Probabilistic assessment of cloud fraction using Bayesian blending of two data sets: a feasibility study

Richard Somerville[1], Samuel Shen[2] , Max Velado[2], and Gabriel Kooperman[1]
[1]Scripps Institution of Oceanography, University of California, San Diego [rsomerville@ucsd.edu]
[2]Deparment of Mathematics and Statistics, San Diego State University [sam.shen@sdsu.edu]

1. Introduction

We describe a probabilistic approach to assess cloud fraction using the Bayesian posterior estimate. The research reported here is a feasibility study designed to explore the method. In this proof-of-concept study, we illustrate the approach using specific observational datasets from the U. S. Department of Energy Atmospheric Radiation Measurement Program's Southern Great Plains (SGP) site in the central United States, but the method is quite general and is readily applicable to other datasets. The total sky imager (TSI) camera observations are used to determine the prior distribution. A regression model and the active remote sensing of clouds (ARSCL) radar/lidar observations are used to determine the likelihood function. The posterior estimate is a probability density function (pdf) of the cloud fraction (CF) whose mean is taken to be the optimal blend of the two observations. The data at hourly, daily, 5-day, monthly, and annual time scales are considered. Some physical and probabilistic properties of the cloud fractions are explored from radar/lidar, camera, and satellite observations and from simulations using the Community Atmosphere Model (CAM5). Our results imply that (a) the Beta distribution is a reasonable model for the cloud fraction for both short- and long-time means, the 5-day data are skewed right, and the annual data are almost normally distributed, and (b) the Bayesian method developed successfully yields a pdf of CF, rather than a deterministic CF value, and it is feasible to blend the TSI and ARSCL data with a capability for bias correction.

Most materials of this poster are contained in a paper recently submitted to *Journal of Geophysical Research* [Shen et al., 2012].

2. Instruments and data

Figure 1. TSI (Total Sky Imager) camera.

Figure 2. ARSCL MMCR (Millimeter-Wave Cloud Radar).

- TSI daytime CF data at the SGP site: TSI has a 352 x 288 pixel resolution and can measure the cloud fraction during daylight. It retrieves CF as the ratio of the number of cloud cover pixels to the 101,376 total field-of-view (FOV) pixels, i.e., 101,376 =352 x 288, when the local solar elevation angle (i.e., the angle between the sun direction and the horizon) is greater than or equal to 10 degrees. Thus, the TSI measures the daytime CF. The daytime length varies according to seasons. The camera sampling rate is one image per 30 seconds. Here we use the 2000-2009 hourly CMBE (climate model best estimate) dataset [Xie et al., 2010].
- ARSCL daytime CF data at the SGP site: ARSCL makes pencil observations along a very narrow FOV of well less than 1º around zenith, compared to the TSI hemispheric dome FOV extending 160º also around zenith. Thus, unlike TSI, the ARSCL CF has to be approximated by temporal cloud averages. The original temporal grid is at 10-second resolution (Table 1 of Xie et al. (2010)). The hourly data is an aggregation of these 10-second data. The ARSCL cloud fraction is defined as the ratio of the number of cloud covering temporal intervals to the total temporal intervals in a given time period [Xi et al., 2010], which is one hour for the data we use here. Thus, the ARSCL cloud fraction definition using the temporal ratio represents the frequency of cloud occurrence [Qian et al., 2012] and is hence different from the CF defined by the TSI instrument. ARSCL CF data for SGP are a function of both height and time. To compare with TSI CF data, here we use the daytime and vertically integrated CF.

3. The Bayesian blending methodology

Step 1: Estimate the prior distribution

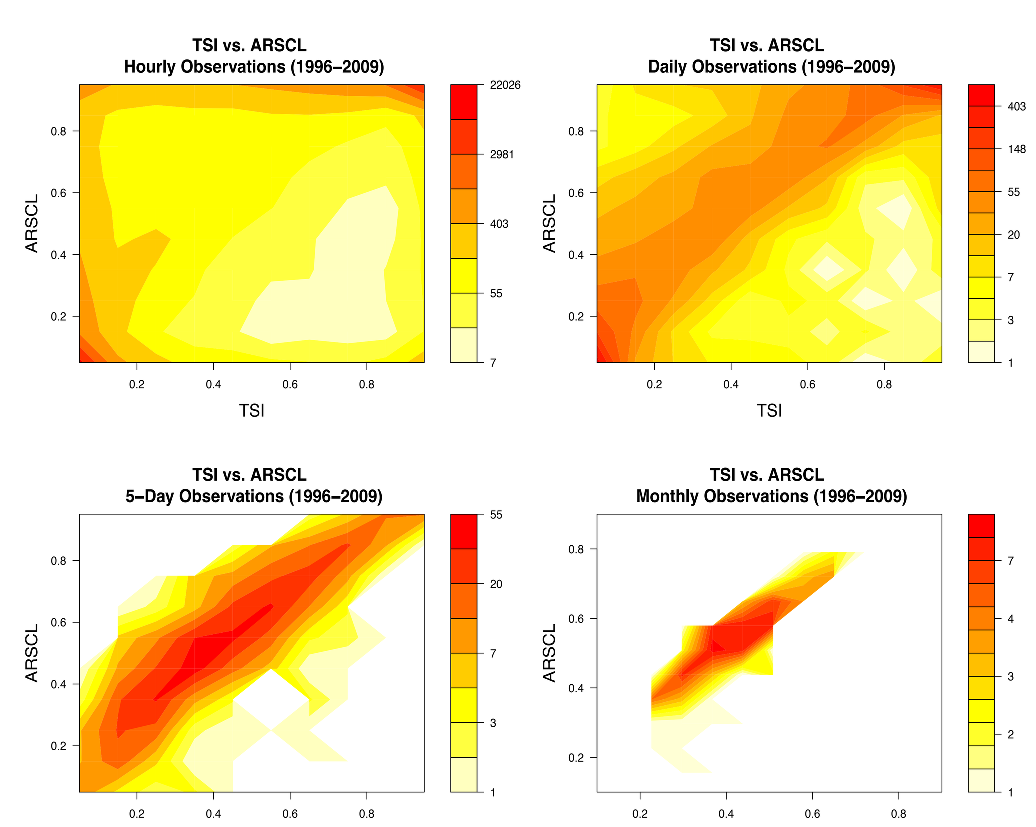

Figure 3. 2D frequency plot between TSI and ARSCL daytime CF data in different time scales

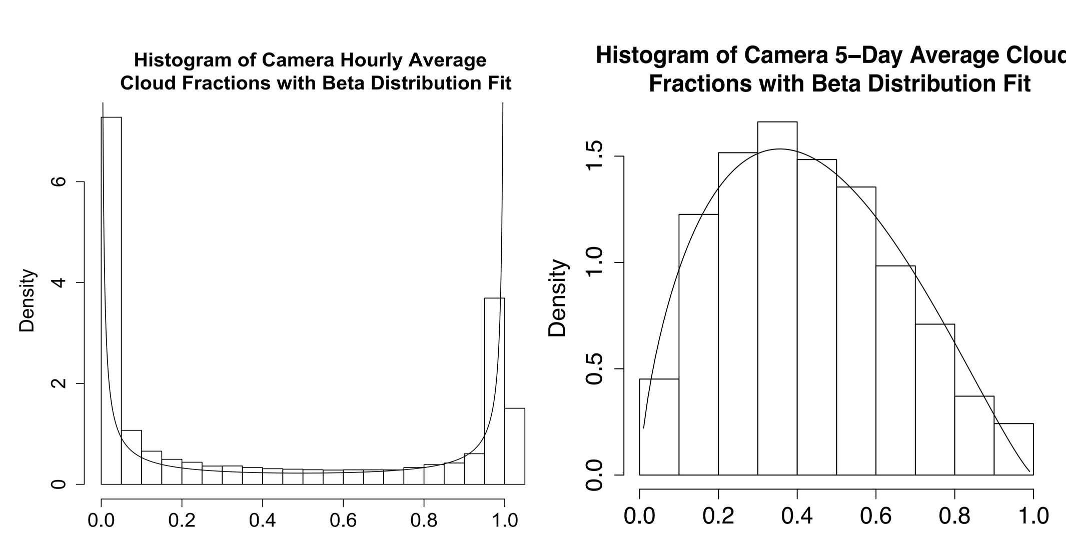

Figure 4. Motivated by the frequent occurrence of zeros and ones in the 2D frequency plots in Fig. 3, we fit the TSI CF data to a Beta distribution as the prior estimate $Beta(\alpha,\beta)(x)=\Gamma(\alpha+\beta)/[\Gamma(\alpha)\Gamma(\beta)]\, x^{\alpha-1}\,(1-x)^{\beta-1}$.

Step 2: Construct a likelihood function using a model.

Here we use a linear regression model between TSI and ARSCL. Physically we may interpret this regression as a mutual correction between the two instruments. Let X represent the ARSCL observation given the prior TSI CF distribution. Regression is used to model X for a given CF θ:

$$x \mid \theta = \beta_0 + \beta_1\theta + \varepsilon, \qquad \varepsilon \sim N(0,\sigma^2)$$

where ε is the regression model error with a standard deviation equal to σ. The explicit expression of a likelihood function $f(x|\theta) \sim N(\beta_0+\beta_1\theta,\, \sigma^2)(x)$ is

$$f(x \mid \theta) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\beta_0-\beta_1\theta)^2}{2\sigma^2}\right]$$

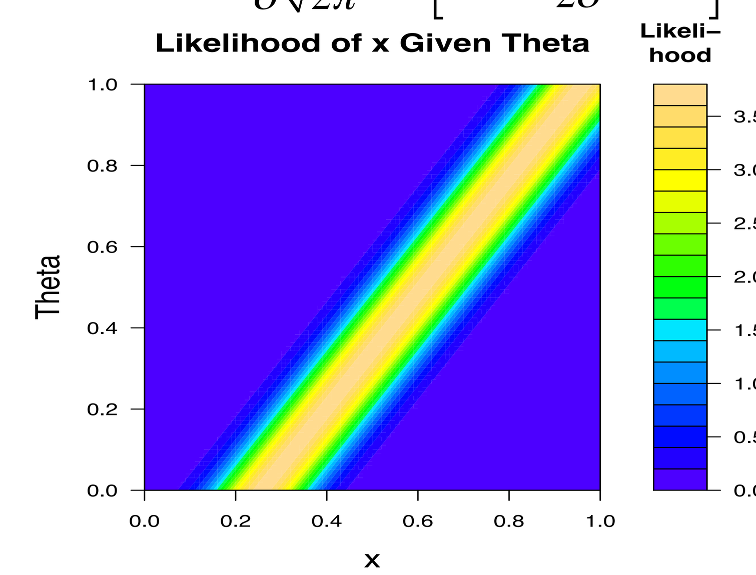

Figure 5. The likelihood function for 5-day data in 2003.

Step 3: Posterior estimate following Bayes' formula

The Bayes' formula yields the Bayesian posterior distribution

$$\pi(\theta \mid x) = \frac{\pi(\theta) f(x \mid \theta)}{m(x)}$$

where $m(x)$ is a normalization factor equal to

$$m(x) = \int_0^1 \pi(\theta) f(x \mid \theta)\, d\theta$$

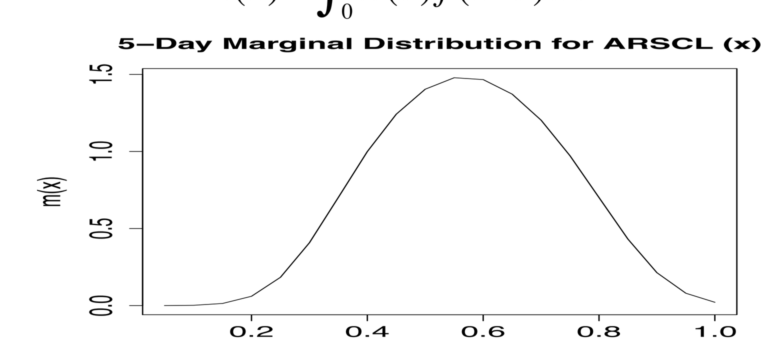

Figure 6. The normalization factor m(x).

This posterior distribution is the pdf of the CF, from which various statistical properties of the CF can be derived, such as mean, variance, skewness, and kurtosis.

4. Results

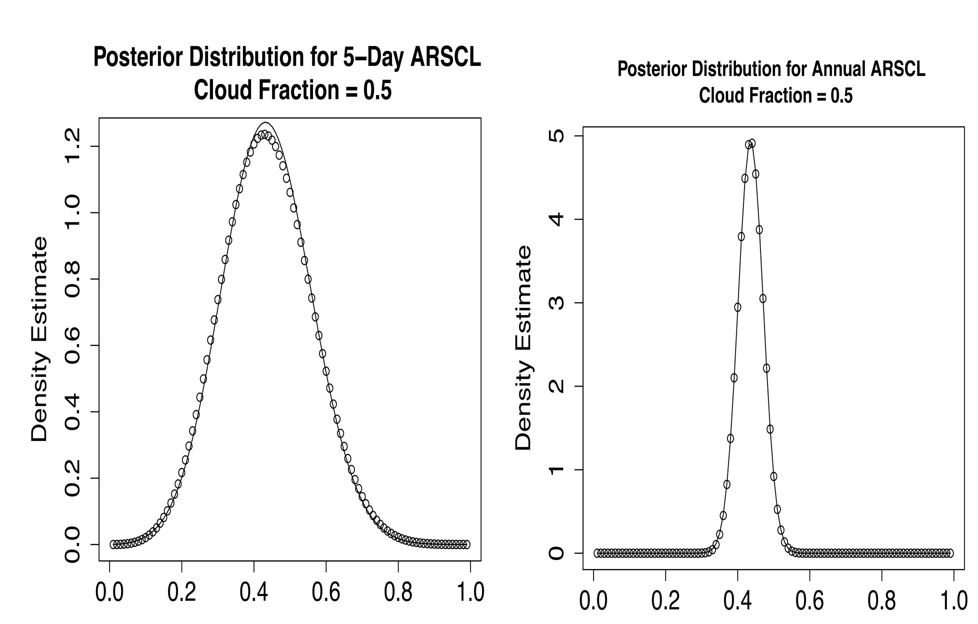

Figure 7. Posterior distribution of CF (left panel) when the ARSCL datum is 0.5 for 5-day data, and (right panel) when the ARSCL datum is 0.5 for annual data.

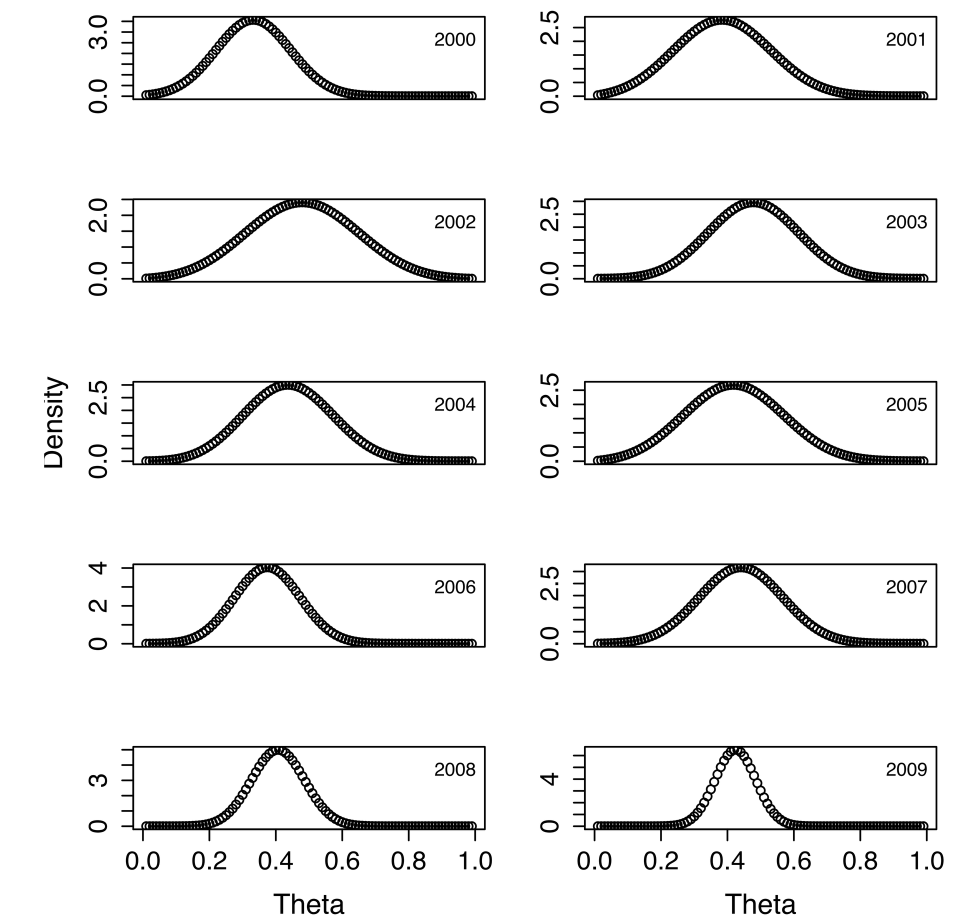

Figure 8. The posterior distribution of each year from 2000 to 2009 at the SGP site for 5-day CF as the mean of the 73 five-day CF values.

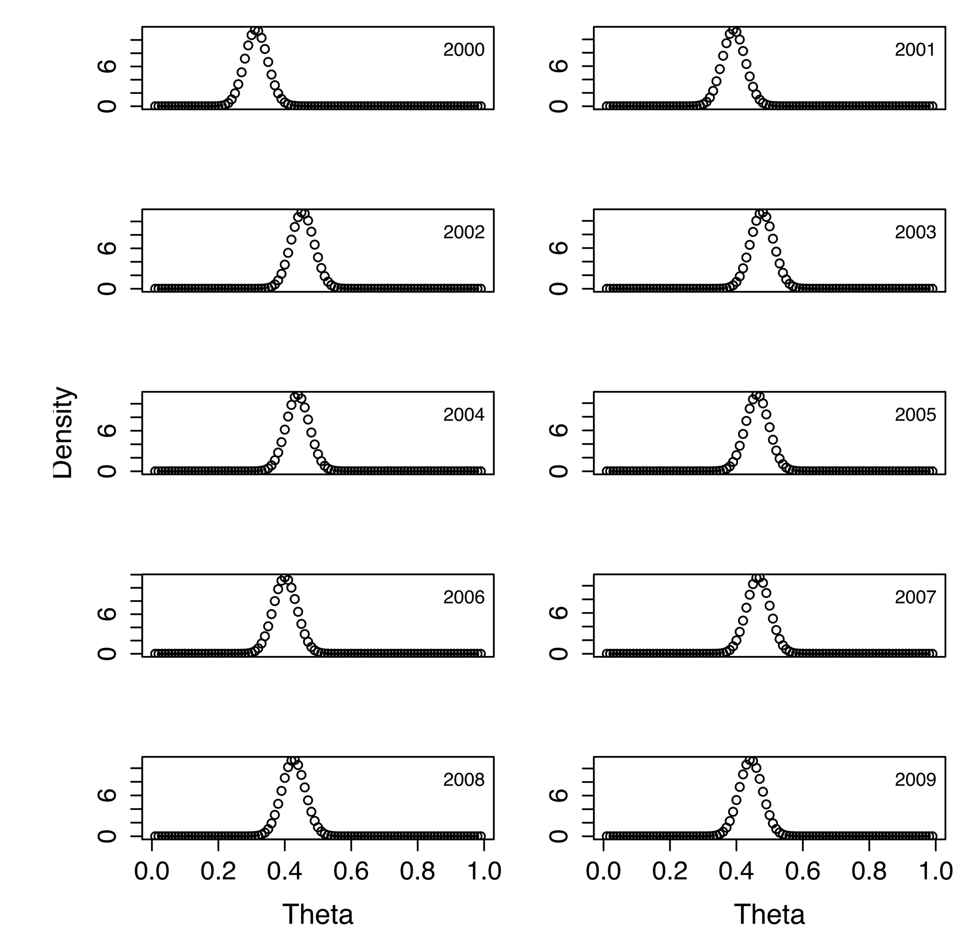

Figure 9. The posterior distribution of the cloud fraction from 2000-2009 at the SGP site for annual CF.

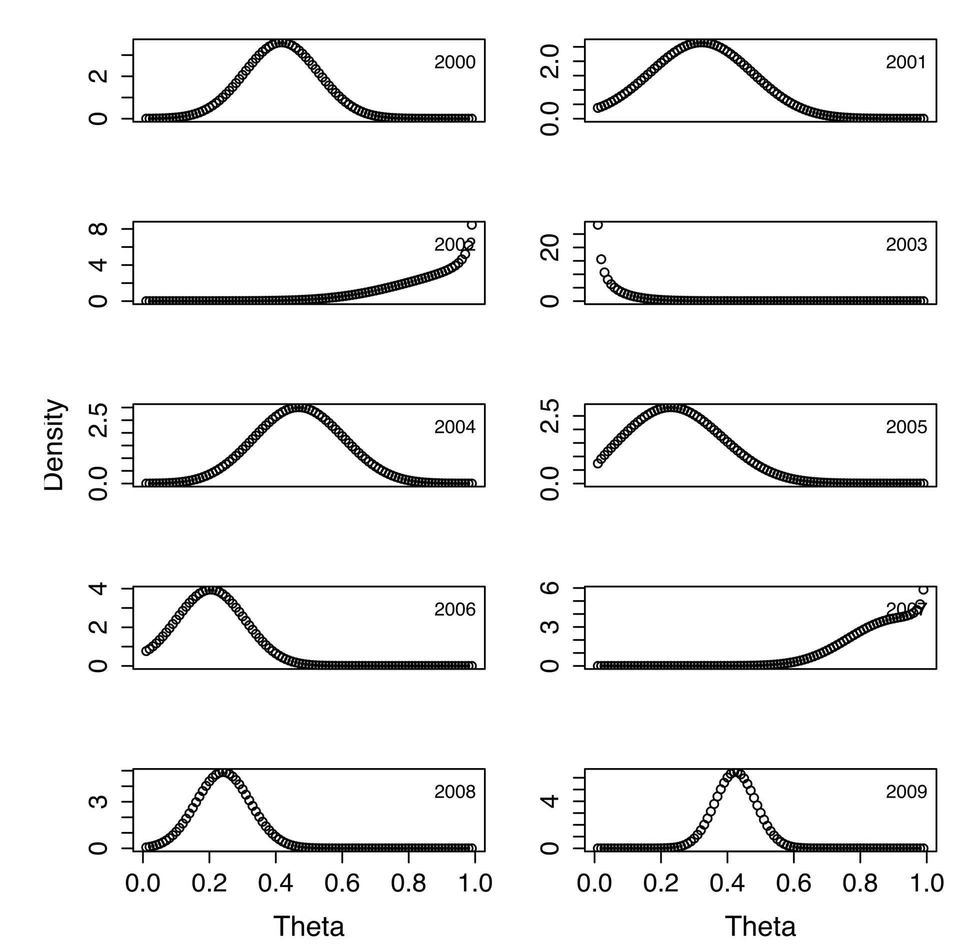

Figure 10. The posterior distribution of the cloud fraction of each year from 2000 to 2009 at the SGP site for the 181st – 185th days mean CF.

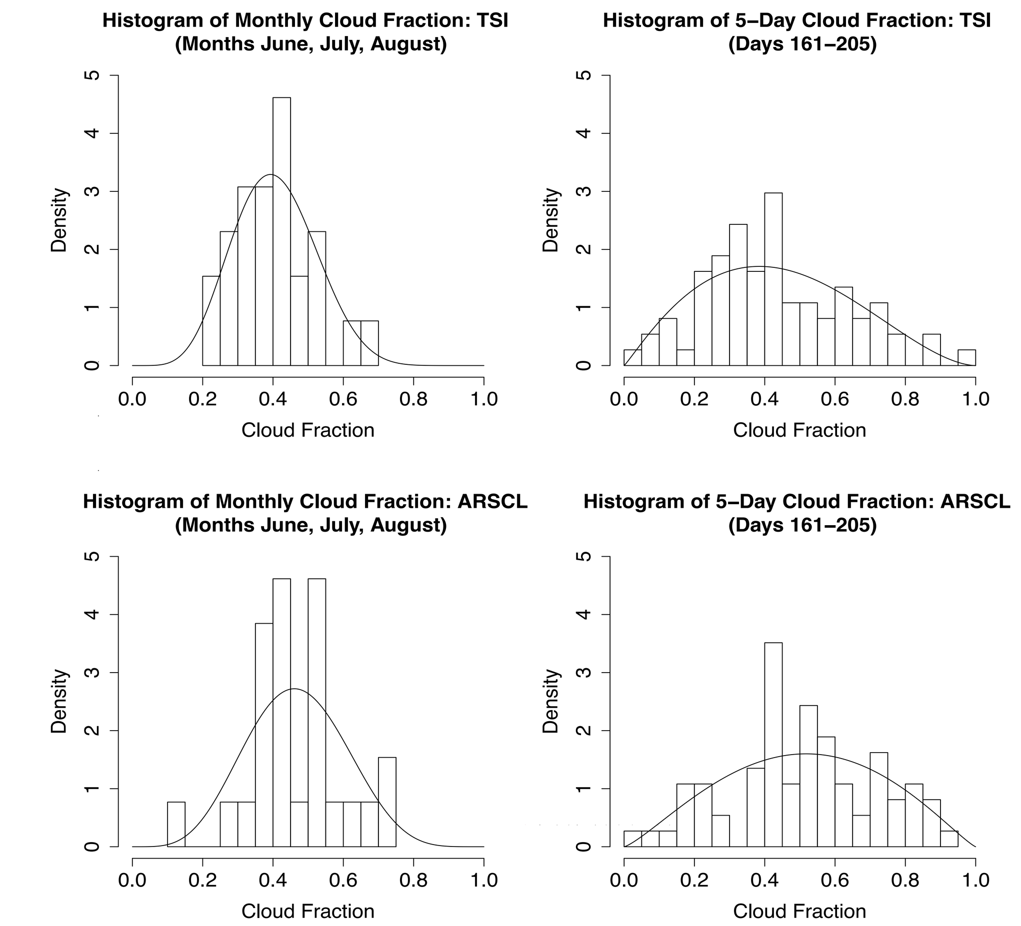

Figure 11. Histograms and their Beta distribution fitting of cloud fractions based on the monthly and 5-day time scales and TSI and ARSCL data.

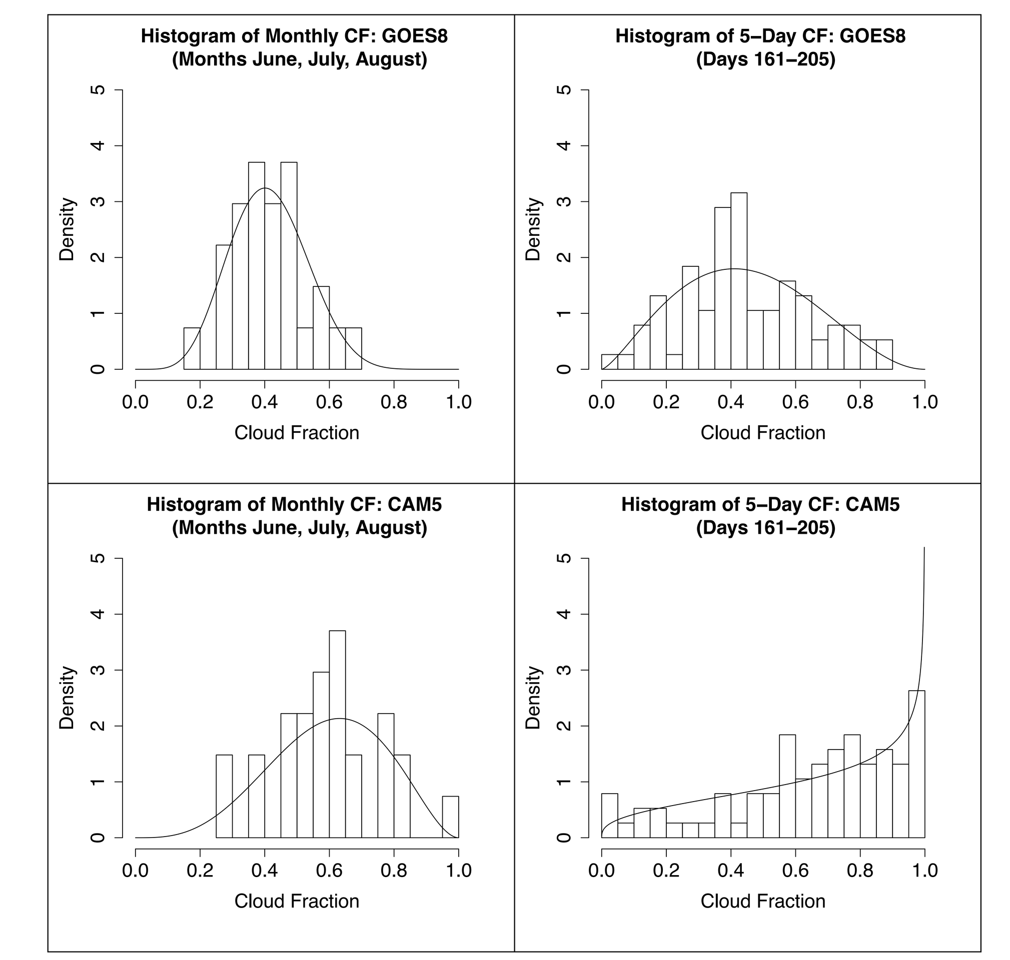

Figure 12. Histograms and their Beta distribution fitting of cloud fractions based on the monthly and 5-day time scales and GOES8 and CAM5 data.

5. Conclusions

- We have introduced a 3-step Bayesian posterior estimate (BPE) approach to optimally blend different cloud fraction datasets.
- The CF is considered intrinsically random and is therefore represented by a pdf rather than a fixed value.
- The summer probabilistic distribution of the SGP CF demonstrates the consistency between the CAM5 model CF and the observed CF from ARSCL, TSI, and GOES8 in the monthly scale and also the inconsistency in the 5-day scale.

6. References

- Qian, Y., C.N, Long, H. Wang, J.M. Comstock, S.A. McFarlane, and S. Xie (2012). *Atmos. Chem. Phys.*, **12**, 1785-1810.
- Shen, S.S.P., M. Velado, R.C. J. Somerville and G. J. Kooperman (2012). *J. Geophysical Res.*, submitted.
- Xi, B., X. Dong, P. Minnis, M. M. Khaiyer (2010). *J. Geophys. Res.*, **115**, D12124, doi:10.1029/2009JD012800, 2010.
- Xie, S. and Co-authors (2010). *Bull. Amer. Meteor. Soc.*, **91**, 13-20.