An Approach for Evaluating Multisite Radiometry Calibration of Sentinel-2B/MSI Using RadCalNet Sites

Caixia Gao, Yaokai Liu, Zhifeng Wu, Lingling Ma, Shi Qiu, Chuanrong Li, Yongguang Zhao, Qijin Han, Enyu Zhao, Yonggang Qian, and Ning Wang

***Abstract*—The Sentinel-2B offers high spatial resolution optical imagery from 13 bands in the visible, near infrared, and shortwave infrared range of the electromagnetic spectrum, providing enhanced continuity to monitoring global terrestrial surfaces and coastal waters. In this article, its radiometry calibration is evaluated using 37 clear-sky observations in 2018 from the four sites of radiometric calibration network for assuring data quality. However, since the single calibration results acquired under different surface and atmospheric conditions have different biases and different uncertainties, it is difficult to determine which calibration sample is much more trustable. In view of this, by assuming that the calibration samples are independent of each other, a state-of-the-art reference value is derived by combining 37 calibration samples using a weighted average method, which has much lower uncertainty and approaches the "true" value. The reference value also could be used to compare each calibration result. The result shows that the reference value of the relative difference between the simulated and observed top-of-atmosphere reflectance is 4.29%, 4.95%, 4.54%, 5.34%, respectively, for bands 2, 3, 4, and 8, and the corresponding uncertainty is 1.09%, 1.10%, 1.10%, and 1.12%, respectively; the degree of equivalence for each sample is calculated by comparing each calibration result with the reference value. It is worth noting that the degrees of equivalence are lower than 5%, and the four samples on July 9, July 13, October 4, and October 11 perform worse than the other sample.**

***Index Terms*—Field calibration, Monte Carlo, radiometric calibration network (RadCalNet), Sentinel-2B/multispectral instrument (MSI), uncertainty analysis.**

Manuscript received April 29, 2021; revised June 24, 2021; accepted July 17, 2021. Date of publication August 10, 2021; date of current version September 7, 2021. This work was supported in part by the National Key R&D Program of China under Grants 2018YFB0504800 and 2018YFB0504804, in part by the Key Project Aerospace Information Research Institute of Chinese Academy of Sciences under Grant E0Z206030F, and in part by the Bureau of International Co-Operation Chinese Academy of Sciences under Grant 181811KYSB20160040. *(Corresponding authors: Caixia Gao; Lingling Ma.)*

Caixia Gao, Yaokai Liu, Lingling Ma, Shi Qiu, Chuanrong Li, Yongguang Zhao, Yonggang Qian, and Ning Wang are with the Key Laboratory of Quantitative Remote Sensing Information Technology, Academy of Opto-Electronics, Chinese Academy of Sciences, Beijing 100094, China (e-mail: gaocaixia@aoe.ac.cn; liuyk@aoe.ac.cn; llma@aoe.ac.cn; sqiu@aoe.ac.cn; lichuanrongaoe@163.com; ygzhao@aoe.ac.cn; qianyg@aoe.ac.cn; wangning@aoe.ac.cn).

Zhifeng Wu is with the National Institute of Metrology, Beijing 100029, China (e-mail: wuzf@nim.ac.cn).

Qijin Han is with the China Centre for Resources Satellite Data and Application, Beijing 100094, China (e-mail: cresda_hanjin@126.com).

Enyu Zhao is with the College of Information Science and Technology, Dalian Maritime University, Dalian 116026, China (e-mail: zhaoenyu@dlmu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2021.3102271

I. Introduction

IN-FLIGHT absolute calibration plays an increasingly important role in remote sensing applications by connecting the response output of a sensor with the real energy it receives. Due to the harsh environmental constraints imposed by the launch and in-orbit environments, the radiometric performance of remote sensors suffers from degradation inevitably in the course of its lifetime. This would not only cause the variation of radiometric performance at individual sensor level but also an overall loss of radiometric consistency between sensors [1]. In view of this, the radiometry vicarious calibration methodologies are widely used by space agencies and private operators to identify the radiometric loss in the sensor lifetime, and thereby, to recover it if/where necessary.

The vicarious calibration makes use of natural or artificial sites on the earth's surface for evaluating at-sensor radiance of remotely sensors using the simulated top-of-atmosphere (TOA) radiance at the time of sensor overpass, which is generated with *in situ* measurements and an inversion of the forward processing algorithm. To support these methodologies, a set of calibration sites, such as Railroad Valley Playa site, La Crau site, Dunhuang site, etc., have been established, and were selected by the Committee on Earth Observation Satellites (CEOS) among 49 sites for constructing a global integrated network of calibration sites for the purpose of tracking sensor performance, conducting cross-sensor comparison and assessing data quality and consistency [2]. Moreover, the CEOS Working Group on Calibration and Validation (WGCV) initiates radiometric calibration network (RadCalNet) activities to collect surface and atmospheric data necessary for the simulation of observations by earth observation optical sensors and thus verify their radiometric calibration. To increase the number of matchups between *in situ* measurements and space sensor observation and reduce the overall uncertainties, four instrumented sites, namely Baotou site, La Crau site, Railroad Valley Playa site, and Gobebaba site, are endorsed by CEOS WGCV as demonstrated standard sites. The ground-based measurements in those sites must be traced to SI standards for ensuring the traceability of the space sensor radiometry and facilitating the establishment of a quality indicator for the data products delivered by instruments.

As a part of the Copernicus programme of the European Commission, the European Space Agency has developed and

is currently operating the Sentinel-2 mission acquiring high spatial resolution (10–60 m) optical imagery from 13 bands in the visible, near infrared, and short-wave infrared (SWIR) range of the electromagnetic spectrum [3]. Nominal radiometric calibration activities of the Multispectral Instrument (MSI) on-board Sentinel-2 are based on the exploitation of the on-board sun diffuser. In fact, the accuracy provided by on-board calibration techniques, however, may not be sufficient for some applications and vicarious techniques are needed to verify the on-board device configuration and health and to achieve the required accuracy. Gascon *et al.* [3] and Lamquin *et al.* [4] validated the Sentinel-2A/MSI using a ground-reflectance-based approach over the RadCalNet sites, such as Railroad Valley Playa site, Gobabeb site, and La Crau site. For each spectral band, the equivalent TOA reflectance was calculated with a known surface reflectance and a known atmosphere, and the ratio between the observed and the simulated equivalent TOA reflectance was derived. However, due to the surface-atmosphere coupled effect, the calibration results and associated uncertainties for each sample were inconsistencies for several reasons, and these inconsistencies would introduce challenges for time series analysis due to calibration artifacts [5]. Thus, the challenge is to reconcile these independent results and provide a comprehensive estimate. To address this issue, a surface-atmosphere invariant reference value should be determined as an intercomparison bridge by combining all the calibration results from multiple sites, and the reference value has much lower uncertainty and could approximate the "true" value as much as possible.

In this article, the radiometry calibration of Sentinel-2/MSI is performed using the four instrumented sites of RadCalNet, and a reference value is derived by the weighted average method taking into account the uncertainty of each sample. The principle of the weighted average method for determining the reference value is presented in the following section, along with a description of the datasets used for the calibration. The implementation of each calibration result and the associated uncertainty are then presented, showing discrepancies among the independent samples. Subsequently, with the aid of the derived reference value, the calibration results of different samples are compared with each other.

II. Datasets

In order to collect much more calibration samples for Sentinel-2B/MSI, the four sites in RadCalNet have been used as a stable ground target for transferring the radiometric values. In this section, satellite data and the corresponding *in situ* measurements are depicted as follows.

A. Satellite Data

Sentinel-2 is a polar-orbiting mission, and it comprises two identical satellites: Sentinel-2A, launched on June 23, 2015, and Sentinel-2B, which followed on March 7, 2017 [6]. Both satellites carry the same imaging payload, MSI, which acquires measurements in 13 spectral bands ranging from the visible and near infrared to the SWIR. Four of these bands (bands 2–4, band 8) are acquired at a spatial resolution of 10 m, six bands (bands 5–7, band 8a, band 11, band 12) at 20 m, and the remaining three bands (band 1, band 9, and band 10) at 60 m. In consideration of the target size and the limitation of paper length, the measurements in bands 2–4, band 8 of Sentinel-2B/MSI are evaluated via RadCalNet four demonstrated sites. Fig. 1 shows the MSI spectral response function.

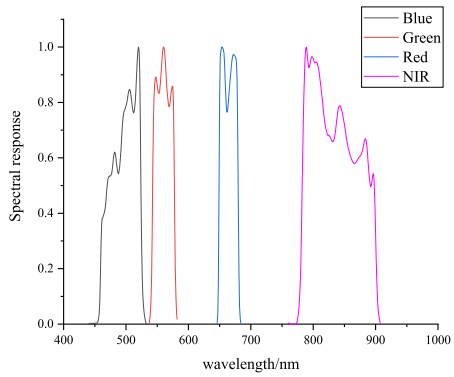

Fig. 1. Spectral response function of Sentinel-2B/MSI.

In this article, 37 scenes of Sentinel-2B/MSI images during 2018 under clear-sky conditions over the four sites are used for performing field calibration. Fig. 2 gives the Sentinel-2B/MSI images over RadCalNet four instrumented sites, and Table I and Fig. 3 give the imaging dates and the corresponding viewing and solar geometries for those 37 matchups.

B. Ground Synchronous Measurements

RadCalNet provides satellite operators with SI-traceable bottom-of-atmosphere (BOA) spectrally-resolved reflectances to aid in the postlaunch radiometric calibration and validation of optical imaging sensor data. It continuously updates nadir-looking BOA reflectances derived over a network of sites, with associated uncertainties, at a 10 nm spectral sampling interval, in the spectral range from 380 to 2500 nm and at 30-min intervals. A portal provides access to all RadCalNet datasets, allowing users to visualize and download data acquired by the four instrumented test sites.[1] The BOA reflectances for the four instrumented sites on the study dates are shown in Fig. 4.

Meanwhile, a National Aeronautics and Space Administration Aerosol Robotic Network standard Cimel CE318 sun photometer is also installed at each RadCalNet site; the aerosol optical thickness (AOT) and water vapor content (WVC) are also provided to the public for generating TOA reflectance. In this article, the AOT at 550 nm and WVC at the time of Sentinel-2B/MSI image acquisition over each site are linearly interpolated, and those matchups combining both concomitant *in situ* measurements and MSI observations, whose AOT at 550 nm are lower than 0.3 under clear-sky condition, are selected to balance the tradeoff between the uncertainty caused by AOT and sample number. The AOT and WVC for each matchup are shown in Fig. 5. It is shown that there are obvious differences in the atmospheric parameters on the study dates. The AOT at 550

[1][Online]. Available: https://www.radcalnet.org/

TABLE I
IMAGING DATE AND GEOMETRIES OF SENTINEL-2B/MSI OVER FOUR SITES DURING 2018

Site	Date	Viewing zenith angle	Viewing azimuth angle	Solar zenith angle	Solar azimuth angle
Baotou	3/25	8.85°	104.70°	42.54°	152.69°
	4/7	6.23°	283.12°	36.65°	155.19°
	5/14	8.84°	104.73°	26.13°	144.57°
	6/13	8.85°	104.70°	22.60°	137.09°
	6/26	6.23°	283.11°	21.58°	140.54°
	7/13	8.83°	104.72°	24.47°	136.25°
	9/21	8.86°	104.70°	42.49°	157.67°
	10/1	8.84°	104.70°	46.00°	160.38°
	10/4	6.23°	283.12°	46.51°	164.42°
	10/11	8.84°	104.73°	49.54°	162.59°
	10/24	6.26°	283.20°	53.58°	167.66°
	10/31	2.46°	172.60°	55.21°	166.73°
La Crau	7/9	7.06°	105.98°	25.46°	141.90°
	7/19	7.06°	105.96°	26.92°	143.02°
	8/11	7.95°	284.68°	30.99°	152.66°
	8/21	7.98	284.67°	33.82°	155.45°
	8/28	7.14°	105.77°	36.73°	153.51°
	8/31	7.97°	284.66°	36.95°	158.32°
	9/7	7.08°	105.94°	39.95°	156.54°
Railroad Valley Playa	4/14	8.34°	284.46°	31.51°	152.09°
	4/21	7.53°	104.58°	30.12°	146.47°
	6/13	8.33°	284.43°	19.44°	136.86°
	6/23	8.33°	284.41°	19.59°	135.34°
	6/30	7.60°	104.65°	21.45°	130.03°
	7/3	8.34°	284.47°	20.27°	135.15°
	7/23	8.33°	284.44°	22.98°	138.61°
	9/16	8.39°	285.38°	37.82°	158.48°
	9/26	8.39°	285.31°	41.33°	161.45°
Gobabeb	3/1	8.27°	284.36°	31.69°	63.39°
	3/11	8.29°	284.39°	33.58°	57.02°
	3/31	8.31°	284.39°	38.05°	45.64°
	4/10	8.32°	284.43°	40.54°	40.99°
	4/20	8.32°	284.39°	43.09°	37.14°
	4/30	8.29°	284.41°	45.59°	34.10°
	5/20	8.28°	284.34°	49.97°	30.33°
	9/27	8.32°	284.39°	31.48°	48.25°
	12/16	8.33°	284.41°	22.03°	93.12°

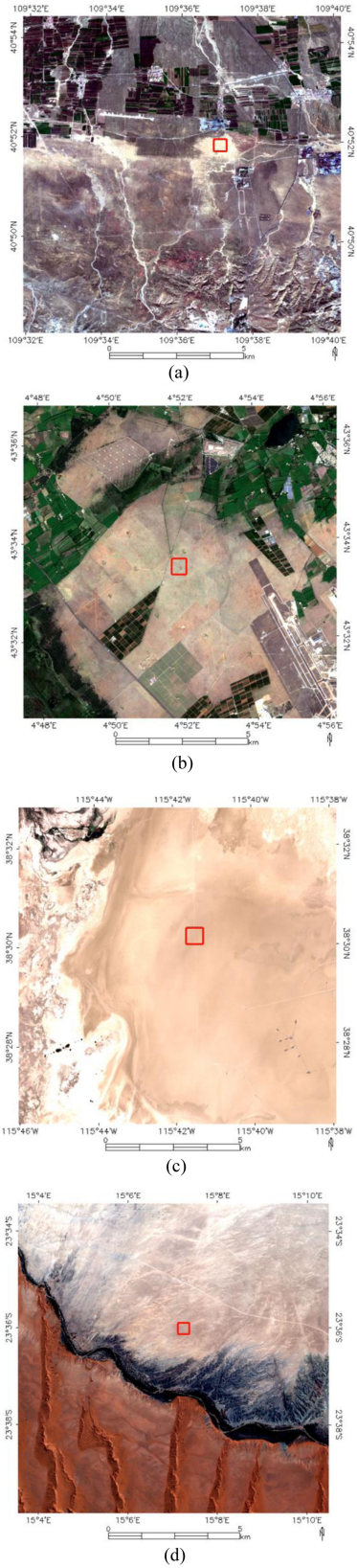

Fig. 2. Sentinel-2B/MSI images over RadCalNet four instrumented sites. (a) Baotou site (9/21). (b) La Crau site (9/7). (c) Railroad Valley Playa site (6/13). (d) Gobebaba site (9/27).

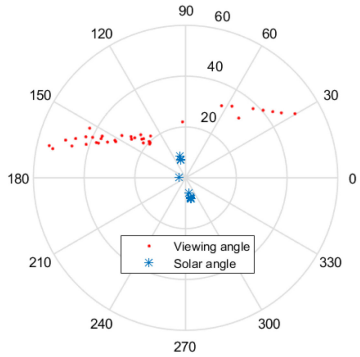

Fig. 3. Viewing and solar geometries at RadCalNet four instrumented sites of Sentinel-2B/MSI.

nm varies from 0.012 to 0.278. The WVC varies from 0.156 to 2.81 g/cm^2.

III. Methodology

This section describes the methodology used to conduct the study. Some of the processing steps are straightforward and require minimal description.

A. Field Calibration Using Single Sample at the Four Demonstrated Sites

For each matchup, with the aid of surface reflectance, atmospheric parameters (AOT and WVC), and the imaging geometries, the simulated TOA spectral reflectance $\rho_\lambda^{T,\mathrm{sim}}$ is generated using MODTRAN 5 atmosphere radiative transfer model, and then is convoluted with the spectral response function of Sentinel-2B/MSI to derive the simulated TOA reflectance in band i [$\rho_i^{T,\mathrm{sim}}$, see (1)]. Subsequently, given the spatial resolution of the sensors and the representative region of the RadCalNet BOA reflectances, 30 × 30 cloud-free pixels (∼300 m × 300 m) region of interest (ROI) is selected, and then the digital number (DN) averaged value of ROI is extracted. After that, the DN value is converted to observed TOA reflectances ($\rho_i^{T,\mathrm{obs}}$) using (2)

$$\rho_i^{T,\mathrm{sim}} = \frac{\int_{\lambda_1}^{\lambda_2} \rho_\lambda^{T,\mathrm{sim}} \times \mathrm{SRF}_\lambda}{\int_{\lambda_1}^{\lambda_2} \mathrm{SRF}_\lambda} \tag{1}$$

where SRF_λ is the relative spectral response function of the sensor, and λ_1 and λ_2 define the spectral range of band i.

$$\rho_i^{T,\mathrm{obs}} = \frac{\mathrm{DN}_i}{10000} \tag{2}$$

where DN_i is the calibrated pixel digital number in band i.

In this article, the relative difference of TOA reflectance (Δ_i) between $\rho_i^{T,\mathrm{sim}}$ and $\rho_i^{T,\mathrm{obs}}$ is used as an index to evaluate the radiometric performance of Sentinel-2B/MSI [see (3)], and a time series of TOA reflectance relative difference is generated.

$$\Delta_i = \frac{\rho_i^{T,\mathrm{sim}}}{\rho_i^{T,\mathrm{obs}}} - 1. \tag{3}$$

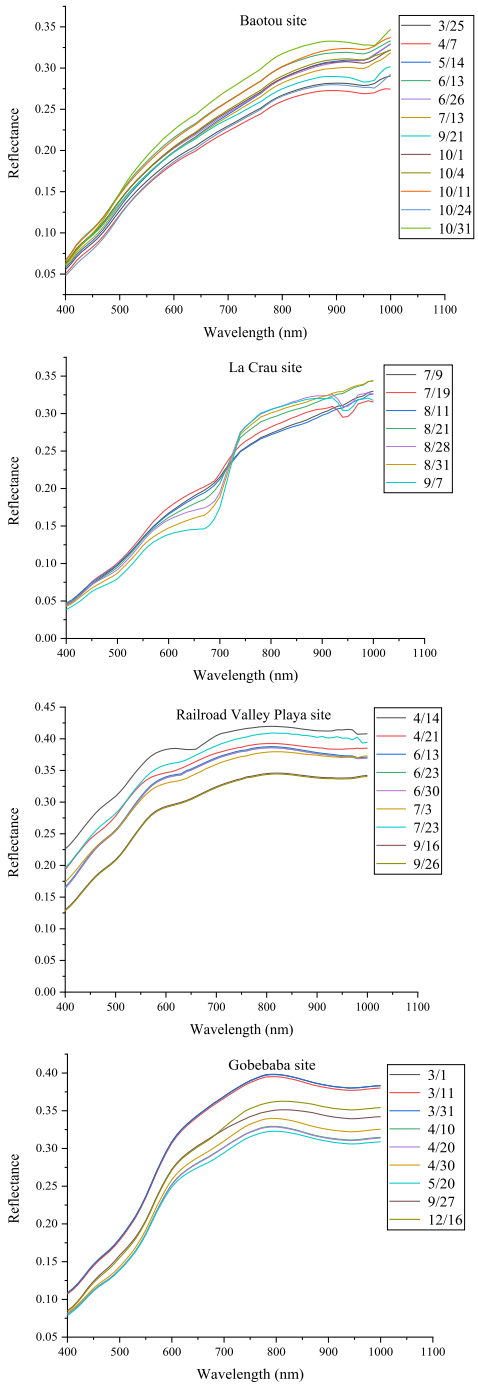

Fig. 4. Surface reflectance at RadCalNet four instrumented sites.

B. Uncertainty Analysis of TOA Reflectance Relative Difference for Each Sample

To quantitatively indicate the comparison results among the sites, an approach is used to analyze the TOA reflectance comparison uncertainties associated with surface characteristics (denoted as σ_1), AOT (denoted as σ_2), WVC (denoted as σ_3), aerosol types (denoted as σ_4), MODTRAN model (denoted as σ_5), solar irradiance (denoted as σ_6), and sensor calibration uncertainties (denoted as σ_7). Among them, the uncertainties associated with surface characteristics for the Baotou site include the uncertainty of the derived BOA surface reflectance, the spatial homogeneity, and the BRDF characteristics of the target, and the uncertainty analysis for each factor is described in detail in the study of Gao *et al.* [7], [8], whereas the uncertainties of BOA reflectance for the other three sites are extracted directly from RadCalNet data, and the BRDF characteristics of the target are also investigated as Baotou site using the Roujean model and MODIS MCD43A1 Collection 6 product [9], [10]. The uncertainties caused by the WVC and AOT are also analyzed using Monte Carlo approaches by assuming an uncertainty of 12% in WVC and 0.1 in AOT [11], [12]. The Monte Carlo simulation is set to run $M = 1000$ trials for generating error-added WVC or AOT, and error-added TOA reflectance is acquired using the MODTRAN 5 model with the other input sources; subsequently, the relative differences between the simulated TOA reflectance with WVC (or AOT) error and those without WVC (or AOT) error could be calculated, and its standard deviation is taken as the TOA reflectance uncertainty caused by WVC (or AOT) uncertainty (see Fig. 6). To analyze the uncertainty caused by the mixed aerosol type of the rural and desert aerosols, two new groups of simulated total reflectance are generated when the desert aerosol model and rural model are used, and their relative difference is calculated as the TOA reflectance uncertainty caused by aerosol type, whereas MODTRAN model, sensor calibration, and solar irradiance uncertainties are propagated independently to the TOA reflectance. Among them, the uncertainty of MODTRAN model is estimated approximately 2% [13], sensor calibration uncertainty is approximately 5% [14], [15], the Thuillier solar irradiance model has a standard uncertainty of 1.5% at 450 nm, 0.9% at 650 nm, 1.1% at 850 nm, and 0.8% at 1550 nm [16], [17], and the value at each wavelength is calculated using an interpolation method. Finally, assuming independence among the uncertainty sources, the final uncertainties for simulated TOA reflectance, $u^2(\rho_i^{T,\mathrm{sim}})$are calculated based on error transfer theory [18].

The corresponding uncertainty for Δ_i, $u(\Delta_{i,j})$ is calculated as

$$u\left(\Delta_i\right) = \sqrt{u^2\left(\rho_i^{T,\mathrm{sim}}\right) + u^2\left(\rho_i^{T,\mathrm{obs}}\right)} \tag{4}$$

where $u(\rho_i^{T,\mathrm{obs}})$ is the uncertainty of $\rho_i^{T,\mathrm{obs}}$, which is estimated approximately 5% for Sentinel-2B/MSI.

C. Evaluation Single-Sample Calibration Result Using a Comprehensive Reference Value

As aforementioned, ground-based *in situ* measurements of surface reflectance and atmospheric parameters, combined with MODTRAN 5.0 model, can provide TOA values for comparison with satellite sensor observations. The relative TOA reflectance difference for each sample is different from each other due to the coupled effects of surface reflectance and atmospheric conditions, and meanwhile, also is associated with different uncertainty. This would result in the fact that some samples have low relative differences with high uncertainties, and this phenomenon is just the opposite of other samples so that it is

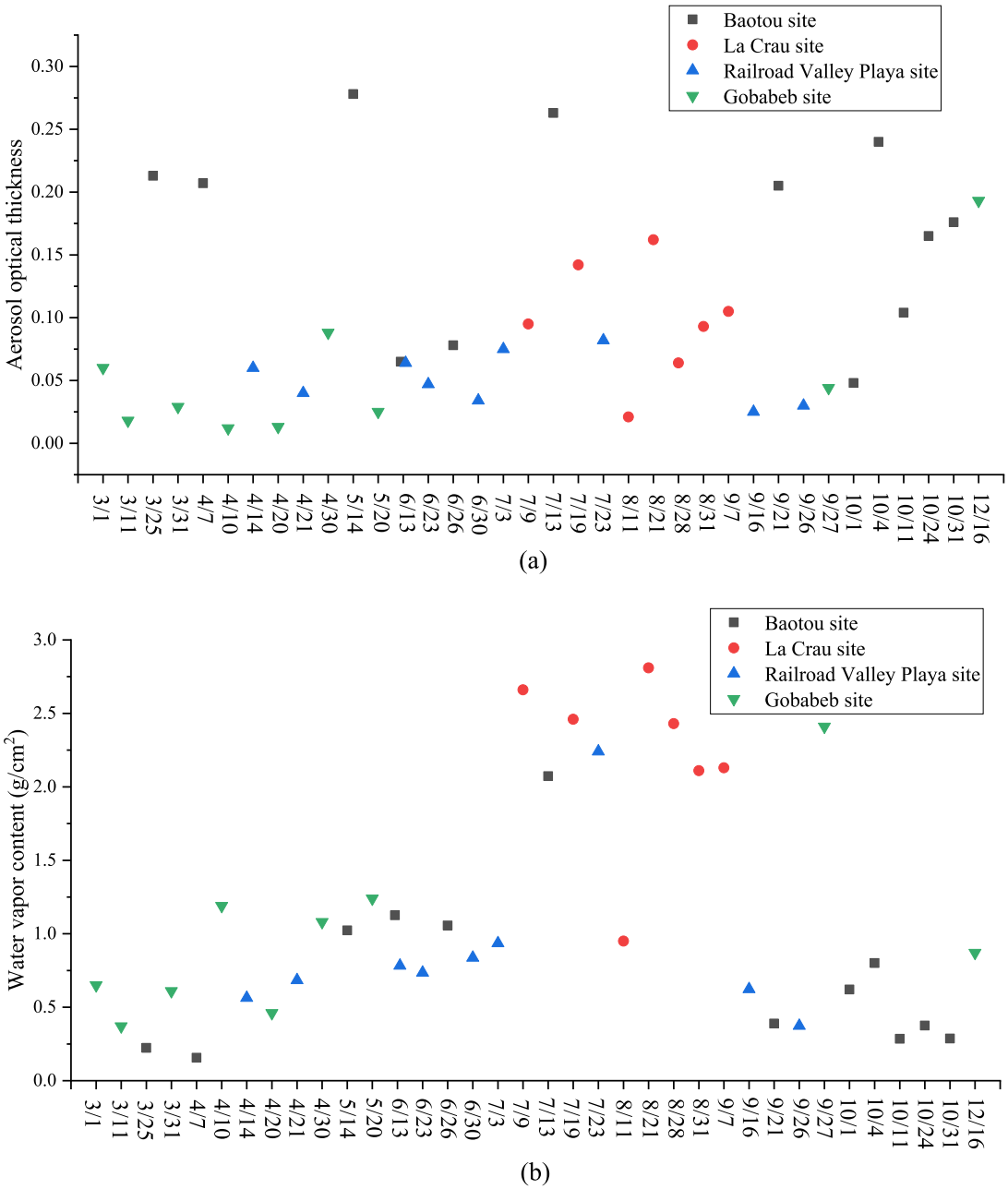

Fig. 5. Ground synchronous atmospheric measurements. (a) AOT. (b) WVC.

difficult to evaluate which one should be better and be trusted. In view of this, a state-of-the-art reference value should be derived based on each calibration result. The reference value could approximate the "true" value with sufficient samples and can be used as a bridge for evaluating each sample.

In this article, under the assumption that each sample is independent of each other, the reference value is calculated as a weighted mean by multiplying the weight since the weighted average depicts where most of the observations fall in and it is more confident than a single-sample result in the metrology field. The weight assigned to each sample is determined according to its calibration uncertainty, which refers to the relative credibility and importance of each data point. Furthermore, in order to treat each sample fairly and avoid a higher contribution made by samples with relatively low uncertainties, the cutoff values for band i, $u_{i,\mathrm{cut-off}}$, are calculated by averaging the calibration uncertainties for those samples whose calibration uncertainties are lower or equal to the median of total samples [see (5)]. And then, the uncertainty for sample j, $u(\Delta_{i,j})$, is adjusted by the cutoff uncertainty, the adjusted uncertainty for sample j in band i, $u_{\mathrm{adj}}(\Delta_{i,j})$, is equal to $u(\Delta_{i,j})$ if $u(\Delta_{i,j}) \geq u_{i,\mathrm{cut-off}}$, otherwise, $u_{\mathrm{adj}}(\Delta_{i,j})$ is equal to $u_{i,\mathrm{cut-off}}$

$$u_{i,\mathrm{cut-off}} = \mathrm{average}\{u(\Delta_{i,j})\} \text{ for } u(\Delta_{i,j}) \leq \mathrm{median}\{u(\Delta_{i,j})\}, j = 1 \text{ to } N. \quad (5)$$

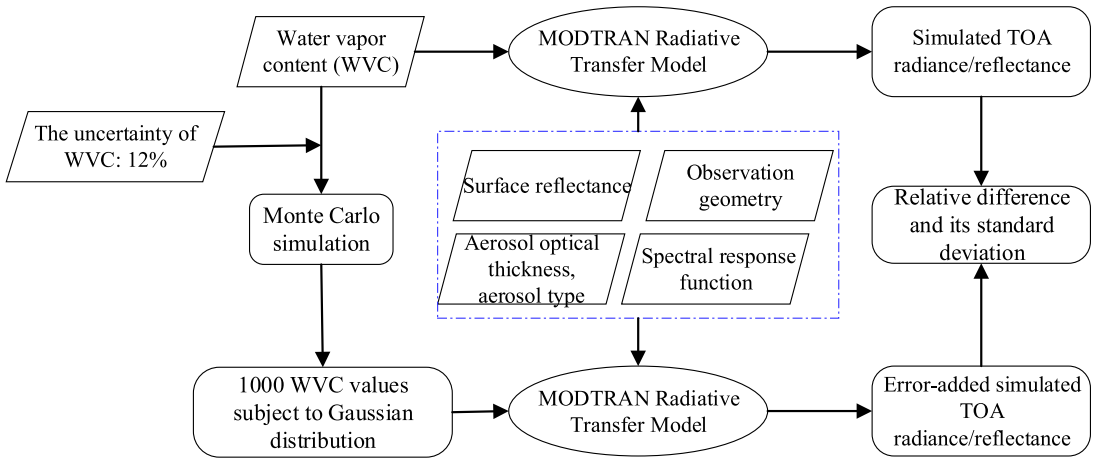

Fig. 6. Flowchart for calculating TOA reflectance uncertainty caused by WVC.

Subsequently, the weight coefficient in band *i* for each sample $w_{i,j}$ is determined by

$$w_{i,j} = \frac{u_{\text{adj}}^{-2}(\Delta_{i,j})}{\sum_{j=1}^{N} u_{\text{adj}}^{-2}(\Delta_{i,j})}. \tag{6}$$

The weighted average value of the relative difference *R* and its uncertainty $u(R)$ are determined according to the Consultative Committee for Photometry and Radiometry (CCPR) by [19]

$$R = \frac{\frac{\Delta_{i,1}}{u_{\text{adj}}^2(\Delta_{i,1})} + \cdots + \frac{\Delta_{i,N}}{u_{\text{adj}}^2(\Delta_{i,N})}}{\frac{1}{u_{\text{adj}}^2(\Delta_{i,1})} + \cdots + \frac{1}{u_{\text{adj}}^2(\Delta_{i,N})}} = \sum_{j=1}^{N} w_{i,j} \cdot \Delta_{i,j} \tag{7}$$

$$u\,(R) = \frac{1}{\sqrt{u_{\text{adj}}^{-2}(\Delta_{i,1}) + \cdots + u_{\text{adj}}^{-2}(\Delta_{i,N})}}. \tag{8}$$

After that, the degree of equivalence for each sample, d_i , is calculated to evaluate the difference between $\Delta_{i,j}$ and *R* using the following equation, and the lower the degree of equivalence is, the more accurate the sample is

$$d_i = \Delta_{i,j} - \text{KCRV}. \tag{9}$$

The uncertainty for d_i, $u(d_i)$ is calculated as

$$u(d_i) = \sqrt{u^2(\Delta_{i,j}) - u^2(\text{KCRV})}. \tag{10}$$

IV. Results

A. Single-Sample Field Calibration

Fig. 7 shows scatterplots of the simulated TOA reflectances and the corresponding sensor-measured TOA reflectances. The blue line in each figure is a standard 1:1 line. It is worth noting that the data points fall in the two sides of 1:1 line, and the TOA reflectance increases with longer wavelength; the TOA reflectance at Railroad Valley Playa site is higher than that of the other three sites, among them, La Crau site has lowest TOA reflectance. Besides, the uncertainties of TOA reflectance caused by BOA reflectance, AOT, WVC, and MODTRAN model and solar irradiance are analyzed using the aforementioned method, and are shown as example in Fig. 8. It is found that the uncertainties from BOA reflectance and MODTRAN model are two major determinants; the uncertainty of TOA reflectance tends to increase with longer wavelength and is within 5% in the spectral range of 400–900 nm.

Fig. 9 shows the relative differences between simulated TOA reflectances and the observed ones. It is noted that the relative differences differ from each other over the four sites at different study dates. In general, the relative differences are within 10%, and exhibit an "up and down" behavior over the four sites. The data points for Railroad Valley Playa site generally perform better than those for the other three sites, and are within 8%. The maximum of the relative difference is approximately 15%. Fig. 10 gives the corresponding uncertainties of those relative differences. Note that the uncertainties range from 6% to 7.4%, and are site- and date-dependent. The maximum value appears on August 31, which is caused by the larger BOA uncertainty.

B. Evaluation Results of Single-Sample Calibration

As shown in Figs. 9 and 10, the TOA relative differences are site-dependent and date-dependent due to the couple effect of atmosphere and surface characteristics, and further investigation into the apparent variation in the Sentinel-2B/MSI TOA reflectances relative difference is extremely necessary. With the aid of the overall uncertainty results of the relative differences, a weight coefficient for each data point is derived using (5) and (6) [see Fig. 11]. It is worth noting that these weight coefficients range from 0.023–0.03, and the weight coefficients have a negative correlation with the adjusted calibration uncertainties; the greater the adjusted uncertainties are, the smaller the weight coefficients are. In this process, the cutoff uncertainty agreed upon by CCPR is given as 6.37%, 6.46%, 6.52%, and 6.61% for bands 2, 3, 4, and 8, respectively, which is used to determine the adjusted calibration uncertainties. Subsequently, the reference value is derived by combining these relative differences of 37 data points. The reference value is 4.29%, 4.95%, 4.54%, and

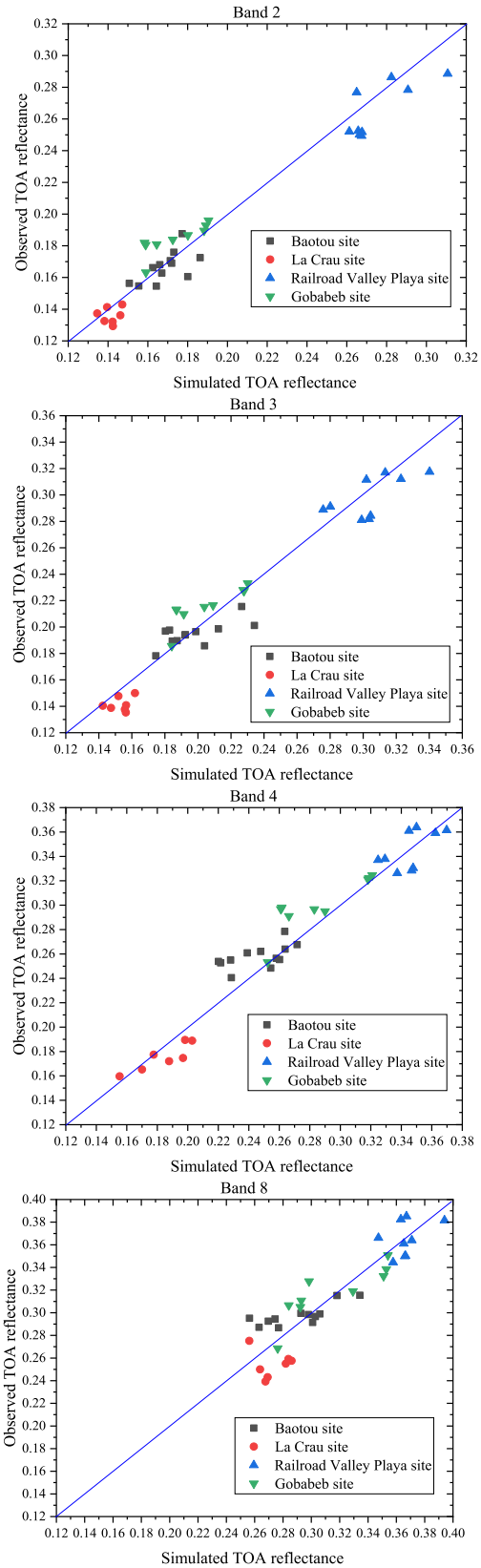

Fig. 7. Scatterplots of the simulated TOA reflectances and the corresponding sensor-measured TOA reflectances.

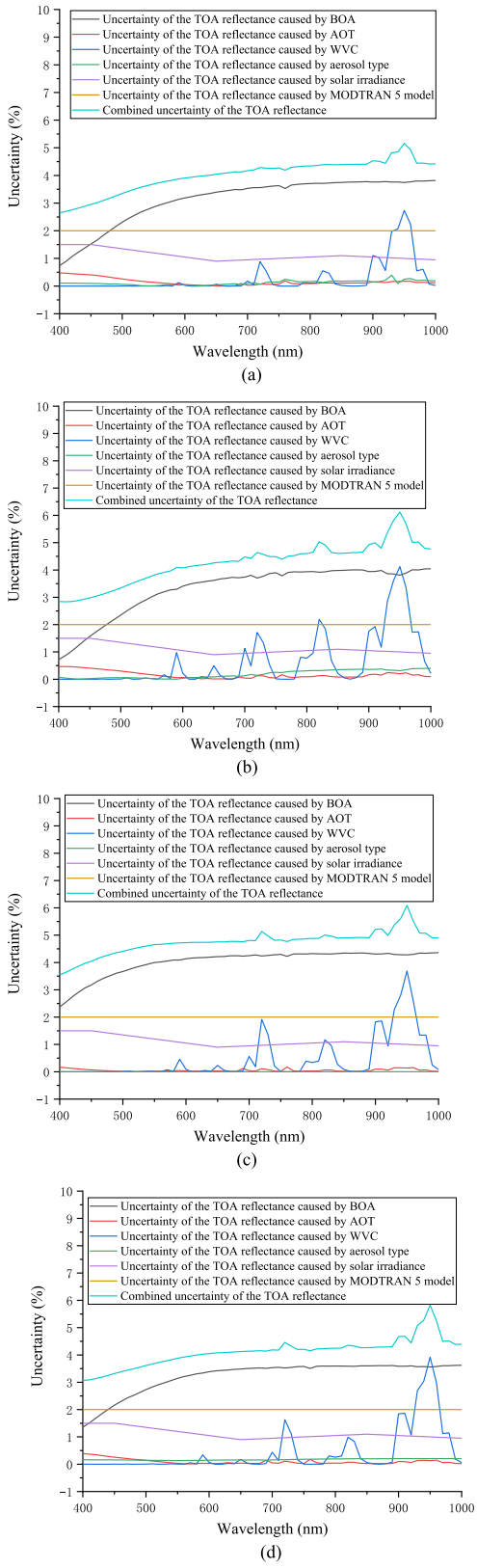

Fig. 8. Uncertainty of simulated TOA reflectances. (a) Baotou site (3/25). (b) La Crau site (7/9). (c) Railroad Valley Playa site (7/3). (d) Gobabeb site (3/1).

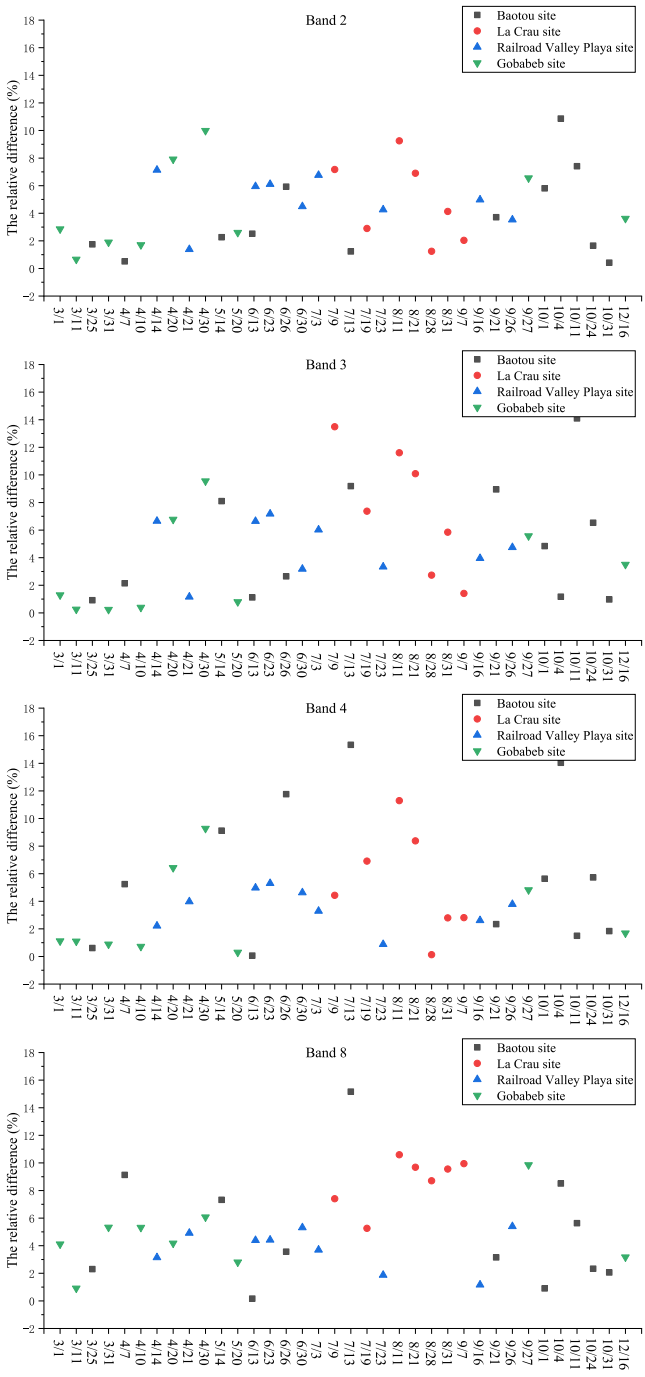

Fig. 9. Relative difference between the simulated TOA reflectances and the sensor-measured TOA reflectances.

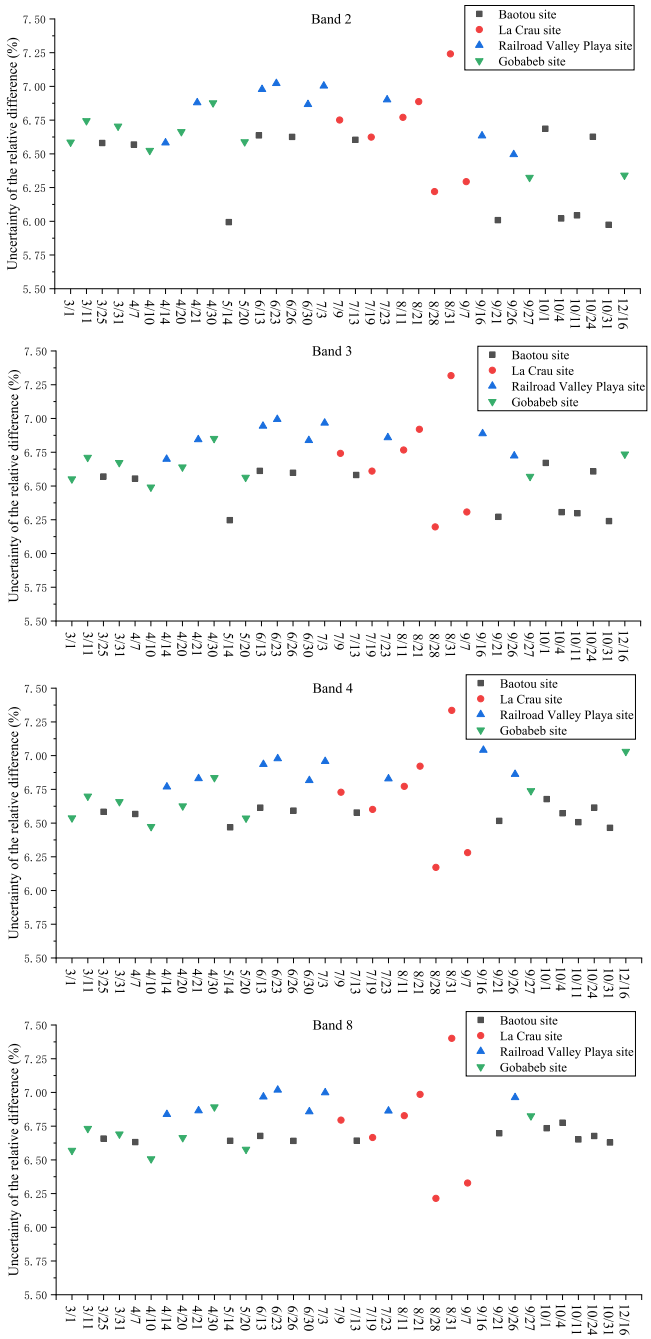

Fig. 10. Uncertainty of relative difference between the simulated TOA reflectances and the sensor-measured TOA reflectances.

5.34% for the bands 2, 3, 4, and 8, respectively, the corresponding uncertainty is also derived using (8), and is 1.09%, 1.10%, 1.10%, and 1.12%, respectively (see Table II).

Since the uncertainty of reference value seriously depends on the number of samples, we randomly select samples to analyze the influence of samples number on the uncertainty of reference value. The results are shown in Fig. 12. Note that the uncertainty of reference value decreases significantly with the increase of

TABLE II
REFERENCE VALUE AND ITS UNCERTAINTY IN BANDS 2, 3, 4, AND 8 OF SENTINEL-2B/MSI

	Band 2	Band 3	Band 4	Band 8
Reference value	4.29%	4.95%	4.54%	5.34%
Uncertainty of reference value	1.09%	1.10%	1.10%	1.12%

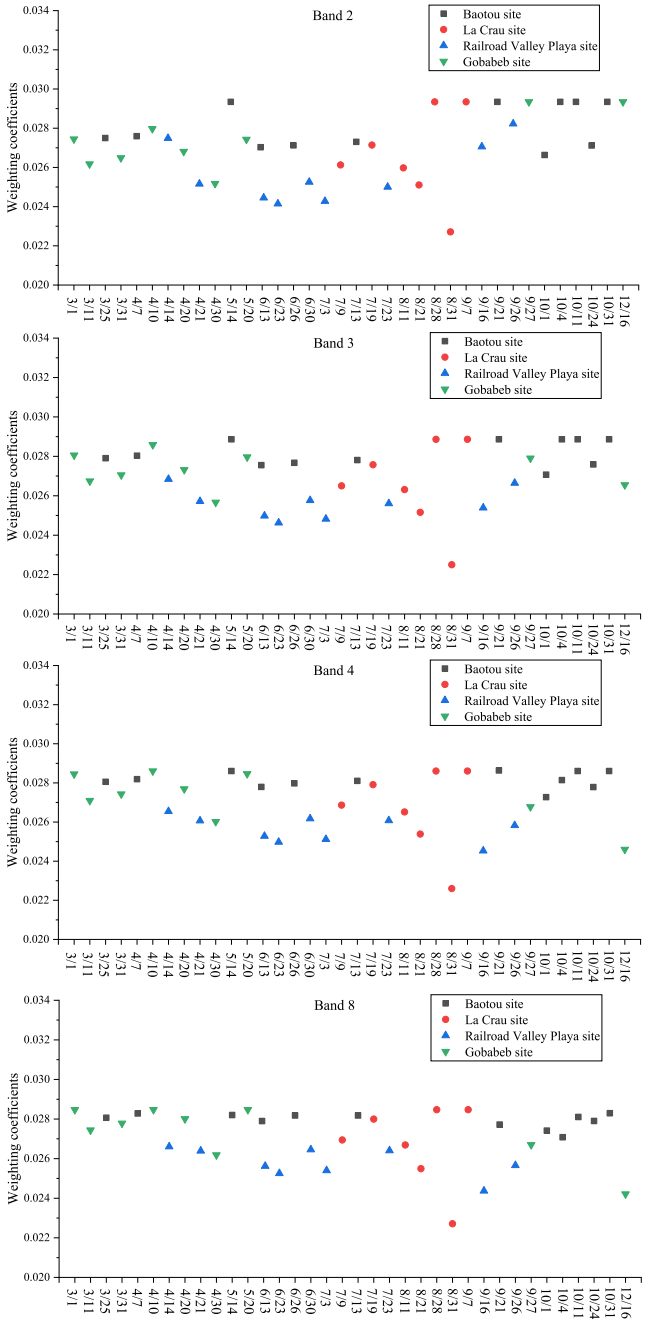

Fig. 11. Weight coefficients for those 37 samples.

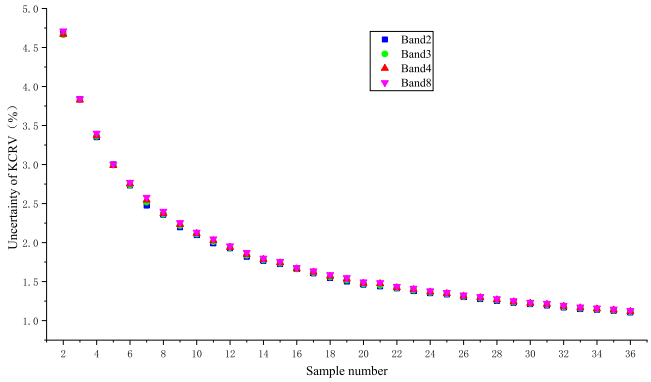

Fig. 12. Uncertainty of reference value derived with randomly selected samples.

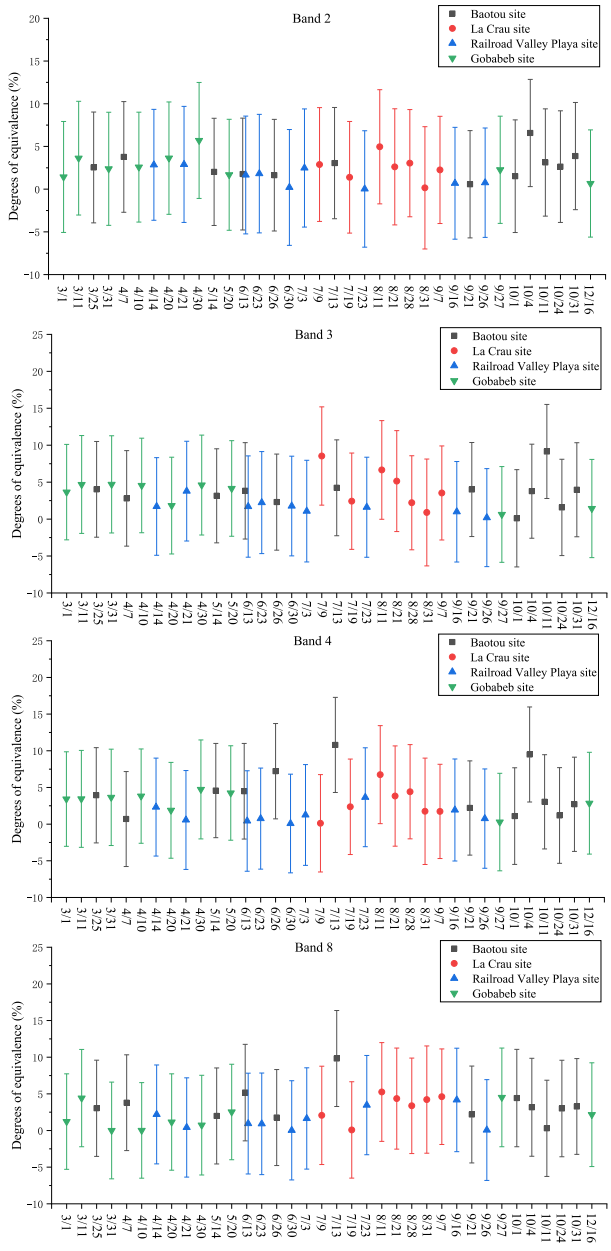

Fig. 13. Degrees of equivalence and the corresponding uncertainty.

the sample number, and its degradation becomes relatively slow when the sample number is 32.

Besides, the degrees of equivalence for the 37 samples and the corresponding uncertainties are given in Fig. 13, where the error bars represent uncertainties of the degrees of equivalence. In general, the degrees of equivalence for most samples are lower than 5%, which indicates that these samples are more accurate, whereas the four samples on July 9, July 13, October 4, and October 11 perform worse.

V. Conclusion

Field calibration is an effective way for evaluating the radiometric performance of in-orbit sensors via ground target. In order to reduce the random uncertainties from a single-sample calibration, multiple samples under different surface and atmospheric conditions are used. However, significant differences in the calibration results and the associated uncertainties for those multiple samples with different surface and atmospheric characteristics are generated, and it is difficult to determine which one is much more trustable. In view of this, a reference value, which approaches the "true" value as much as possible, should be derived.

This article evaluates the sensors' radiometric performances using multiple samples from RadCalNet four demonstrated sites, and a state-of-the-art reference value close to the corresponding "true" value is calculated by assuming those samples are independent of each other. In this article, 37 clear-sky samples over Baotou site, La Crau site, Railroad Valley Playa site, and Gobebaba site during 2018 are used, and the relative differences between simulated TOA reflectance and the observed one are derived. In general, the relative differences are within 10%, and exhibit an "up and down" behavior. Subsequently, a weight coefficient is assigned to each sample according to its calibration uncertainty so that a reference value is calculated for bands 2, 3, 4, and 8 of Sentinel-2B/MSI, respectively, using a linear weighted method. The reference value of the relative difference between the simulated and observed TOA reflectance is 4.29%, 4.95%, 4.54%, and 5.34%, respectively for bands 2, 3, 4, and 8, and the corresponding uncertainty is 1.09%, 1.10%, 1.10%, and 1.12%, respectively. The uncertainty is reduced dramatically due to the increase in sample number. After that, the relative difference of each sample is compared with the reference value, and the degrees of equivalence for each sample are calculated. It is worth noting that the degrees of equivalence are lower than 5%, and the four samples on July 9, July 13, October 4, and October 11 perform worse than the other samples.

Due to the limitation on the matchup of satellite observations and *in situ* measures, 37 validation samples during 2018 are used. In the future, many more validation samples over different surface types and synchronized ground measurements will be required by adding other test sites. Meanwhile, this method also is valuable for the validation of remote sensing products, such as surface reflectance and land surface temperature.

Acknowledgment

The authors would like to thank the European Space Agency for providing Sentinel-2B/MSI satellite data, and CEOS WGCV RadCalNet for providing *in situ* measurements, and the anonymous reviewers.

References

[1] M. Bouvet, "Radiometric comparison of multispectral imagers over a pseudo-invariant calibration site using a reference radiometric model," *Remote Sens. Environ.*, vol. 140, pp. 141–154, 2014.

[2] A. Wu, X. Xiong, and A. Angal, "Derive a MODIS-based calibration for the AVHRR reflective solar channels of the NOAA KLM operational satellites," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 3, pp. 1405–1413, Mar. 2013.

[3] F. Gascon *et al.*, "Copernicus Sentinel-2A calibration and products validation status," *Remote Sens.*, vol. 9, pp. 584–665, 2017.

[4] N. Lamquin *et al.*, "An inter-comparison exercise of Sentinel-2 radiometric validations assessed by independent expert groups," *Remote Sens. Environ.*, vol. 233, 2019, Art. no. 111369.

[5] C. Y. Cao *et al.*, "Mission-long recalibrated science quality Suomi NPP VIIRS radiometric dataset using advanced algorithms for time series studies," *Remote Sens.*, vol. 13, 2021, Art. no. 1075.

[6] R. Fernandez-Beltran, F. Pla, and A. Plaza, "Sentinel-2 and Sentinel-3 intersensor vegetation estimation via constrained topic modeling," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 10, pp. 1531–1535, Oct. 2019.

[7] C. X. Gao *et al.*, "Determination of the key comparison reference value from multiple field calibration of Sentinel-2B/MSI over the Baotou site," *Remote Sens.*, vol. 12, 2020, Art. no. 2404.

[8] L. L. Ma *et al.*, "Uncertainty analysis for RadCalNet instrumented test sites using the Baotou sites BTCN and BSCN as examples," *Remote Sens.*, vol. 12, 2020, Art. no. 1696.

[9] J. Roujean, "A bidirectional reflectance model of the earth's surface," *J. Geophys. Res.*, vol. 97, pp. 455–468, 1992.

[10] X. H. Che, M. Feng, J. O. Sexton, S. Channan, Y. P. Yang, and Q. Sun, "Assessment of MODIS BRDF/Albedo model parameters (MCD43A1 collection 6) for directional reflectance retrieval," *Remote Sens.*, vol. 9, pp. 1123–1139, 2017.

[11] B. N. Holben *et al.*, "AERONET—A federated instrument network and data archive for aerosol characterization," *Remote Sens. Environ.*, vol. 66, pp. 1–16, 1998.

[12] S. Aliaksandr, B. N. Holben, A. Smirnov, and F. E. Thomas, "Assessment of error in aerosol optical depth measured by AERONET due to aerosol forward scattering," *Geophys. Res. Lett.*, vol. 39, no. 23, pp. 23806–23811, 2012.

[13] A. Berk, G. Anderson, P. Acharya, and E. Shettle, "MODTRAN5 5.2.1 user's manual," Spectral Sci. Inc., Burlington, MA, USA, 2011.

[14] Z. Szantoi and P. Strobl, "Copernicus Sentinel-2 calibration and validation," *Eur. J. Remote Sens.*, vol. 52, no. 1, pp. 253–255, 2019.

[15] F. T. Z. Tuli, C. T. Pinto, A. Angal, X. X. Xiong, and D. Helder, "New approach for temporal stability evaluation of pseudo-invariant calibration sites (PICS)," *Remote Sens.*, vol. 11, no. 12, 2019, Art. no. 1502.

[16] G. Thuillier *et al.*, "The solar spectral irradiance from 200 to 2400 nm as measured by the SOLSPEC spectrometer from the ATLAS and EURECA missions," *Sol. Phys.*, vol. 214, pp. 1–22, 2003.

[17] G. Thuillier *et al.*, "The visible solar spectral irradiance from 350 to 850 nm as measured by the SOLSPEC spectrometer during the Atlas 1 mission," *Sol. Phys.*, vol. 177, pp. 41–61, 1998.

[18] M. Cox and P. Harris, "An outline of supplement 1 to the guide to the expression of uncertainty in measurement on numerical methods for the propagation of distributions," *Meas. Techn.*, vol. 48, pp. 336–345, 2005.

[19] K. Boris *et al.*, "CCPR-S1 supplementary comparison for spectral radiance in the range of 220nm to 2500nm," *Metrologia*, vol. 46, pp. S174–S180, 2009.