Remote Sensing of Environment 181 (2016) 198–206

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Representativeness errors of point-scale ground-based solar radiation measurements in the validation of remote sensing products

Guanghui Huang [a,b], Xin Li [b,a,*], Chunlin Huang [b,c], Shaomin Liu [d], Yanfei Ma [d,e], Hao Chen [f]

[a] Cold and Arid Regions Remote Sensing Observation System Experimental Station, CAREERI, CAS, Lanzhou 730000, China
[b] Key Laboratory of Remote Sensing of Gansu Province, CAREERI, CAS, Lanzhou 730000, China
[c] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, 210023, China
[d] State Key Laboratory of Remote Sensing Science, School of Geography, Beijing Normal University, Beijing 100875, China
[e] Department of Geography, Handan College, Hebei 056005, China
[f] Key Laboratory of Land Surface Process and Climate Change in Cold and Arid Regions, CAREERI, CAS, Lanzhou 730000, China

ARTICLE INFO

Article history:
Received 30 July 2015
Received in revised form 17 March 2016
Accepted 1 April 2016
Available online 26 April 2016

Keywords:
Representativeness error
Solar radiation
Ground-based measurements
Remote sensing products
Validation

ABSTRACT

We usually use ground-based solar radiation measurements to validate satellite-derived solar radiation products from kilometer to grid scales. Questions such as, how large is the representativeness error of surface measurements in the validation and how much of the product-measurement difference can be attributed to their inherent differing spatial scales, cast doubts on the suitability of this direct validation approach. In this paper, we will investigate and quantify the representativeness errors of point-scale ground-based measurements using the surface flux-observation matrix from HiWATER (Li et al., 2013) and the solar radiation data retrieved from geostationary meteorological satellite (Huang, Li, Ma, & Li, 2016). The current study demonstrates that wildly fluctuating representativeness errors exist which are strongly contingent on the time and space scales of remote sensing products, as well as instant atmospheric conditions. For example, for an area of 5×5 km^2 1.4~8.1% of representativeness errors are found from monthly to "instantaneous" timescales; while for an area of $1° \times 1°$ grid 3.1~8.1% of representativeness errors are seen. Such scale-dependent representativeness errors offer some implications for validations of remote sensing products. On timescales longer than or equal to one day, representativeness errors do not need to be considered for validations of kilometer-level products, but on shorter timescales representativeness errors will affect the validation results to some extent. For instantaneous products with 5 km resolution, our study indicates over 13% of errors can be attributed to the inherent representativeness error, and 30-minute surface measurements are recommended for a routine validation. However, for validations of grid-level products, representativeness errors basically cannot be neglected regardless of timescales. The errors caused by the poor representativeness of surface sites, likely significantly contribute to the large differences between measurements and products.

1. Introduction

Monitoring surface radiation budget by remote sensing technique is a very important use of satellite data. Especially for surface solar radiation (SSR), the earliest studies can date back to the 60s of the last century (Fritz, Rao, & Weinstein, 1964). Besides large and continuous spatio-temporal coverage, another main advantage of the satellite remote sensing technique over other approaches (e.g., discrete surface observation or reanalysis modelled products) is that it can accurately capture the spatial distributions and dynamic changes of cloud, which is regarded as a superior modulator of SSR (Forman & Margulis, 2009). Up to now, over decades of rapid development, estimation of SSR based on satellite remote sensing technique has become increasingly mature, and many products have already been produced (Liang, Wang, Zhang, & Wild, 2010; Pinker & Laszlo, 1992; Pinker et al., 2003; Posselt, Mueller, Stockli, & Trentmann, 2012; Zhang, Liang, Zhou, Wu, & Zhao, 2014; Zhang, Rossow, Lacis, Oinas, & Mishchenko, 2004).

Approximately, these products can be separated into two classes: 1) kilometer-level or pixel-level products such as the SSR data (Posselt et al., 2012) from European Satellite Application Facility on Climate Monitoring (CM SAF) and the downward surface shortwave radiation data (Huang et al., 2013; Zhang et al., 2014) from Global Land Surface Satellite (GLASS); 2) grid-level products (it should be noted that "grid" in the paper exclusively refers to larger geographic latitude-longitude grids that are generally ≥0.25°) such as the surface radiation budget data (Pinker & Laszlo, 1992) of Global Energy and Water Cycle Experiment (GEWEX) and the flux data (Zhang et al., 2004) of International Satellite Cloud Climatology Project (ISCCP). The kilometer-level products are typically derived from satellite top-of-atmosphere (TOA) radiance or reflectance directly, and primary high spatial resolution is

* Corresponding author at: Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou 73000, China.
E-mail addresses: lixin@lzb.ac.cn, luckhgh@lzb.ac.cn (X. Li).

http://dx.doi.org/10.1016/j.rse.2016.04.001
0034-4257/

preserved; whereas the grid-level products are usually calculated either using coarse satellite and reanalysis atmospheric data as the model inputs or through simply averaging higher spatial resolution products, and mainly used for the studies of global water cycle and climate change.

In general, the validation of products, whether at the kilometer-level or grid-level, is conducted via a direct comparison of collocated estimates from products with ground-based SSR measurements (Huang et al., 2013; Sanchez-Lorenzo, Wild, & Trentmann, 2013; Yang, Koike, Stackhouse, Mikovitz, & Cox, 2006; Yang et al., 2008). However, the spatial nature of satellite remote sensing products and that of ground-based measurements are totally different. Essentially, what kilometer-level or pixel-level products provide are the areal retrieved values over the footprints of satellite pixels; what grid-level products give are the means within latitude-longitude grids; whereas ground-based measurements are point-specific. They are three distinct spatial samplings on surface radiation fields and embody the space attributes of kilometer-scale, grid-scale, and point-scale, respectively. Such scale diversity poses the following question: how well do surface point measurements represent the larger scale surrounding means characterized by satellite remote sensing products? If the error originating from their unique spatial sampling scales is defined as representativeness error (Li, 2014), how large is it in the routine validation of remote sensing products?

To our knowledge, an in-depth and comprehensive study on this issue has not been performed to this date. Sporadic, related researches mainly focus on discussions of sampling errors between the kilometer-scale and the grid-scale spaces. For example, Li, Cribb, and Chang (2005) used SSR retrieved from Geostationary Operational Environmental Satellite (GOES) to mimic ground measurements and quantify the sampling errors within grid-level surroundings; Hakuba, Folini, Sanchez-Lorenzo, and Wild (2013) calculated the representativeness errors with respect to the standard 1° grid on climatological mean conditions in Europe with the help of the SSR from CM SAF. Therefore, it is necessary to thoroughly investigate and quantify the representativeness errors from the point-scale to the kilometer-scale, to the grid-scale.

A surface flux-observation matrix from the Multi-Scale Observation Experiment on Evapotranspiration over heterogeneous land surface of Heihe Watershed Allied Telemetry Experimental Research (HiWATER-MUSOEXE) provided an opportunity for us to study and address this issue (Li et al., 2013). HiWATER is a comprehensive ecohydrological experiment within the framework of the Heihe Plan, which was launched by the National Natural Science Foundation of China (NSFC) in 2010. MUSOEXE is the first thematic experiment launched by HiWATER, and the flux-observation matrix in the middle reach of the Heihe river basin is the major content of MUSOEXE. Although this observation matrix was initially designed to monitor the spatio-temporal variation of evapotranspiration, dense multi-point radiation measurements can also be used to derive representativeness errors of point-scale measurements with respect to their kilometer-level surroundings. Similarly, sampling errors from kilometer to grid scales can be obtained by means of high resolution SSR satellite products (Hakuba et al., 2013; Li et al., 2005). Furthermore, based on the above two types of errors, representativeness errors of point-scale measurements within grid-level domains can be quantified. By analyzing representativeness errors of point-scale ground-based SSR measurements with respect to different levels of areas (kilometer-level or grid-level), implications for the routine validation of remote sensing products are drawn and presented finally.

2. Data

2.1. Ground-based observations

As mentioned earlier, surface observation data in this study are only provided by the surface flux-observation matrix from HiWATER-MUSOEXE (Li et al., 2013). This observation matrix includes 17 sets of eddy covariance (EC) system and automatic meteorological station and 4 pairs of large aperture scintillometer (LAS) systems, and covers approximately 5×5 km^2 spatial area located in the midstream area of Heihe river basin, northwest China. The experiment was carried out from June to September 2012, and its purpose was to capture the variability of evapotranspiration over heterogeneous land surfaces (Liu et al., 2011).

Solar radiation in HiWATER-MUSOEXE was observed with routine meteorological parameters. The detailed distribution of meteorological sites equipped with radiometers is presented in Fig. 1. In a ~5×5 km^2 spatial area, 17 radiometers were installed and formed a dense radiation observation matrix. Most of them were CNR1 and CNR4 manufactured by Kipp & Zonen (Netherlands) except Sites 3, 15 and 16 (see Table 1 for details). At Sites 3 and 15 NR01 radiometer (Hukseflux, Netherlands) and PSP/PIR radiometer (Eppley, U.S.) were equipped respectively, and at Site 16 the instrument is Q7 produced by REBS (U.S.). Because only net radiation was output by Q7 radiometer, Site 16 was excluded. Among these radiometers, the pyranometer of PSP/PIR belongs to the 1st class of the World Meteorological Organization (WMO) classification, and the others can only be qualified as the 2nd class of WMO classification. In order to identify and reduce the calibration error among instruments, an instrumental intercomparison test was conducted over the Gobi desert between May 16 and 22, 2012 (Xu et al., 2013) before the matrix experiment. This test indicates some pyranometers have relatively significant calibration error up to ~3%. Hence, with the PSP/PIR pyranometer as the reference, a series of linear regression fits were performed to reduce the calibration error, and the resulting regression equations in Table 1 would be used into the practical field calibrations. After corrections, the discrepancy among instruments is trivial, and the root mean squared deviation (RMSD) all are <8 W/m^2 at 10-min timescale.

During the matrix experiment, all measurements were carried out carefully with continuous supervision. Spirit levels and glass domes of the 17 radiometers were checked weekly to guard against any perceptible instrument tilting and possible soiling of the sensors. In spite of this, compared to other routine meteorological parameters, the measurements of surface radiation components are more prone to all kinds of errors (Moradi, 2009). Before further work was proceeded with, a series of error corrections need to be done firstly. Following the study of Vuilleumier et al. (2014), a thermal offset correction, a calibration correction, a leveling correction and a correction on soiling error were initially devised to perform in order. Adopting the method suggested by Dutton et al. (2001) the thermal offset error was first analyzed and corrected for each radiometer separately. Next, the correction on calibration error was conducted by using the linear regressions tabulated in Table 1. By checking azimuth-wise irradiance in some typical clear days (Menyhart, Anda, & Nagy, 2015), we found that most of the sites were well-leveled and horizontal except Site 9. Site 9 may have very slight tilt because the peak of interpolated azimuth-wise irradiance appeared at the azimuth of 182° but not 180°. However, due to the limitation of our observed data (direct beam radiation and diffuse radiation were not measured independently in the experiment), it is very difficult to find a valid approach to correct the leveling-induced error. As for the soiling error, in view of our cautious maintenance, this kind of error should be able to be avoided. That is, latter two kinds of corrections in practice were not performed. The recording cycle of raw SSR was 10 min, and data gaps nearly did not exist during the experiment. The corrected data at the 16 sites would be utilized for the following analyses.

2.2. Satellite retrievals

Previous SSR satellite products (Huang et al., 2016) over Heihe river basin in northwest China are used in the current study. The products were based on the look-up table algorithm of Huang, Ma, Liang, Liu,

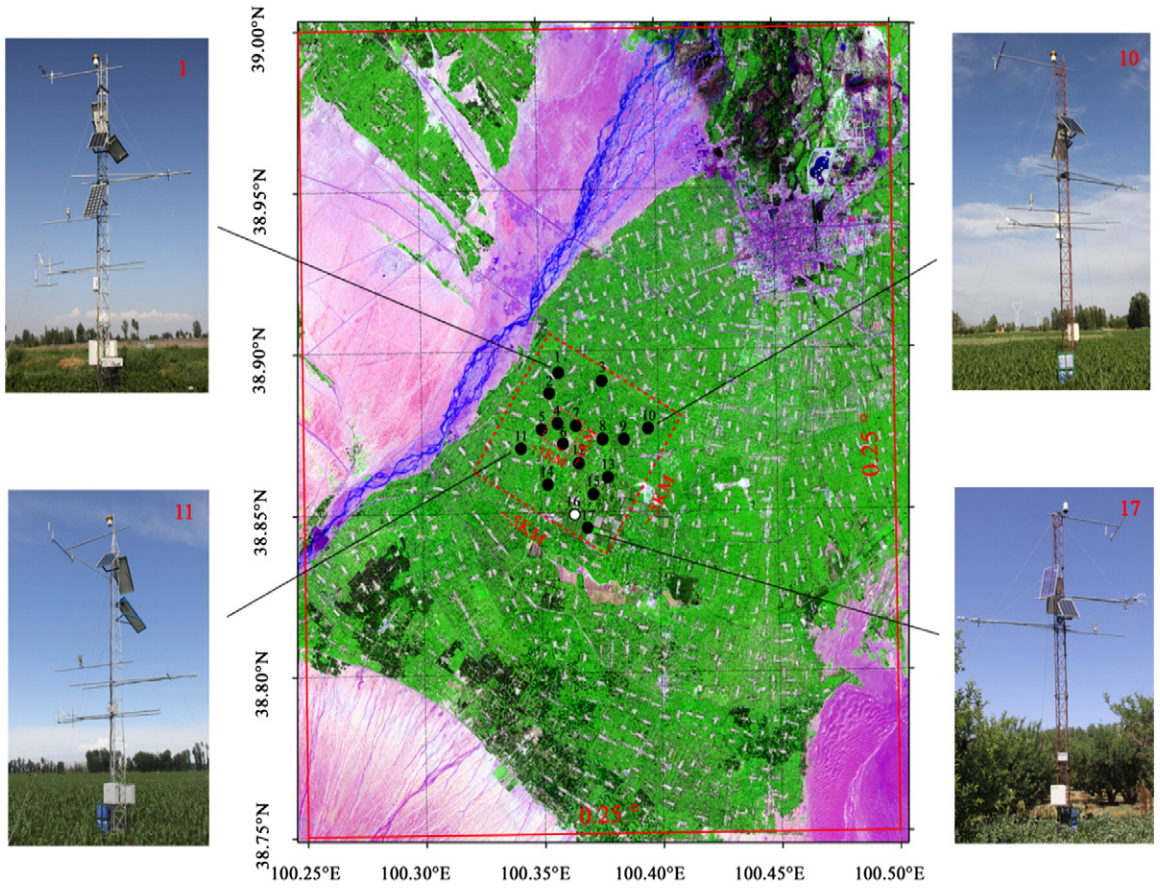

Fig. 1. Distribution of meteorological sites equipped with a radiometer in the flux-observation matrix of HiWATER-MUSOEXE.

and Li (2011), and the Japanese second generation geostationary meteorological satellite, Multifunctional Transport Satellite (MTSAT).

The original spatial resolution of the products is 0.05° (~5 km), and the temporal resolution is half hour. A previous routine validation demonstrated for instantaneous and hourly SSR a root mean squared deviation of 93 W/m^2 and 88 W/m^2, respectively; while for daily SSR it further dropped to 34 W/m^2. All linear correlations between measurements and products were higher than 0.9, and mean absolute deviations were controlled within 18%. Such high resolution and accuracy can be used to characterize the variability of SSR caused by clouds on the grid scale. The graticules in Fig. 1 visually depict the actual spatial locations of the collocated product pixels.

3. Methodology

Our primary purpose is investigating the representativeness errors of point-scale ground-based SSR measurements with respect to kilometer-level and grid-level surroundings. Concretely, for the kilometer-level domain here 1 × 1 km^2 and 5 × 5 km^2 areas are

Table 1
Basic information related to surface radiation measurements in HiWATER-MUSOEXE and regression fits between PSP/PIR pyranometer and other pyranometers.

Site	Instrument	Manufacturer	Regression functions	RMSD before calibration	RMSD after calibration
NO. 1	CNR4	Kipp & Zonen	y = 1.003x − 0.432	9.74	4.58
NO. 2	CNR4	Kipp & Zonen	y = 1.001x − 0.368	5.17	4.62
NO. 3	NR01	Hukseflux	y = 0.981x + 0.922	7.81	3.67
NO. 4	CNR1	Kipp & Zonen	y = 0.980x + 1.418	7.40	2.84
NO. 5	CNR1	Kipp & Zonen	y = 0.999x + 3.654	5.08	3.74
NO. 6	CNR4	Kipp & Zonen	y = 1.012x − 5.695	7.61	6.93
NO. 7	CNR4	Kipp & Zonen	y = 0.956x − 7.954	26.77	7.63
NO. 8	CNR4	Kipp & Zonen	y = 1.002x − 3.146	5.02	4.32
NO. 9	CNR1	Kipp & Zonen	y = 0.979x − 1.583	11.47	5.75
NO. 10	CNR1	Kipp & Zonen	y = 0.987x + 0.908	6.63	4.80
NO. 11	CNR1	Kipp & Zonen	y = 1.012x − 3.113	5.24	4.66
NO. 12	CNR4	Kipp & Zonen	y = 0.999x − 1.143	4.47	4.12
NO. 13	CNR4	Kipp & Zonen	y = 1.000x − 1.151	4.07	3.60
NO. 14	CNR4	Kipp & Zonen	y = 0.997x + 0.689	3.46	3.11
NO. 15	PSP/PIR	Eppley	–	–	–
NO. 17	CNR1	Kipp & Zonen	y = 0.959x + 3.528	15.19	6.34

selected; while for the grid-level domain standard 0.25°, 0.50° and 1° equal-angle grid areas are chosen.

3.1. Representativeness error within the kilometer-level domain

In the main graphs of Fig. 1, Sites 4, 6 and 7 clustered together in a small group covering an area of ~1 × 1 km^2. Therefore, SSR measurements from these sites would be used to reveal the representativeness error within the 1 × 1 km^2 domain, and all measurements in the flux-observation matrix were exploited to draw the representativeness error within the 5 × 5 km^2 domain.

First, "areal SSR" on the kilometer-scale was estimated by a simple multi-point averaging method. That is, the "areal SSR" over the 1 × 1 km^2 and the 5 × 5 km^2 domains were separately estimated by averaging the measurements from Sites 4, 6 and 7, and all sites except Site 16. Second, measured SSR from each individual site were compared with the estimates of areal SSR, and differences between them were deemed the representativeness error from point to kilometer scales. Since the means of all sites are seen as the estimates of areal SSR, the mean of representativeness error consequentially is zero. Namely, there is no overall bias for the representativeness error. It should be acknowledged that such a definition of representativeness error certainly contains the contamination from measurement uncertainties. But in view of the fact that our measurements had been checked and processed seriously, this contamination should be marginal (a rough estimate of it is given by the "RMSD after calibration" of Table 1). In this study, the original 10-min SSR is regarded as the "instantaneous" measurements and other timescales include 1 and 3 hours; 1, 5 and 10 days; and 1 month. Fig. 2 presents a period of typical "instantaneous" (10-min) measurements over 5 × 5 km^2 domain in July 2, 2012. In the figure, the red line indicates the estimates of areal SSR, and gray regions denote the "instantaneous" measurements from each individual site, in which the dark gray region encloses 50% of the sites. One can visualize the discrepancy between the "areal SSR" and individual SSR observed at each site. Due to strongly increasing uncertainty both in measurements and radiative transfer simulations at low solar elevations (Deneke, Feijt, & Roebeling, 2008), our study restricts solar zenith angles to a maximum of 80° for "instantaneous", hourly and 3-hour measurements.

3.2. Representativeness error within the grid-level domain

Around the flux-observation matrix of HiWATER-MUSOEXE in Fig. 1, standard grid domains of 0.25° × 0.25°, 0.50° × 0.50° and 1° × 1° (only the 0.25° × 0.25° domain is shown) were delineated to study the representativeness errors in the grid-level surroundings.

Representativeness error of point-scale measurements with respect to grid-level surrounding include two levels of errors, namely detailed error from point to kilometer scales and macroscopic error from kilometer to grid scales. Section 3.1 describes what we consider as the first type of error. Thus, here the second type of error is briefly introduced. Following the lead of Li et al. (2005) and Hakuba et al. (2013), we also attempt to address this issue by means of high resolution satellite radiation products. The mean of satellite products within one grid is assumed be the surrogate of grid area SSR, and the differences between it and each pixel values are the macroscopic representativeness error from kilometer to grid scales. Accordingly, the macroscopic representativeness errors for the 0.25° × 0.25°, 0.50° × 0.50° and 1° × 1° grid domains were separately calculated using the satellite products of 0.05° resolution produced by us before (Huang et al., 2016). It should be noted that such approaches to determine the kilometer-to-grid representativeness error also have limitations. Our SSR products tend to be smoothed because the coarser aerosol products are used in the retrieving algorithm. This eventually will lead to an underestimation on the representativeness error.

Finally, representativeness error from the point-scale to grid-scale is estimated by combining the point-to-kilometer sampling error with the kilometer-to-grid sampling error. In fact, here the kilometer scale refers to 5 km spatial scale and the underlying assumption is the representativeness error determined for the 5 × 5 km^2 area can be generalized to the rest regions within the grid. The detailed computation on the composite representativeness error from point to grid scales is discussed in Section 3.3.

3.3. Measures of representativeness error

To measure the representativeness error within a given area, a statistic, mean squared deviation (MSD) in Eq. (1) is defined firstly. In this study, the MSD is the variance of representativeness error in time-space domain, which measures the overall variability of

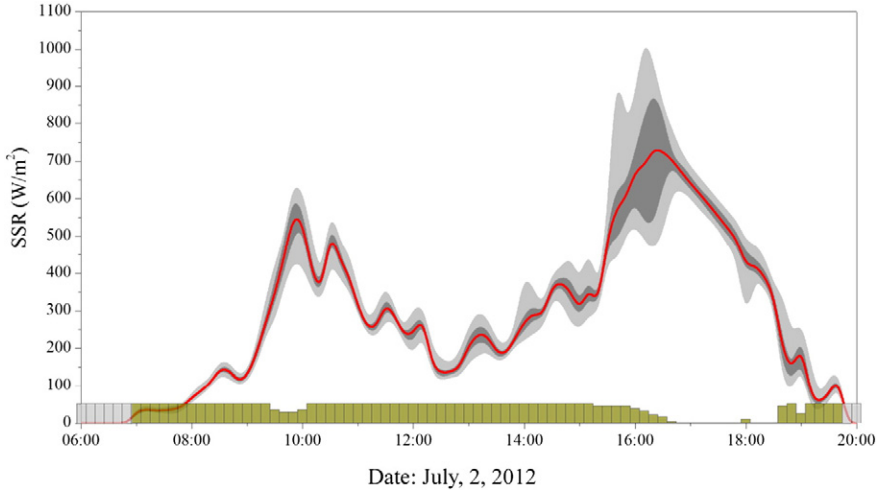

Fig. 2. "Instantaneous" (10-min) SSR measurements from 16 HiWATER-MUSOEXE sites and "instantaneous areal SSR" (red line) over the 5 × 5 km^2 domain on July 2, 2012 (the light gray region denotes fluctuations of measurements among sites, and the dark gray region encloses 50% of sites; the columns at the bottom represent cloud cover fraction (CCF), while the gray ones indicate nonexistent CCF).

representativeness error. Then, root mean squared deviation (RMSD) and relative mean deviation (RMD) are shown in Eq. (2). In nature, RMSD is the standard deviation (SD) of representativeness error in time-space domain, and describes the average dispersion degree of representativeness error, while RMD gives the spatial variability of representativeness error relative to the area mean in percent. The detailed definitions are listed below,

$$\mathrm{MSD} = \frac{1}{n \times m} \sum_{i=1}^{n} \sum_{j=1}^{m} \left(X_{i,j} - \overline{X_i}\right)^2 \tag{1}$$

$$\mathrm{RMSD} = \sqrt{MSD}, \quad \mathrm{RMD} = \frac{RMSD}{\overline{X}} \times 100 \tag{2}$$

and,

$$\overline{X_i} = \frac{1}{m} \sum_{j=1}^{m} X_{i,j} \tag{3}$$

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} \overline{X_i} \tag{4}$$

in which, $X_{i,j}$ denotes one sample or measurement in the space domain of j (total m) and the time series of i (total n), $\overline{X_i}$ is the means over a certain space domain (namely estimates of areal SSR), and $\overline{X}$is the average of all samples or measurements.

In Section 3.2, we pointed out that the representativeness error from the point-scale to the grid-scale was the combination of point-to-5km sampling error and 5km-to-grid sampling error. Assuming the two levels of sampling errors are normally distributed, due to no overall bias (means of two levels of sampling errors both are 0), the MSD of the composite representativeness error in numerical value is the arithmetic sum of their respective MSD (Deneke, Feijt, van Lammeren, & Simmer, 2005). Thus, the statistic on the point-to-grid representativeness error is given by

$$RMSD_{PtoG}^2 = RMSD_{Pto5KM}^2 + RMSD_{5KMtoG}^2 \tag{5}$$

where $RMSD_{PtoG}$,$RMSD_{Pto5KM}$and $RMSD_{5KMtoG}$ respectively are the composite RMSD, the corresponding RMSD from the point to 5-kilometer scales and the corresponding RMSD from the 5-kilometer to grid scales. Furthermore, approximately the composite RMD is equal to the percentage ratio of the composite RMSD to the mean SSR.

4. Results and discussion

As presented in Section 3, the representativeness errors of point-scale measurements with respect to kilometer-level surroundings were investigated from “instantaneous” to monthly timescales by analyzing the matrix SSR measurements from HiWATER-MUSOEXE. For the 1 × 1 km^2 domain, 3.2~21.9 W/m^2 of RMSD and 1.2~4.4% of RMD are found with the smaller value for monthly timescale and larger one for “instantaneous” timescale; whereas for the 5 × 5 km^2 domain, 3.2~40.2 W/m^2 of RMSD and 1.2~8.1% of RMD are seen, respectively. Meanwhile, representativeness errors of point-scale measurements with respect to the grid-level areas were discussed based on the combined efforts from HiWATER-MUSOEXE and the previous satellite products. Within the 0.25° × 0.25° grid, the composite RMSD and RMD respectively are 5.1~53.9 W/m^2 and 1.9~9.8% corresponding to the monthly till the instantaneous averaging intervals; within the 0.50° × 0.50° grid, the range of composite RMSD and RMD are 6.3~70.3 W/m^2 and 2.2~12.7%, respectively; while within the 1° × 1° grid, they are 8.8~91.7 W/m^2 and 3.3~16.6%, respectively.

Compared with the results of Li et al. (2005), representativeness errors obtained here are significantly larger than those reported over Southern Great Plains (SGP) in the United States. One possible reason is the high dependence of representativeness errors on cloud cover and type (Barnett, Ritchie, Foat, & Stokes, 1998; Hakuba et al., 2013), which may have different climatology in the two study areas. Another potential reason is that Li et al. (2005) only considered the variability of the high resolution satellite products for deriving the representativeness, without including the contribution from the high frequency SSR variability within satellite pixels.

4.1. Dependence on timescale

No matter on which level of spacescales, the dependences of representativeness errors on timescales are both very strong. Time-dependent representativeness errors indicate the time cycle of cloud variation.

Fig. 3 shows the representativeness errors of point-scale measurements within two kilometer-level domains. The left vertical axis represents RMSD, the right vertical axis represents RMD, and the abscissa axis is the averaging timescale. The representativeness errors rapidly decrease as the averaging timescale increases to a day. On the daily timescale, RMSD of the representativeness errors within 1 × 1 km^2 and 5 × 5 km^2 domains are reduced to about 4.3 W/m^2 and 5.1 W/m^2, respectively; the corresponding RMD are 1.7% and 2.0%. Beyond one day, they tend to level off to stable values. On timescales less than one day, representativeness errors also are strongly contingent on the specific size of domains, whereas on longer timescales representativeness errors have a convergence between the two domains. Obviously, the ability of ground-based point observation representing larger surroundings would be significantly strengthened as the timescale increases to a day or longer.

Similar features are also observed for grid-level domains. However, as described in Fig. 4, representativeness errors here are clearly greater that those within the kilometer-level domains. Especially on the instantaneous timescale, the RMSD of the representativeness error within the 1° × 1° grid is as high as 92 W/m^2, and the corresponding RMD is 16.1%. Even on the month scale, the inherent representativeness error still is close to 10 W/m^2, a magnitude that is only slightly less than the goal accuracy of global surface radiative flux (Suttles & Ohring, 1986) set by the World Climate Research Program (WCRP). It seems that for any standard 1° product it is very challenging to achieve this accuracy in a routine validation if representativeness error is not considered.

4.2. Influence of cloud cover fraction

Another evident property regarding representativeness error is that it is apparently affected by cloud cover fraction (CCF), regardless of time and space scales. Let us focus on Fig. 2 again. In the figure, the columns at the bottom denote the varying CCF with time. Before 15:20 Beijing time,

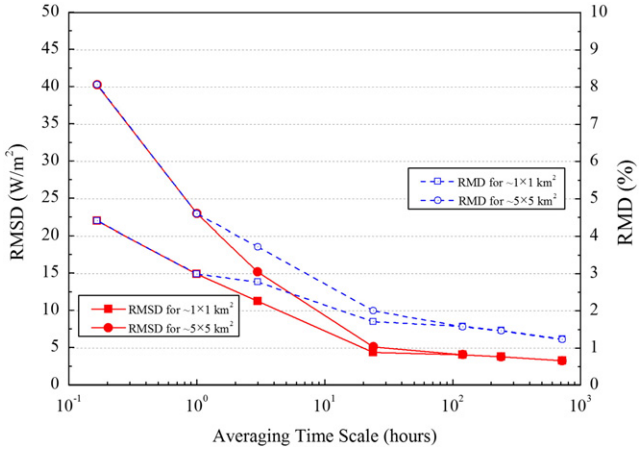

Fig. 3. Representativeness errors of point-scale SSR measurements within two kilometer-level domains (the left is the RMSD axis; the right is the RMD axis).

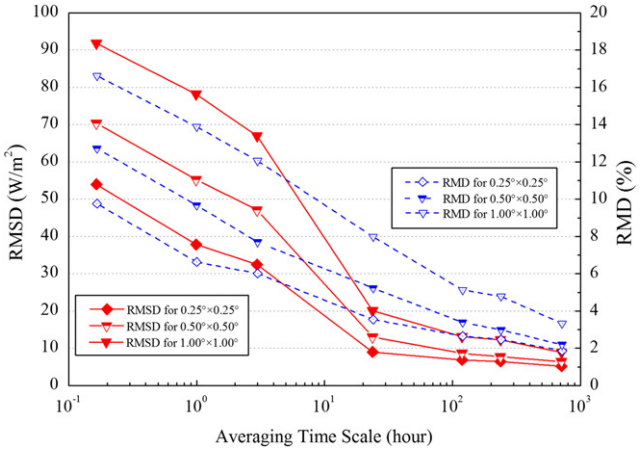

Fig. 4. Representativeness errors of point-scale SSR measurements within three grid-level domains (explanation is same as for Fig. 3).

it was substantially cloudy; then it began to gradually transition into partial cloud cover; finally after ~90 min the skies became totally clear. Drastic variability among the sites' measurements can be discerned during the transitional stage. Accordingly, the representativeness error in this period is very large.

Fig. 5 further gives the fluctuations of the representativeness errors from "instantaneous" to daily timescales with varying CCF, also using the 5 × 5 km^2 domain as an example. Here, calculations of CCF on any timescales all are based on the 10-min measurements. Using the matched aerosol observations, "instantaneous" atmospheric flux transmittance (AFT) in clear-sky are modelled firstly, and then measured site-specific AFT are compared with modelled AFT to determine whether clouds are present or not. Subsequently, all "instantaneous" atmospheric conditions at each site and in one temporal interval are judged, and CCF on one timescale is the percentage of cloudy cases in all atmospheric conditions. One phenomenon visible on Fig. 5 is that the magnitudes of representativeness errors quickly increase and then decrease with ascending CCF. Thus, when skies are partly covered by clouds, the representativeness errors are the largest. In contrast, under completely clear or cloudy conditions, representativeness errors are relatively small.

A similar feature occurs on other space and time scales, too. Fig. 6 depicts the fluctuations of hourly representativeness errors within the 0.25° × 0.25°, 0.50° × 0.50°and 1° × 1° grid domains as CCF increases. This is not a surprising phenomenon. The reason for this behavior can be simply explained by the fact that cloud are usually the main driver

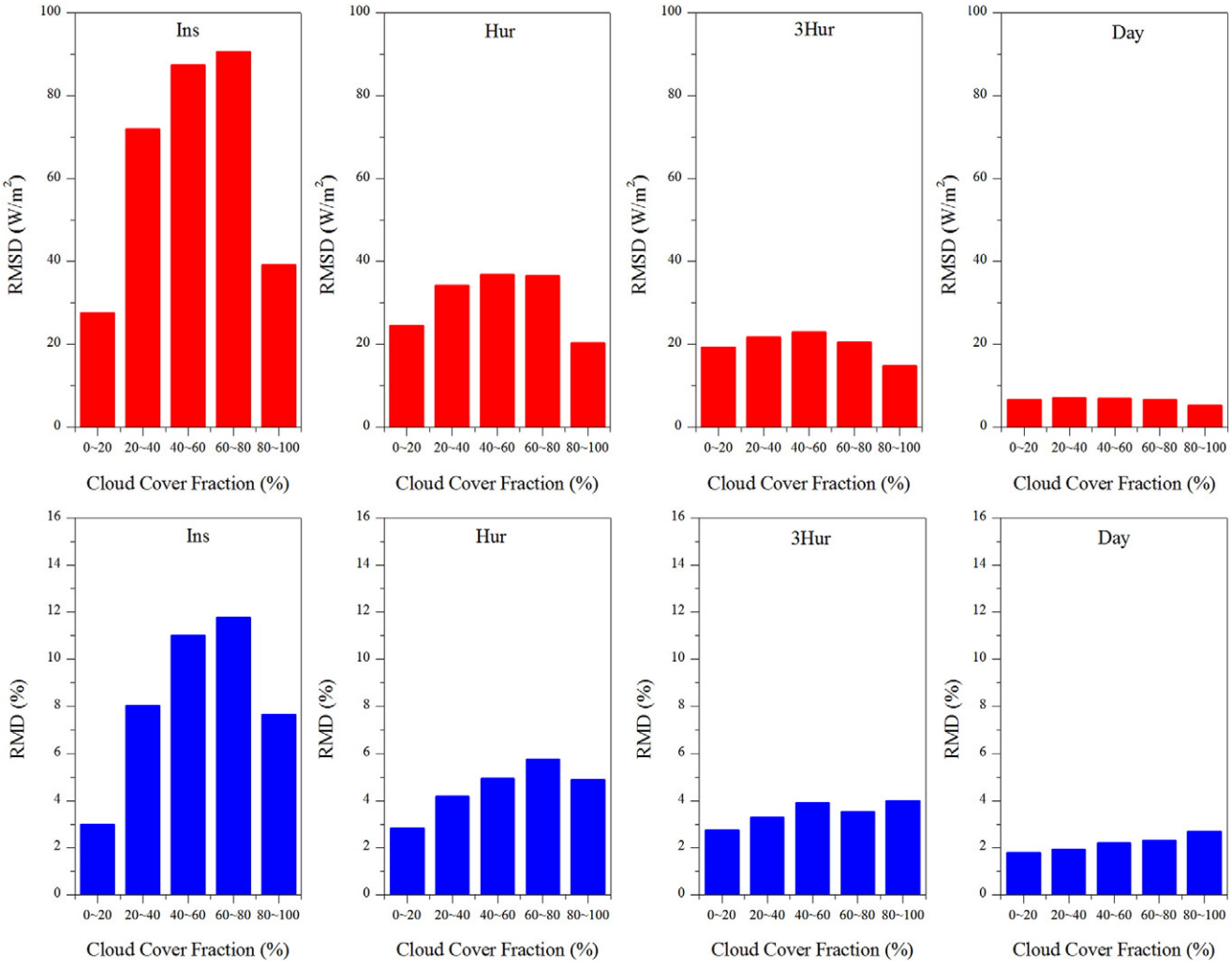

Fig. 5. Influence of cloud cover fraction (CCF) on representativeness errors within the 5 × 5 km^2 domain.

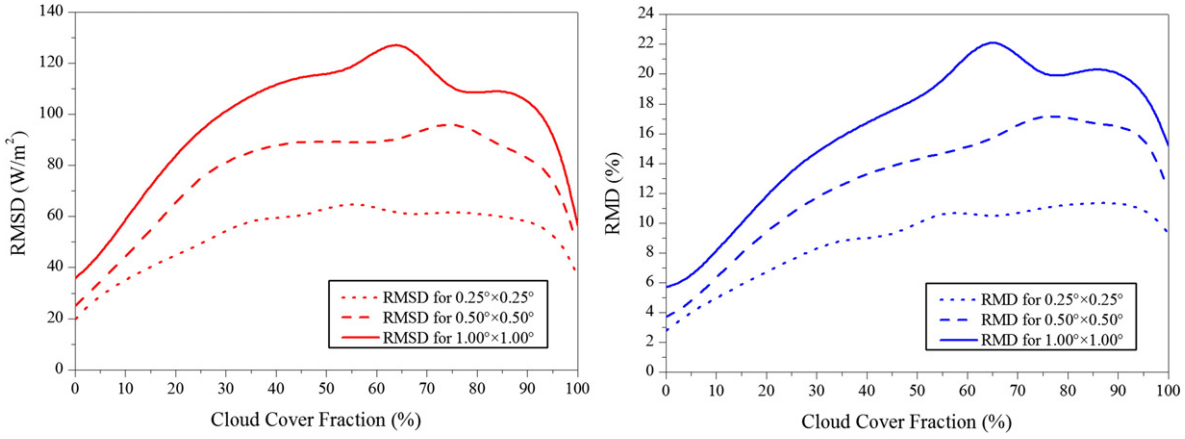

Fig. 6. Fluctuations of hourly representativeness errors within the 0.25°, 0.50° and 1° equal-angle grid domains as CCF increases.

of the radiation field variability, and the most spatially homogenous cloud cover situations are complete cloud cover or clear sky situations.

4.3. Implications for the validation of remote sensing products

Our study on representativeness error has important implications for the validation of SSR remote sensing products via a direct comparison to ground-based single site's measurements. The quality of SSR products usually is quantified by metrics such as the root mean square error (RMSE) between products and ground-based measurements. The results of this study indicate that such metrics may include a component that is not due to the product error but to the difference in their spatial scales.

For kilometer-level SSR products, on a large timescale longer than or equal to one day, representativeness error indeed could be ignored in the validation because the representativeness error is close to the instrumental observation error. On daily timescale the RMSD of representativeness error is 5.1 W/m^2 (~2%), while Michel, Philipona, Ruckstuhl, Vogt, and Vuilleumier (2008) confirmed for widely used commercial pyranometers such as Kipp & Zonen CM3, the RMSE for daily SSR from the instrumental observation error and uncertainties is as high as 5 W/m^2 (~2%). However, on a shorter timescale, especially for instantaneous products, considerable representativeness error is noted. For example, for the 5 × 5 km^2 domain, RMSD of the "instantaneous" representativeness error is as high as 40 W/m^2. So, in a practical validation, to what extent does the representativeness error affect the validation results, and what proportion of the product-measurement differences can be attributed to the inherent sampling error of radiation fields?

We attempt to answer these questions by using the surface matrix observations to actually validate the instantaneous satellite products introduced in Section 2.2 during the matrix experiment. In Fig. 1, it is noted that most sites, except Sites 11 and 17, all fall into the same pixel implied by the satellite products. Consequently, their means of multiple sites (excluding Sites 11 and 17) are used to evaluate the satellite products rather than the measurements of any single site. Fig. 7 provides the validation results of the instantaneous products which are classified into three groups by the cloud cover fraction. In the figure, the completely clear or cloudy cases are displayed at both sides, and the middle is the partly cloudy case in which the representativeness error is considered to be relatively large. Here, overall the RMSE is reduced to 87.6 W/m^2, while the mean RMSE from the validations based on each individual site is 101.2 W/m^2. Specifically, for the partly cloudy case the RMSE is effectively limited to 65.2 W/m^2 from an averaged RMSE of 86.7 W/m^2 derived from the site by site validations, while for the completely clear or cloudy cases the improvements are small. Moreover, this validation result apparently is better than any validation result which is conducted using single site measurements. For this reason, to a certain extent the product-measurement differences in routine validations indeed are caused by their differing spatial scales, and our results

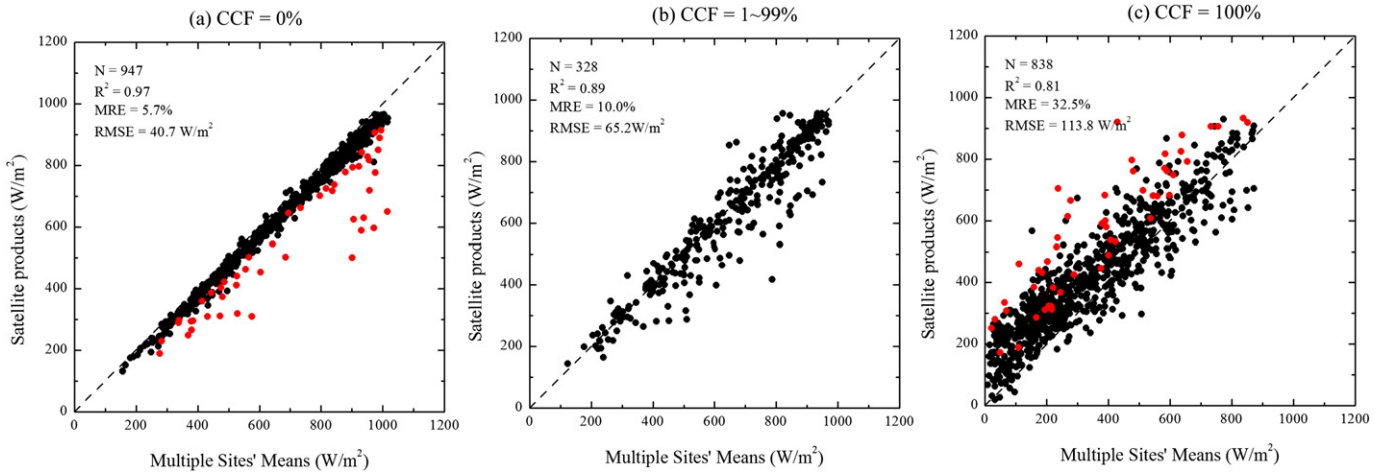

Fig. 7. Validation of instantaneous 0.05° satellite remote sensing products using multi-site measurements (validation results are classified into three groups by the cloud cover fraction: completely clear (a), partly cloudy (b) and completely cloudy (c); and the red points denote contradictory atmospheric conditions suggested by all sites' measurements and the satellite products).

indicate that for instantaneous products with a frequent resolution of 0.05°, the error due to the inadequate representativeness of point-scale measurements increases the final RMSE by approximately 13.4%. However, even though multi-site measurements are used to validate the satellite products, there is still a considerable discordance between measurements and satellite products for the cloudy case (the right in Fig. 7). It seems that the retrieval quality of satellite products still needs further refinement under cloudy skies, but in practice it is very challenging in view of the style of satellite observing and the ubiquitous cloud inhomogeneity (Deneke et al., 2005; Nunez, Fienberg, & Kuchinke, 2005). In Fig. 7, points noted with red represent the contradictions where clear skies are suggested by all sites' measurements but the satellite products suggest cloudy skies, or the opposite occurs. Obviously, these cases would contaminate the validation results heavily. One natural challenge for current high spatial resolution SSR satellite products is the cloud inhomogeneity given the fact that 3D radiative transfer effects are neglected in the retrieval scheme. Accordingly, the inherent shortage of the kilometer-level SSR satellite products also significantly contributes the product-measurement differences. In addition, impacts from representativeness errors are markedly reduced when longer timescales are assigned. For hourly products, representativeness error increases the final RMSE by ~5%. For daily products, only ~3% of increase in the final RMSE is due to the representativeness error. Incidentally, Fig. 7 also demonstrates that in routine validations we should be consciously aware of the cases that atmospheric conditions indicated by satellite products and surface measurements are distinct.

The preceding discussion reveals that in the routine validation of instantaneous kilometer-level products a significant product-measurement discrepancy (over 10%) is caused by the inadequate spatial representativeness of ground measurements. A method of enhancing spatial representativeness of ground measurements is lowering their temporal resolution (Deneke, Knap, & Simmer, 2009; Hakuba et al., 2013). Namely, inadequate spatial representation can be partly compensated with temporal averaging on point measurements (Deneke et al., 2008). Likewise, the matrix SSR measurements were used to obtain the "instantaneous areal SSR" over the 5×5 km^2 domain. Through comparing with the different temporal resolution measurements at each individual site, we investigated on which timescale individual site measurements and "instantaneous areal SSR" matched best. Table 2 gives the statistics on comparison results. We see that their discrepancy will slightly decrease at the beginning, and then rapidly increase as the temporal resolutions descend. The best matching timescale is 30 min. This finding suggests that for the validation of instantaneous kilometer-level products, most optimal timescale of ground-based measurements is 30 min instead of smaller time intervals. It can be explained because the higher frequency ground-based measurements perhaps include more detailed cloud information on sub-pixel scale, which cannot be captured by satellite pixels.

However, for the grid-level domains, the representativeness error is at a greater order of magnitude than the instrument error, regardless of timescales. For instance, for the standard grid of 1° resolution used by such as GEWEX-SRB, CEREX (Wielicki et al., 1996) etc., even on the monthly timescale, the representativeness error still is up to 3.3%, which is much larger than common instrumental observation error of ~1% (Michel et al., 2008). Thus in the validation of the grid-level SSR products, even on large timescales longer than one day, the representativeness errors between ground-based measurements and remote sensing products still cannot be ignored. Furthermore, unlike kilometer-level satellite products which are susceptible to cloud inhomogeneity and 3D radiative transfer effects, grid-level products themselves are more robust because for large domains of $>25 \times 25$ km^2, the influences from the 3D effects of clouds would be mitigated dramatically (Wyser, O'Hirok, & Gautier, 2005; Wyser, O'Hirok, Gautier, & Jones, 2002). Therefore, grid-level products are insensitive to some errors, and except for the potential accuracy problems the product-measurement differences more likely are caused by the inadequate representativeness of measurements. These facts both imply that for routine validations of grid-level products, the lack of spatial representativeness is probably a dominant factor at many sites, which results in increased product-measurement differences.

It should be said that the inadequate spatial representation of point observations within a larger grid cell had been noticed early (Li, Whitlock, & Charlock, 1995; Zhang et al., 2013). Li et al. (2005) suggested this question can be resolved by the use of multiple sites to approximate a larger grid cell's mean and just 2–3 sites are needed. Yet, taking into account the operability, level-by-level or hierarchical validation approach seems to be more appropriate. That is, ground-based measurements are used to validate the kilometer-level SSR products, and then the validated products are further exploited to evaluate grid-level SSR products.

Table 2
Comparison results of "instantaneous areal SSR" and different temporal resolution individual site's measurements over the 5×5 km^2 domain.

Statistics	10 min (Ins)	30 min	50 min	70 min	90 min
RMD (%)	8.76	8.07	10.34	11.23	12.12
RMSD (W/m^2)	43.24	39.51	50.25	55.84	61.23

5. Summary

Ground-based solar radiation measurements are frequently used for directly validating satellite-derived surface solar radiation (SSR) products and even correcting any biases in the remote sensing products (Zhang, Liang, Wild, & Jiang, 2015). Nevertheless, we realize that their spatial representativeness is per se different, and differences of observations and products not only are caused by the product-generating algorithms themselves but also by poor representativeness of surface sites.

In this paper, we first investigate the inadequate spatial representativeness of ground-based solar radiation measurements and quantify the representativeness errors of point measurements with respect to larger surroundings. Subsequently, features on representativeness errors are extracted, and finally their influences on the routine validation of remote sensing products are discussed. Our study demonstrates that wildly fluctuating representativeness errors indeed exist in the routine validation of remote sensing products, which strongly depend on their time and space scales as well as instant atmospheric conditions. For the most common 5×5 km^2 domain, the RMSD of the "instantaneous" representativeness error is about 40 W/m^2 (~8.1% of the mean SSR); whereas for the standard $1° \times 1°$ grid domain, the RMSD of the instantaneous representativeness error is as great as 91 W/m^2 (~16.1% of the mean SSR). Such scale-dependent representativeness errors conversely provide critical constraints on the direct comparisons of surface measurements and satellite products, and thus are valuable for narrowing down the uncertainty range between them.

For a kilometer-level SSR product, if its timescale is longer than or even equal to one day, in validation the error originating from their different spatial footprints can be completely neglected. However, if its timescale is less than one day, representativeness error has a certain influence on the validation result, especially for instantaneous products. Our practical validation via utilizing the high-density SSR matrix observations further points out that on the instantaneous and 0.05° time-space scales the error due to the inadequate representativeness of point-scale measurements increases the final RMSE by approximately 13.4%. From the perspective of reducing the representativeness error, for the similar time-space resolution products, the best matching timescale of surface single site's measurements is 30 min instead of shorter time intervals. This is similar to the conclusion of 40 min drawn by Deneke et al. (2009). In addition, in routine validations, the cases for which distinct atmospheric conditions are indicated by satellite products and surface measurements, should be particularly noted and even excluded.

Different conclusions are drawn for grid-level products. Though representativeness errors similarly rapidly decrease with increasing timescales, for the standard 1° grid even on the monthly timescale representativeness error still is much larger than the instrument error. As spatial coverage increases, a single observation site really increasingly can not represent the areal means of the grid, even if the site is located at the center of the grid. Here representativeness error probably has a great influence on routine validation results because compared with kilometer-level products grid-level products are insensitive to some errors (e.g., the error from cloud inhomogeneity and 3D radiative transfer effects). Therefore, for grid-level products hierarchical validation approach seems to be more desirable.

Moreover, we note that representativeness errors of point-scale ground-based solar radiation measurements are also related to the local climate and weather regimes, especially cloud climatology. Thus, representativeness of surface measurements is region-specific. Nonetheless, our study still offers some in-depth insights into how representativeness error influences a routine validation, and some conclusions may be used to guide the actual validation work. Our discussion on representativeness errors will promote a better assessment for different types of remote sensing products using ground-based observations, and thereby is also beneficial to the further improvements of the quality of satellite-derived products.

Acknowledgements

The authors would like to thank the investigators of HiWATER-MUSOEXE for their quality data and anonymous reviewers for their valuable comments. This work was jointly supported by the Chinese Academy of Sciences Action Plan for West Development Program Project – "Remote Sensing Data Products in the Heihe River Basin: Algorithm Development, Data Products Generation and Application Experiments" (grant: KZCX2-XB3-15), the Opening Fund of Key Laboratory for Land Surface Process and Climate Change in Cold and Arid Regions (grant: LPCC201302), the National Natural Science Foundation of China (grant: 41571358), the interdisciplinary Innovation Team of the Chinese Academy of Science, and the One Hundred Person Project of the Chinese Academy of Sciences (grant: 29Y127D01).

References

Barnett, T. P., Ritchie, J., Foat, J., & Stokes, G. (1998). On the space-time scales of the surface solar radiation field. *Journal of Climate*, *11*, 88–96.

Deneke, H. M., Feijt, A. J., & Roebeling, R. A. (2008). Estimating surface solar irradiance from METEOSAT SEVIRI-derived cloud properties. *Remote Sensing of Environment*, *112*, 3131–3141.

Deneke, H., Feijt, A., van Lammeren, A., & Simmer, C. (2005). Validation of a physical retrieval scheme of solar surface irradiances from narrowband satellite radiances. *Journal of Applied Meteorology*, *44*, 1453–1466.

Deneke, H. M., Knap, W. H., & Simmer, C. (2009). Multiresolution analysis of the temporal variance and correlation of transmittance and reflectance of an atmospheric column. *Journal of Geophysical Research-Atmospheres*, *114*.

Dutton, E. G., Michalsky, J. J., Stoffel, T., Forgan, B. W., Hickey, J., Nelson, D. W., ... Reda, I. (2001). Measurement of broadband diffuse solar irradiance using current commercial instrumentation with a correction for thermal offset errors. *Journal of Atmospheric and Oceanic Technology*, *18*, 297–314.

Forman, B. A., & Margulis, S. A. (2009). High-resolution satellite-based cloud-coupled estimates of total downwelling surface radiation for hydrologic modelling applications. *Hydrology and Earth System Sciences*, *13*, 969–986.

Fritz, S., Rao, P. K., & Weinstein, M. (1964). Satellite measurements of reflected solar energy and the energy received at the ground. *Journal of the Atmospheric Sciences*, *21*, 141–151.

Hakuba, M. Z., Folini, D., Sanchez-Lorenzo, A., & Wild, M. (2013). Spatial representativeness of ground-based solar radiation measurements. *Journal of Geophysical Research-Atmospheres*, *118*, 8585–8597.

Huang, G. H., Ma, M. G., Liang, S. L., Liu, S. M., & Li, X. (2011). A LUT-based approach to estimate surface solar irradiance by combining MODIS and MTSAT data. *Journal of Geophysical Research-Atmospheres*, 116.

Huang, G. H., Wang, W. Z., Zhang, X. T., Liang, S. L., Liu, S. M., Zhao, T. B., ... Ma, Z. G. (2013). Preliminary validation of GLASS-DSSR products using surface measurements collected in arid and semi-arid regions of China. *International Journal of Digital Earth*, *6*, 50–68.

Huang, G. H., Li, X., Ma, M. G., & Li, H. Y. (2016). High resolution surface radiation products for studies of regional energy, hydrologic and ecological processes over Heihe river basin, northwest China. *Agricultural and Forest Meteorology*. http://dx.doi.org/10.1016/j.agrformet.2016.04.007 in press.

Li, X. (2014). Characterization, controlling, and reduction of uncertainties in the modeling and observation of land-surface systems. *Science China-Earth Sciences*, *57*, 80–87.

Li, X., Cheng, G. D., Liu, S. M., Xiao, Q., Ma, M. G., Jin, R., ... Xu, Z. W. (2013). Heihe Watershed Allied Telemetry Experimental Research (HiWATER): Scientific objectives and experimental design. *Bulletin of the American Meteorological Society*, *94*, 1145–1160.

Li, Z. Q., Cribb, M. C., & Chang, F. L. (2005). Natural variability and sampling errors in solar radiation measurements for model validation over the Atmospheric Radiation Measurement Southern Great Plains region. *Journal of Geophysical Research-Atmospheres*, *110*.

Li, Z. Q., Whitlock, C. H., & Charlock, T. P. (1995). Assessment of the monthly mean surface insolation estimated from satellite measurements using global energy-balance archive data (vol 8, pg 315, 1995). *Journal of Climate*, *8*, 1245.

Liang, S. L., Wang, K. C., Zhang, X. T., & Wild, M. (2010). Review on estimation of land surface radiation and energy budgets from ground measurement, remote sensing and model simulations. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *3*, 225–240.

Liu, S. M., Xu, Z. W., Wang, W. Z., Jia, Z. Z., Zhu, M. J., Bai, J., & Wang, J. M. (2011). A comparison of eddy-covariance and large aperture scintillometer measurements with respect to the energy balance closure problem. *Hydrology and Earth System Sciences*, *15*, 1291–1306.

Menyhart, L., Anda, A., & Nagy, Z. (2015). A new method for checking the leveling of pyranometers. *Solar Energy*, *120*, 25–34.

Michel, D., Philipona, R., Ruckstuhl, C., Vogt, R., & Vuilleumier, L. (2008). Performance and uncertainty of CNR1 net radiometers during a one-year field comparison. *Journal of Atmospheric and Oceanic Technology*, *25*, 442–451.

Moradi, I. (2009). Quality control of global solar radiation using sunshine duration hours. *Energy*, *34*, 1–6.

Nunez, M., Fienberg, K., & Kuchinke, C. (2005). Temporal structure of the solar radiation field in cloudy conditions: Are retrievals of hourly averages from space possible? *Journal of Applied Meteorology*, *44*, 167–178.

Pinker, R. T., & Laszlo, I. (1992). Modeling surface solar irradiance for satellite applications on a global scale. *Journal of Applied Meteorology*, *31*, 194–211.

Pinker, R. T., Tarpley, J. D., Laszlo, I., Mitchell, K. E., Houser, P. R., Wood, E. F., ... Higgins, R. W. (2003). Surface radiation budgets in support of the GEWEX Continental-Scale International Project (GCIP) and the GEWEX Americas Prediction Project (GAPP), including the North American Land Data Assimilation System (NLDAS) Project. *Journal of Geophysical Research-Atmospheres*, *108*.

Posselt, R., Mueller, R. W., Stockli, R., & Trentmann, J. (2012). Remote sensing of solar surface radiation for climate monitoring - the CM-SAF retrieval in international comparison. *Remote Sensing of Environment*, *118*, 186–198.

Sanchez-Lorenzo, A., Wild, M., & Trentmann, J. (2013). Validation and stability assessment of the monthly mean CM SAF surface solar radiation dataset over Europe against a homogenized surface dataset (1983-2005). *Remote Sensing of Environment*, *134*, 355–366.

Suttles, J., & Ohring, G. (1986). Report of the workshop on surface radiation budget for climate applications. *WCRP WC-119, World Meteorological Organization Tech. Doc. WCRP WC-119, World Meteorological Organization Tech. Doc.,WMO/TD 109*.

Vuilleumier, L., Hauser, M., Felix, C., Vignola, F., Blanc, P., Kazantzidis, A., & Calpini, B. (2014). Accuracy of ground surface broadband shortwave radiation monitoring. *Journal of Geophysical Research-Atmospheres*, *119*, 13838–13860.

Wielicki, B. A., Barkstrom, B. R., Harrison, E. F., Lee, R. B., Smith, G. L., & Cooper, J. E. (1996). Clouds and the earth's radiant energy system (CERES): An earth observing system experiment. *Bulletin of the American Meteorological Society*, *77*, 853–868.

Wyser, K., O'Hirok, W., & Gautier, C. (2005). A simple method for removing 3-D radiative effects in satellite retrievals of surface irradiance. *Remote Sensing of Environment*, *94*, 335–342.

Wyser, K., O'Hirok, W., Gautier, C., & Jones, C. (2002). Remote sensing of surface solar irradiance with corrections for 3-D cloud effects. *Remote Sensing of Environment*, *80*, 272–284.

Xu, Z. W., Liu, S. M., Li, X., Shi, S. J., Wang, J. M., Zhu, Z. L., ... Ma, M. G. (2013). Intercomparison of surface energy flux measurement systems used during the HiWATER-MUSOEXE. *Journal of Geophysical Research-Atmospheres*, *118*, 13140–13157.

Yang, K., Koike, T., Stackhouse, P., Mikovitz, C., & Cox, S. J. (2006). An assessment of satellite surface radiation products for highlands with Tibet instrumental data. *Geophysical Research Letters*, *33*.

Yang, K., Pinker, R. T., Ma, Y., Koike, T., Wonsick, M. M., Cox, S. J., ... Stackhouse, P. (2008). Evaluation of satellite estimates of downward shortwave radiation over the Tibetan Plateau. *Journal of Geophysical Research-Atmospheres*, *113*.

Zhang, X. T., Liang, S. L., Wild, M., & Jiang, B. (2015). Analysis of surface incident shortwave radiation from four satellite products. *Remote Sensing of Environment*, *165*, 186–202.

Zhang, X. T., Liang, S. L., Zhou, G. Q., Wu, H. R., & Zhao, X. (2014). Generating Global LAnd Surface Satellite incident shortwave radiation and photosynthetically active radiation products from multiple satellite data. *Remote Sensing of Environment*, *152*, 318–332.

Zhang, Y. C., Rossow, W. B., Lacis, A. A., Oinas, V., & Mishchenko, M. I. (2004). Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global data sets: Refinements of the radiative transfer model and the input data. *Journal of Geophysical Research-Atmospheres*, 109.

Zhang, T. P., Stackhouse, P. W., Gupta, S. K., Cox, S. J., Mikovitz, J. C., & Hinkelman, L. M. (2013). The validation of the GEWEX SRB surface shortwave flux data products using BSRN measurements: A systematic quality control, production and application approach. *Journal of Quantitative Spectroscopy & Radiative Transfer*, *122*, 127–140.