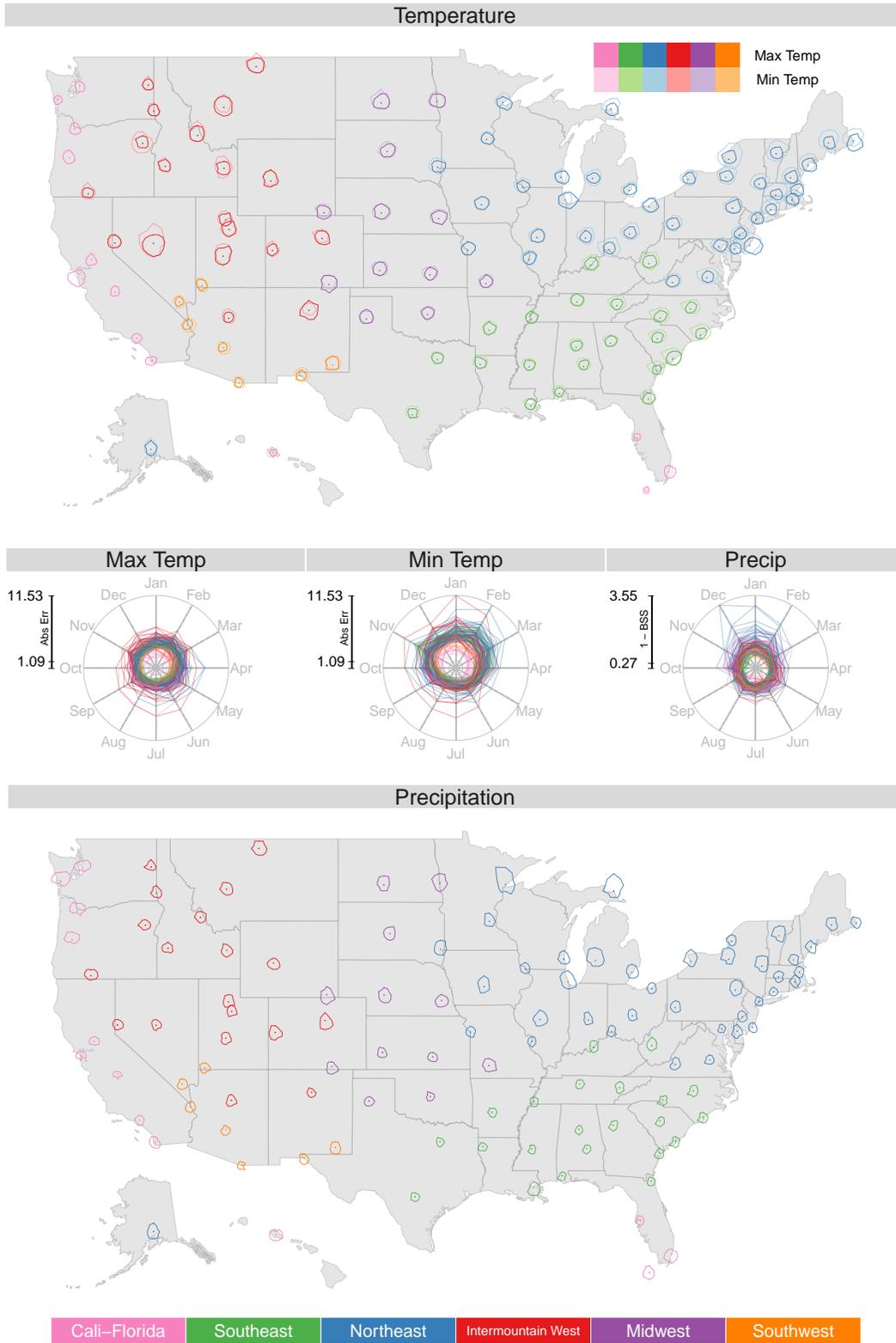
Figure 4: Glyph plots of weather forecast accuracy averaged by month. The error is represented as the scaled distance from a center point to the edge of a polygon beginning with January at the 12:00 position and proceeding clockwise.
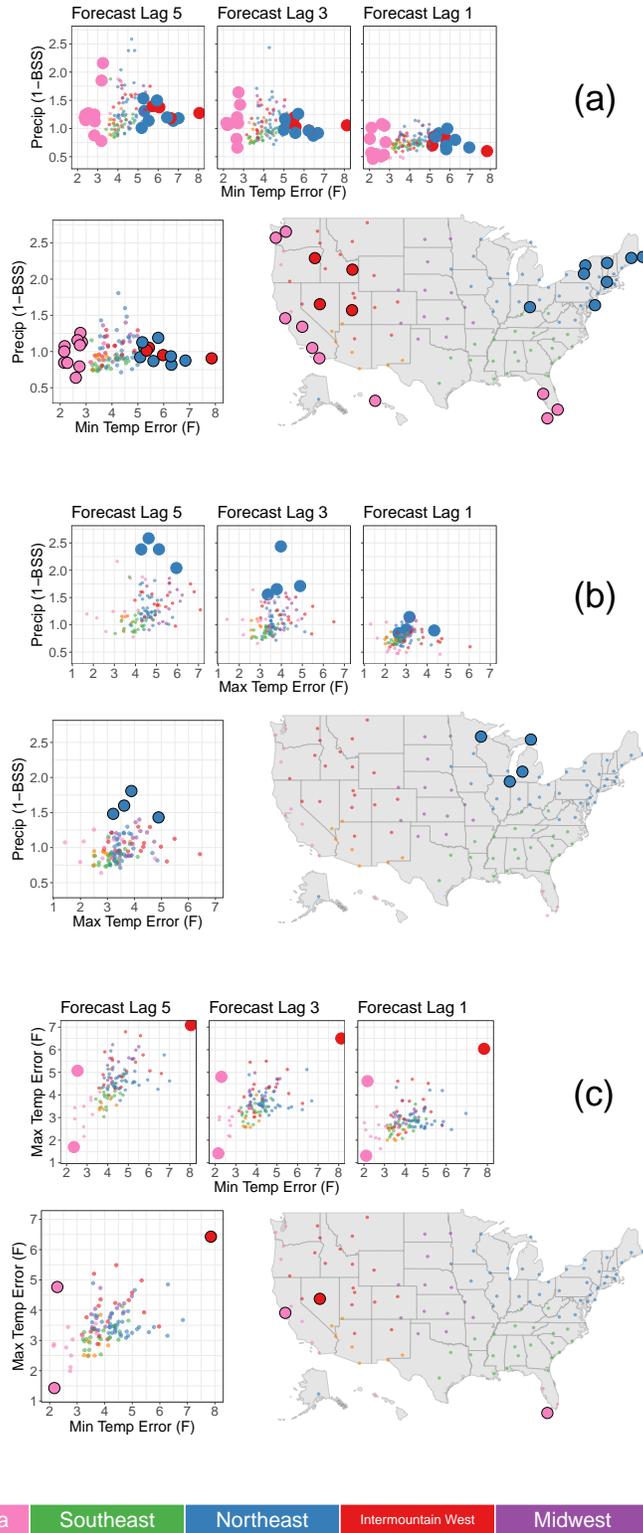
Figure 5: Scatterplots comparing the three forecast error variables. Points of interest discussed in the text are highlighted in the respective plots.
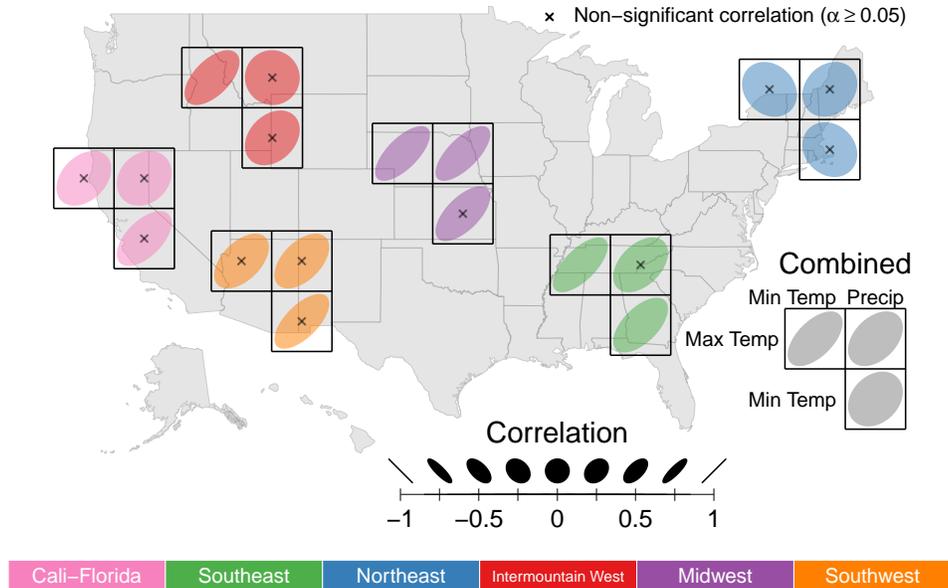
Figure 6: Spearman correlations between forecast error variables represented as ellipses superimposed on a map of the U.S.

3.3 Error Correlations

The scatterplots provide some sense of the correlation between the different error types among the different city locations. We explore such correlations further through the use of correlation ellipses (Murdoch and Chow, 1996) superimposed on a map of the U.S. as observed in Figure 6. We calculate Spearman correlations between each pair of measurements for the locations within each cluster. The sign of the correlation coefficient is denoted by the slope of the ellipse and the strength of correlation is denoted by the width of the ellipse. All of the correlations between error variables are positive except for correlations between minimum temperature and the other two variables in the Northeast. Only a few cluster-specific correlations are significant. This is likely due to the small number of stations in many of the weather regions. However, the overall correlations for the 113 weather stations are all positive and significant. The observations made using this correlation ellipse map illustrate how this plot style facilitates multi-dimensional comparisons across space.

3.4 Variable Importance

We used random forests (Friedman et al., 2001, pp. 593-594) to determine which weather variables had the greatest impact on the forecast errors. The data were aggregated over forecast lag and month. Three random forest models were generated for each weather region using forecast lag and the means and standard deviations for each of the weather variables listed in Table 1. Each of the forecast error variables was used as a response. Figure 7 contains three parallel coordinate plots that show the variable importance measures of each region for each forecast error variable. The importance measures obtained from random forests were recentered by subtracting the minimum importance measure and then rescaled by dividing by the maximum importance measure of the recentered values for each weather cluster and forecast error variable combination. Thus, the most important variable within each cluster has a value of 100 and the least important has a value of 0 for each error measure. This allows direct comparisons of importance between weather regions and
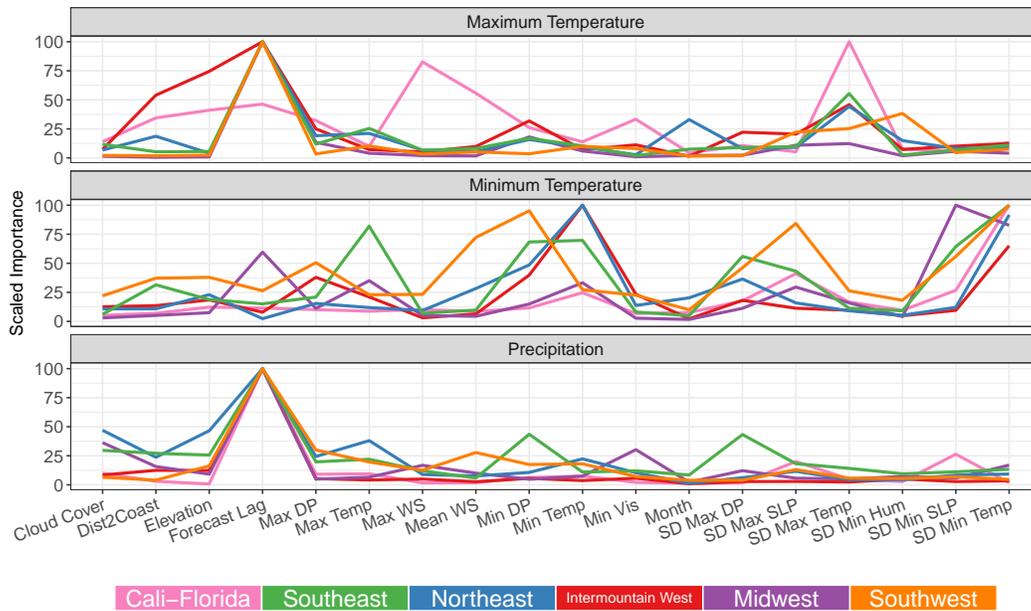
Figure 7: Variable importance for each of the three forecast accuracy measurements. Variable importance measures have been rescaled to make the measures directly comparable between weather regions and accuracy measures.

across error measures.

Figure 7 shows that the most important variable for the precipitation error is forecast lag regardless of weather region. None of the other variables are very important relative to lag. The Southeast shows minimum dew point (DP) and the standard deviation of maximum dew point as being somewhat important. Cloud cover is important for the precipitation error in the Northeast.

Forecast lag is also the most important variable for the maximum temperature error for all weather regions except Cali-Florida. Maximum wind speed (WS) and the standard deviation of maximum temperature are more important than lag for the maximum temperature error in Cali-Florida. The variability in maximum temperature is also important for the Southeast, Northeast, and the Intermountain West. Distance to coast (Dist2Coast) and elevation are important for the maximum temperature error in the Intermountain West.

Variables that are important for the minimum temperature error varied substantially across weather regions. The variability in minimum temperatures is important for all regions, but otherwise the important variables vary widely from region to region. Minimum temperature is the most important for the Northeast and Intermountain West, but maximum temperature is important for the Southeast. Minimum dew point and the variability in the maximum sea level pressure (SLP) are important in the Southwest while variability in minimum sea level pressure is the most important for the Midwest. Forecast lag is not particularly important for any of the regions.

4. Interactive Explorations

We developed an interactive Shiny app to enhance our weather data explorations. This app can be accessed at

```
https://jilllundell.shinyapps.io/finaldataexpoapp/.
```

The first tab of the app is an interactive version of the parallel coordinate plot introduced in Figure 3. The app allows the user to select a weather region which is highlighted on the graph. Characteristics of the selected region can be easily seen and compared to all other observations.

The second tab of the app is an interactive scatterplot. Figure 5 (a-c) shows examples of the graphs from this tab. The user can select up to two of the three forecast error variables to be on the axes. The forecast lag can also be selected. Points on the scatterplot can be brushed or clicked and the selected points show up on a map of the U.S. Information about selected stations is listed in a table under the graph. This app allows for a more complete exploration of outliers and trends in the data across forecast lags and between error variables than a static graph.

5. Conclusions

Climate patterns in the United States cleanly separate into six recognizable regions through a cluster analysis using the means and standard deviations of the weather variables provided in Table 1. We explored seasonal differences of forecast errors in Figure 4 and observed that seasonal differences in forecast errors tend to be more pronounced in northern, inland clusters than southern clusters. We also showed that location specific anomalies, such as the asymmetry in seasonal maximum temperature forecast errors in San Francisco and the precipitation forecast errors near the Great Lakes, have plausible explanations in the literature.

We visualized the pairwise relationship between forecasts errors through a series of scatterplots across all forecast lags in Figure 5. These plots highlight the superiority of locations in the Cali-Florida region for predicting minimum temperature across all lags, and also show that the poor precipitation predictions of the Great Lakes region are mostly confined to forecasts greater than lag 2. Lastly, the abnormally high errors in Austin, Nevada, are likely a product of the large distance between forecast and measurement locations. These scatterplots show signs of correlations between the different error measurements which we rigorously explored with a series of correlation ellipses superimposed on a map of the U.S. in Figure 6. We found that all clusters show signs of positive correlations among the error variables with the exception of the Northeast cluster.

Next, we compared the important variables in determining forecast errors across clusters using scaled random forest variable importance measures in Figure 7. These measures demonstrate that forecast lag is most important in determining the maximum temperature and the precipitation forecast errors, but not important in predicting the minimum temperature forecast errors. Many clusters place similar importance on a few variables, but there are some variables that are important only in a single cluster, such as the importance of maximum wind speed in predicting the maximum temperature forecast error in Cali-Florida.

For further insight regarding the nature of forecast errors across these six clusters, we refer readers to our R shiny app described in the previous section. A current version of the app can be found at the following URL:

```
https://jilllundell.shinyapps.io/finaldataexpoapp/
```

This app, in conjunction with the visualizations presented in this article, reinforces the idea that the U.S. cleanly clusters into well defined weather regions and patterns in forecast errors are closely related to the unique climates that characterize each region.

6. Tools and Acknowledgements

References

Bivand, R., Keitt, T., and Rowlingson, B. (2018). rgdal: Bindings for the 'geospatial' data abstraction library. `https://cran.R-project.org/package=rgdal`. R package version 1.2-18.

Bivand, R. S., Pebesma, E., and Gomez-Rubio, V. (2013). *Applied spatial data analysis with R*. Springer, New York, NY, second edition.

Carr, D. B., Olsen, A. R., and White, D. (1992). Hexagon mosaic maps for display of univariate and bivariate geographical data. *Cartography and Geographic Information Systems*, 19(4):228–236.

Chamberlain, S. (2017). rgbif: Interface to the global 'biodiversity' information facility api. `https://cran.R-project.org/package=rgbif`. R package version 0.9.9.

Cohen, J., Pfeiffer, K., and Francis, J. A. (2018). Warm Arctic episodes linked with increased frequency of extreme winter weather in the United States. *Nature Communications*, 9(1):869.

Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The elements of statistical learning*, volume 1. Springer, New York, NY.

GACC (2018). Austin, Nevada: So much to do. The Greater Austin Chamber of Commerce, `http://austinnevada.com`.

Google (2018). Get started. `https://developers.google.com/maps/documentation/elevation/start`.

Hijmans, R. J. (2016). geosphere: Spherical trigonometry. `https://cran.R-project.org/package=geosphere`. R package version 1.5-5.

Liaw, A. and Wiener, M. (2002). Classification and regression by randomforest. *R News*, 2(3):18–22. `https://cran.R-project.org/doc/Rnews/`.

McIlroy, D., Brownrigg, R., Minka, T. P., and Bivand., R. (2017). mapproj: Map projections. `https://cran.R-project.org/package=mapproj`. R package version 1.2-5.

Meschiari, S. (2015). latex2exp: Use latex expressions in plots. `https://cran.R-project.org/package=latex2exp`. R package version 0.4.0.

Murdoch, D. and Chow, E. (1996). A graphical display of large correlation matrices. *The American Statistician*, 50(2):178–180.

Murtagh, F. and Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *Journal of Classification*, 31(3):274–295.

Narasimhan, R. (2017). weatherData: Get weather data from the web. `https://cran.R-project.org/package=weatherData`. R package version 0.5.0.

Neuwirth, E. (2014). Rcolorbrewer: Colorbrewer palettes. `https://cran.R-project.org/package=RColorBrewer`. R package version 1.1-2.

NGS (2018). NOAA shoreline data explorer. National Geodetic Survey, `https://www.ngs.noaa.gov/CUSP/`.

NOAA (2018). Climate data online. `https://www.ncdc.noaa.gov/cdo-web/`.

NOAA (2018). NOAA medium resolution shoreline. `https://shoreline.noaa.gov/data/datasheets/medres.html`.

Nolte, C. (2016). The story of the San Francisco summer is a bit foggy. `https://www.sfchronicle.com/bayarea/nativeson/article/Summer-fog-can-be-here-today-gone-tomorrow-9141028.php`.

Nychka, D., Furrer, R., Paige, J., and Sain, S. (2015). fields: Tools for spatial data. `https://cran.R-project.org/package=fields`. R package version 9.0.

Pedersen, T. L. (2018). ggforce: Accelerating 'ggplot2'. `https://cran.R-project.org/package=ggforce`. R package version 0.1.2.

Scott, R. W. and Huff, F. A. (1996). Impacts of the Great Lakes on regional climate conditions. *Journal of Great Lakes Research*, 22(4):845–863.

Silver, N. and Fischer-Baum, R. (2014). Which city has the most unpredictable weather? `https://fivethirtyeight.com/features/which-city-has-the-most-unpredictable-weather/`.

Tufte, E. R. (2002). *The visual display of quantitative information*. Graphics Press, Cheshire, CT, second edition.

Weigel, A. P., Liniger, M. A., and Appenzeller, C. (2007). The discrete Brier and ranked probability skill scores. *Monthly Weather Review*, 135(1):118–124.

Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, 21(12):1–20. `http://www.jstatsoft.org/v21/i12/`.

Wickham, H. (2017). tidyverse: Easily install and load the 'tidyverse'. `https://cran.R-project.org/package=tidyverse`. R package version 1.2.1.

Wickham, H., Hofmann, H., Wickham, C., and Cook, D. (2012). Glyph-maps for visually exploring temporal patterns in climate data and models. *Environmetrics*, 23(5):382–393.