Noname manuscript No.
(will be inserted by the editor)

Accuracy of Precipitation Forecasts:
Finding the Right Threshold for What is Considered Rain

Darren Keeley · Eric A. Suess

Received: date / Accepted: date

Abstract Accurately predicting the rain is a fundamental component of weather forecasting. However, looking at the data provided for the 2018 ASA Data Expo Challenge, forecasts were consistently underpredicting the proportion of rainy days. The default threshold in inches of rain for what is considered a rainy day is 0.01 inches or more as defined by the National Weather Service. We found that adjusting the threshold for each city dramatically increases probability of precipitation forecast accuracies, and that generally across the United States a threshold of 0.07 inches is better than 0.01.

Keywords 2018 ASA Data Expo [2] · weather forecasting · threshold for rain · data science

1 Introduction

The United States has regions of vastly different climates. The weather varies to such a degree that it may be beneficial to judge and measure it differently for each region. Indeed, rainfall is one of the most variable characteristics, and what the residents of one city may consider a rainy day may not reach the threshold residents of another city would consider a rainy day.

In this paper different thresholds for rainy days are explored across 108 cities in the United Sates. The impact this has on the accuracy of precipitation forecasts is analyzed. The similarities amongst the least accurate cities are

Darren Keeley
Department of Statistics and Biostatistics, California State University, East Bay,
25800 Carlos Bee Boulevard, Hayward, CA 94542
E-mail: darrenkeeley@gmail.com

Eric A. Suess
Department of Statistics and Biostatistics, California State University, East Bay,
25800 Carlos Bee Boulevard, Hayward, CA 94542
E-mail: eric.suess@csueastbay.edu

explored, as well as those for the most accurate [3]. R was utilized for the data wrangling, data analysis, and visualizations for this paper [5] [7].

2 Measuring Precipitation Forecast Accuracy

The original data set contained 3,191,972 daily weather forecasts and 130,457 historical data points from July 2014 to September 2017. Forecasts consisted of two types, temperature and probability of precipitation, and only the latter is discussed in this paper. The weather forecasts ranged from same day to seven days out. The historical dataset contained many more metrics than those in the forecast dataset, including dew point, humidity and inches of precipitation.

Forecasts for rain were given by the Probability of Precipitation (`Pop`), defined by the National Weather Service (NWS) as the probability that at least an average of .01 inches of rain will fall within an area in a given day [4]. The `Pop` is computed as follows:

$$\texttt{Pop} = C * A, \tag{1}$$

where C is the probability that any rain will fall somewhere in the area, and A is the percentage of the area covered by the rain. For example, if there is an 80% chance that rain will fall upon 50% of a town and 0% chance upon the other half, then the `Pop` for that area is 40%.

An original accuracy metric was defined for this paper since the NWS gives no guidelines on how to assess `Pop` accuracy, called *Rain Error* (RE). This metric first bins all the days by their `Pop` forecasts; the `Pop` forecasts given were incremental discrete values of 10 between 0 and 100, yielding 11 bins altogether [6]. For instance, all days in a period of interest with 10% `Pop` were binned together, all days with 20% `Pop` were binned, etc. Then each `Pop` increment is compared to the proportion of days in their bin that it actually rained. Written mathematically,

$$Rain\ Error = Proportion\ of\ Rainy\ Days - \texttt{Pop}. \tag{2}$$

Examples:

- For days that had 40% `Pop` forecasts, if it rained on 40% of those days, then the forecasts were accurate with an error of 0.
- If on 60% of those days it rained, then the forecasts under-predicted the rain with an error of 20%.
- And if only on 10% of those days it rained, then the forecast over-predicted the rain with an error of -30%.

It is important to note that **positive errors** reflect forecasts that were **under-predicted**, with the weather yielding more rainy days than expected (in other words, wetter than expected). And that **negative errors** reflect forecasts that were **over-predicted**, with the weather yielding fewer rainy days than expected (or dryer than expected).

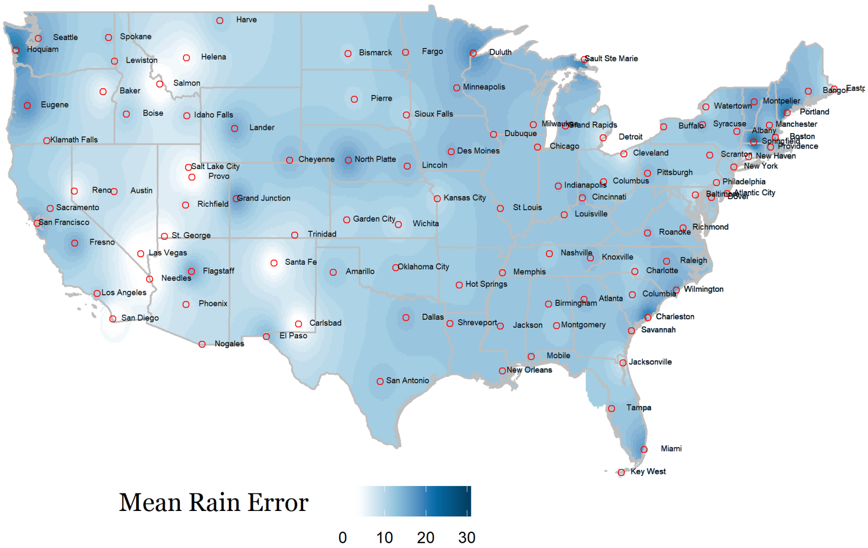

Fig. 1 Weather is usually wetter than predicted. Blue indicates more rainy days than predicted (positive Mean Rain Error). White is relatively equal numbers of days.

3 The Weather is Usually Wetter than Predicted

For each city the *Mean Rain Error* was computed to determine the cities with the least accurate forecasts, see Table 1. The *Mean Rain Error*, for each city was computed, as follows:

$$Mean\ Rain\ Error_{city} = \frac{1}{11}\sum_{i=1}^{11} Rain\ Error_i \tag{3}$$

where the 11 bins are the percentiles the *Rain Error* is split into, as described above. If certain bins do not contain any forcasts the mean is computed over the reduced number of bins that contained forcasts. For example, the city of Albany did not have any 10% `Pop` forcasts, so only 10 bins were used and averaged over.

Precipitation forecasts generally under-predict the rain, with the weather being wetter than expected across nearly the entirety of the United States, except for the Rocky Mountains and cities adjacent to them. In Figure 1, blue indicates more rainy days than predicted (positive Mean Rain Error). White is relatively equal numbers of days.

The map shows that a large majority of cities are underpredicting the number of rainy days, with the rest being a bit more accurate [8]. Since there was only data on the cities, the precipitation errors in the spaces between them were interpolated to aid with the illustration [1].

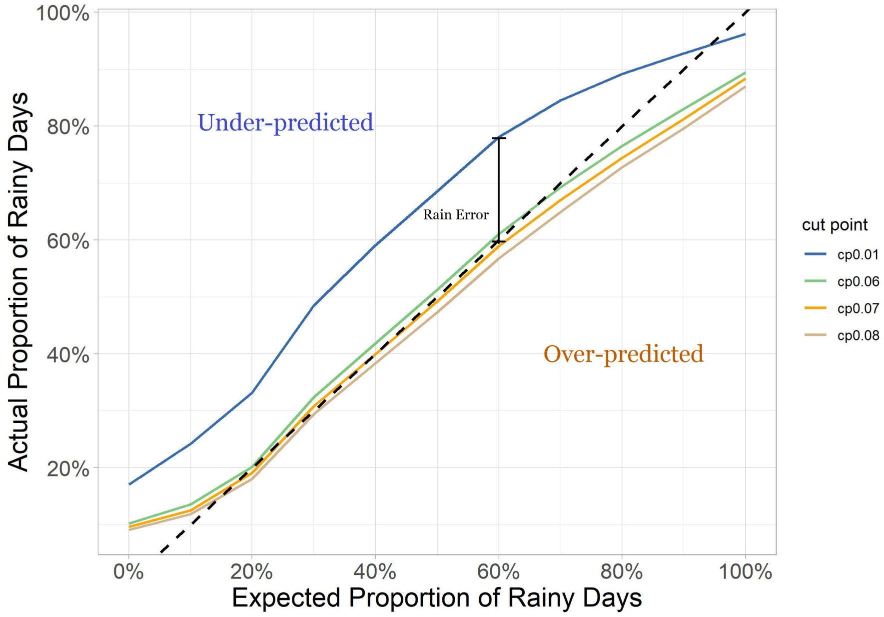

Fig. 2 Above the dashed line is more rain than predicted, thus **under-predicted**. Using higher cut points give better accuracy.

Part of the issue may be that the NWS sets the bar for a rainy day as 0.01 inches of precipitation or more [4]. Such small amounts of rain could be too small to predict or even be reasonably considered rain. If these very light rainy days are safe to ignore, then perhaps it would be better to not predict them at all. Raising the threshold for how many inches of rain constitutes a rainy day would accomplish this, and is what is explored in the rest of the paper.

Figure 2 shows a visualization of the *Rain Error* of all cities by `Pop`, where the x-axis *Expected Proportion of Rainy Days* is the `Pop`, and the y-axis is the proportion of days that it actually rained. The black dashed line indicates perfect forecasts (indicated by the 45 degree line in the plots), where `Pop` equals the proportion of rainy days. Being vertically above the line is rainier (under-predicted), and vertically below is drier (over-predicted). Looking at the graph showing different cut points, or thresholds, 0.07 or more inches seems to yield smaller forecast errors: most of the time the proportion of rainy days is accurately predicted or otherwise over-predicted. Over-predicting the rain may be a less egregious error than under-predicting it, given the latter would cause outdoor plans to be canceled, and in the former case plans would never be made at all.

The two maps that follow, see Figure 3, illustrate the overall improvement in forecast accuracy after adjusting the threshold. Here *Mean Absolute Rain Error (MARE)* is used, so there is no distinction between cities that are drier or wetter than forecasted. This is also a better metric for accuracy since negative and positive errors do not cancel each other out. Raising the cut point to 0.07 shows improvement in most areas, though cities in the Rocky Mountains worsened because they were relatively drier and already accurate. The higher threshold of .07 may be reasonable across most cities if one can accept unannounced light drizzle. The *Mean Absolute Rain Error* is defined as follows:

$$MARE_{city} = \frac{1}{11} \sum_{i=1}^{11} |Rain\ Error_i| \tag{4}$$

where the 11 bins are the percentiles the *Rain Error* is split into, as described above.

The ten cities with the highest *MARE* using the 0.01 inch threshold are labeled in the Figure 3. Most of them are located on the coasts. The exact values for the *MARE* for each of the cities in the data set can be found in Table 1.

4 The Effects of Forecast Length

Longer forecast lengths have an adverse effect on accuracy. Beyond two days the *Mean Absolute Rain Error* tends to rise as the length increases, as shown in Table 2. These *MARE* values are computed for all of the US cities, available in the dataset, by forecast horizon.

However, the initial problem remains: rain is still consistently underpredicted, no matter the forecast length, see Figure 4. As such, forecasts of all lengths will continue to be included in the calculation of forecast accuracy.

5 What are the Characteristics of the Least (and Most) Accurate Cities?

The correlation between the number of rainy days per year and MARE is 0.6291, p-value $< .00001$. As shown in Figure 5, this value may be heavily influenced by the ten cities with very high errors and number of rainy days. Without them, the correlation is only 0.4551, p-value $< .00001$.

The 10 least accurate cities are shown in Figure 6. These error-prone cities experience rain often, but not in large amounts. They are the rainiest cities in terms of days but not necessarily in inches. Springfield, the city with the highest MARE, experiences an average of 209.74 rainy days per year with only 0.20 inches per day, corresponding to normalized (or standardized) values of 2.87 and -0.85, respectively. The formula used for normalization was

$$Z = \frac{X - \mu}{\sigma} \tag{5}$$

Figure 7 displays the 10 most accurate cities. All are dry but unremarkably so, with none being more than 2 standard deviations from either mean.

Table 1 Mean Absolute Rain Error (MARE) by city, ranked by highest errors to lowest. Entire 3-year dataset was used.

Rank	City	MARE	Rank	City	MARE
1	Springfield	29.608	55	Columbus	14.015
2	Portland	25.962	56	San Antonio	13.999
3	Hoquiam	24.613	57	St Louis	13.979
4	Charleston	23.560	58	Shreveport	13.905
5	Sault Ste Marie	23.152	59	Birmingham	13.887
6	Duluth	21.557	60	Sacramento	13.791
7	Eugene	20.435	61	Nashville	13.716
8	Miami	19.212	62	Las Vegas	13.678
9	Montpelier	19.184	63	Scranton	13.666
10	North Platte	18.976	64	Buffalo	13.553
11	Minneapolis	18.587	65	Amarillo	13.533
12	Grand Junction	18.523	66	Tampa	13.528
13	Wilmington	18.231	67	Manchester	13.520
14	San Francisco	17.534	68	Montgomery	13.517
15	Raleigh	17.403	69	New York	13.326
16	Cheyenne	17.343	70	Austin	13.276
17	Des Moines	17.249	71	Oklahoma City	13.215
18	Flagstaff	17.208	72	Idaho Falls	13.210
19	Lander	17.185	73	Columbia	13.152
20	Cincinnati	16.864	74	Kansas City	13.145
21	Seattle	16.768	75	Hot Springs	13.126
22	Fresno	16.763	76	Nogales	13.014
23	Pittsburgh	16.437	77	Harve	12.961
24	Boston	16.073	78	Richfield	12.804
25	Syracuse	15.858	79	Philadelphia	12.549
26	Knoxville	15.671	80	Garden City	12.525
27	Fargo	15.654	81	Klamath Falls	12.396
28	Lincoln	15.380	82	Los Angeles	12.181
29	Atlanta	15.363	83	Watertown	12.138
30	Chicago	15.337	84	Charlotte	12.014
31	Grand Rapids	15.337	85	Sioux Falls	11.947
32	Louisville	15.325	86	Key West	11.943
33	El Paso	15.309	87	Bismarck	11.825
34	Providence	15.292	88	Trinidad	11.698
35	Dallas	15.181	89	Wichita	11.610
36	Bangor	15.175	90	Baker	11.041
37	Honolulu	15.101	91	Salmon	10.824
38	Savannah	15.077	92	San Diego	10.582
39	Indianapolis	15.075	93	Eastport	10.547
40	Jackson	15.055	94	Cleveland	10.101
41	Atlantic City	14.991	95	Detroit	10.014
42	Dubuque	14.945	96	Baltimore	10.005
43	Dover	14.890	97	Jacksonville	9.698
44	Roanoke	14.850	98	Salt Lake City	9.366
45	Albany	14.829	99	Phoenix	8.758
46	Richmond	14.810	100	Needles	8.569
47	Mobile	14.575	101	Helena	8.287
48	New Orleans	14.487	102	Santa Fe	8.155
49	Boise	14.460	103	Carlsbad	8.105
50	Pierre	14.354	104	St. George	7.611
51	Milwaukee	14.269	105	Provo	6.128
52	Lewiston	14.235	106	Anchorage	5.503
53	Memphis	14.214	107	Reno	5.476
54	Spokane	14.085	108	New Haven	5.467

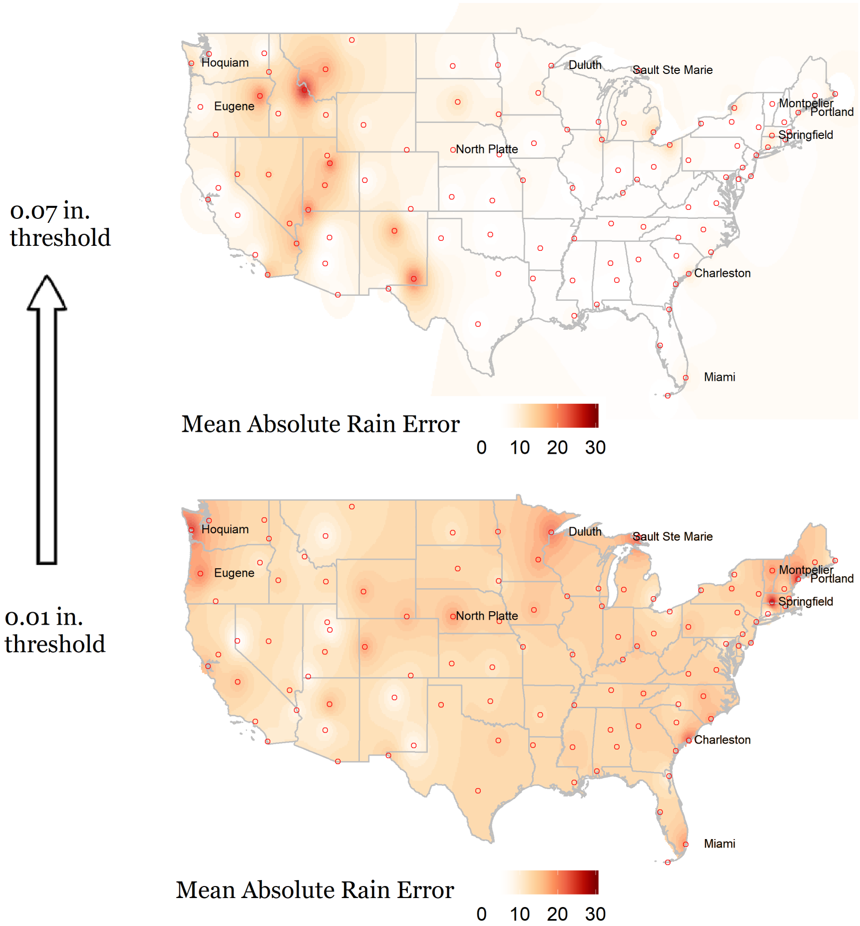

Fig. 3 When the cut point is raised from 0.01 to 0.07 in., days with barely any rain are not counted as rainy. Mean Absolute Rain Error (MARE) falls except for cities in the Rocky Mountains which were dry or already accurate. The top 10 least accurate cities are labeled.

Table 2 Errors in precipitation forecasts of the United States tend to increase as the forecast horizon increases.

Forecast Length	MARE
0 days out	13.58
1 day out	12.18
2 days out	13.20
3 days out	14.11
4 days out	17.26
5 days out	14.33
6 days out	15.30

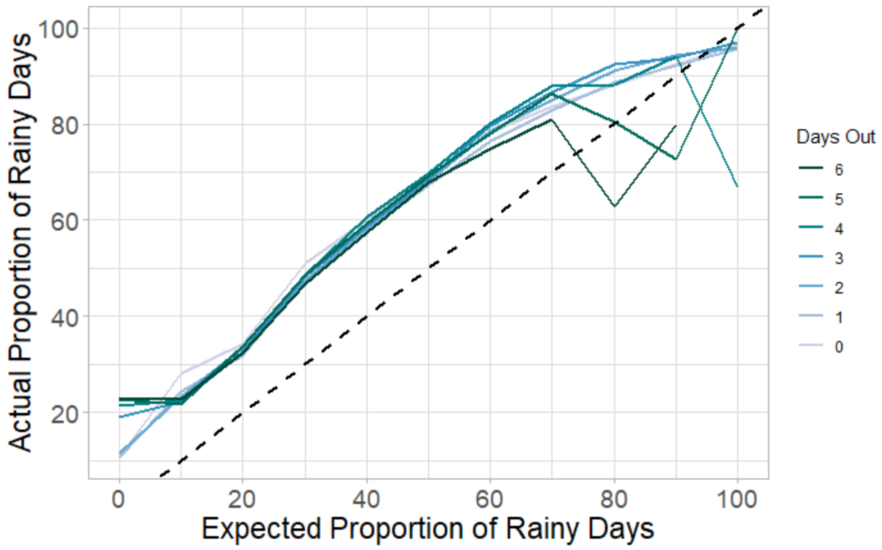

Fig. 4 All forecast lengths follow the same pattern of underpredicting precipitation. The threshold shown is 0.01 inches.

6 A Unique Threshold for Each City

Changing the cut point from 0.01 to 0.07 helps most cities. Forecasts from Springfield and Portland, the two least accurate cities, become much more precise when 0.06 inches of rain and below are ignored. See Figure 8. Seattle and Sacramento are two cities that do not rank highly in terms of errors, and they also benefit from the change. See Figure 9.

However, changing the cut point does not help every city. Honolulu has an exceptional number of trace rain days, making it seemingly drier by these metrics. Honolulu has an exceptional number of *trace* rain days, which are days that it rained less than 0.01 inches and not counted as rainy days. An average of 29.2% of days per year in Honolulu were trace rain days, making it seemingly drier by these metrics. The median proportion of trace days amongst all cities is 9.6%. Salmon, located in the Rocky Mountains, is a very dry city with few rainy days. See Figure 10.

The cut off for what constitutes rain can drastically alter the accuracy of precipitation forecasts. Local meteorologists should consider what cut point makes sense for their area, as some people may consider 0.01 inches of rain as rainy and others not.

7 Summary

Changing the precipitation threshold for rainy days is a simple and actionable way to improve rain forecasts. For cities that often experience very light rain,

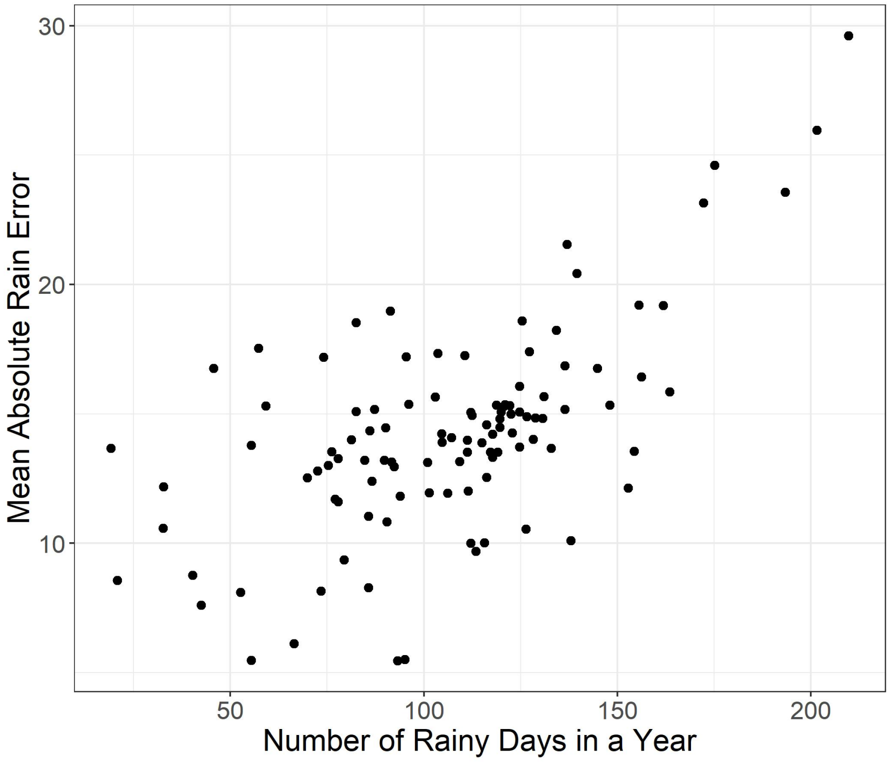

Fig. 5 There is moderate correlation between how often it rains in a city and its Mean Absolute Rain Error. Each point is a city.

discounting those days may not adversely affect the populace and would save rain notifications for heavier drizzles. Not every city needs an adjustment, and drier areas could benefit from greater precision in precipitation detection.

8 Further Work

The weather analyses conducted here could be applied to identifying, forecasting and mitigating wildfire risk. Data on wind and humidity were present in the dataset; the intersection of these factors plus the extremes of precipitation (droughts and downpours) could be combined to produce a useful metric for fire risk. This metric could then be displayed using the interpolated mapping technique utilized in this paper to identify high fire risk regions and hopefully spur better preparedness and policy.

Acknowledgements The authors are grateful to the organizers of the 2018 Joint Statistical Meetings, and the members of the ASA section on Statistical Graphics and the section on Statistical Computing for the opportunity to create and share our work.

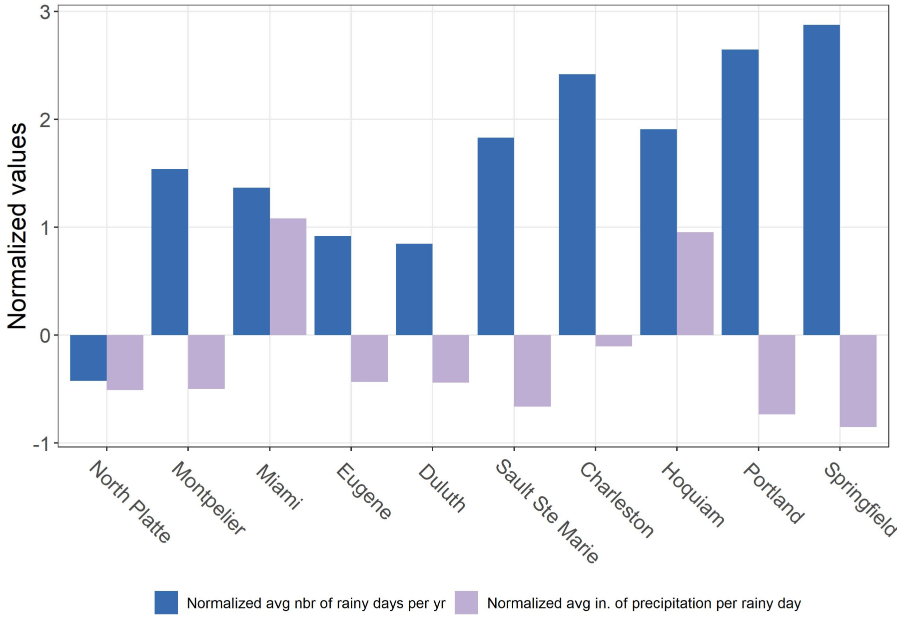

Fig. 6 The cities with the highest errors experience rain often, but not in necessarily high amounts. The blue bars represent the normalized MARE and the purple bars represent the normalized inches of rain.

References

1. Aasa, Anto, Spatial interpolation in R, `http://aasa.ut.ee/LOOM02331/` (2016)
2. Cetinkaya-Rundel, Mine, Martinez, Wendy, The 2018 Data Challenge Expo of the American Statistical Association, Computational Statistics, Vancouver, BC, Canada (2023)
3. Dubner, Stephen, How Valid Are T.V. Weather Forecasts?, `http://freakonomics.com/2008/04/21/how-valid-are-tv-weather-forecasts` (2008)
4. Marsh, Duke, What Does Probability of Precipitation Mean?, `https://www.weather.gov/lmk/pops` (2015)
5. R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2014) `http://www.R-project.org/`
6. Suess, Eric A., Trumbo, Bruce E., Introduction to Probability, Simulation and Gibbs sampling with R, Springer, New York, NY (2010)
7. Wickham, Hadley, ggplot2: elegant graphics for data analysis, Springer, New York, NY (2009)
8. Yau, Nathan, Visualize This: The FlowingData Guide to Design, Visualization, and Statistics, Wiley, Indianapolis, IN (2011)

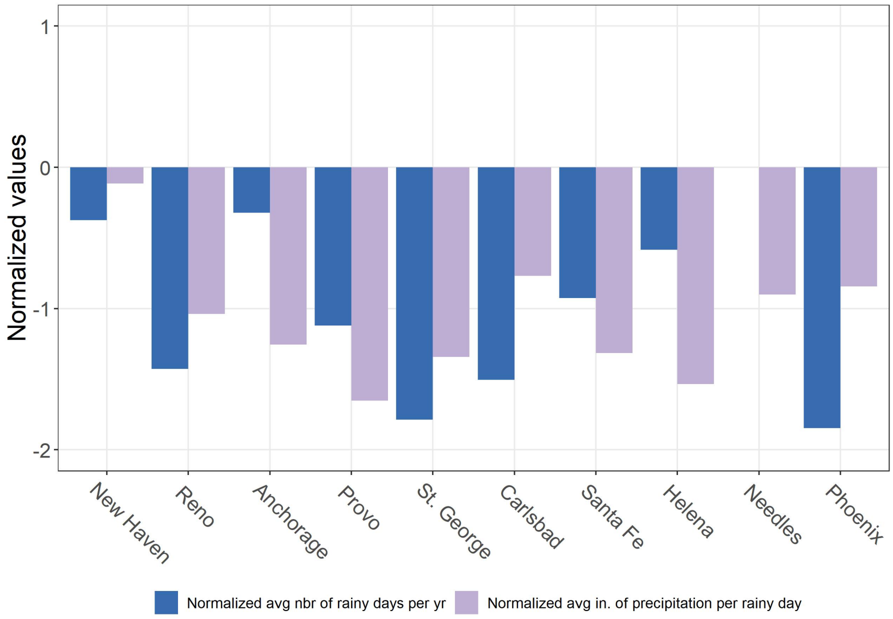

Fig. 7 The cities with the lowest errors experience rain less often and in smaller amounts. The blue bars represent the normalized MARE and the purple bars represent the normalized inches of rain.

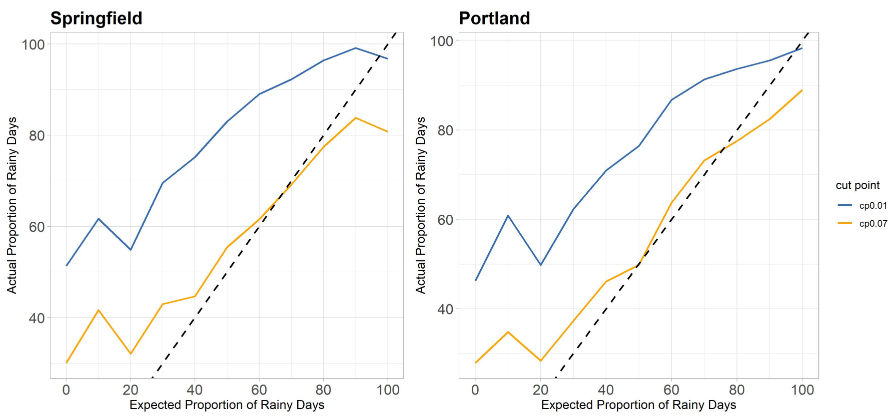

Fig. 8 The higher cut point of 0.07 (yellow line) improves the accuracy over the 0.01 cut point (blue line).

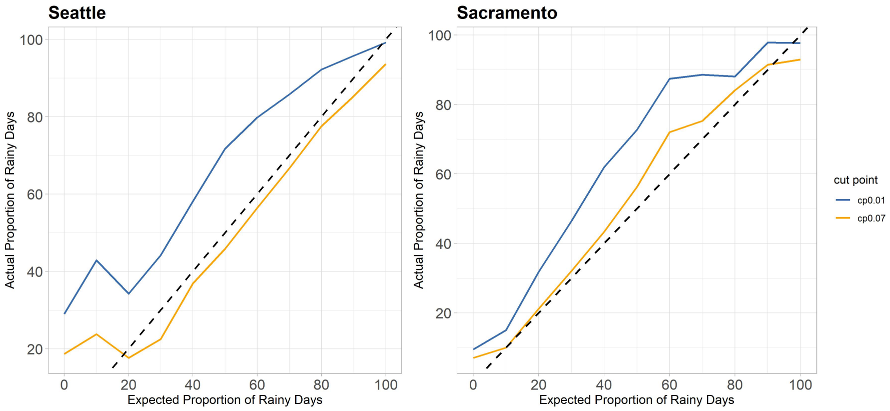

Fig. 9 The higher cut point of 0.07 (yellow line) improves the accuracy over the 0.01 cut point (blue line).

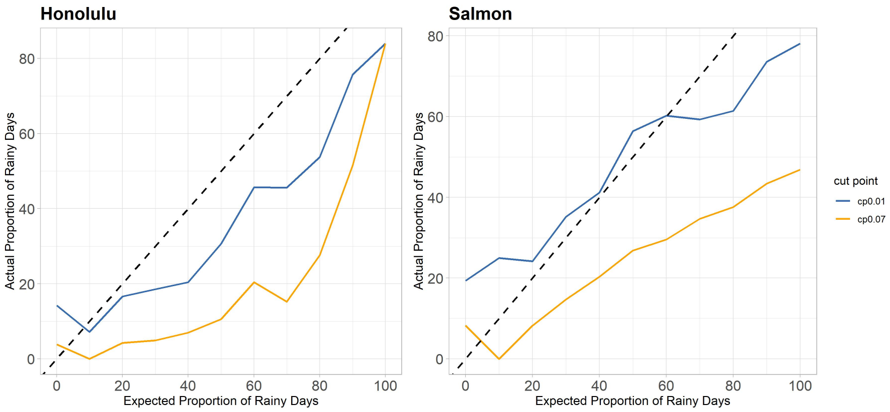

Fig. 10 For some cities, the original cut point should be kept. The higher cut point of 0.07 (yellow line) does not improve the accuracy over the 0.01 cut point (blue line).