Randomization method for comparing station and satellite observations of precipitation, temperature and cloud

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

1. Introduction

This poster will describe several randomization approaches to quantify uncertainties in climate data from both observation and model. Five basic error and uncertainty problems are listed below.

Problem 1: Compare in situ pencil observation with satellite borne remote sensing: How to define the ground truth and quantify the spatial or temporal sampling error?

Problem 2: Climate models are defined from continuum mechanics and processes, while the equations are solved for grid points: How to define the modeling data error relative to in situ observations?

Problem 3: Errors of the CMBE products need to be quantified: How to define and assess the CMBE product errors using advanced statistical methods?

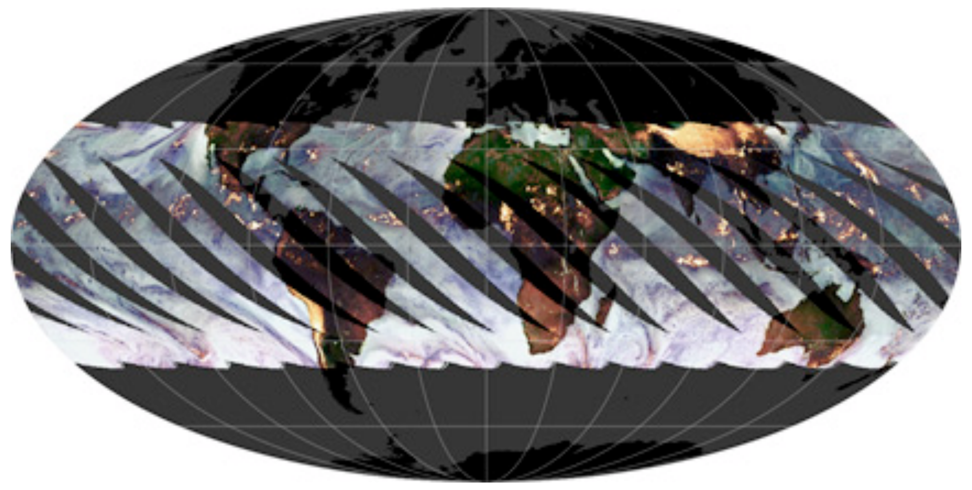
Problem 4: The mean square error (MSE) of the first moment, i.e., mean, is estimated from the properties of the second moment, i.e., covariance functions. How can climate models be used to calculate the covariance function? What are the correct formulations of the errors?

Problem 5: Due to observational and modeling errors, a climate parameter has intrinsic uncertainty and hence are described by a probability density function (pdf) rather than a single deterministic value. How to estimate the pdf?

This poster will outline mathematical formulations to address the above problems with examples for precipitation, surface air temperature, and cloud fraction.

2. Data

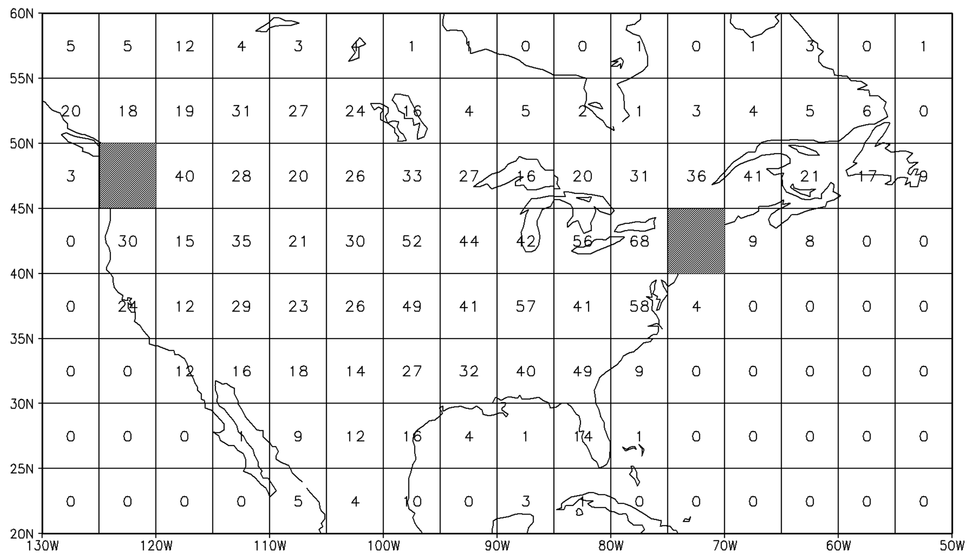
- TRMM precipitation data: TRMM satellite orbits with an inclination of 35 degrees to the Equator. The orbit return period is 46 days. This sampling time is randomized to address the issue of diurnal cycle.



TRMM footprints.

- Global Historical Climatology Network (GHCN) station data since 1861. A resampling randomization is used to assess the grid box data error.

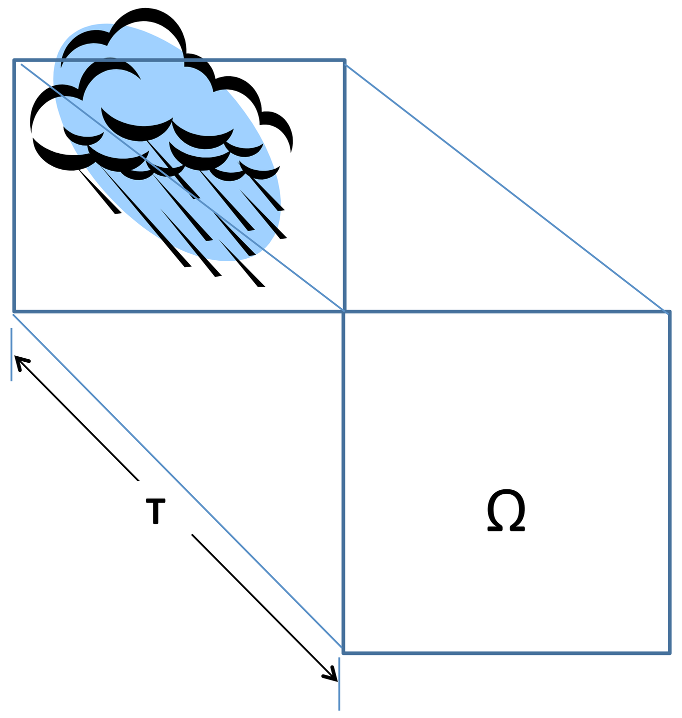
US grid boxes and number of GHCN stations.



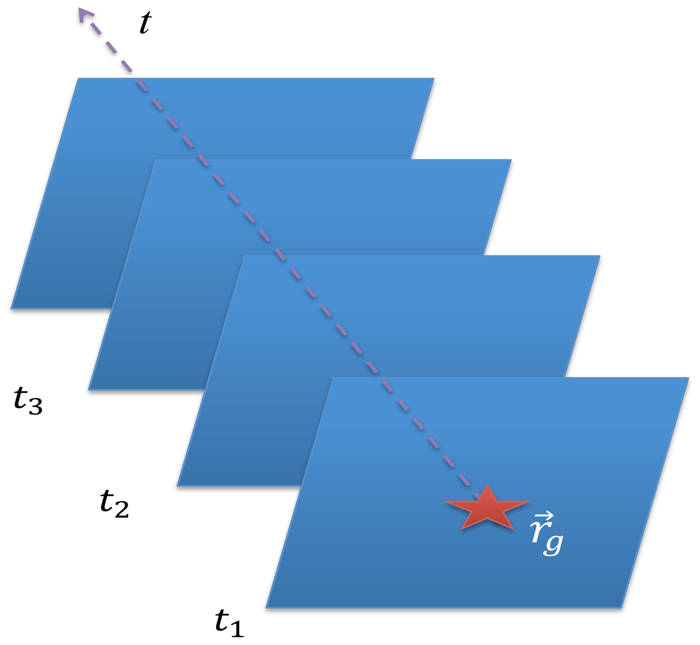
- ARM SGP site ARSCL radar and TSI camera data. We use Bayesian randomization to evaluate the cloud fraction pdf.

3. Methodology

◆ A atmospheric process in a space-time box *B=Ω×[0,T]* .



◆ A single station over a grid box, MSE formulation, and randomization of the station location (North et al., 1994).



Mean Square Error for station over region Ω :

$$\varepsilon^2(\vec{r}_g) = E_R\left[\left(\int_\Omega R(\vec{r})(1-\delta(\vec{r}-\vec{r}_g))d\Omega\right)^2\right]$$

Randomize the station location $\vec{r}_g$:

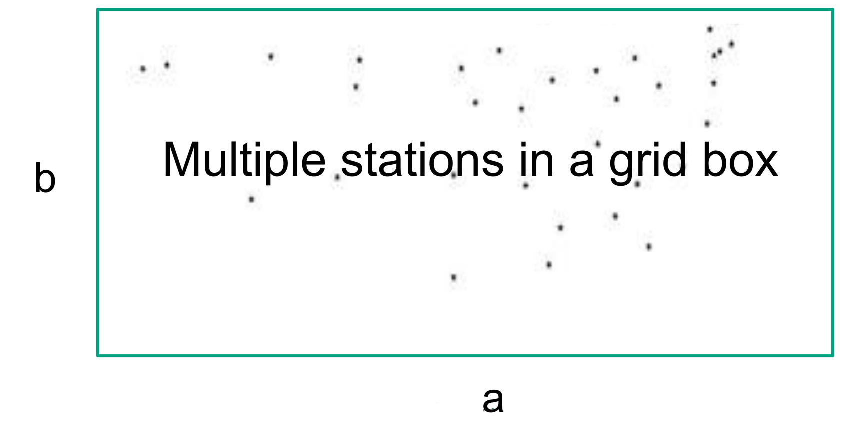
$$E_g^2 = E_g[\varepsilon^2(\vec{r}_g)]$$

◆ Randomize satellite visiting time (t_1, t_2, t_3, …, t_N) in a space-time box *B=Ω×[0,T]* (North et al., 1993).

$$\varepsilon_s^2(t_1,t_2,\ldots,t_N) = E_s\left[\left(\frac{1}{AT}\int_B R(\vec{r},t)\left(1-\sum_{n=1}^{N}\delta(t-t_n)\Delta t\right)d\Omega\right)^2\right]$$

$$E_s^2 = E_t[\varepsilon_s^2(t_1,t_2,\ldots,t_N)]$$

◆ Multiple stations in a grid box (Shen et al., 2007; Shen et al., 2011). We randomize sub-samplings and use Monte Carlo method.



$$\varepsilon_g^2 = \left\langle\left(\bar{T}_g - \hat{\bar{T}}_g\right)^2\right\rangle = \alpha_s \times \frac{\sigma_s^2}{N_g}$$

Correlation factor and spatial variances are estimated by randomizing sub-sampling and using a regression.

$$\alpha_s = 1 + \frac{1}{N}\sum_{\substack{i,j=1\\ i\neq j}}^{N}\left\langle\frac{(T_i-\bar{T})}{\sigma_s}\frac{(T_j-\bar{T})}{\sigma_s}\right\rangle$$

$$\sigma_s^2 = \left\langle\frac{1}{N}\sum_{j=1}^{N}\left(T_j(t)-\bar{T}(t)\right)^2\right\rangle$$

4. Results

4.1: TRMM ground truth results with a single station in a field of view (FOV) and with the spectra and covariance calculated from a simple model (North et al., 1994). We consider a rectangular FOV with length a and width b. The rainrate is modeled by a simple convection-diffusion process:

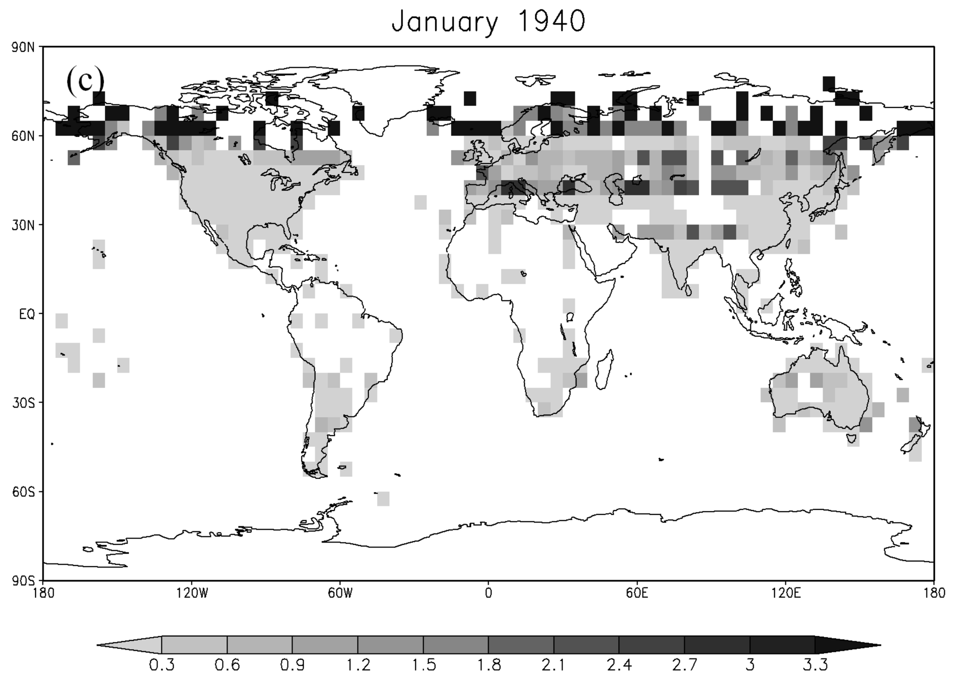
$$\tau_0\frac{\partial\psi}{\partial t} - \lambda_0^2\nabla^2\psi + \psi = F(\vec{r},t)$$

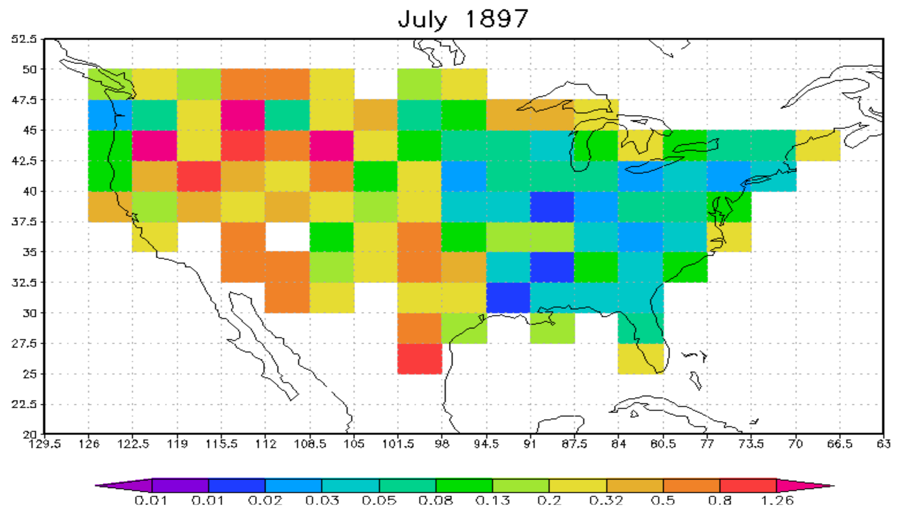
Here the temporal and spatial scales are 12 hours and 40 km. Table 1 shows the dimensionless root mean square errors (DRMSE) for half day and 60 days samplings.

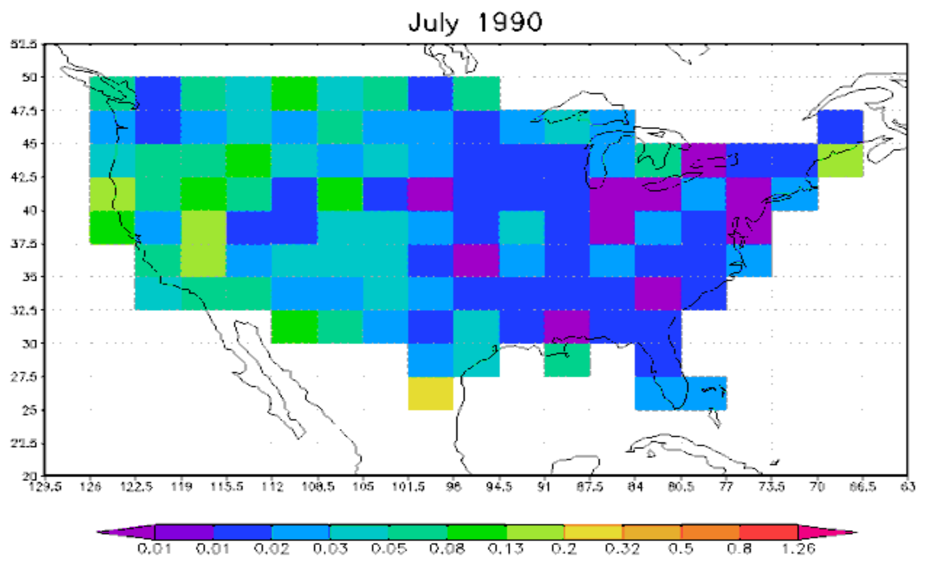
Table 1. DRMSE for a ground truth station over a rectangular FOV.

Length a [km]	Width b [km]	DRMSE with 12 hours [%]	DRMSE with 2 mons [%]
10	10	46.0	5.9
20	20	63.0	8.1
30	10	63.3	8.2
30	20	68.1	8.8
30	30	72.1	9.3

4.2: Sampling error variances of the surface air temperature station data for the GHCN over the globe and the USHCN over the United States (Shen et al., 2007, and 2011) [units: Celsius deg sq.].

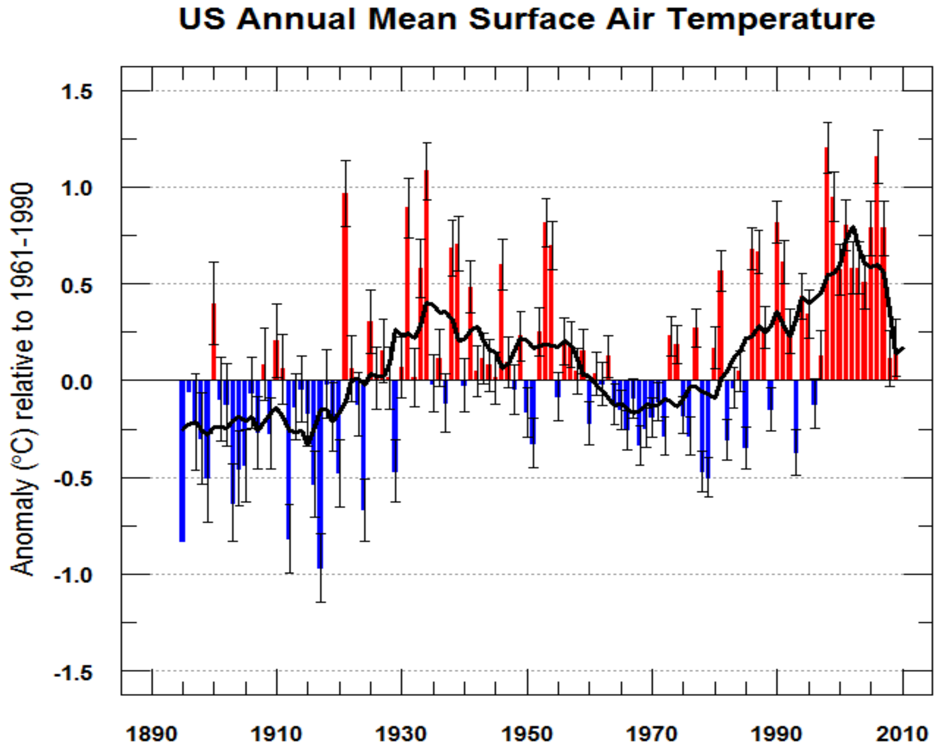




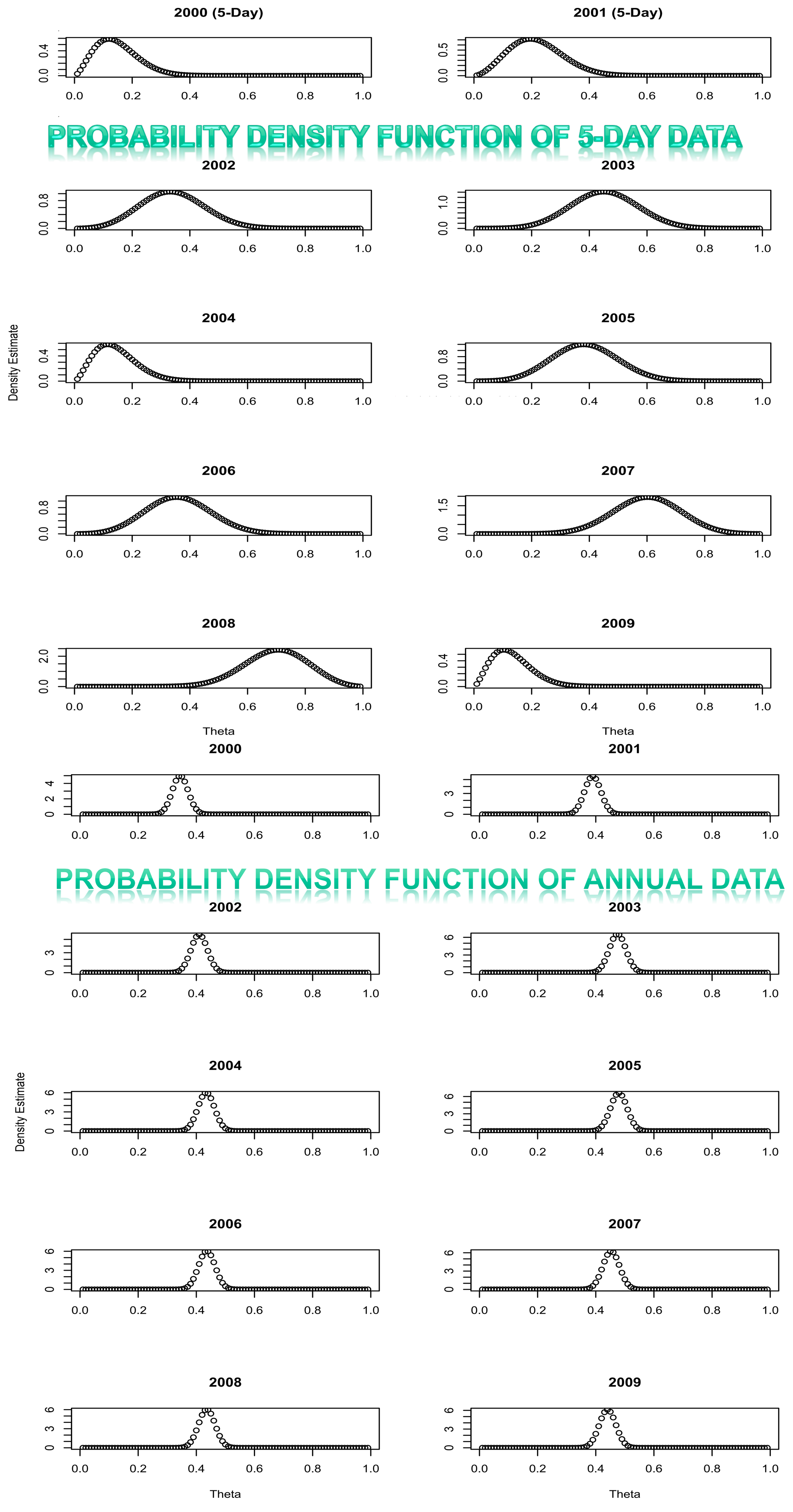


From the sampling error variance on each grid box, we can systematically estimate the error bars for the US average annual mean surface air temperature when the degrees of freedom are available. These error bars will help with statistical inference of hottest years and coldest years from the historical data. The figure below shows the results for the US average annual mean SAT (Tmean) and its uncertainty.

US Annual Mean Surface Air Temperature



4.3: Bayesian blending and randomization by a probabilistic model. The pdfs calculated from the TSI and ARSCL data via the Bayesian posterior estimate (BPE) are shown below for the 5-day data (181st-185th day of a year excluding Feb 29) and the annual data of cloud fractions at the SGP site from 2000-2009. The 5-day pdfs are skewed, while the annual data are almost normally distributed..



5. Conclusions

- Randomization approach provides a systematic way to compare point measurement with areal measurement in both observations and models.
- Results can be expressed in terms of MSE or pdf and can be evaluated through randomization of station locations, satellite visiting time, resampling, and Bayesian estimate.

6. References

- North, G.R., S.S.P. Shen, and R.B. Upson, 1993: Sampling errors in rainfall measurements by multiple satellites. *J. Appl. Meteor.* **32**, 399-410.
- North, G.R., J.B. Valdes, E. Ha, and S.S.P. Shen, 1994: The ground truth problem for satellite estimates of rain rate. *J. Atmos. Ocean. Tech* **11**, 1035-1041.
- Shen, S.S.P., H. Yin, and T.M. Smith, 2007: An estimate of the error variance of the gridded GHCN monthly surface air temperature data. *J. Climate* **20**, 2321-2331.
- Shen, S.S.P., C.K. Lee, and J. Lawrimore, 2011: Uncertainties, trends, and hottest and coldest years of US surface air temperature since 1895: an update based on the USHCN V2 data. *J. Climate*, submitted.