Capstone :Prediction of Hourly Rainfall

Author: Venkatesh Viswanathan

Mentor Dr.Christos Papadopoulos

Date: 12/26/2015

Introduction

Rainfall is highly variable across space and time, making it notoriously tricky to measure. Rain gauges can be an effective measurement tool for a specific location, but it is impossible to have them everywhere. In order to have widespread coverage, data from weather radars is used to estimate rainfall nationwide. Unfortunately, these predictions never exactly match the measurements taken using rain gauges. In order to improve the efficiency of the rain fall measurements the weather department has installed dual polarimetry. Following report describes the problem stated in the project and proposes a solution. The data is used is from kaggle competition and relevant link is provided below.

https://www.kaggle.com/c/how-much-did-it-rain-ii

Problem

The key problem is to predict the amount of rainfall for remaining 10/11 days given the rainfall information for the 1st 20 days from April to August in the corn growing states. The existing data consists of Reflectivity, Reflectivity Composite, expected values and other data.

Data Selection

The following two variables are of importance as they are related to amount of rainfall

Ref	RefComposite	Expected Rainfall	ZDR	Kdp
Variable 1	Variable 2	Target	Positive values indicate that the targets are larger horizontally than they are vertically, while negative values indicate that the targets are larger vertically than they are horizontally. Values near zero suggest that the target is spherical, with the horizontal and vertical size being nearly the same. It is best used to detect heavy rain. Not Used	The specific differential p hase is defined as the range derivative of the differential phase shift between the horizontal and vertical pulse phases Used to delineate between rain, hail and snow Not Used

Table 1 : Defines the data that is selected for analysis

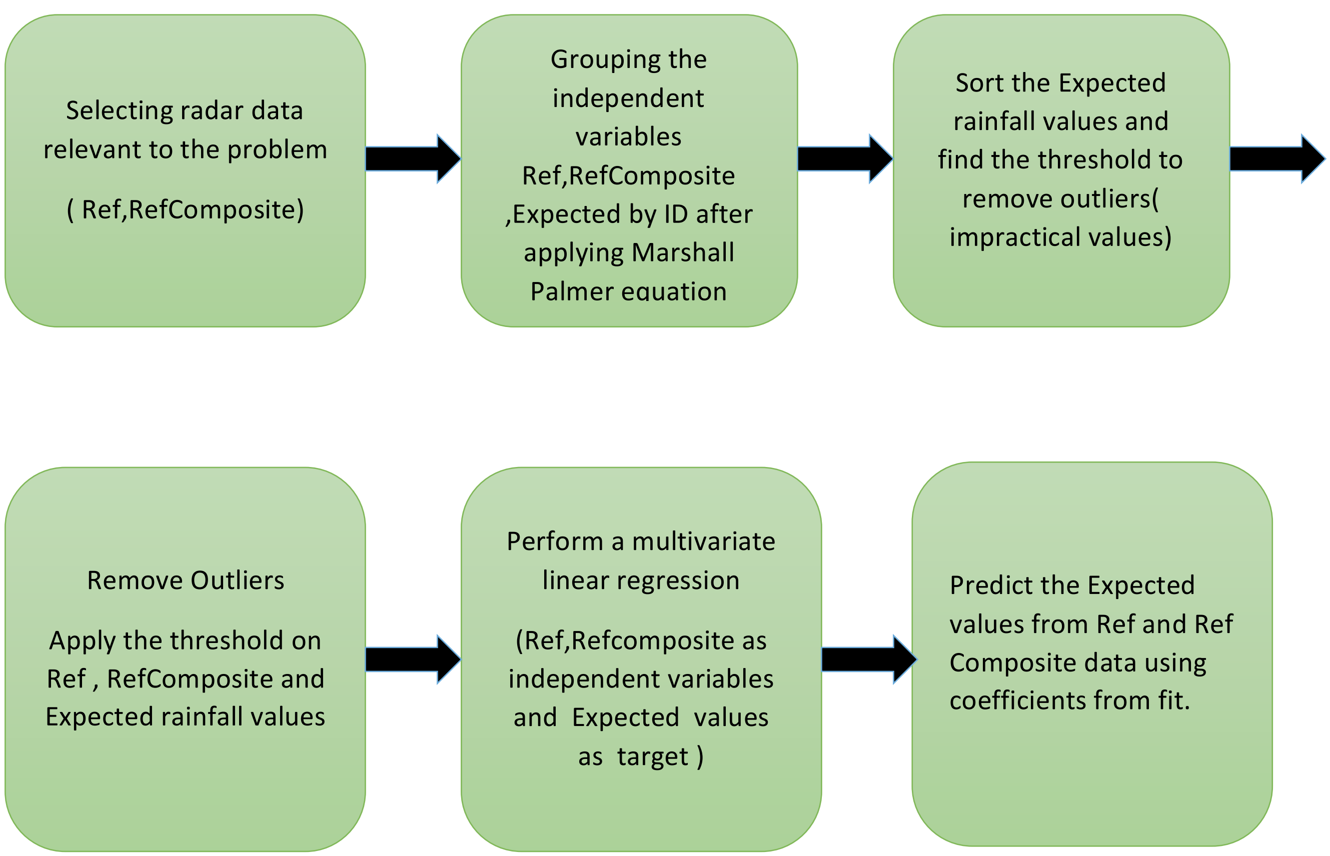
Reflectivity

Return echoes from targets ("reflectivity") are analyzed for their intensities to establish the precipitation rate in the scanned volume. The *Z*–*R* relationship developed by J. S. Marshall and W. M. Palmer (1948) consistent with an exponential distribution. The relationship is $Z = 200R^{1.6}$, where Z ($mm^6\ m^{-3}$) is the reflectivity factor and R ($mm\ h^{-1}$) is the rainfall rate. The relationship is sometimes generalized to the form $Z = a'R^b$**, where *a* and *b* are adjustable parameters.** In general it represents the rainfall at the gauges and its related to amount of rainfall per hour

RefComposite

The **composite reflectivity** is the maximum dBZ reflectivity from any of the reflectivity angles of the NEXRAD weather radar.[1] The reflectivity on individual PPI angles show the precipitation intensity at that specific angle above the horizon

Solution Design & Experiments



Two regression models were evaluated

1. Single variable (Ref) with Expected values as the target
2. Two variables (Ref,RefComposite) with Expected values as the target

The R2 scores were analyzed .

Observations

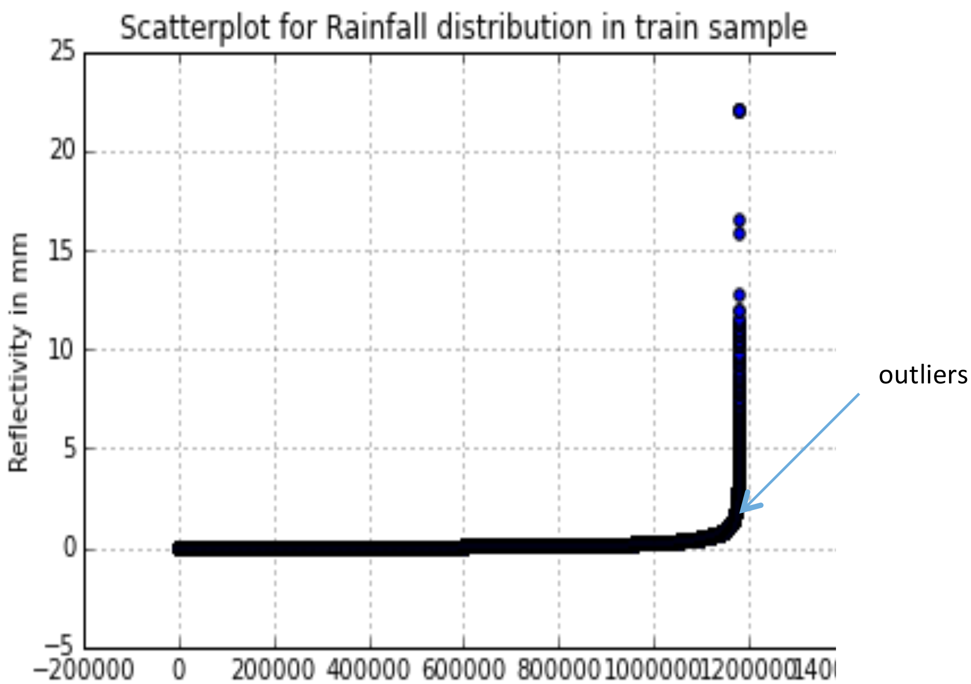


Fig 1 : Expected rainfall sorted to illustrate outliers

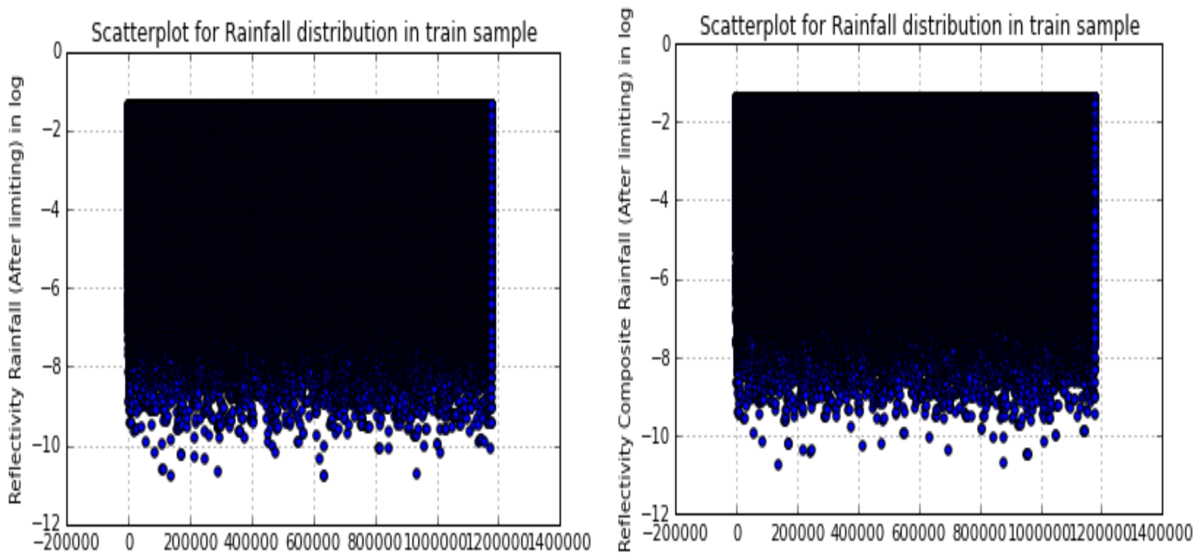


Fig 2 : Reflectivity and Composite Reflectivity in log after thresholding

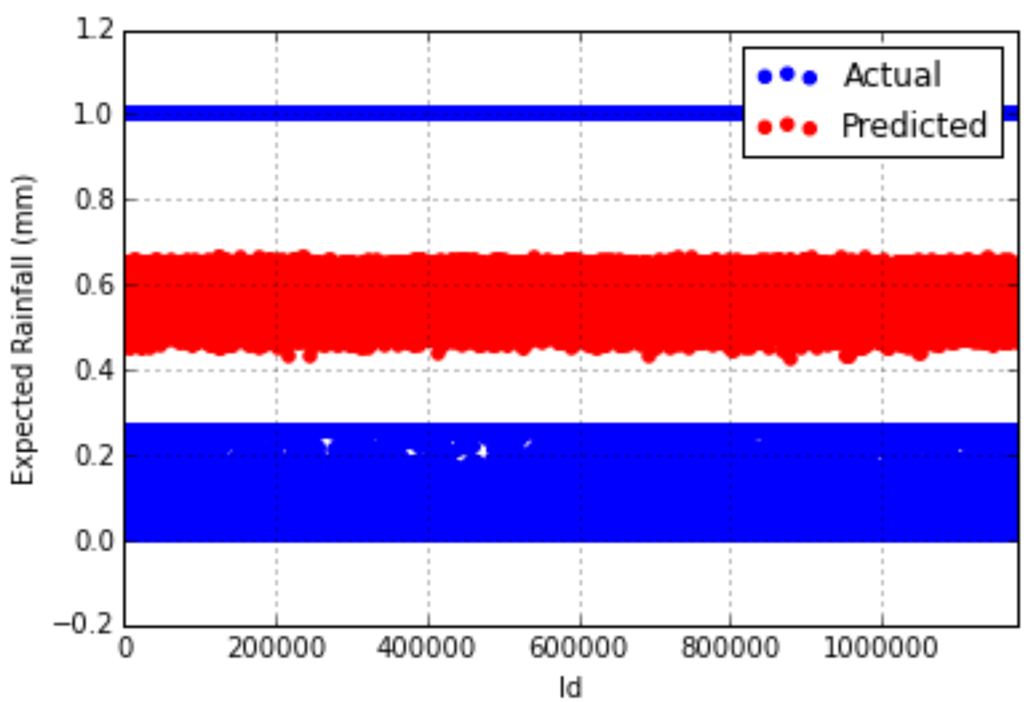


Fig 3 : Expected rainfall (Predicted) from training set(*Actual in blue and Predicted in Red*)

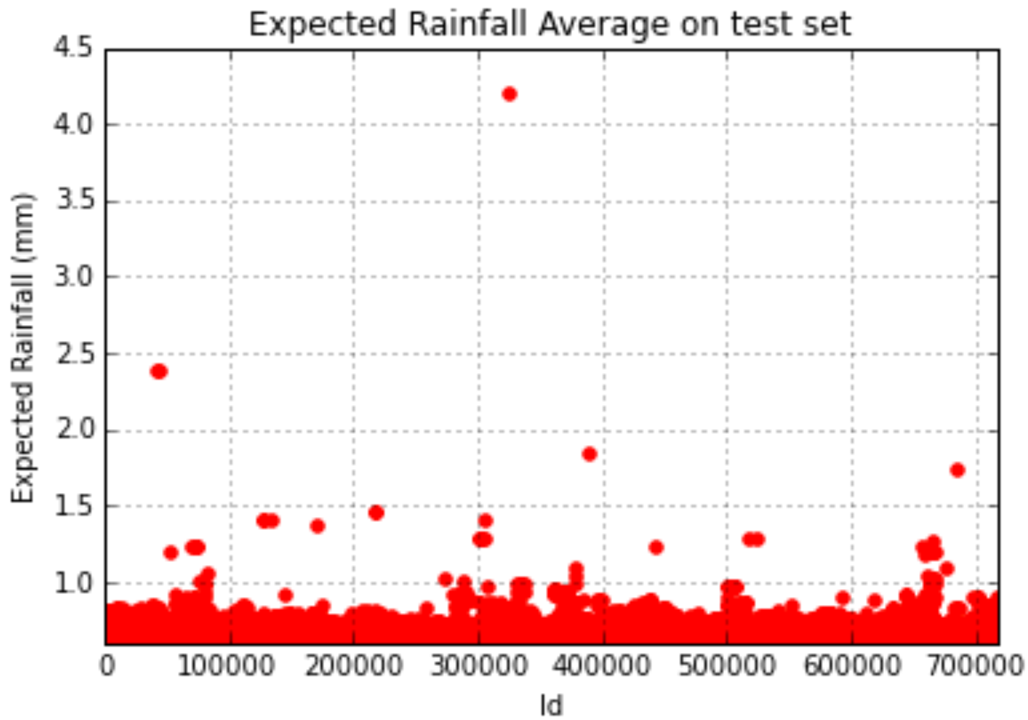


Fig 4 : Average rainfall(expected) in mm on the test set

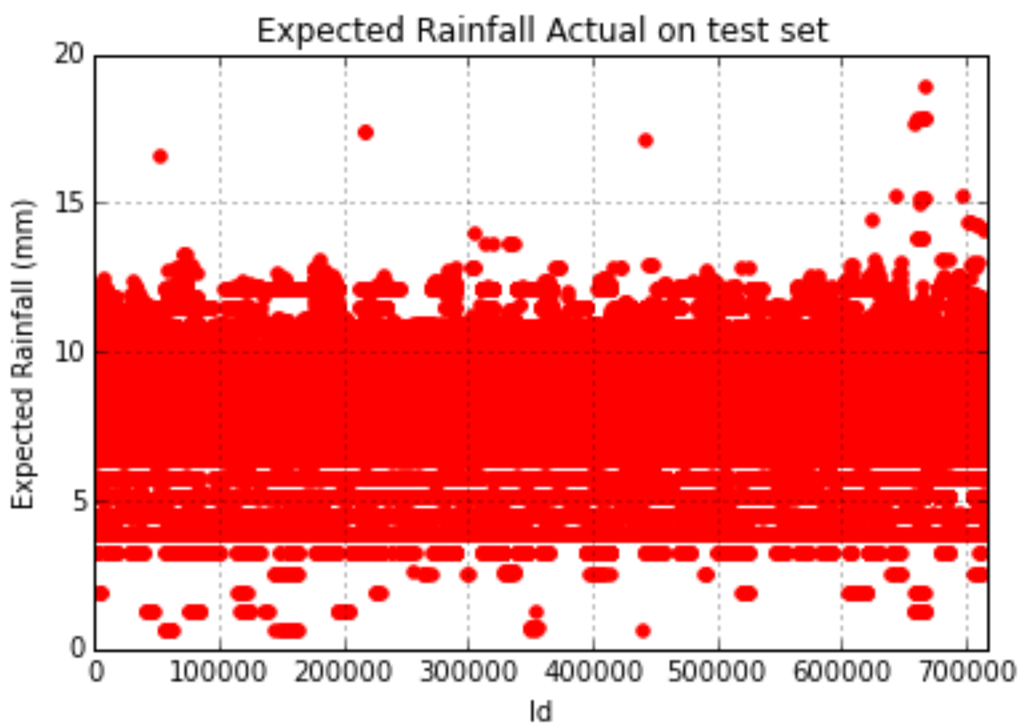


Fig 5 : Actual rainfall (Expected) on test set

Model	R2 Score	Comments
Single variable linear regression	0.00334163451724	Expected vs Ref
Bi variate linear regression	0.00409765657753	Expected vs Ref, RefComposite

Table 2: As the features increase the prediction score also increases

Conclusion

This project analyzes the data and identifies Ref and RefComposite as independent variables. As an important step it also removes outliers from the data by setting thresholds. A bi variate linear fit is done with Expected values as target on the training set resulting in coeffients. These coefficients are applied on the test set. The project also compares R2 scores on bivariate and single variate regression.

Future Scope

An alternate formula based on Kdp and ZDR is also a potential candidate that can be used to obtain the rain rate(based on Sachidananda and Zrnic (1987) given below.)

```
RATE_KDP = sign(KDP) * kdp_aa * (|KDP| ** kdp_bb).
```

References

1.http://radarscope.tv/hrf_faq/specific-differential-phase-kdp/

2. http://www.desktopdoppler.com/help/nws-nexrad.htm