Evaluating radar reflectivity measurements as predictors of rainfall

Marisa Akers and Meera Venkataraman
Department of Statistics, North Carolina State University

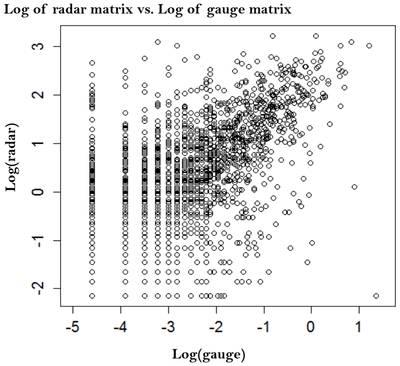
Abstract

To improve predictions of weather system models, it is important to have accurate measurements of precipitation at all locations. Actual amounts of rainfall have high variability across space and time, and patterns are generally unpredictable. Gauges measure rainfall, but only at specific locations. Therefore, a reliable prediction method for all locations in a given region is needed. One common method of predicting rainfall is to use measurements of reflectivity from radars. However, radar data is not directly comparable to gauge data because they measure reflectivity and actual precipitation amounts, respectively. The data analyzed contains 406 radar measurements covering about 62,000 square miles in Kansas for August 2004. We match these hourly readings to the 180 gauge stations in this region by the day and hour of measurement. Our main goal is to evaluate how radar reflectivity measurements can be used to predict precipitation. To address this goal, we examine zero-inflated regression models with precipitation as the response variable and radar reflectivity readings as a covariate.

Problem Definition

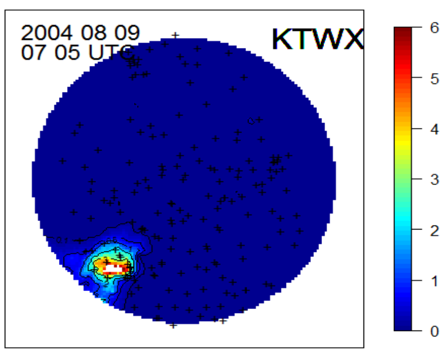
-Radar and gauge data, due to the high percentage of days lacking precipitation, is zero-inflated. We expect to see a positive correlation between radar and gauge values, and using a log transformation helps improve clarity when plotting.

Log of radar matrix vs. Log of gauge matrix



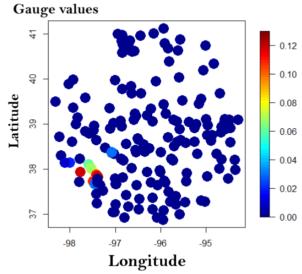
-Radar provides a 360°, continuous set of reflectivity measurements. Gauge readings only provide information for a single location per time.

Radar reflectivity values, with marked gauge locations

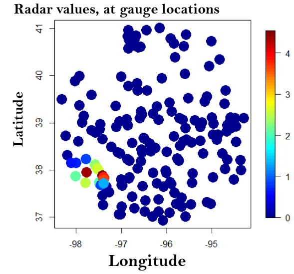


-To make a comparison between collected radar and gauge data, spatial and temporal information must be collected and values must be sorted and paired accordingly.

Gauge values



Radar values, at gauge locations



Method

Modeling precipitation data is notoriously difficult because of the high proportion of zeros. Therefore, we separate the model into two components: (1) a logistic regression for the presence of rain and (2) a linear regression for the log rainfall amount, given there is rain.

Data:
Collected August 1st-31st, 2004
Topeka, Kansas

180 Gauges measured precipitation 406 times

The same180 locations were pinpointed on the radar at the 406 specified times in order to extract coordinated data.

9 Covariates:

- -Gauge: collected precipitation, in inches
- -Gauge mask: set of 1's (if gauge value>0) and 0's (if gauge value=0)
- -Radar: reflectivity at the locations of the gauges, in decibels of Z (dBZ)
- -Radar mask: set of 1's (if radar value>0) and 0's (if radar value=0)
- -Radar² and $\sqrt{\text{Radar}}$: radar data, transformed
- -Longitude/Latitude: location of each gauge
- -Distance: calculated distance of each gauge location from the radar

Both the logistic and linear models started with all covariates included. Backwards elimination was used to remove covariates, until all included were statistically significant.

Logistic Regression: Calculates probability of precipitation occurring

$$\text{Model: } Probability\ of\ rain = e^{z}/[1+e^{z}]$$

where $z = \alpha_0 + \alpha_1 radar + \alpha_2 radar^2 + \alpha_3\sqrt{radar} + \alpha_4 longitude + \alpha_5 longitude^2 + \alpha_6 radar\ mask$

Linear Regression: Predicts gauge value, given a certain radar value and location

$$\text{Model: } log(rain\ value) = \beta_0 + \beta_1 log(radar) + \beta_3 radar^2 + \beta_4\sqrt{radar} + \beta_5 dist.v + \beta_6 radar.mask + error$$

Results

Signif. codes: 0 '***' 0.001 '**' 0.01 '*'

Logistic Regression

	Estimate	Std. Error	z value	Pr(>|z|)
radar	-1.806	0.239	-7.555	4.20e-14 ***
radar²	0.033	0.009	3.510	0.000448 ***
$\sqrt{\text{radar}}$	6.07	0.4956	12.247	< 2e-16 ***
longitude	-27.75	7.566	-3.668	0.000245 ***
longitude²	-0.144	0.039	-3.664	0.000249 ***
radar mask	1.299	0.231	5.628	1.82e-08 ***

AIC: 5085.4

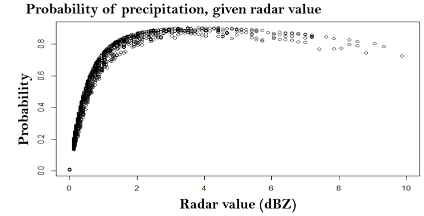
Linear Regression

	Estimate	Std. Error	t value	Pr(>|t|)
radar.v	0.089	0.002	41.367	< 2e-16 ***
radar2	-0.002	0.00008	-29.022	< 2e-16 ***
sqrtradar	-0.077	0.005	-16.990	< 2e-16 ***
dist.v	-0.00001	0.000004	-3.072	0.002128 **
radar.mask	0.025	0.002	12.381	< 2e-16 ***

Adjusted R-squared: 0.482

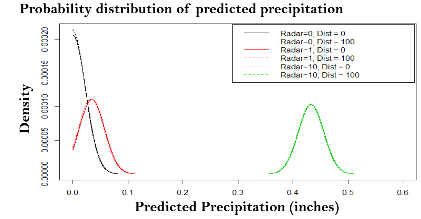
Precipitation is likely zero if the radar value is zero, and probability generally increases with the radar value.

Probability of precipitation, given radar value



Distance from radar, though significant, has a small effect on predicted precipitation.
As radar values increase, predicted precipitation amount does as well.

Probability distribution of predicted precipitation



Future Work

- Apply spatial statistics methods to produce a map of estimated precipitation values throughout the spatial domain.
- Explore various transformations to our zero-inflated data that will allow the use of standard statistical methods that assume data are normally distributed.
- Use the results for quality control by flagging observations that are very different from our spatial predictions.

Acknowledgments

Dr. Brian Reich and Gina-Maria Pomann, NC State Department of Statistics
Dr. Dongsoo Kim, National Climatic Data Center