

# Natural variability and sampling errors in solar radiation measurements for model validation over the Atmospheric Radiation Measurement Southern Great Plains region

Zhanqing Li,<sup>1</sup> M. C. Cribb, and Fu-Lung Chang

Department of Meteorology, Earth System Science Interdisciplinary Center, University of Maryland at College Park, College Park, Maryland, USA

Alexander Trishchenko and Yi Luo

Applications Division, Canada Centre for Remote Sensing, Ottawa, Ontario, Canada

Received 17 May 2004; revised 29 November 2004; accepted 21 December 2004; published 26 April 2005.

[1] Ground-based radiation measurements are frequently used for validating the performance of a model in simulating clouds. Such important questions are often raised as: (1) How well do the measurements represent model grid mean values?; (2) How much of model-observation differences can be attributed to inherent sampling errors?; and (3) What scale does modeling need to be performed in order to capture the cloud variation? We attempt to address these questions using surface solar irradiance data retrieved from the Geostationary Operational Environmental Satellite (GOES) and measured at the Atmospheric Radiation Measurement (ARM) Southern Great Plains (SGP) site. The satellite retrievals are used to mimic ground measurements with various spatial densities and temporal frequencies, from which the sampling errors of the ground observations are quantified and characterized. Most of the differences between point-specific measurements and area-mean satellite retrievals originate from ground sampling errors. We quantify these errors for different months, model grid sizes, and integration intervals. In March 2000, for example, the sampling error is  $16 \text{ W m}^{-2}$  for instantaneous irradiances averaged over an area of  $10 \times 10 \text{ km}^2$ . It increases to 46 and  $64 \text{ W m}^{-2}$  if the model grid size is enlarged to  $200 \times 200 \text{ km}^2$  and  $400 \times 400 \text{ km}^2$ , respectively. The sampling uncertainties decrease rapidly as the time-averaging interval increases up to 24 hours and then level off to a relatively small and stable value. Averaging over periods greater than 5 days reduces the error to a magnitude of less than  $15 \text{ W m}^{-2}$  over all grid sizes. The sampling error also decreases as the number of ground stations increases inside a grid, but the most substantial reduction occurs as the number of ground sites increases from 1 to 2 or 3 for a grid size of  $200 \times 200 \text{ km}^2$ . This means that for computing grid-mean surface solar irradiance, there is no need for an overly dense network of observation stations.

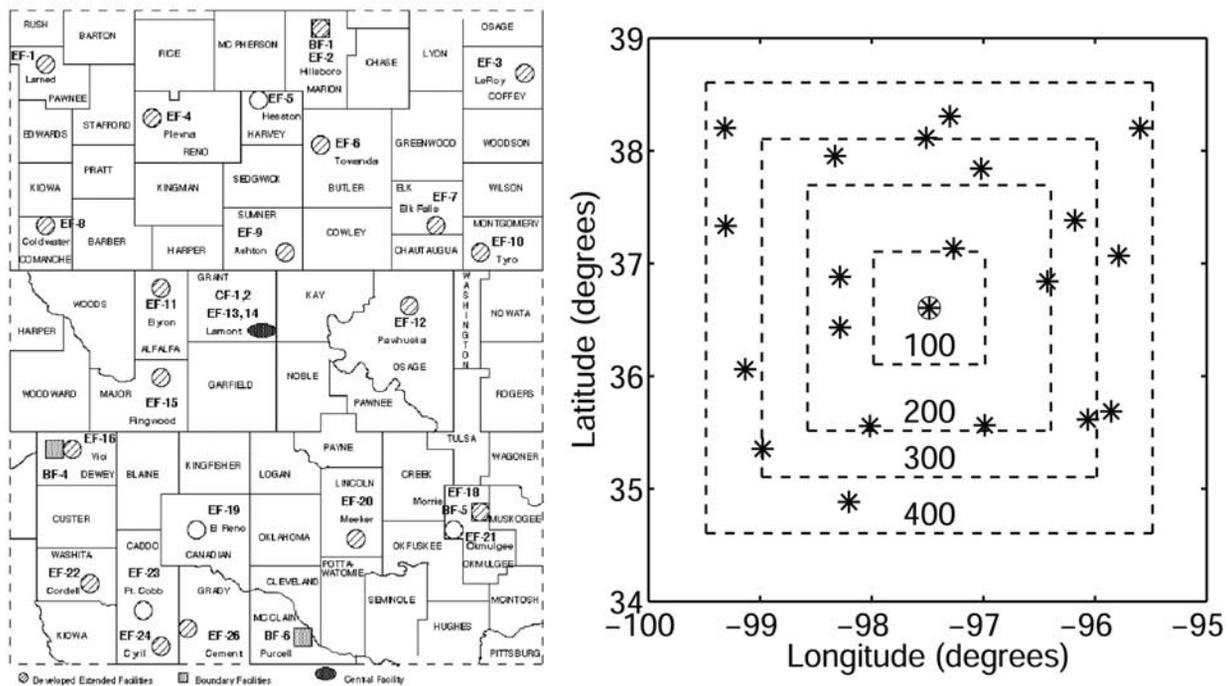
**Citation:** Li, Z., M. C. Cribb, F.-L. Chang, A. Trishchenko, and Y. Luo (2005), Natural variability and sampling errors in solar radiation measurements for model validation over the Atmospheric Radiation Measurement Southern Great Plains region, *J. Geophys. Res.*, *110*, D15S19, doi:10.1029/2004JD005028.

## 1. Introduction

[2] Radiation measurements have been widely used for evaluating cloud parameterization schemes and model simulation results [Morcrette, 2002]. Radiative quantities are affected by numerous atmospheric (e.g., cloud) and surface (albedo) variables on a range of scales from thousands of kilometers to hundreds of meters [Rossow *et al.*, 2002]. General circulation models (GCMs) typically have fixed grid cell sizes of tens to hundreds of kilometers.

Such scale differences pose a serious challenge in comparing observation data and modeling results [Randall *et al.*, 2003]. For computational efficiency and accuracy, important subgrid details in cloud and radiation processes must be parameterized. Improving cloud and radiation parameterizations in GCMs is of paramount importance and has been the central goal of the Atmospheric Radiation Measurement (ARM) program of the United States Department of Energy [Ackerman and Stokes, 2003; Stokes and Schwartz, 1994]. At present, the most critical requirements to advance the parameterization of radiative processes for use in GCMs are (1) the quantitative description of radiative energy under a wide range of meteorological conditions; (2) the identification and the investigation of the processes controlling the radiative balance, with a special emphasis on the role of

<sup>1</sup>Also at Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China.



**Figure 1.** Distribution of the ARM extended facilities (left panel) and different model grids used in the study ( $100 \times 100 \text{ km}^2$ ,  $\sim 200 \times 200 \text{ km}^2$ ,  $300 \times 300 \text{ km}^2$  and  $400 \times 400 \text{ km}^2$ , right panel).

clouds in modulating the radiative balance; and (3) intensive observations of cloud and radiation at a variety of temporal and physical scales.

[3] To meet these needs, ARM has spearheaded major efforts to acquire measurements on the scale of GCM grid boxes, in order to define the physics underlying some parameterizations used in GCMs. In the ARM Vision 2000 report, the modeling working group (MWG) recommended that more cloud and radiation data sets, incorporating more detailed information on their spatial and temporal distributions, be produced for direct use by the modeling community [Randall *et al.*, 2000]. The inability of the MWG to test parameterizations of subgrid microphysical processes was attributed to insufficient subgrid measurements of cloud, surface and radiative quantities. This view is also echoed by the ARM Instantaneous Radiative Flux working group, which calls for more emphasis on resolving spatial variability rather than focusing on point-specific measurements [Ellingson, 2000]. Given the long history of ground-based measurements in the ARM program and in other field programs, a major task is how to connect these local point-specific data with global models under a wide range of large-scale conditions [Randall *et al.*, 2000].

[4] Single-column modeling has been the primary tool connecting GCMs and data collected in the field [Randall *et al.*, 2003]. A single column model (SCM) includes the bulk of the physics of a GCM and can be run in isolation from the rest of the GCM. In a SCM, a grid column of a climate model is isolated and supplemented with more detailed models simulating cloud and radiation processes. Data collected in the field can then be used to evaluate the performance of the improved parameterizations. However, the SCM suffers from the same coarse-resolution problem as a GCM when it comes to resolving cloud variability. A cloud system

resolving model (CSRMs) has high spatial and temporal resolutions and is useful for testing some radiative parameterizations over a wide range of space and time domains [Browning *et al.*, 1993]. For example, several CSRMs and SCMs were compared for multiday simulations of convection during the TOGA-COARE experiment [Krueger and Lazarus, 2000]. Some CSRMs are now running at very high resolutions (down to a kilometer or less). As illustrated by Randall *et al.* [2003], field data can be used to combine CSRMs and SCMs in order to develop improved GCM parameterizations. Krueger and Burks [1998] developed and used an empirical relation between TOA-pixel and surface-pixel solar fluxes using satellite and point surface measurements to estimate 3-hourly area-averaged surface fluxes.

[5] The ARM program provides test beds for different climate regimes. Under study here is the Southern Great Plains (SGP) Cloud and Radiation Test Bed region spanning central Oklahoma northward into southern Kansas (Figure 1). This region includes the Central Facility (CF) where the bulk of the instrumentation is located and 23 extended facilities with basic instrumentation such as flux radiometers spread nonuniformly throughout the SGP region. Despite the high density of this network, several outstanding questions still confront us: (1) Can cloud behavior be understood from the deployed instruments?; (2) What type of measurements, and at what density, are required to specify the boundary conditions for SCMs?; (3) What, if any, guideline should be followed for adding observation stations (permanent and mobile) and for deciding their locations?; and (4) At what scale does modeling need to occur in order to capture the physical properties that drive the system? To answer these questions, we need to understand system-scaling behavior due to inhomogeneous cloud and surface cover.

[6] Radiation observation networks that cover the areas of typical mesoscale weather systems (e.g., the Oklahoma Mesonet) may resolve the large-scale variability of the solar radiation field, but are insufficient to acquire statistics for comparison with high-resolution models such as CSRMs. Satellite observations offer an alternative source of information that can be used to study problems such as the spacing of ground-based observation stations [Perez *et al.*, 1997]. In this study, we take advantage of the high spatial and temporal resolution of the Geostationary Operational Environmental Satellites (GOES) data set from which surface solar net irradiance (SSNI) is retrieved. The retrieval can be used to (1) mimic ground-based measurements made at varying spatial densities and temporal frequencies, and (2) characterize the uncertainties of the simulated SSNI caused by cloud variability at different scales in different seasons. Such scale-dependent statistics of observation uncertainties provide constraints on model-observation comparisons and are thus valuable for improving and validating cloud parameterization schemes.

[7] The following section introduces the retrieval models and input data used to estimate SSNI on different spatial and temporal scales. Section 3 presents the results concerning the observation uncertainties and their ramifications on determining the scales upon which informative validation can be performed for testing cloud parameterization schemes and model simulation results.

## 2. Models and Input Data

### 2.1. Satellite Inversion Model

[8] The SSNI was calculated using the models of Li *et al.* [1993] and Masuda *et al.* [1995]. The model of Li *et al.* [1993] essentially connects the net solar radiative fluxes at the top of the atmosphere (TOA) and at the surface through a linear relationship that is a function of the TOA albedo, solar zenith angle and precipitable water amount. This relationship is independent of cloud optical thickness and surface albedo so it can be used under both clear and cloudy sky conditions. The Li *et al.* [1993] model has been used to generate a global climatology of SSNI [Li and Leighton, 1993] with an accuracy of  $5 \text{ W m}^{-2}$  for monthly mean fluxes [Li *et al.*, 1995]; good accuracies were also reported in other studies [Conant *et al.*, 1997; Ramanathan *et al.*, 1995; Waliser *et al.*, 1996, 1999; Jing and Cess, 1998]. The model is currently used as one of the Cloud and the Earth's Radiant Energy System algorithms generating the global SSNI [Wielicki *et al.*, 1996].

[9] For instantaneous and daily mean SSNI, larger errors are expected to occur due to a lack of the explicit inclusion of factors such as cloud top altitude and heavy aerosol loading [Li, 1998; Feng and Leighton, 2003]. Masuda *et al.* [1995] introduced corrections to the Li *et al.* [1993] algorithm to explicitly account for these two particular influences. Clouds at different altitudes affect atmospheric absorption, hence the SSNI, in different ways [Li and Moreau, 1996]. Aerosols also play a role in determining the absorption of solar radiation in the atmosphere. The Li *et al.* [1993] algorithm includes a nominal aerosol with an optical depth of 0.05 at 550 nm, which is inaccurate for heavy aerosol loading or where the aerosol is strongly

absorbing. Correction terms accounting for these two important atmospheric influences on the SSNI were implemented in this study.

### 2.2. Input Data

[10] TOA albedo data are from the high-resolution (approximately 4-km pixel-level data) GOES-8 data set. This satellite data set, generated by the NASA Langley Cloud and Radiation Research Group, uses visible and infrared radiances to derive various cloud and radiation products [Minnis and Smith, 1998]. The lack of onboard calibration requires a careful postlaunch calibration, followed by a narrowband-to-broadband conversion, in order to obtain broadband fluxes from the narrowband uncalibrated digital readings. Both data processes are subject to uncertainties [Trishchenko and Li, 1998]. As a result, the GOES-8 products have been revised several times. Good accuracy in radiation quantities was demonstrated in the later versions as described in the work of Minnis *et al.* [2002]. Employed here were data sets covering the periods of March, May, July, August, September, October, and December of the year 2000 (<ftp://angler.larc.nasa.gov/public/products/armsgp/visst-pixel-bin/goes8/>). They encompass an area of approximately  $400 \times 400 \text{ km}^2$  centered on the SGP CF site in north-central Oklahoma with a time resolution of about every half-hour on a daily basis. In addition to TOA broadband albedo, the data set provides cloud macrophysical properties such as cloud top height and pixel cloudiness and cloud microphysical properties such as effective radius.

[11] Another input parameter to the satellite inversion models is precipitable water (PW) and this quantity was interpolated from the  $2.5^\circ \times 2.5^\circ$  global National Centers for Environmental Prediction (NCEP) reanalysis [Kalnay *et al.*, 1996]. This data set covers the period from 1948 to the present and each file contains data for one year with values of precipitable water given at four times during each day (0, 6, 12, 18 UTC). For the simulations performed here, the subset of data encompassing the SGP region for the year 2000 was bilinearly interpolated to determine PW at the GOES data grids. A more extensive PW data set is available from the SGP where 5-min averages of microwave radiometer (MWR) retrievals have been made on a continual daily basis since about 1994 [Liljegren and Lesht, 1996] at the CF and four facilities located at the perimeter of the SGP region. Although these data are more accurate, they do not have the spatial coverage needed to determine the PW over the entire SGP region. Another source of PW data is the network of Global Positioning System stations located within and around the SGP grid [Rocken *et al.*, 1995]. These data have a better spatial coverage than the MWR data but are still inadequate. For the year 2000, for example, many days of data are missing. To evaluate uncertainties in the SSNI due to PW, we derived the SSNI using PW data from the NCEP reanalysis and the SSNI from the MWR. For March 2000 at grid sizes of  $4 \times 4 \text{ km}^2$  and  $400 \times 400 \text{ km}^2$  (where the CF value for PW was used for all pixels), the standard deviations between the two retrievals are small ( $2.2$  and  $2.8 \text{ W m}^{-2}$ , respectively).

[12] Aerosol optical depths at 550 nm were derived from the Aerosol Robotic Network (AERONET) [Holben *et al.*,

2001] atmospheric extinction measurements taken at wavelengths of 340, 380, 440, 500, 670, 870 and 1020 nm at the CF. Data from 1994 until September 2001 were used to derive the Ångström exponent, allowing us to interpolate the aerosol optical depths to 550 nm. The mean value of the derived aerosol optical depths at 550 nm is 0.138. The values of aerosol optical depth derived from AERONET measurements taken at the CF were assumed to be representative of the entire SGP region, given that aerosol loading is not expected to vary greatly over this generally rural area. There may be exceptions during the spring season when aerosol loading could be increased by agricultural burning in preparation for reseeded.

### 2.3. Methodology

[13] Unlike ground-based point measurements, satellite-measured radiances are areal means over the footprints of the satellite pixels. Averaging the pixel-level data produces mean fluxes over grids of any size. Therefore, through use of high-resolution satellite data, one can gain further insight into the statistics of the spatial variability of the solar radiation field. Also, frequent observations, such as those provided by a geostationary satellite, facilitate the examination of the temporal variability of the solar radiation field.

[14] To that end, we calculated means of SSNI over grids of different size and over different time intervals. The number of pixels within a grid can range from a single pixel to several thousand pixels. The model of *Li et al.* [1993] was first applied to each individual pixel and the correction factors accounting for the effects of aerosols and cloud top height on atmospheric absorption were added, depending on whether the pixel was identified as clear or cloudy. If the pixel was identified as clear, the aerosol correction factor was added. If the pixel was identified as cloudy, the cloud top height correction factor was applied. Different grid sizes centered on the CF were selected. The SSNI was computed for every half-hour during the daytime for all the days when satellite data were available. Grid sizes were chosen to be representative of various SCM and GCM modeling schemes ( $10 \times 10$ ,  $20 \times 20$ ,  $50 \times 50$ ,  $100 \times 100$ ,  $200 \times 200$ ,  $300 \times 300$ , and  $400 \times 400$  km<sup>2</sup>). From the daytime SSNI calculated for each half-hour, temporally averaged SSNI were computed for various scales. Time-averaging intervals chosen in this study were 1, 2, 4, 8 hours, daily, 5 and 10-day means, and monthly means.

## 3. Results

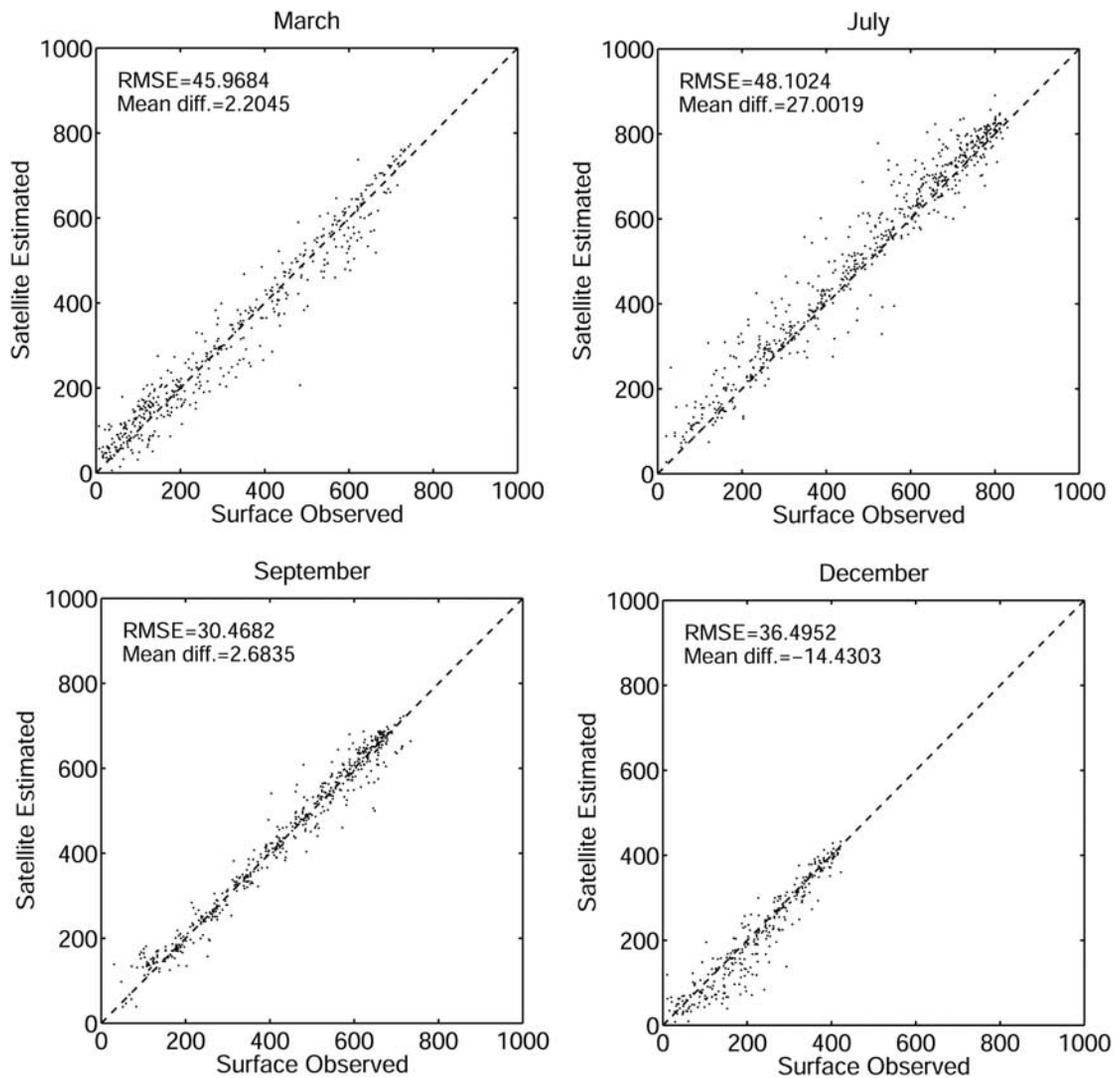
### 3.1. Validation

[15] The values of satellite-retrieved SSNI were first validated against ground-based observations. From satellite retrievals at half-hour intervals, a mean SSNI was computed over an area of  $4 \times 4$  km<sup>2</sup> centered on the CF and averaged over one hour using data from three time points (the time point closest to the satellite observation and the time points before and after). Although the size of the region was selected in a somewhat ad hoc manner, it represents the bulk of the sky area exposed to a radiometer. The mean satellite-retrieved fluxes over this  $4 \times 4$  km<sup>2</sup> area were compared to the average of the ground measurements taken within the 1-hour interval. The observed SSNI was computed from surface downwelling and upwelling irradiances

(corrected for the thermal offset) measured at the Solar Infrared Radiation Station at the CF. To reduce uncertainties resulting from surface albedo, the average values of the surface albedo measurements made from the 10-m and 60-m towers were used. The surfaces below the towers are covered by grass and wheat. Note that surface albedo measurements generally have a very limited spatial coverage, which may introduce significant uncertainties in solar radiative transfer calculations [*Li et al.*, 2002]. The mean of the albedo measurements from the two towers better represents the surface reflective characteristics surrounding the CF [*Michalsky et al.*, 2003].

[16] In Figure 2, comparisons of the SSNI estimated from satellite data and surface observations are presented. In general, the two sets of data agree fairly well, especially in terms of relative differences. Biases range from  $-14$  W m<sup>-2</sup> to  $27.0$  W m<sup>-2</sup> and the root mean square errors (RMSE) range from  $30$  W m<sup>-2</sup> to  $48$  W m<sup>-2</sup> with the smallest biases occurring in March and the smallest RMSE in September. Seasonal changes in surface albedo may play an important role in the biases, while the RMSE is dictated primarily by cloud variability. In September 2000, uniform cloud systems dominate the region [*Dong et al.*, 2000] so there is less variability in the solar radiation field hence less deviation from the point measurements. To a large extent, the scatter in the plots is caused by mismatches between satellite estimates and ground measurements in time and space. Satellite estimates were averaged over a  $4 \times 4$  km<sup>2</sup> area including the CF. They are not necessarily centered on the exact location of the instrument. Also, the frequencies of the satellite estimates and the surface observations differ significantly. The one-hour averages of satellite-estimated SSNI were calculated from three data points separated by 30 min. Observations of downwelling surface fluxes have a temporal resolution of 1 min and surface albedo measurements have a temporal resolution of 15 min. Thus the ground data contain more information concerning the temporal variation of the SSNI. On the other hand, satellite estimates better describe the spatial variation of the SSNI.

[17] The scatter seen in the comparisons thus stems from both physical causes and statistical sampling errors. The latter is linked primarily to cloud variability, with possible minor contributions from other factors such as surface inhomogeneity. To gain further insight into the discrepancies related to data sampling, two comparisons are made over a grid of  $400 \times 400$  km<sup>2</sup>. Table 1 presents the mean and RMSE differences for the comparisons between the means of satellite retrievals averaged over the entire grid and surface observations made at all radiation measurement stations within this grid. While the biases are generally compatible with those shown in Figure 2, the RMSEs are reduced substantially. The reduced RMSEs, albeit still significant, are attributed almost exclusively to the sampling uncertainties. This is clearly seen from Table 2, which gives the statistics of comparisons between two satellite retrievals only (no ground measurements are involved). One retrieval is the mean of all pixels falling inside the  $400 \times 400$  km<sup>2</sup> grid, while the other is the mean of the retrievals over  $4 \times 4$  km<sup>2</sup> smaller regions surrounding all ground stations located within the grid. Note that the latter data may be regarded as simulated surface observations. Since the two sets of data were all retrieved from satellite data using the



**Figure 2.** Comparisons of surface solar net irradiance observed at the CF and estimated from satellite (averaged over a  $4 \times 4 \text{ km}^2$  area centered on the CF) for the months of March, July, September, and December. Units are  $\text{W m}^{-2}$ .

same algorithm, their differences attest to sampling uncertainties. Note that the RMSE values given in Tables 1 and 2 are rather close to each other in general. This implies that the bulk of the differences are caused by sampling errors that are associated with cloud variability. This finding is helpful in devising a useful tool to investigate observation sampling errors, as described in the following section.

### 3.2. Spatial Sampling

[18] The sampling uncertainties incurred by using single-point data to represent the means over grids of varying size

were first examined. Such single-point measurements have been widely used in validating GCMs with grid box sizes of  $\sim 200 \times 200 \text{ km}^2$  [e.g., Wild *et al.*, 1995].

[19] Point measurements of SSNI at the CF were simulated and compared with areal means of SSNI in order to evaluate the representativeness of single-point data with respect to regions of varying sizes typical of model grid cells. Table 3 gives the means and RMSE differences for the March 2000 comparisons between single-site and areal mean SSNI data estimated from the GOES data. The single-site estimates were simulated by averaging SSNI

**Table 1.** The Mean and RMSE Differences Between the Surface Solar Net Irradiances Estimated From All Pixels Over the  $400 \times 400 \text{ km}^2$  Grid and Those Averaged From the 21 Ground-Based Radiation Observation Stations Located Within This Grid

Month	RMSE, $\text{W m}^{-2}$	Mean Difference, $\text{W m}^{-2}$
March	23.9	-4.9
July	24.2	30.7
September	19.3	21.3
December	34.4	-19.0

**Table 2.** Same as Table 1 Except That Values of the Ground Observations Are Replaced by Satellite Retrievals Over  $21 \times 4 \text{ km}^2$  Areas Surrounding Each Radiation Observation Station

Month	RMSE, $\text{W m}^{-2}$	Mean Difference, $\text{W m}^{-2}$
March	28.3	-8.0
July	24.0	-3.7
September	10.6	1.3
December	15.7	-8.3

**Table 3.** The RMSE and Mean Differences From Comparisons of Surface Solar Net Irradiance Estimated at the CF (Averaged Over a  $4 \times 4 \text{ km}^2$  Area) and Over Larger Grids of Varying Size From  $10 \times 10 \text{ km}^2$  to  $400 \times 400 \text{ km}^2$  for March 2000<sup>a</sup>

Grid Size, $\text{km}^2$	RMSE, $\text{W m}^{-2}$	Mean Difference, $\text{W m}^{-2}$
$10 \times 10$	16.2	-0.4
$20 \times 20$	19.8	0.9
$50 \times 50$	27.4	0.1
$100 \times 100$	34.4	-0.7
$200 \times 200$	46.5	0.6
$400 \times 400$	64.2	1.6

<sup>a</sup>The surface solar net irradiances were averaged over an hour.

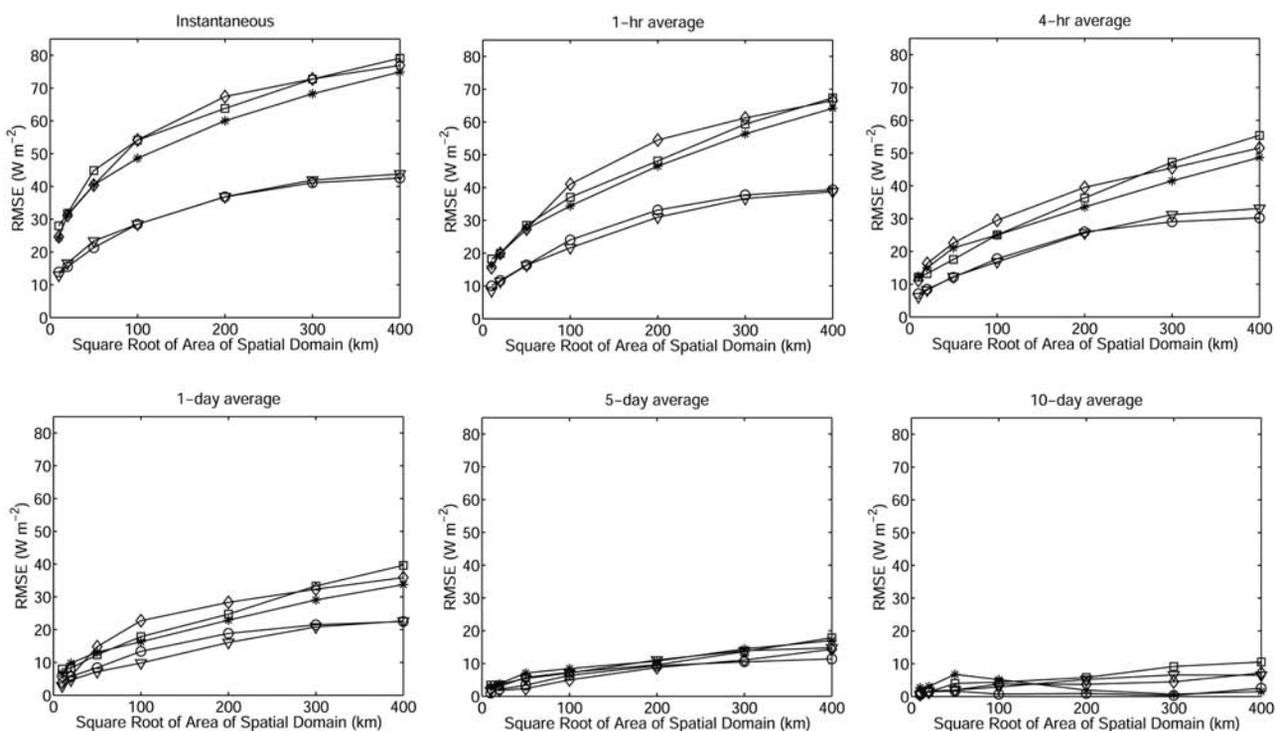
retrieved at individual pixels over a  $4 \times 4 \text{ km}^2$  region surrounding the CF (a proxy for ground observations at the CF). The areal mean SSNI values are averaged over areas of varying size ( $10 \times 10$ ,  $20 \times 20$ ,  $50 \times 50$ ,  $100 \times 100$ ,  $200 \times 200$  and  $400 \times 400 \text{ km}^2$ ), typical of grid cells for cloud resolving models up to GCMs. Since any errors in the single-site and areal mean estimates caused by the retrieval algorithm are virtually identical, their comparison should not suffer from any large systematic difference. This explains why the mean differences are small and exhibit little trend. As the grid size increases, the RMSE increases dramatically, suggesting that more variability in the SSNI is captured.

[20] A similar result was shown in the work of Long *et al.* [2002] using cloud amount data from their Surface Cloud Grid product. A  $0.25^\circ \times 0.25^\circ$  grid of cloud amount over the ARM SGP area was generated using the shortwave flux

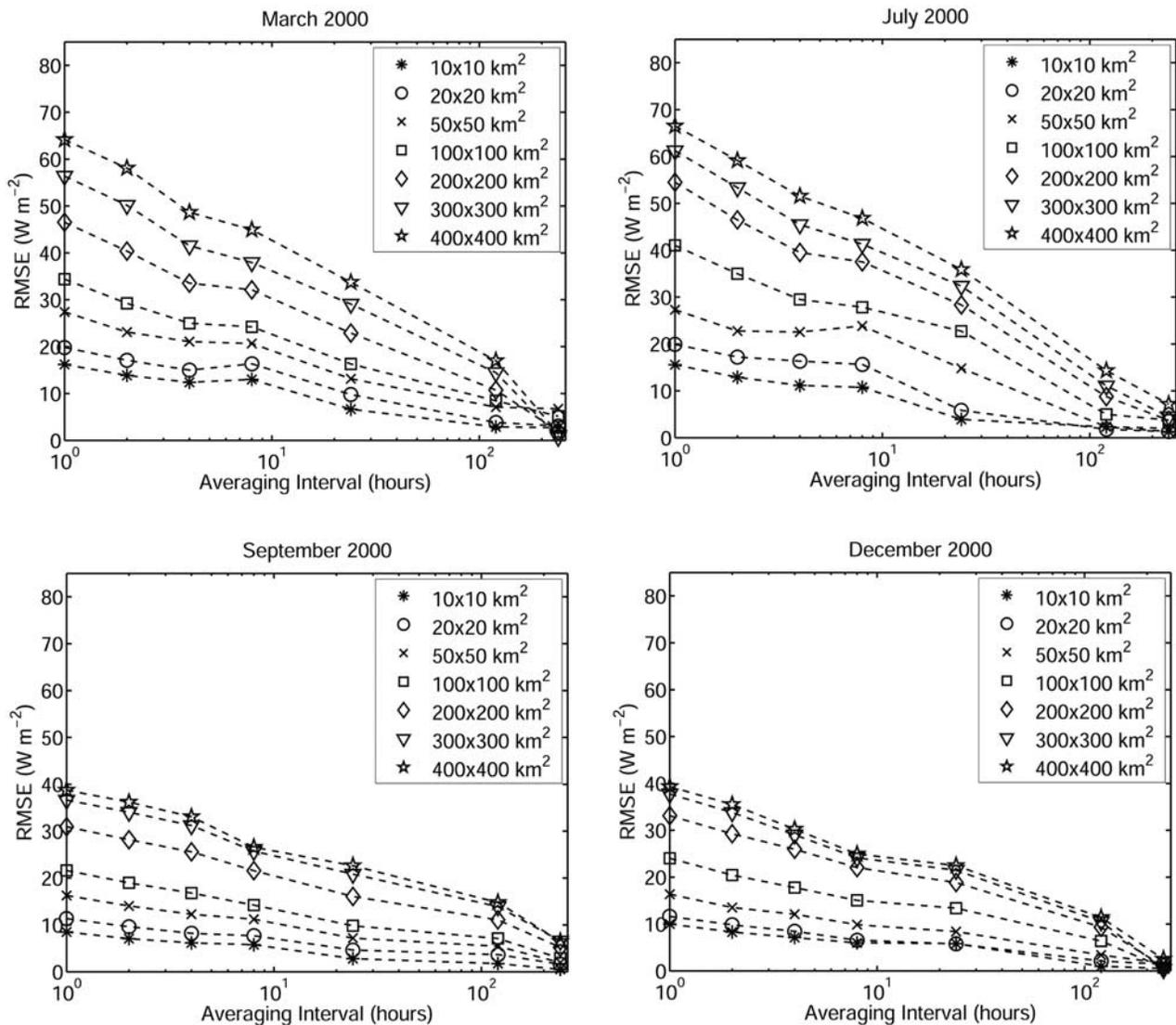
value-added product from all radiation measurements sites within the SGP region as input. An analytic approximation interpolation technique developed by Caracena [1987] was used to create the  $0.25^\circ \times 0.25^\circ$  grid extending over the SGP region. From comparisons of daily averages of the cloud amount derived from the shortwave flux analysis at the CF with the daily averages of cloud amount from the Surface Cloud Grid product over areas of increasing size, they found that the CF data became less representative of the larger area.

[21] Note that the RMSE values given in Tables 1 to 3 represent the inherent uncertainties of ground-based observations due to spatial sampling. As a result, one may not reduce model observation differences below these ranges. For example, the results presented in Table 3 for March 2000 imply that the best accuracy one may achieve with a model of  $10 \times 10 \text{ km}^2$  resolution would be  $16 \text{ W m}^{-2}$ , which increases to 46 and  $64 \text{ W m}^{-2}$  as the model grid increases to  $200 \times 200 \text{ km}^2$  and  $400 \times 400 \text{ km}^2$ , respectively.

[22] The RMSE computed for any fixed grid is contingent upon the temporal averaging intervals. Figure 3 shows the RMSE as a function of grid size for different averaging intervals (instantaneous, 1 hour, 4 hours, 1 day, 5 days and 10 days) in March, May, July, September, and December. In general, the RMSE varies considerably from month to month, indicating that the variability of SSNI has a strong dependence on season. The least variability occurred in September and December over all averaging intervals and grids. For a grid of  $400 \times 400 \text{ km}^2$ , the RMSEs calculated from instantaneous data and data averaged over 4 hours for



**Figure 3.** The standard deviation of the SSNI differences between satellite-simulated “point measurements” and mean values averaged over grids of varying size up to  $400 \times 400 \text{ km}^2$  for instantaneous values and values averaged over different time intervals in March (stars), May (squares), July (diamonds), September (triangles), and December (circles).



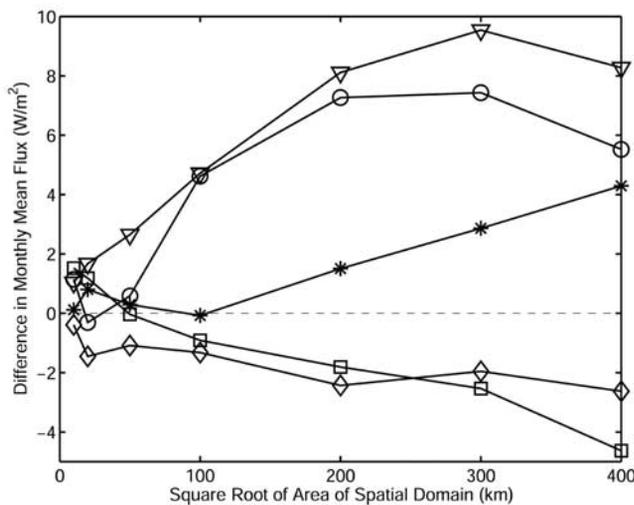
**Figure 4.** Similar to Figure 3 but plotted as a function of the averaging period (1 hour, 2 hours, 4 hours, 8 hours, 1 day, 5 days, and 10 days) for different grids and months.

these two months are less than those for other months by more than  $25 \text{ W m}^{-2}$ . July shows the most change in the magnitude of the RMSE as the averaging interval increases, presumably caused by the prevalence of small convective clouds. However, as the averaging interval increases, the difference among different months diminishes. This is because cloud systems have certain temporal scales. As the averaging interval is increased, more changes in the cloud field, hence the SSNI field, are captured. For a one-day averaging interval, the RMSE decreases substantially relative to its value averaged over hours, but the differences among various months are still significant. For typical GCM grid boxes of  $200 \times 200 \text{ km}^2$ , the daily sampling errors range from 16 to  $28 \text{ W m}^{-2}$  in September and July, respectively. They decrease to  $10 \text{ W m}^{-2}$  or less if the model grid is reduced to  $10 \times 10 \text{ km}^2$ . For the 5-day average, however, the differences among all the months almost vanish.

### 3.3. Temporal Averaging

[23] Physical processes in the atmosphere occur at various scales in both space and time domains. The physical

state of the atmospheric system can change substantially over time. It is thus important to determine appropriate timescales upon which solar radiative quantities are averaged so that errors incurred in matching model estimates and measurements are minimized. Figure 4 shows the variation of the sampling errors in the SSNI as a function of the averaging interval for different grid sizes and months. A sharp decrease in the sampling error occurs as the averaging interval increases to a day. Beyond one day, the error tends to level off to a stable value of small magnitude. Note that in Figure 3, an interesting feature appears with regard to how the dependence of the RMSE on grid size varies with the averaging interval. For averaging intervals less than a day, the dependence follows a log function, which gradually transforms into a linear function for intervals longer than a day. This is probably associated with cloud scales. Small-scale clouds exhibit high-frequency variations that last for relatively short periods of time, whereas large-scale clouds, which influence longer-term averaging, show much slower temporal variation. For intervals of 10 days or longer, this dependence is very



**Figure 5.** Difference between monthly mean SSNI simulated for “point measurements” and areal mean values as a function of grid size varying from  $10 \times 10 \text{ km}^2$  to  $400 \times 400 \text{ km}^2$  in March (stars), May (squares), July (diamonds), September (triangles), and December (circles).

weak, as most cloud systems have a lifetime shorter than 10 days.

[24] It follows from the above discussions that the sampling error depends on the grid size and the month chosen. The statistics shown in Figures 4 and 5 provide objective measures of what would be an acceptable error for climate modeling purposes. For example, for a grid of  $100 \times 100 \text{ km}^2$  in March, increasing the averaging period from one hour to two hours reduces the sampling error from 34 to  $29 \text{ W m}^{-2}$ . In September, for the same grid size and averaging period, the sampling error would reduce from 22 to  $19 \text{ W m}^{-2}$ . If the time interval used for averaging is greater than five days, the errors remain steady and very small (less than  $10 \text{ W m}^{-2}$ ) for all grid sizes. Overall, the highest magnitudes in the RMSE tend to occur in summer (July) when cloud systems are more complex and variable in both time and space [Lazarus *et al.*, 2000; Dong *et al.*, 2000]. Thus use of the point measurements for validating a model would result in large uncertainties unless the measurements are averaged over a long time interval. So temporal averaging of the SSNI is an effective means of minimizing matching errors if the physical process under study is not unduly sensitive to time passages.

[25] Figure 5 shows the difference between the monthly mean SSNI over different grid sizes and the monthly mean SSNI simulated at the CF for the months of March, May, July, September, and December. The magnitude of the sampling errors is less than  $10 \text{ W m}^{-2}$  for all months shown and grid sizes chosen and the magnitude of the error diminishes to less than  $3 \text{ W m}^{-2}$  for typical GCM grid boxes with areas of  $200 \times 200 \text{ km}^2$  for March, May and July. While the magnitude of the errors agrees well with the general requirement of an accuracy of  $5 \text{ W m}^{-2}$  for climate studies [Suttles and Ohring, 1986], it would be an unrealistic goal for certain regions/months, such as September 2000 over the  $400 \times 400 \text{ km}^2$  SGP area. In this case, the inherent sampling error in monthly mean surface observa-

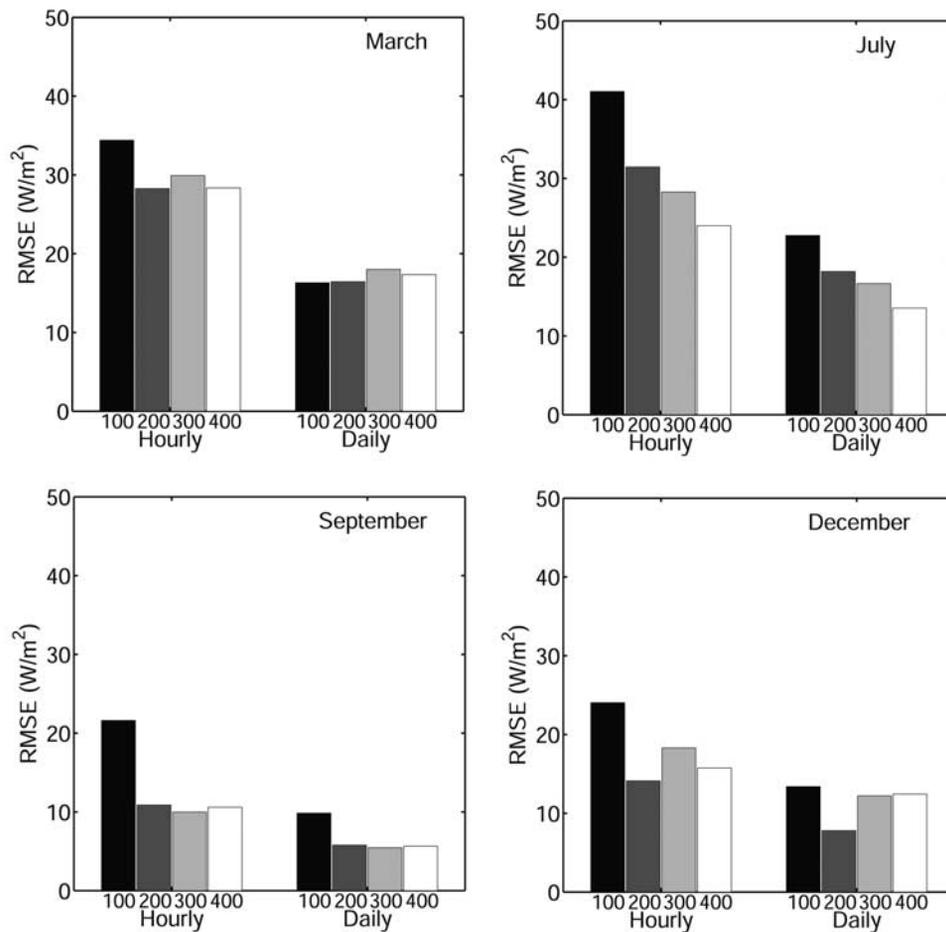
tions is  $8 \text{ W m}^{-2}$ , exceeding the required accuracy for climate studies.

### 3.4. Multiple Ground Stations

[26] Since single-point measurements do not well represent areal means over a large grid, multiple radiation stations distributed over a large grid around the CF improve the spatial representation of the SSNI, as they capture more of the variability in the SSNI field. Questions remain, however, as to how much improvement in accuracy is gained as the number of stations increases, and how many stations are really needed to meet certain accuracy requirements. To address these issues, grids of different sizes centered at the CF were selected ( $100 \times 100$ ,  $\sim 200 \times 200$ ,  $300 \times 300$ , and  $400 \times 400 \text{ km}^2$ ), each containing 1, 7, 12, and 21 observation sites, respectively. These grids and the locations of the observation sites are illustrated in Figure 1.

[27] For each particular grid size, satellite-estimated SSNI was calculated over  $4 \times 4 \text{ km}^2$  regions centered on each site (a proxy for actual surface measurements). The mean values averaged over all sites inside a particular grid are compared to the satellite-estimated areal means over the entire grid. Figure 6 presents the RMSEs calculated for 1-hour and daily means in March, July, September, and December for the grid sizes of  $100 \times 100$ ,  $\sim 200 \times 200$ ,  $300 \times 300$ , and  $400 \times 400 \text{ km}^2$ . The sampling error is the largest for the  $100 \times 100 \text{ km}^2$  grid in which there is only one station (the CF). For other grid sizes, the magnitudes of the RMSE vary significantly with month, from about  $6 \text{ W m}^{-2}$  (September) to more than  $17 \text{ W m}^{-2}$  (March) for the daily means, and from  $10 \text{ W m}^{-2}$  (September) to  $30 \text{ W m}^{-2}$  (March) for the hourly means. The December values fall somewhere in between.

[28] Similar calculations were also conducted for a fixed grid size of about  $200 \times 200 \text{ km}^2$  using data from different combinations of the radiation measurement stations contained within this grid. In addition to the CF, the grid encompasses radiation measurement stations located in Ashton, Kansas, and Byron, Pawhuska, Ringwood, El Reno, and Meeker in the state of Oklahoma. The number of observation stations used ranges from 1 (the CF) to 7 (all stations). Intermediate combinations of the stations used consist of the CF plus an increasing number of the other stations (for 2 stations, CF and Byron; for 3 stations, CF, Byron and Ashton; for 4 stations, CF, Byron, Ashton, and Pawhuska; for 5 stations, CF, Byron, Ashton, Pawhuska, and Ringwood; for 6 stations, CF, Byron, Ashton, Pawhuska, Ringwood, and El Reno). The RMSE for March 2000 was calculated and plotted as a function of the number of surface stations for instantaneous, 1-hour, 4-hour, 1-day, 5-day and 10-day means (Figure 7). In general, the RMSE decreases as the number of observation sites increases. However, the reduction is more significant when using 2 or 3 sites, after which the RMSE more or less levels off (and sometimes increases a bit due to random sampling). This means that for computing grid mean SSNI, there is no need for an overly dense network of observation stations. The number of stations required is contingent upon the averaging period and month. The shorter the averaging interval, the more stations are needed to reduce the RMSE of the SSNI. Mean SSNI values averaged over one day remain



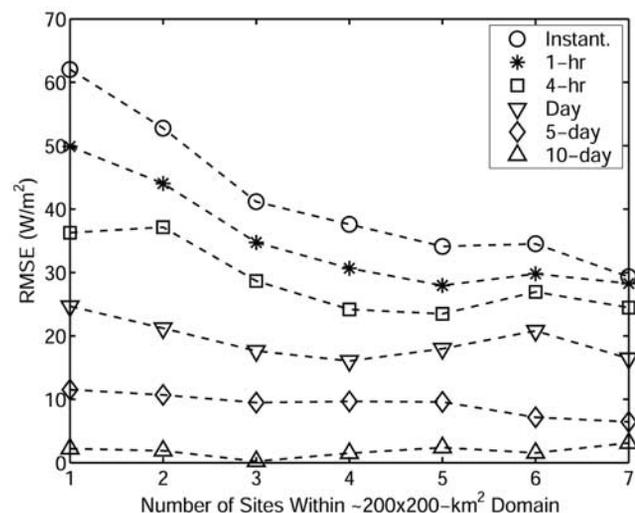
**Figure 6.** The standard deviations of the SSNI differences between satellite-estimated areal means and simulated point-specific values averaged over all stations falling within the grids of  $100 \times 100$ ,  $\sim 200 \times 200$ ,  $300 \times 300$ , and  $400 \times 400$  km<sup>2</sup> containing 1, 4, 7, and 21 sites, respectively. Only hourly and daily averaged values are given.

generally stable as the number of stations increases from 1 to 7. Similar analysis for other months (not shown) reveals a more significant reduction in the magnitude of the RMSE as the number of sites increases.

#### 4. Summary

[29] Radiation measurements have been widely employed for evaluating cloud parameterization schemes and model simulation results. Different data have been employed to obtain the estimates of area-averaged surface solar radiation for time intervals ranging from an hour to a few days in order to validate model results. The data include (1) temporal averages of ground-based measurements, (2) spatial means of satellite-based estimates, and (3) a combination of the two, in which the point measurements are used to correct any biases in the satellite estimates.

[30] In this study, we take advantage of the high spatial and temporal resolution of a recently processed Geostationary Operational Environmental Satellite data set of cloud properties to simulate ground-based measurements of the surface solar net irradiance (SSNI). By averaging the GOES retrievals in space and time, we can characterize the observation uncertainties of SSNI caused by cloud variability at



**Figure 7.** The standard deviation of the differences between grid-averaged SSNI and the means of simulated surface observations at 1–7 stations falling within a grid of  $200 \times 200$  km<sup>2</sup> for March 2000.

different scales for different months. Such scale-dependent statistics of observation uncertainties provide critical constraints on model observation comparisons, and are thus valuable for improving and validating cloud parameterization schemes. The Department of Energy Atmospheric Radiation Measurement program's observation network in Oklahoma and Kansas provided a useful test bed for this study. In terms of spatial averaging, a single observation site (the Central Facility) does an increasingly poor job of representing areal means of SSNI as the grid size increases. Averaging the SSNI at more observation sites results in a decrease in error as the grid size (and number of observation sites) increases. As for temporal averaging, increasing the time interval also leads to a general decrease in the sampling error. Averaging over periods greater than 5 days smooths out the error to a generally stable magnitude of less than  $15 \text{ W m}^{-2}$ . Use of multiple observation sites helps capture the variability of the SSNI field.

[31] When modeled radiation quantities are compared to ground observations, the inherent uncertainties due to sampling errors must be taken into consideration. Such inherent uncertainties are simulated from our satellite retrievals, which are given as functions of model grid size, averaging period, number of observation stations, and month. If the differences between modeled and observed radiation quantities are comparable to or less than the corresponding inherent uncertainty, no further insight may be gained with regard to any model's deficiencies. Such statistics are thus valuable for validating models when testing their parameterization schemes.

[32] In evaluating the performance of a model using surface or satellite data, one should bear in mind the pros and cons of each data set, noting that ground observations have the highest accuracy and frequency, while satellite estimates provide the most extensive spatial coverage and uniform quality. A combination of the two types of data set is thus recommended for model evaluation. Not only are the ground observations useful for validating satellite retrievals, they may also be used to remove any biases caused by infrequent sampling, especially for regions outside the coverage of the geostationary satellites.

[33] **Acknowledgments.** The Cloud and Radiation group led by Pat Minnis at the NASA Langley Research Center generated the GOES products used in the study. Comments by three reviewers and Warren Wiscombe are greatly appreciated. NCEP reanalysis data were obtained from the Web site of the NOAA-CIRES Climate Diagnostics Center in Boulder, Colorado (<http://www.cdc.noaa.gov/>). Aerosol data were downloaded from the Web site of the Aerosol Robotic Network (<http://aeronet.gsfc.nasa.gov/>). This study has been supported by the DOE ARM grants DE-FG02-01ER63166 and DE-FG02-02ER63351.

## References

- Ackerman, T., and G. Stokes (2003), The Atmospheric Radiation Measurement Program, *Phys. Today*, *56*, 38–45.
- Browning, K. A., et al. (1993), GEWEX cloud system study (GCSS), *Bull. Am. Meteorol. Soc.*, *74*, 387–399.
- Caracena, F. (1987), Analytic approximation of discrete field samples with weighted sums and the gridless computation of field derivatives, *J. Appl. Meteorol.*, *3*, 396–409.
- Conant, W. C., V. Ramanathan, F. P. J. Valero, and J. Meywerk (1997), An examination of the clear-sky solar absorption over the central equatorial Pacific: Observation versus models, *J. Clim.*, *10*, 1874–1884.
- Dong, X., P. Minnis, T. P. Ackerman, E. E. Clothiaux, G. G. Mace, C. N. Long, and J. C. Liljegren (2000), A 25-month database of stratus cloud properties generated from ground-based measurements at the Atmospheric Radiation Measurement Southern Great Plains Site, *J. Geophys. Res.*, *105*, 4529–4537.
- Ellingson, R. G. (2000), The ARM radiative flux (IRF) working group in 2000: A summary of accomplishment, strengths, weaknesses, and ideas for future activities, *ARM Vision 2000 Rep.*, 15 pp., Atmos. Radiat. Measure. Prog., U.S. Dept. of Energy, Washington, D. C.
- Feng, J., and H. G. Leighton (2003), Effects of absorbing aerosols on the determination of the surface solar radiation, *J. Geophys. Res.*, *108*(D7), 4227, doi:10.1029/2002JD002852.
- Holben, B. N., et al. (2001), An emerging ground-based aerosol climatology: Aerosol optical depth from AERONET, *J. Geophys. Res.*, *106*, 12,067–12,097.
- Jing, X., and R. Cess (1998), Comparison of atmospheric clear-sky radiation models to collocated satellite surface measurements in Canada, *J. Geophys. Res.*, *103*, 28,817–28,824.
- Kalnay, E., et al. (1996), The NCEP/NCAR reanalysis 40-year project, *Bull. Am. Meteorol. Soc.*, *77*, 437–471.
- Krueger, S. K., and J. E. Burks (1998), Radiative fluxes and heating rates during TOGA COARE over the Intensive Flux Array, in *Proceedings of the Eighth Atmospheric Radiation Measurement (ARM) Science Team Meeting*, pp. 399–402, U.S. Dept. of Energy, Washington, D. C.
- Krueger, S. K., and S. M. Lazarus (2000), Intercomparison of multi-day simulations of convection during TOGA COARE with several cloud resolving and single-column models, in *Proceedings of the Workshop on Cloud Processes and Cloud Feedbacks in Large-Scale Models, Rep. WCRP-110, WMO/TD Rep. 993*, pp. 84–91, Eur. Cent. for Medium-Range Weather Forecast., Reading, UK.
- Lazarus, S. M., S. K. Krueger, and G. G. Mace (2000), A cloud climatology of the Southern Great Plains ARM CART, *J. Clim.*, *13*, 1762–1775.
- Li, Z. (1998), Influence of absorbing aerosols on the inference of solar surface radiation budget and cloud absorption, *J. Clim.*, *11*, 5–17.
- Li, Z., and H. G. Leighton (1993), Global climatologies of solar radiation budgets at the surface and in the atmosphere from 5 years of ERBE data, *J. Geophys. Res.*, *98*, 4919–4930.
- Li, Z., and L. Moreau (1996), Alteration of atmospheric solar absorption by clouds: Simulation and observation, *J. Appl. Meteorol.*, *35*, 653–670.
- Li, Z., H. G. Leighton, K. Masuda, and T. Takashima (1993), Estimation of SW flux absorbed at the surface from TOA reflected flux, *J. Clim.*, *6*, 317–330.
- Li, Z., T. Charlock, and C. Whitlock (1995), Assessment of the global monthly mean surface insolation estimated from satellite measurements using global energy balance archive data, *J. Clim.*, *8*, 315–328.
- Li, Z., M. C. Cribb, and A. P. Trishchenko (2002), Impact of surface inhomogeneity on solar radiative transfer under overcast conditions, *J. Geophys. Res.*, *107*(D16), 4294, doi:10.1029/2001JD000976.
- Liljegren, J. C., and B. M. Lesht (1996), Measurements of integrated water vapor and cloud liquid water from microwave radiometers at the DOE ARM Cloud and Radiation Test Bed in the U.S. Southern Great Plains, paper presented at the IEEE International Geosciences and Remote Sensing Symposium, Atmos. Radiat. Measure. Prog., U.S. Dept. of Energy, Lincoln, Neb.
- Long, C. N., T. P. Ackerman, and J. E. Christy (2002), Variability across the ARM SGP area by temporal and spatial scale, paper presented at Twelfth Atmospheric Radiation Measurement Science Team Meeting, Atmos. Radiat. Measure. Prog., U.S. Dept. of Energy, St. Petersburg, Fla.
- Masuda, K., H. G. Leighton, and Z. Li (1995), A new parameterization for the determination of solar flux absorbed at the surface from satellite measurements, *J. Clim.*, *8*, 1615–1629.
- Michalsky, J., Q. Min, J. Barnard, R. Marchand, and P. Pilewskie (2003), Simultaneous spectral albedo measurements near the Atmospheric Radiation Measurement Southern Great Plains (ARM SGP) central facility, *J. Geophys. Res.*, *108*(D8), 4254, doi:10.1029/2002JD002906.
- Minnis, P., and W. L. Smith Jr. (1998), Cloud and radiative fields derived from GOES-8 during SUCCESS and the ARM-UAV spring: 1996 Flight series, *Geophys. Res. Lett.*, *25*, 1113–1116.
- Minnis, P., L. Nguyen, D. R. Doelling, D. F. Young, W. F. Miller, and D. P. Kratz (2002), Rapid calibration of operational and research meteorological satellite imagers: part I: Evaluation of research satellite visible channels as references, *J. Atmos. Oceanic Technol.*, *19*, 1233–1249.
- Morcrette, J.-J. (2002), Assessment of the ECMWF model cloudiness and surface radiation fields at the ARM SGP site, *Mon. Weather Rev.*, *130*, 257–277.
- Perez, R., R. Seals, and A. Zelenka (1997), Comparing satellite remote sensing and ground network measurements for the production of site/time specific irradiance data, *Sol. Energy*, *60*, 89–96.
- Ramanathan, V., B. Subasilar, G. J. Zhang, W. Conant, R. D. Cess, J. T. Kiehl, H. Grassl, and L. Shi (1995), Warm pool heat budget and shortwave cloud forcing: A missing physics, *Science*, *267*, 499–503.

- Randall, D., S. Krueger, T. Del Genio, S. Ghan, M. Zhang, and R. Cederwall (2000), ARM cloud parameterization and modeling working group, *ARM Vision 2000 Rep.*, 16 pp.
- Randall, D., et al. (2003), Confronting models with data: The GEWEX cloud system study, *Bull. Am. Meteorol. Soc.*, *84*, 455–469.
- Rocken, C., T. Van Hove, J. Johnson, F. Solheim, R. H. Ware, M. Bevis, S. Businger, and S. Chiswell (1995), GPS/STORM—GPS sensing of atmospheric water vapor for meteorology, *J. Atmos. Oceanic Technol.*, *12*, 468–478.
- Rossow, W. B., C. Delo, and B. Cairns (2002), Implications of the observed mesoscale variations of clouds for the Earth's radiation budget, *J. Clim.*, *15*, 557, doi:10.1175/1520-0442.
- Stokes, G. M., and S. E. Schwartz (1994), The Atmospheric Radiation Measurement (ARM) program: Programmatic background and design of the cloud and radiation test bed, *Bull. Am. Meteorol. Soc.*, *75*, 1201–1221.
- Suttles, J. T., and G. Ohring (1986), Report of the workshop on surface radiation budget for climate applications, *Tech. Rep. WCP-115*, World Meteorol. Org., Geneva.
- Trishchenko, A., and Z. Li (1998), Use of ScaRaB measurements for validating a GOES-based TOA radiation product, *J. Appl. Meteorol.*, *37*, 591–605.
- Waliser, D. E., W. D. Collins, and S. P. Anderson (1996), An estimate of the surface shortwave cloud forcing over the western Pacific during TOGA COARE, *Geophys. Res. Lett.*, *23*, 519–522.
- Waliser, D. E., R. A. Weller, and R. D. Cess (1999), Comparisons between buoy-observed, satellite-derived, and modeled surface shortwave flux over the subtropical North Atlantic during the subduction experiment, *J. Geophys. Res.*, *104*, 31,301–31,320.
- Wielicki, B. A., et al. (1996), Clouds and the Earth's Radiant Energy System (CERES): An Earth Observing System experiment, *Bull. Am. Meteorol. Soc.*, *77*, 853–868.
- Wild, M., A. Ohmura, H. Gilgen, and E. Roeckner (1995), Validation of general circulation model radiative fluxes using surface observations, *J. Clim.*, *8*, 1309–1324.

---

F.-L. Chang, M. C. Cribb, and Z. Li, Department of Meteorology and Earth System Science, Interdisciplinary Center, University of Maryland, College Park, MD 20742, USA. (zli@atmos.umd.edu)

Y. Luo and A. Trishchenko, Applications Division, Canada Centre for Remote Sensing, Ottawa, Ontario, Canada K1A 0Y7.