Bayesian Data Blending for Measuring Cloud Fraction

Sam Shen and Jeff Ledahl, Department of Mathematics and Statistics, San Diego State University
Richard Somerville, Scripps Institution of Oceanography, University of California, San Diego

INTRODUCTION

Cloud fraction is the fraction of a sky area which is covered by clouds. The Bayesian posterior estimation method is used in this research to blend cloud fraction data from multiple sources. The prior distribution may be constructed from one source of observation, while the another observation helps build the likelihood function via a linear regression procedure. The posterior estimate is a probability density function which combines the two observations. We have applied this approach to blend the cloud fraction data from the Atmospheric Radiation Measurement (ARM) Program. The site under investigation is Lamont, Oklahoma. Our first data source of cloud fraction is obtained from a Total-Sky-Image (TSI) camera (Figure 1). Our second data source is from Millimeter-wave Cloud Radars (Figure 2), Micropulse Lidars, and laser Ceilometers.

Figure 1: TSI

Figure 2: MMCR

DATA

Our radar and camera data in 2000-2009 are considered here. Both datasets consist of hourly observations. Figures 3 and 4 show scatterplots of radar vs. camera for hourly and daily averages. Figure 5 show the scatterplot for 5-day averages with an estimated linear regression line, used later for modeling the likelihood function. See Figure 6 for a histogram of hourly camera observations.

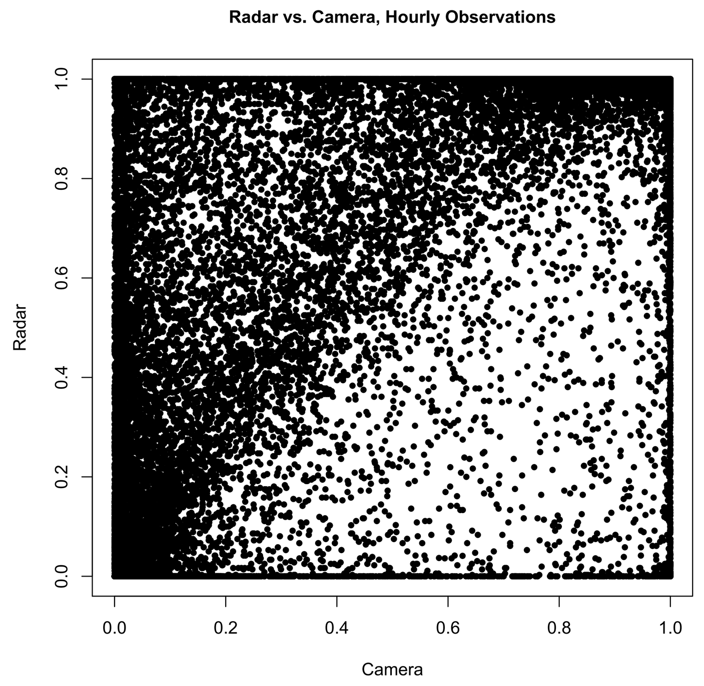

Figure 3: Radar vs. Camera, Hourly Observations

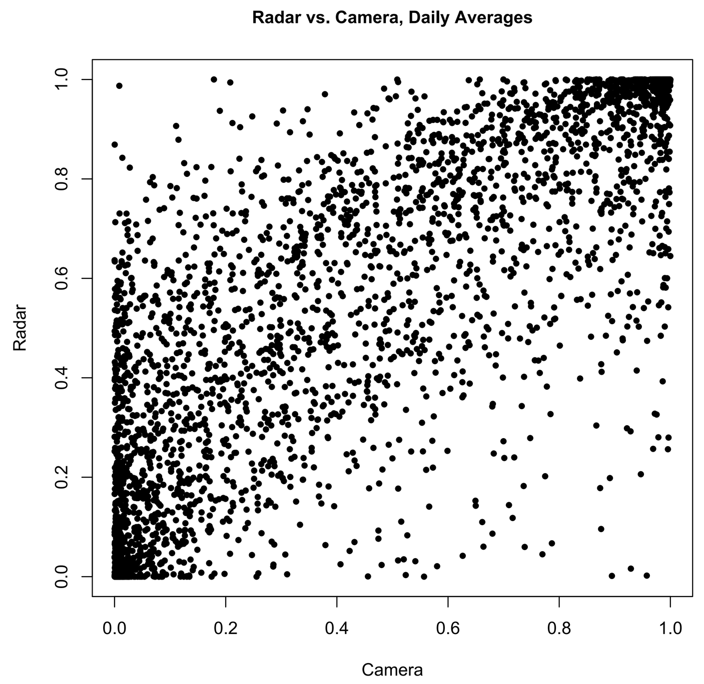

Figure 4: Radar vs. Camera, Daily Averages

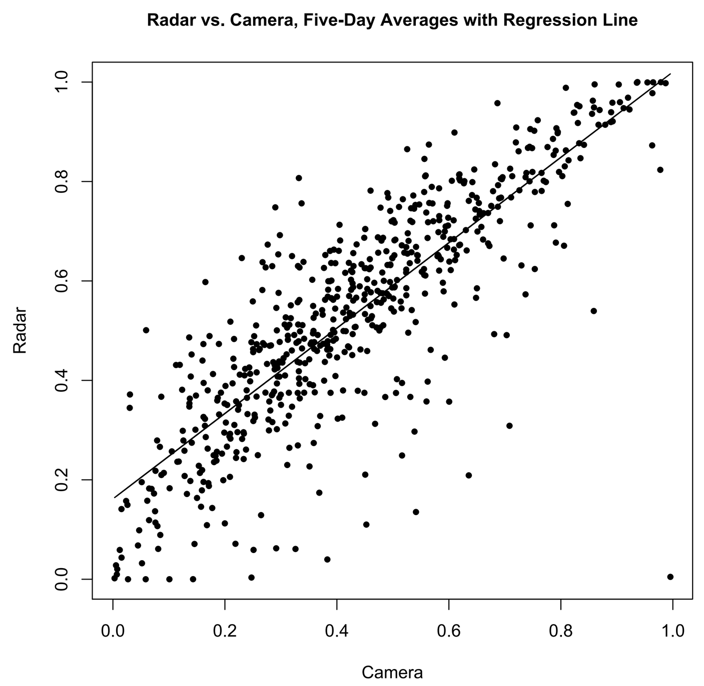

Figure 5: Radar vs. Camera, 5-day Averages

DATA

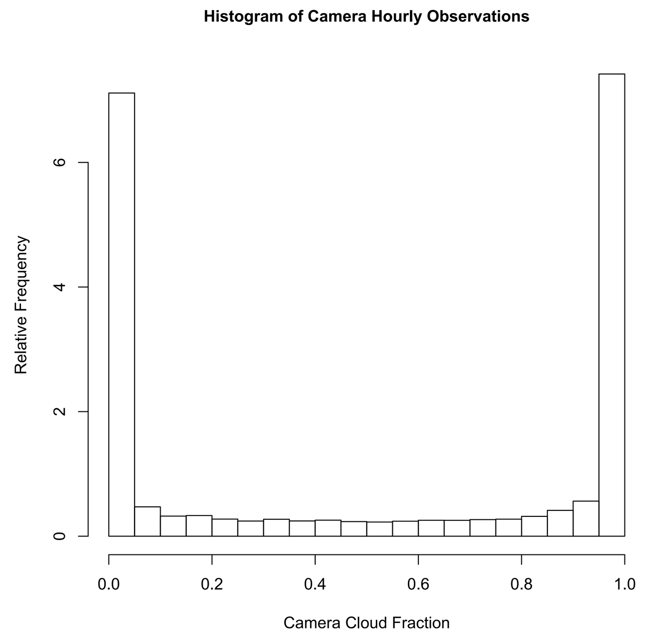

Figure 6: Histogram of Hourly Camera Observations

METHOD

Consider cloud fraction for 5-day averages as an example. Let θ be a random variable, representing cloud fraction from TSI. We estimate a Beta probability density function for θ using the method of moments. Figure 6 shows a histogram of TSI 5-day averages with our fitted Beta density function, which is the prior distribution, $\pi(\theta)$. Let X be the radar measurements. Regression is used to model X for a given θ:

$X \mid \theta = \beta_0 + \beta_1\theta + \epsilon,\ \epsilon \sim N(0, \sigma)$

Hence, the probability density function is

$f(X \mid \theta) \sim N(\beta_0 + \beta_1\theta, \sigma)$

Bayes theorem implies

$\pi(\theta \mid X) = \pi(\theta)\pi(X \mid \theta)/m(X)$

where m(X) is a normalization factor

Our final estimate of cloud fraction is obtained through the posterior mean $\int_0^1 \theta\, \pi(\theta \mid X)$ which is estimated through numerical integration

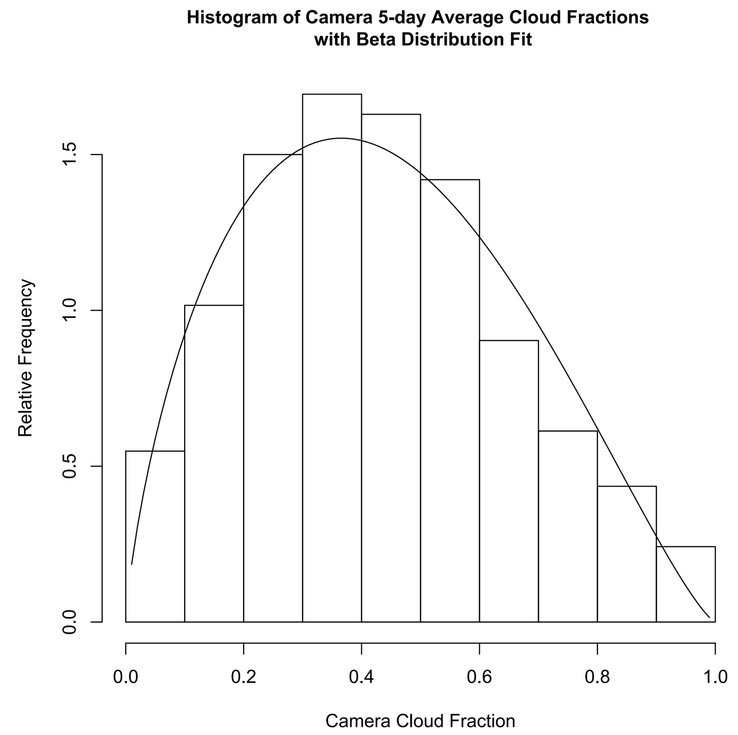

Figure 7: Camera 5-day Averages

RESULTS

Using this method, for any observed radar X we create a probability distribution, $\pi(\theta \mid X)$. Figure 8 shows the $\pi(\theta \mid X)$ for an observed X of 0.5

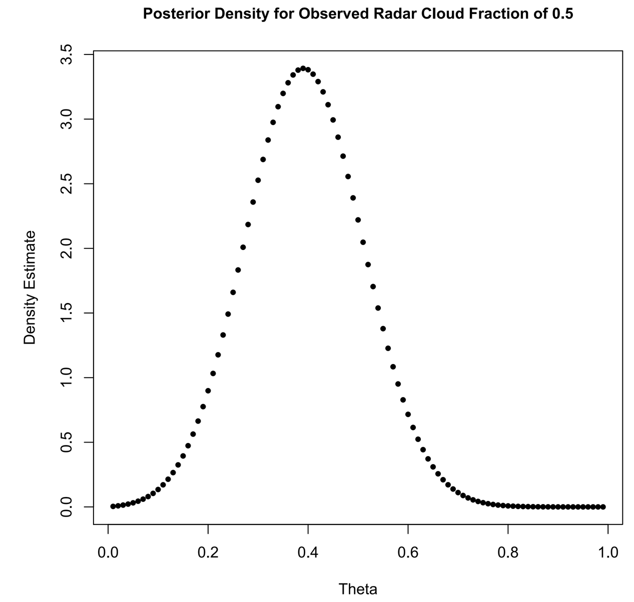

Figure 8: Posterior Distribution for X = 0.5

Our estimates of the posterior means were then used to create annual estimates of cloud fractions (Figure 9).

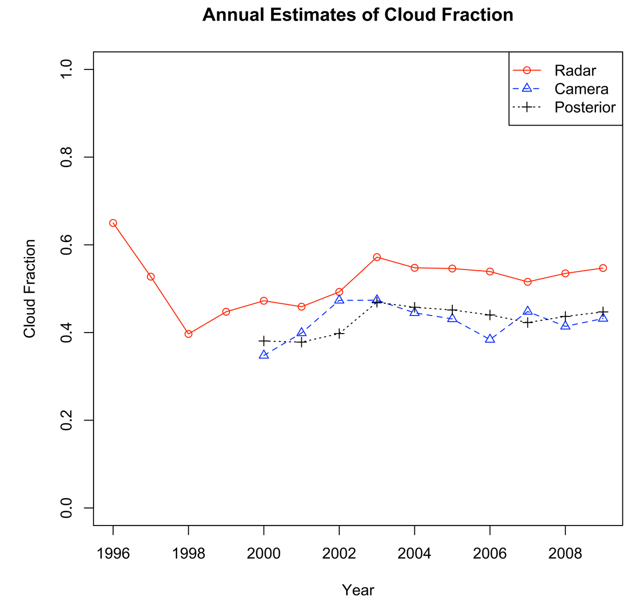

Figure 9: Annual Cloud Fraction Estimates

CONCLUSIONS

It was found that two observed datasets can be combined through the Bayesian posterior estimation, created with an estimated prior distribution and a likelihood function modeled by a linear regression procedure. Further work will focus on examining the variance properties of our estimated cloud fraction, and using the posterior density to create credible sets.

REFERENCES

- Coelho, C.A.S. and Co-authors, 2004: Forescast calibration and combination: A simple Bayesian approach for ENSO. J. Climate, 17, 1504-1516.
- Xie, S and Co-authors, 2010: ARM Climate Modeling Best Estimate Data. Bull. Amer. Meteor. Soc., 91, 13-20.

rsomerville@ucsd.edu
shen@math.sdsu.edu