ESTIMATING RAINFALL FOR INDEX-BASED AGRICULTURAL INSURANCE USING SATELLITE DATA

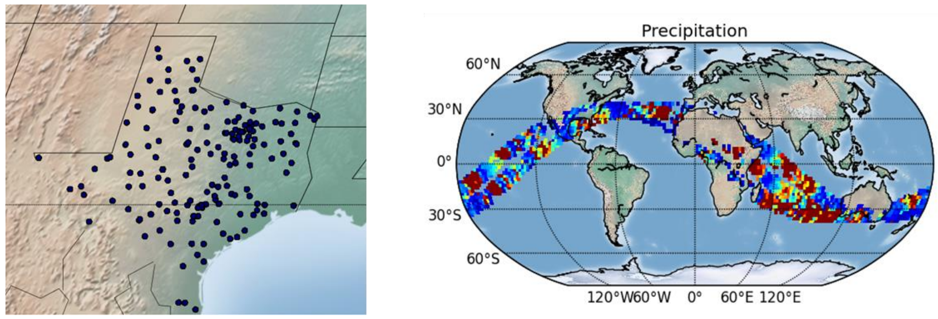
Arif Albayrak and William Teng
ADNET Systems, Inc.
NASA Goddard Earth Sciences Data and Information Services Center
Code 610.2
Greenbelt, MD 20771

In previous work on estimating rainfall for the purpose of indexed-based agricultural insurance, we used Neural Network to estimate rainfall amount for a specific geographical location, based on the nearest available station data (Albayrak and Teng, 2014; Bryle-Tressler at al., 2011; Dick and Stoppa, 2011). We focused on optimizing the NN as a supervised mathematical estimator, by characterizing rainfall data from three weather stations in Iowa. We used machine learning algorithms to obtain overall improvements on the rainfall estimates, as information content of inputs (e.g., number of stations) increased. Our results, while generally promising for application to weather-indexed agricultural insurance, showed the restrictions of the Neural Network approach in areas with limited number of weather stations, where the location of interest is too far away from these stations. Because of the sparsity of station data, variations in rain regimes were not accurately represented, resulting in spatially incomplete precipitation statistics and, therefore, an unbalanced training data set.

In our current work, we increased the amount of input data from station only to a combination of station and satellite data (gridded and swath). With the inclusion of Tropical Rainfall Measuring Mission (TRMM) satellite precipitation data in the training data set, we increased the spatial precipitation coverage, which enabled us to build a more complete spatial and temporal statistics of the region of interest. We investigated the biases between station and satellite data, and we used a machine learning approach to remove the biases. Our overall result, with the additional input data, showed an improvement and better correlation between the estimates of the stations.

For this study, we collected one year's worth of hourly precipitation data from the National Climatic Data Center (NCDC), using the web interface, "Climate Data Online: Data Discovery Tool (http://ncdc.noaa.gov)." We used Texas as our working area, for easy prototyping. There were over 300 stations offering data for the January-December 2010 period, and, from these, we selected 169 stations that offered full coverage. TRMM data were obtained from the Goddard Earth Sciences Data and Information Services Center (GES DISC). For this study, we used the TRMM Microwave Imager Hydrometeor Profile (2A12) product (http://disc.sci.gsfc.nasa.gov/daac-bin/DataHoldingsPDISC.pl?LOOKUPID_List=2A12), which includes surface rainfall, at the original sensor swath resolution. We also used the NASA Giovanni visualization tool for preliminary precipitation analysis. Co-location of the weather station and 2A12 swath data was performed using a tool developed at the GES DISC by one of the authors (Albayrak), called *Optimum Swath Reader Library for Science Data of Earth Observing System Mission* (OpSEEDat). OpSEEDat is designed specifically for handling multi-sensor satellite data (Fig. 1).

The major difference in our current vs. previous work is that the problem was modeled as a bias adjustment between stations and satellite data, rather than as an interpolation between stations. After the input data were prepared and reformatted, we used the Feed Forward Neural Networks method to estimate the precipitation (Blackwell and Chen, 2009) and the Gradient Boosting approach to estimate the regressors' importance levels (Friedman, 2001).



a b

Figure 1. OpSEEDat plots of (a) station locations in Texas and (b) TRMM Microwave Imager orbits showing retrieved precipitation from 2010-07-09 (red=heavy rain).

We evaluated the Neural Network in two stages, similar to what was described in Albayrak and Teng (2014). In the first stage, we set the base case for comparison, by including all data from 169 stations and the satellite. We obtained a correlation coefficient of 0.80 between the target station values and the TRMM estimates. After Neural Network was applied, the correlation coefficient increased by nearly 5%. Compared with the previous interpolation approach (Albayrak and Teng, 2014), using only station data, the inclusion of satellite data resulted in increase in correlation coefficient over 0.2. In the second stage, we excluded from the training process some of the stations, to represent target farm locations where rain is to be estimated. We were able to obtain results similar to those of the base case indicating minimum loss of information with the new approach estimations.

We are planning two follow-up efforts in this ongoing study. One is to capture the low and high frequency patterns by including additional regressors, such as humidity and soil moisture. The other is to incorporate data from the Global Precipitation Measurement (GPM) and Soil Moisture Active Passive (SMAP) missions.

REFERENCES

Albayrak A. and W. Teng, 2014, Estimating Rainfall for Index-Based Agricultural Insurance, in Proc. ASPRS 2014 Annual Conference, Louisville, KY.

Bryla-Tressler D. et al., 2011, Weather index insurance for agriculture: Guidance for development practitioners, *Agriculture and Rural Development Discussion Paper 50*, The International Bank for Reconstruction and Development / The World Bank.

Dick W. and A. Stoppa, 2011, Weather Index-based Insurance in Agricultural Development: A Technical Guide, International Fund for Agricultural Development (IFAD), ISBN 9789290722762.

Blackwell W. and F. Chen, 2009, Neural Networks in Atmospheric Remote Sensing, Lincoln Lab. ISBN 978-1-59693-372-9.

Friedman J. H., 2001, Greedy function approximation: A gradient boosting machine, The Annals of Statistics, 29(5), 1189-1232.