

Article

Determination of the Key Comparison Reference Value from Multiple Field Calibration of Sentinel-2B/MSI over the Baotou Site

Caixia Gao ¹, Yaokai Liu ^{1,*}, Jinru Liu ¹, Lingling Ma ¹, Zhifeng Wu ², Shi Qiu ¹, Chuanrong Li ¹, Yongguang Zhao ¹, Qijin Han ³, Enyu Zhao ⁴, Yonggang Qian ¹ and Ning Wang ¹

- ¹ Key Laboratory of Quantitative Remote Sensing Information Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China; gaocaixia@aoe.ac.cn (C.G.); liujingru17@mails.ucas.ac.cn (J.L.); mall@aircas.ac.cn (L.M.); qiushi@aircas.ac.cn (S.Q.); licr@aircas.ac.cn (C.L.); ygzha@aoe.ac.cn (Y.Z.); qianyg@aoe.ac.cn (Y.Q.); wangning@aoe.ac.cn (N.W.)
- ² National Institute of Metrology, Beijing 100029, China; wuzf@nim.ac.cn
- ³ China Centre for Resources Satellite Data and Application, Beijing 100094, China; hanqijin@chinasivei.com
- ⁴ College of Information Science and Technology, Dalian Maritime University, Dalian 116026, China; zhaoenyu@dlmu.edu.cn
- * Correspondence: liuyk@aoe.ac.cn

Received: 29 May 2020; Accepted: 23 July 2020; Published: 27 July 2020



Abstract: Field calibration is a feasible way to evaluate space-borne optical sensor observations via natural or artificial sites on Earth's surface with the aid of synchronous surface and atmospheric characteristic data. Since field calibration is affected by the coupled effects of surface and atmospheric characteristics, the single calibration results acquired under different surface and atmospheric conditions have different biases and different uncertainties, making it difficult to determine the consistency of these multiple calibration results. In view of this, by assuming that the radiometric performance is invariant during field calibration and the calibration samples are independent of each other, the surface-atmosphere invariant Key Comparison Reference Value (KCRV) is essentially derived from various calibration results. As the number of calibration samples increases, the uncertainty in the KCRV should decrease, and the KCRV should approach the "true" value. This paper addresses a novel method for estimating a weighted average value from multiple calibration results that can be used to compare each calibration result, and this value is accepted as the KCRV. Furthermore, this method is preliminarily applied to the field calibration of the Multispectral Instrument (MSI) onboard the Sentinel-2B satellite via the desert target at the Baotou site, China. After employing a chi-squared test to verify that 12 calibration samples are independent from each other, the KCRV of the 12 calibration samples at the Baotou site is derived, which exhibits much lower uncertainty than a single sample. The results show that the KCRVs of the relative differences between the simulated and observed at-sensor reflectance are 3.75%, 5.11%, 6.09%, and 5.03% for the four bands of Sentinel-2B/MSI, respectively, and the corresponding uncertainties are 1.84%, 1.87%, 1.90%, and 1.93%. It is noted that the KCRV uncertainty obtained with only 12 calibration samples is reduced significantly, and in the future, more samples in other instrumented sites will be used to validate this method thoroughly.

Keywords: field calibration; key comparison reference value; Monte Carlo; uncertainty analysis

1. Introduction

The ability to detect and quantify changes in Earth's environment and its global energy balance depends on satellite sensors that can provide calibrated, consistent measurements of Earth's surface

features [1]. Periodic evaluation of the radiometric performance of space-borne optical sensors is becoming mandatory to ensure the availability of consistent and accurate data products to the user community [2–4]. Field calibration is a meaningful approach for monitoring the radiometric performances of sensors, and this process utilizes natural or artificial sites on Earth’s surface to evaluate the observations by the satellite sensors.

In recent years, there has been increasing interest in establishing a global integrated network of calibration sites for monitoring sensor performance. To increase the dynamic range of calibration and validation samples, eight instrumented sites were proposed by the Committee on Earth Observation Satellites (CEOS) to collect the surface reflectance through field campaigns [5]. However, in traditional field campaigns, the surface reflectance is collected by personnel, and the measured data are used in atmospheric radiative transfer codes (e.g., MODTRAN or 6S) to simulate the top-of-atmosphere (TOA) radiance at the time of satellite overpass. Then, the simulated TOA radiance is compared with the satellite sensor observations to evaluate its radiometric performance. This process is labor intensive and complex and can be prone to errors because of human measurement errors and data processing. In 2014, the CEOS Working Group on Calibration & Validation (WGCV) initiated Radiometric Calibration Network (RadCalNet) activities to provide satellite operators with SI-traceable TOA spectrally resolved reflectance to aid in the radiometric calibration of optical imaging sensors from a coordinated network of instrumented land-based test sites [6]. This project aims to reduce the error using automated instruments and nominal standard data processing systems to increase the number of matchups between in situ measurements and space sensor observations and to reduce the overall uncertainties. RadCalNet provides a feasible way to perform multiple calibrations of optical sensors in visible and near-infrared (VNIR) spectra. However, the field radiometric calibration results for specific satellite sensors differ when different samples are used. This difference may be due to the uncertainties associated with the differences in the surface and atmospheric characteristics. Although the uncertainties of each field radiometric calibration result associated with the surface and atmosphere are usually given, it remains a problem for satellite data users whose results should be trusted and used. Taking two samples, for example, the relative difference between simulated TOA radiance and the observed one is calculated for evaluating the radiometric performance of sensors. One sample has much lower relative difference than the other one, but has higher uncertainty, and it is difficult to evaluate which one should be better and be trusted. Therefore, we need to find a reference value that approximates the “true” value based on each radiometric calibration result. Then, combined field radiometric calibration results could be expressed in the form of a single radiometric calibration.

The Mutual Recognition Arrangement (MRA) was signed in 1999 with the objectives of establishing equivalent national measurement standards and providing mutual recognition of calibration and measurement certificates issued by National Metrology Institutes (NMIs). The effectiveness of the MRA depends critically on key comparisons: measurement comparisons conducted among the NMIs that test their key skills and provide an experimental confirmation of their capability. In the key comparisons, the most significant difference arises because there is usually no “high echelon laboratory” capable of providing reference values with negligible uncertainty, and the uncertainties associated with the drift and variations in the attributes of the artifacts are different for each NMI, and the differences are rarely negligible [7]. In view of this, the technical supplement to the MRA encourages the calculation of a Key Comparison Reference Value (KCRV), with the expectation that the KCRV should be “... a close approximation to the corresponding SI value”. From this background, the Consultative Committee of Photometry and Radiometry (CCPR) identified several key comparisons at its meeting in March 1997. One of these comparisons was the CCPR Key Comparison K1-a for Spectral Irradiance in the spectral region from 250 to 2500 nm [8]. This comparison was carried out to meet the requirements of the mutual recognition arrangement by 13 participating NMIs. In this comparison, a method for determining a weighted mean with a cutoff as the KCRV was proposed and agreed upon by the CCPR [8,9]. In view of this, by drawing on the experience of the method in CCPR Key Comparison K1-a, a novel method for determining the weight coefficient for each calibration sample and subsequently a single KCRV is

proposed as a weighted average value from multiple calibration results, 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 derived KCRV is much closer to the “true” value than a single-sample result and, besides, it could be used as a bridge for comparing the consistency of these multiple calibration results. Moreover, this method is tested by applying it to the field calibration of the Multispectral Instrument (MSI) onboard the Sentinel-2B satellite using the desert target at the Baotou site, and a surface–atmosphere invariant KCRV of the 12 calibration samples is derived as the weighted average value of these samples.

This paper is organized as follows. The test site and datasets are described in Section 2. Section 3 provides the method for determining the KCRV, and the results are presented in Section 4. The discussion and the conclusions are drawn in Sections 5 and 6, respectively.

2. Site Description and Dataset

In this study, the desert target at the Baotou site is used to perform the field calibration of Sentinel-2B/MSI. The characteristics of the Baotou site and the corresponding datasets used are described as follows.

2.1. Baotou Site Description

The Baotou site was developed under the support of the Ministry of Science and Technology of China in response to the need to collect data for the ever-increasing number of high-resolution on-orbit sensors without the restrictive requirement of ground-based personnel. This site is located in Inner Mongolia, China, with geographical coordinates of 109.6°N, 40.85°E, 50 km away from Baotou city, covering a flat area of approximately 300 km² with an average altitude of 1270 m. The Baotou site features a semidry continental climate in the southern region of the Mongolian Plateau. The area is dominated by various land surfaces, including lakes, sand, bare soil, maize, and grass, and is equipped with various artificial targets, including portable and permanent targets. Among them, in consideration of the spatial resolution of Sentinel-2B/MSI, the uniform desert target is used to carry out the field calibration in this study. The desert target has a size of 300 m × 300 m to meet the needs of the radiometric calibration of moderate spatial resolution optical sensors (see Figure 1).

The Baotou site is one of four demonstrated sites that took part in the prototyping phase of RadCalNet along with sites at La Crau (France), Railroad Valley (USA), and Gobabeb (Namibia) [10]. This site currently provides a demonstrated global standard automated radiometric calibration service.



Figure 1. Desert target used for field calibration at the Baotou site in this study.

2.2. Dataset

We choose 2018 as the study year for all the datasets in this research. To avoid the impact of clouds on satellite observations, a total of 12 clear-sky calibration samples are used in the analysis. For clarity, the satellite image, surface reflectance, bidirectional reflectance distribution function (BRDF) data and the corresponding atmospheric parameters are described as follows.

2.2.1. Satellite Image Datasets

As part of the Copernicus Program of the European Commission (EC), the European Space Agency (ESA) is currently operating the Sentinel-2 constellation of two polar orbiting satellites Sentinel-2A and Sentinel-2B, and each one is equipped with an optical imaging sensor MSI. Sentinel-2A was launched on 23 June 2015 and Sentinel-2B followed on 7 March 2017. The MSI acquires high spatial resolution (band 2–4, band 8, at 10 m; band 5–7, band 8A, band 11, band 12, at 20 m; band 1, band 9, and band 10, at 60 m) optical imagery from 13 bands in the visible, near infrared, and short wave infrared range of the electromagnetic spectrum [11]. In consideration of the target size, bands 2, 3, 4, and 8 with a resolution of 10m are used in the field calibration. Figure 2 shows the spectral response function of the four bands of Sentinel-2B/MSI.

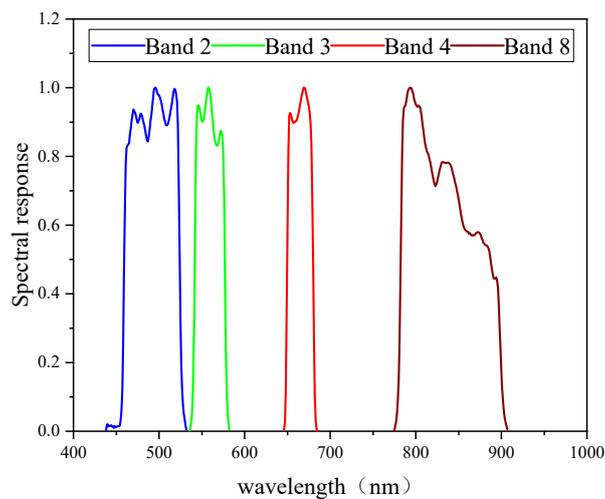


Figure 2. Spectral response function of Multispectral Instrument (MSI) onboard Sentinel-2B satellite.

Twelve scenes of Sentinel-2B/MSI images under clear-sky conditions over the Baotou site are used in this study for performing field calibration. The image acquisition time (IAT) in Coordinated Universal Time (UTC) and local time and the corresponding sun-observation geometry information (viewing zenith angle (VZA), viewing azimuth angle (VAA), sun zenith angle (SZA), and solar azimuth angle (SAA)) are listed in Table 1 (columns 3–8).

Table 1. Data acquisition information of Sentinel-2B/MSI over the Baotou site.

Sample No.	Date	Local Time	Coordinated Universal Time	Viewing Zenith Angle (°)	Viewing Azimuth Angle (°)	Sun Zenith Angle (°)	Solar Azimuth Angle (°)	Aerosol Optical Thickness	Water Vapor Content (g/cm ²)
1	3/25/2018	11:25:29	3:25:29	8.85	104.7	42.54	152.69	0.213	0.223
2	4/7/2018	11:35:29	3:35:29	6.23	283.12	36.65	155.19	0.207	0.156
3	5/14/2018	11:25:39	3:25:39	8.84	104.73	26.13	144.57	0.278	1.023
4	6/13/2018	11:25:39	3:25:39	8.85	104.7	22.6	137.09	0.065	1.126
5	6/26/2018	11:35:39	3:35:39	6.23	283.11	21.58	140.54	0.078	1.056
6	7/13/2018	11:25:39	3:25:39	8.83	104.72	24.47	136.25	0.263	2.073
7	9/21/2018	11:25:29	3:25:29	8.86	104.7	42.49	157.67	0.205	0.389
8	10/1/2018	11:25:29	3:25:29	8.84	104.7	46	160.38	0.048	0.621
9	10/4/2018	11:35:59	3:35:59	6.23	283.12	46.51	164.42	0.24	0.801
10	10/11/2018	11:26:39	3:26:39	8.84	104.73	49.54	162.59	0.104	0.285
11	10/24/2018	11:38:09	3:38:09	6.26	283.2	53.58	167.66	0.165	0.375
12	10/31/2018	11:28:49	3:28:49	2.46	172.6	55.21	166.73	0.176	0.287

2.2.2. Ground Synchronous Measurements

Under the framework of RadCalNet, two sets of automatic measurement systems are developed and installed in the desert target at the Baotou site (see Figure 3), which is well calibrated and SI traceable for assuring the accuracy of its measurements. The system could continuously observe the

surface-reflected spectral radiance at nadir (L_g), covering a spectral range of VNIR (400–1000 nm) every 2 min [12]. Meanwhile, a Cimel CE318 sun photometer was also installed at the Baotou site to acquire the aerosol optical thickness (AOT) and water vapor content (WVC), which can be used to simulate solar direct irradiance (E_{solar}^{direct}) and diffuse sky irradiance ($E_{sky}^{diffuse}$) for retrieving surface reflectance and then to generate the TOA spectral radiance. Since the Baotou site has joined the Aerosol Robotic Network (AERONET), the AOT values in the 440, 670, 870, and 1020 nm channels are retrieved by the AERONET portal from the measured direct solar irradiance every 15 min. Then, the AOT at 550 nm used in this study is calculated via logarithmic interpolation from the AOT in the 440, 670, 870, and 1020 nm channels. In addition, the measured direct solar irradiance in the 936 nm channel of the solar radiometer is used to retrieve the WVC using a modified Langley algorithm via the AERONET portal [12,13]. The AOT at 550 nm and WVC at the time of Sentinel-2B/MSI image acquisition over the Baotou site are linearly interpolated and shown in columns 9 and 10 of Table 1. It is shown that there are obvious differences in the atmospheric parameters on the study dates. The AOT at 550 nm varies from 0.048 to 0.278. The WVC varies from 0.156 g/cm² to 2.073 g/cm².



Figure 3. The automatic surface-reflected radiance measurement system over the desert target at the Baotou site.

With the radiance of the desert target collected by the automatic measurement system, the surface reflectance (ρ) can then be retrieved using Equation (1) [12].

$$\rho = \frac{\pi L_g}{E_{solar}^{direct} + E_{sky}^{diffuse}}, \quad (1)$$

where E_{solar}^{direct} and $E_{sky}^{diffuse}$ are the solar direct irradiance and sky scattering irradiance reaching the ground, respectively, which can be simulated by the MODTRAN 5 atmospheric radiative transfer model with the aid of the synchronous AOT at 550 nm and WVC.

Figure 4 shows the retrieved surface reflectances of the desert target at the acquisition time of the Sentinel-2B/MSI satellite images and the standard deviation of surface reflectance during the study dates. The standard deviation is generally less than 0.02.

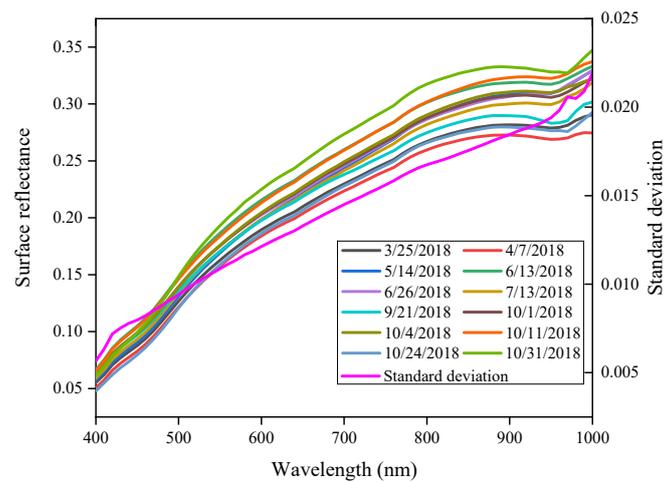
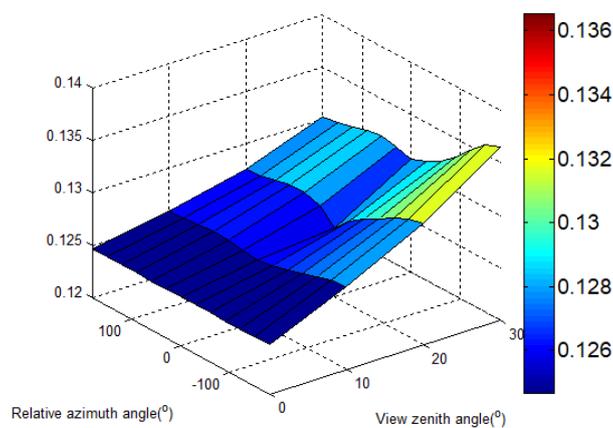


Figure 4. Retrieved surface reflectance of desert target covering the spectral range of Sentinel-2B/MSI on the study dates.

In addition, the surface bidirectional reflectance factor (BRF) of the desert target at different viewing angles (VZA: from 0 degrees to 40 degrees with 10-degree intervals; VAA: from 0 degrees to 360 degrees with a 30-degree interval) from 10 a.m. to 3 p.m. are collected using an SVC HR-2014 spectrometer, which is controlled by a multi-angle automatic observation system. Then, the Roujean BRDF model [14], as expressed by Equation (2), is used to model the BRDF characteristics of the desert target.

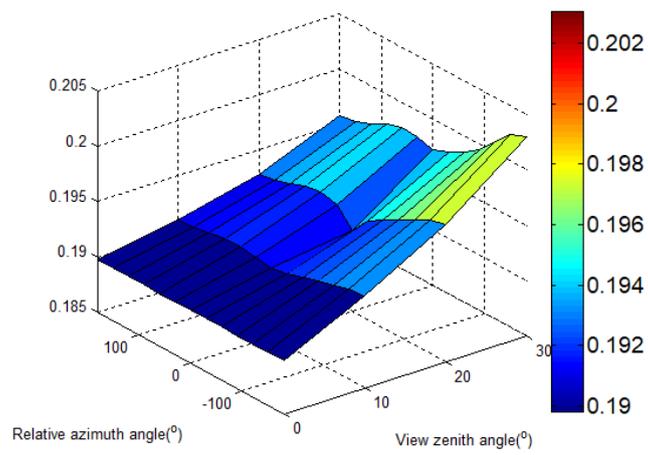
$$R(\theta_i, \theta_v, \varphi, \lambda) = f_{iso}(\lambda) + f_{vol}(\lambda)K_{vol}(\theta_i, \theta_v, \varphi, \lambda) + f_{geo}(\lambda)K_{geo}(\theta_i, \theta_v, \varphi, \lambda) \quad (2)$$

where θ_i is the SZA, θ_v is the VZA of the spectrometer, φ is the relative azimuth angle, λ is the wavelength, and K_{vol} and K_{geo} are the volumetric and geometric scattering kernels, respectively. Both the volumetric and geometric scattering kernels can be calculated from the sun and spectrometer viewing angles as reported by Roujean in reference [14]. f_{iso} , f_{vol} , and f_{geo} are the isotropic, volumetric, and geometric kernel coefficients, respectively, which can be determined with the aid of in situ BRF measurements using a linear least square fit method. Figure 5 shows the simulated BRF in bands 2, 3, 4, and 8 of Sentinel-2B/MSI using our constructed BRDF model when the solar zenith angle is set to 20 degrees.

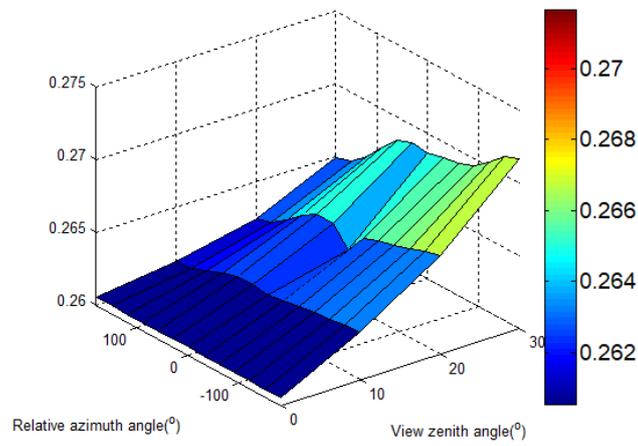


(a)

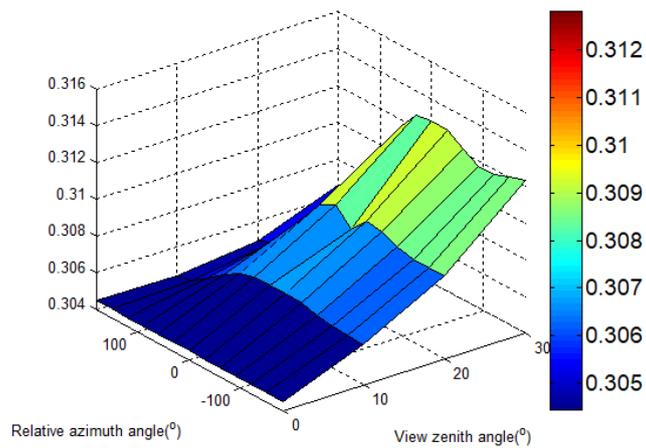
Figure 5. Cont.



(b)



(c)



(d)

Figure 5. Surface bidirectional reflectance factor of the desert target. (a) Band 2; (b) band 3; (c) band 4; (d) band 8.

3. Method for KCRV Determination

The field calibration method makes use of natural or artificial targets on the Earth's surface to evaluate satellite sensor observations. In the spectral range of 400–2500 nm, the surface reflectance is measured by a calibrated spectrophotometer; subsequently, the corresponding simulated TOA radiance or reflectance in band i for sample j ($L_{i,j}^{TOA,sim}$ or $\rho_{i,j}^{TOA,sim}$) is generated using MODTRAN 5 model driven by synchronous surface reflectance and atmosphere parameters. It is compared with the radiance or reflectance observed by the sensor ($L_{i,j}^{TOA,obs}$ or $\rho_{i,j}^{TOA,obs}$), and the relative difference is calculated. In this process, each simulated TOA radiance (or reflectance) is related to input uncertainties from instrumentation used to make surface reflectance and atmospheric measurements, the exoatmospheric solar irradiance model, spatial sampling of the site, and the MODTRAN 5 radiative transfer code. Therefore, an end-to-end uncertainty budget and the overall uncertainty of the simulated TOA radiance (or reflectance) is determined using a combination of analytical and Monte Carlo approaches. Then, the total uncertainty for each calibration result can be determined. On the basis of this, a KCRV is determined and, meanwhile, a chi-squared test is applied to verify the assumption that the calibration samples are independent from each other, and then the degree of equivalence is calculated to evaluate the consistency of the calibration results in multiple calibrations (see Figure 6).

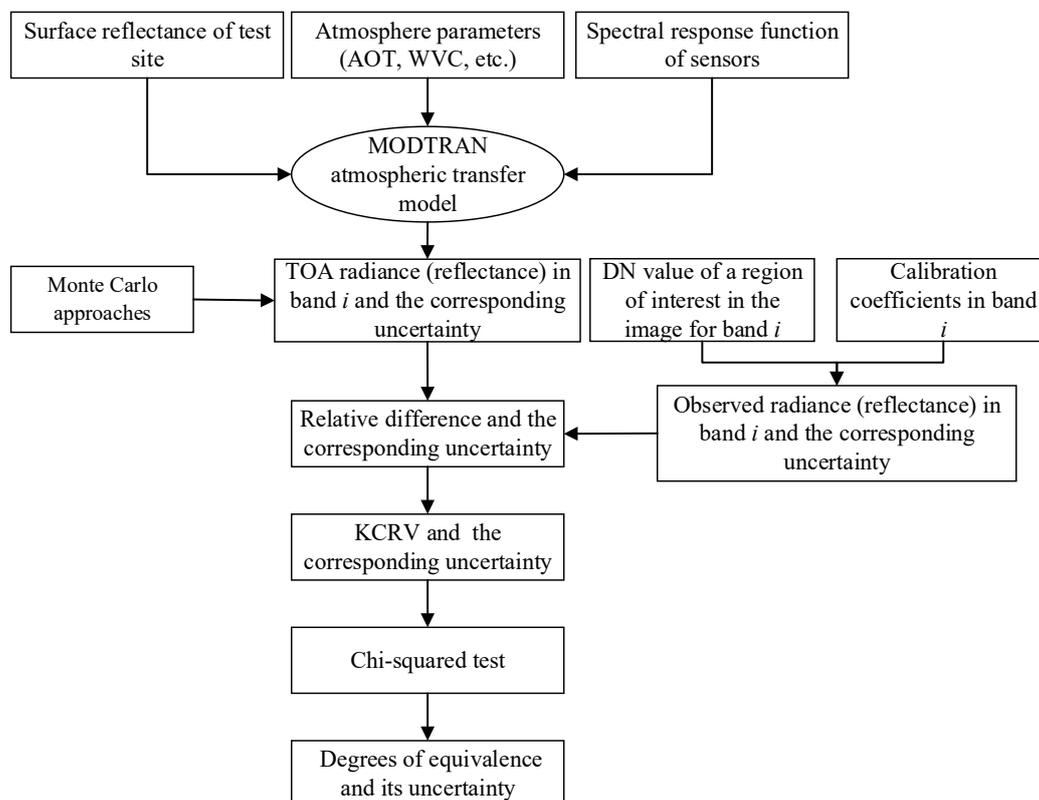


Figure 6. The flowchart of the method for determining the Key Comparison Reference Value (KCRV).

3.1. Uncertainty Analysis Method Using the Monte Carlo Approach

The Monte Carlo approach, as presented by the Guide to Uncertainty Measurement (GUM) Supplement 1, involves the propagation of the distributions of the input sources of uncertainty by using the model to provide the distribution of the output [15–17]. This process is illustrated in Figure 7. In this process, the most appropriate probability density functions (PDFs) for each of the input quantities should be defined, and then a number of Monte Carlo trials should be selected. Generally, the greater the number of simulation trials, the greater the convergence of the results. In this study, this number is chosen by using an adaptive methodology, and by checking the stabilization of the standard deviation

of the output quantity, this number is set to 1000. The last stage is to summarize and express the results. According to GUM Supplement 1, the standard uncertainty, taken as the standard deviation of these generated values, is reported as the results [15].

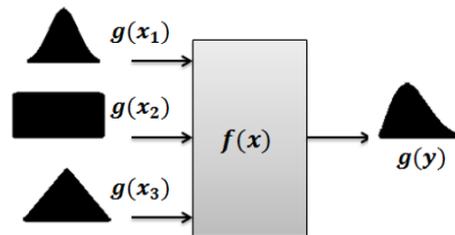


Figure 7. The process of the Monte Carlo methodology. x_1 , x_2 , and x_3 are input quantities, $g(x_1)$, $g(x_2)$, and $g(x_3)$ are the distribution functions of the input quantities, and $g(y)$ is the distribution function of the measurand.

To better understand the application of Monte Carlo simulations to the estimation of measurement uncertainty, a case study of estimating the simulated TOA radiance uncertainty caused by AOT uncertainty is presented. For each calibration sample, first, the simulated TOA radiance (or reflectance) of Sentinel-2/MSI is generated using the MODTRAN 5 model driven by measured surface reflectance, AOT, WVC, and sun-viewing geometries. Then, by assuming the AOT uncertainty is approximately 0.01 and its PDF is assumed to be a Gaussian distribution, a Monte Carlo simulation is set to run $M = 1000$ trials of the MODTRAN 5 model using the error-added AOT and other input sources, and a new group of TOA radiances (or reflectances) is generated. Furthermore, the relative differences between the simulated TOA radiances (or reflectances) with AOT error and those without AOT error could be calculated, and its standard deviation is taken as the TOA radiance uncertainty caused by AOT uncertainty [18].

3.2. Determining the KCRV Using the Weighted Average Approach

First, the differences in band i $\Delta_{i,j}$ between the simulated TOA reflectance $\rho_{i,j}^{TOA,sim}$ and the observed reflectance $\rho_{i,j}^{TOA,obs}$ for calibration sample j are determined, and it is represented as Equation (3):

$$\Delta_{i,j} = \frac{\rho_{i,j}^{TOA,sim}}{\rho_{i,j}^{TOA,obs}} - 1. \quad (3)$$

Its corresponding uncertainty $u(\Delta_{i,j})$ is calculated as:

$$u(\Delta_{i,j}) = \sqrt{u^2(\rho_{i,j}^{TOA,sim}) + u^2(\rho_{i,j}^{TOA,obs})}, \quad (4)$$

where $u(\rho_{i,j}^{TOA,sim})$ and $u(\rho_{i,j}^{TOA,obs})$ are the uncertainties of $\rho_{i,j}^{TOA,sim}$ and $\rho_{i,j}^{TOA,obs}$, respectively.

Then, to avoid a higher contribution of samples with relatively low uncertainties, the KCRV is calculated as a weighted mean with a cutoff agreed upon by CCPR. The cutoff values for band i , $u_{i,cut-off}$ are calculated by [19,20]:

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

Subsequently, the uncertainty $u(\Delta_{i,j})$ is adjusted by the cutoff, and the adjusted uncertainty for sample j in band i $u_{adj}(\Delta_{i,j})$ can be calculated as follows:

If $u(\Delta_{i,j}) \geq u_{i,cut-off}$ then $u_{adj}(\Delta_{i,j}) = u(\Delta_{i,j})$
 else $u(\Delta_{i,j}) < u_{i,cut-off}$ then :

$$u_{adj}(\Delta_{i,j}) = u_{i,cut-off} \quad (6)$$

Then, by assuming that the calibration samples are independent from each other, the weight coefficient in band i for each sample $w_{i,j}$ is determined by [19,20]:

$$w_{i,j} = \frac{u_{adj}^{-2}(\Delta_{i,j})}{\sum_{j=1}^N u_{adj}^{-2}(\Delta_{i,j})} \quad (7)$$

The weighted average value of the relative difference y is determined by [18,19]:

$$y = \frac{\frac{\Delta_{i,1}}{u_{adj}^2(\Delta_{i,1})} + \dots + \frac{\Delta_{i,N}}{u_{adj}^2(\Delta_{i,N})}}{\frac{1}{u_{adj}^2(\Delta_{i,1})} + \dots + \frac{1}{u_{adj}^2(\Delta_{i,N})}} = \sum_{j=1}^N w_{i,j} \cdot \Delta_{i,j} \quad (8)$$

The uncertainty associated with y is calculated as:

$$u(y) = \frac{1}{\sqrt{u_{adj}^{-2}(\Delta_{i,1}) + \dots + u_{adj}^{-2}(\Delta_{i,N})}} \quad (9)$$

Then, a chi-squared test is applied to check whether the assumption of independence of these calibration samples is justified. If the test does not fail, y is accepted as the KCRV; otherwise, the KCRV is null. Thus, the KCRV values represent the average scale of the comparison. The observed chi-squared value, χ_{obs}^2 , is expressed as Equation (9), and the degrees of freedom are assigned as $N - 1$. The greater the observed chi-squared value, the stronger the calibration sample correlation.

$$\chi_{obs}^2 = \frac{(\Delta_{i,1} - y)^2}{u_{adj}^2(\Delta_{i,1})} + \dots + \frac{(\Delta_{i,N} - y)^2}{u_{adj}^2(\Delta_{i,N})} \quad (10)$$

The chi-squared test is regarded as failing if

$$probability\{\chi^2(\gamma) > \chi_{obs}^2(\gamma)\} < 0.05 \quad (11)$$

where $\chi^2(\gamma)$ is the critical value of the chi-squared distribution, which can be acquired in reference [21], when the degrees of freedom are equal to $N - 1$.

Finally, in order to evaluate the calibration result, the difference between the calibration results of each sample $\Delta_{i,j}$ and the KCRV d_i is denoted as degrees of equivalence, and is calculated by Equation (12). The lower degrees of equivalence are, the more accurate the sample is.

$$d_i = \Delta_{i,j} - KCRV \quad (12)$$

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

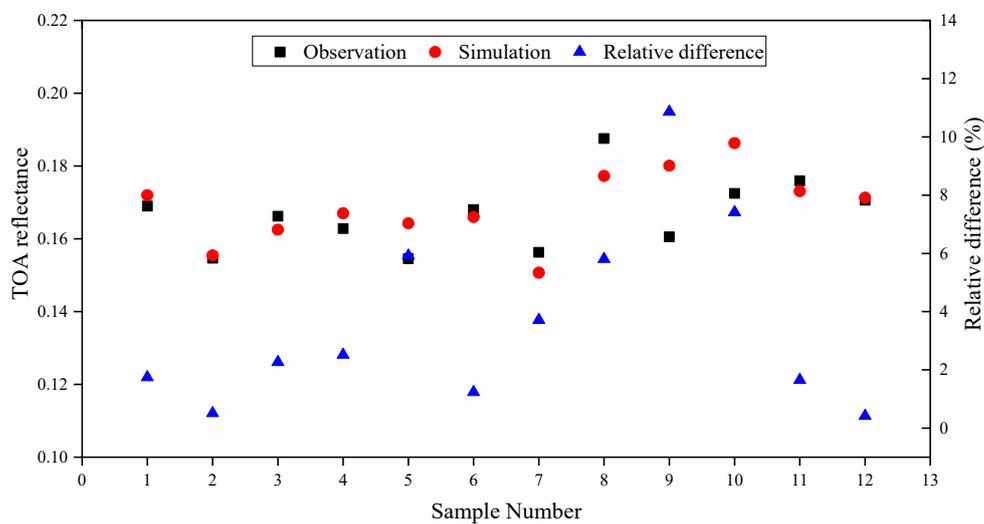
$$u(d_i) = \sqrt{u^2(\Delta_{i,j}) - u^2(KCRV)} \quad (13)$$

4. Results

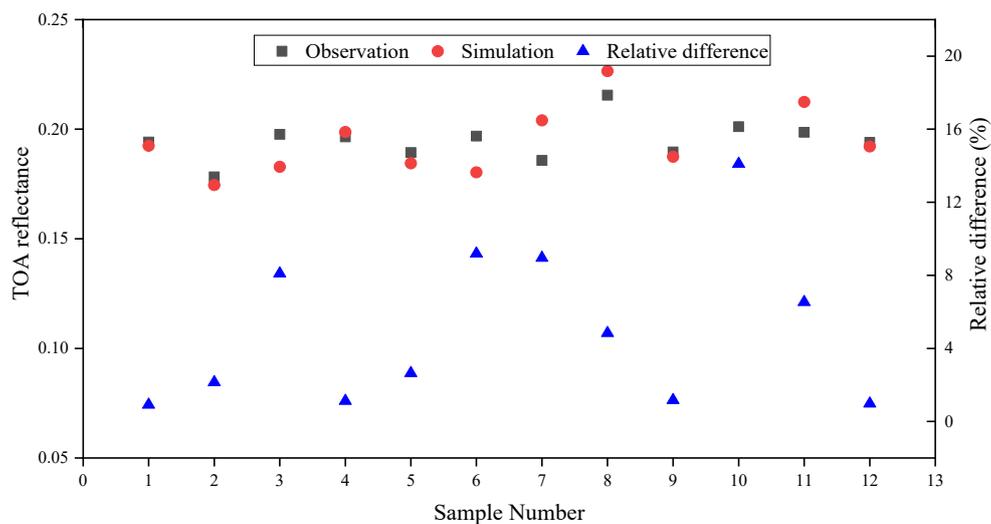
In this study, the 12 samples in 2018 are used to calibrate Sentinel-2B/MSI, and a KCRV weighted by uncertainty is derived to evaluate the consistency of these samples.

4.1. Field Calibration Results for Each Sample

With the aid of the retrieved surface reflectance and synchronous atmospheric parameters (WVC, AOT), the simulated TOA radiances in the Sentinel-2B/MSI bands over the desert target at the Baotou site on the study dates are generated using the MODTRAN 5 model and then converted to TOA reflectances. Subsequently, these TOA reflectances for the 12 samples are compared with the observed values, which are calculated according to the digital number (DN) values extracted from the Sentinel-2B/MSI images over the ground target and the corresponding calibration coefficients published by the ESA. Then, the relative differences between the simulated TOA reflectances and the observed reflectances are also acquired using Equation (3). The compared results for the 12 calibration samples are shown in Figure 8. Each sample has a different relative difference due to the different surface and atmosphere characteristics. In general, the simulated TOA reflectances for samples 1–5, 7, 8, 11, and 12 are much closer to the observed reflectances than those for the other samples, and the corresponding relative differences are within 10% for bands 2, 3, 4, and 8. In addition, large relative differences of approximately 16% are also presented.

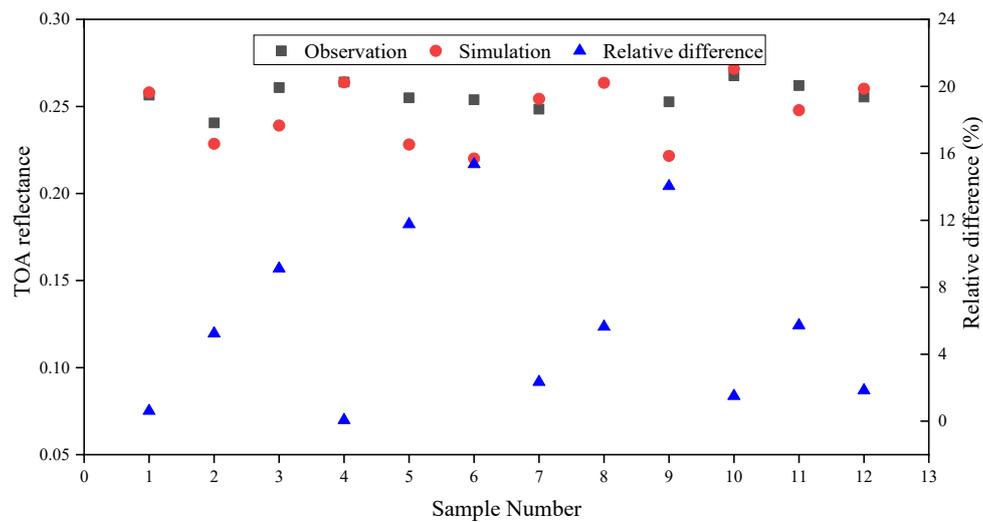


(a)

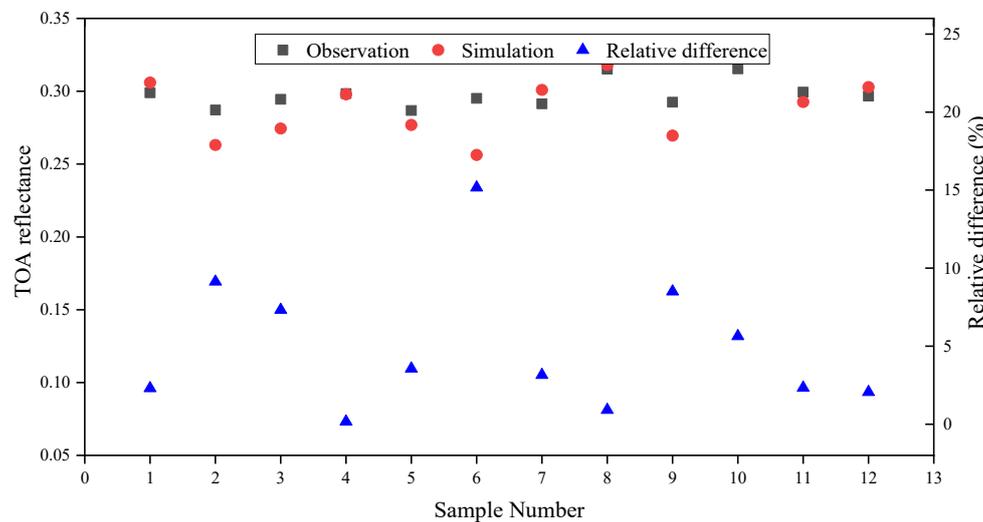


(b)

Figure 8. Cont.



(c)



(d)

Figure 8. The compared results for the 12 calibration samples. (a) Band 2; (b) band 3; (c) band 4; (d) band 8.

4.2. Uncertainty Analysis Results for the Simulated TOA Reflectance

Uncertainty is an estimation of the magnitude of error, typically expressed in terms of a confidence interval within which the error lies. In this field calibration process, the simulated TOA reflectance relies on a set of representative input parameters, namely, the surface reflectance of the calibration target, atmospheric parameters, sun-sensor geometry, etc. The error budget process requires understanding the uncertainty of the TOA reflectance used for the field calibration associated with input parameter uncertainty. In this study, the uncertainty of the simulated TOA reflectance associated with uncertainty sources, including the surface reflectance, the spatial homogeneity, and the BRDF characteristics of the target and atmosphere parameters (WVC, AOT, and aerosol type), are analyzed based on the Monte Carlo simulation method.

4.2.1. TOA Reflectance Uncertainty Caused by Surface Characteristics (Denoted as $\delta_{TOA,ref}$)

The uncertainty of the retrieved surface reflectance, the spatial homogeneity, and the BRDF characteristics significantly contribute to the uncertainty of the TOA reflectance. The three terms of uncertainties are quantified and clarified as follows.

1. Uncertainty analysis of the retrieved surface reflectance at the bottom of the atmosphere (BOA) (denoted as $\delta_{BOA,ref}$).

As described in Equation (1), the uncertainty of the retrieved surface reflectance is determined by the uncertainties of the measured surface-reflected radiance and the estimated total irradiance arriving at the surface. The measured surface-reflected radiance is affected by the laboratory calibration accuracy (denoted as $\delta_{1,1}$), the stability (denoted as $\delta_{1,2}$), and the temperature sensitivity (denoted as $\delta_{1,3}$) of the spectrometer during in situ measurement. Both the calibration accuracy and temperature sensitivity of the spectrometer are calibrated by the NMI [22]. The $\delta_{1,1}$ is calculated by accounting for the uncertainty factors including the lamp-diffuser panel radiance source, source nonuniformity, spectrometer nonlinearity, dark signal, external and internal stray light, and spectrometer noise during calibration when determining the calibration coefficients of the spectrometer in the NMI laboratory. All the uncertainty factors are measured in the NMI laboratory under the ambient temperature of 25 °C and are given by the NMI. Moreover, the calibration coefficients of the spectrometer could change due to temperature sensitivity and the different environmental conditions at the test site compared to those in the laboratory. Therefore, the spectrometer is calibrated with a stable source when the ambient temperature T varies from 10 °C to 40 °C, and derived calibration coefficients at an ambient temperature of 0 °C–40 °C are divided by those at an ambient temperature of 25 °C to calculate a series of temperature correction coefficients $f(\lambda, T)$. Then, an empirical temperature correction model is fitted with these experimental data, as shown in Equation (14). It is noted that $\delta_{1,3}$ includes the uncertainty associated with the temperature correction model itself $\delta_{1,3}^m$ and the uncertainty associated with the gain correction due to the uncertainty in instrument temperature $\delta_{1,3}^{temp}$. The $\delta_{1,3}^m$ could be calculated from the residual to the model. The $\delta_{1,3}^{temp}$, associated with the gain correction due to an estimated 2 °C uncertainty, could be calculated by applying the law of the propagation of uncertainty. In addition, $\delta_{1,2}$ is calculated from the standard deviation of 15 successive field measurements at intervals of 2 min over 15 min before and after the image acquisition time of Sentinel-2/MSI.

$$f(\lambda, T) = (a_0 + a_1\lambda + a_2\lambda^2)T + (b_0 + b_1\lambda + b_2\lambda^2) \quad (14)$$

where a_i and b_i ($i = 1, 2, 3$) are the coefficients, which could be determined with those experimental data using a linear least square fit method. a_1 , a_2 , and a_3 are equal to $1.3579 \times 10^{-2} \text{ }^\circ\text{C}^{-1}$, $-4.6143 \times 10^{-5} \text{ nm}^{-1} \text{ }^\circ\text{C}^{-1}$, and $4.0289 \times 10^{-8} \text{ nm}^{-2} \text{ }^\circ\text{C}^{-1}$, respectively. b_1 , b_2 , and b_3 are equal to 6.4813×10^{-1} , $1.1923 \times 10^{-3} \text{ nm}^{-1}$, and $-1.0507 \times 10^{-6} \text{ nm}^{-2}$, respectively.

The combined uncertainty result of the measured surface-reflected radiance covering 400 nm to 1000 nm is calculated based on error transfer theory (see Figure 9). It is noted that the values of $\delta_{1,1}$ and $\delta_{1,3}$ decrease rapidly with wavelength in the spectral range of 450–500 nm, while $\delta_{1,3}$ increases smoothly with wavelength in the spectral range of 500–1000 nm. The combined uncertainty results of the measured surface-reflected radiance in bands 2, 3, 4, and 8 of Sentinel-2B/MSI are 2.46%, 1.23%, 1.22%, and 1.45%, respectively.

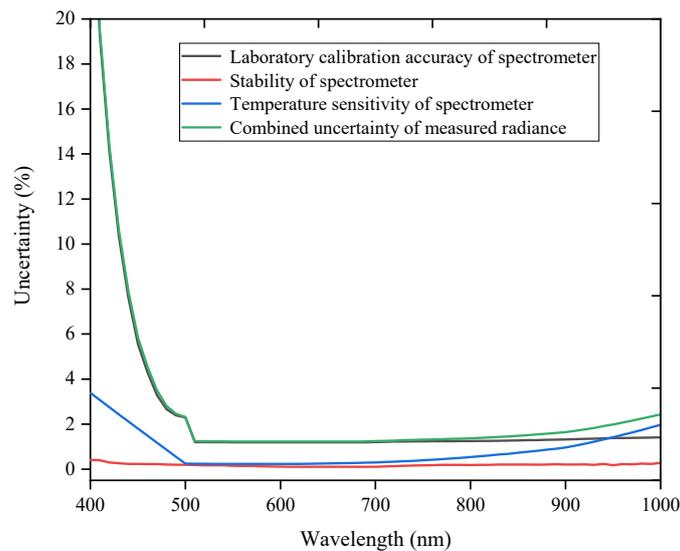
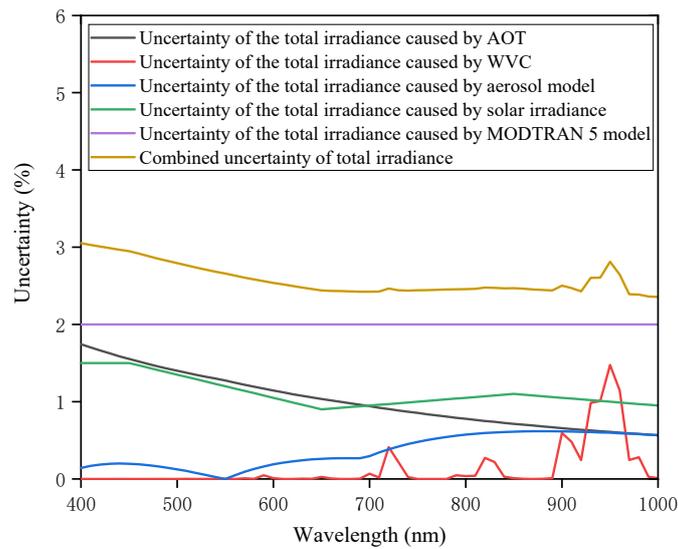


Figure 9. The combined uncertainty of the measured radiance covering 400 nm to 1000 nm due to the laboratory calibration accuracy, the stability, and temperature sensitivity of the spectrometer during in situ measurement.

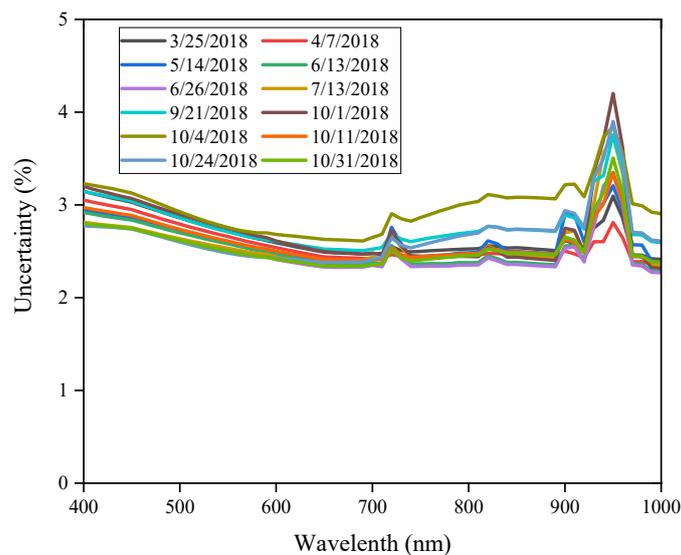
The uncertainty of the total irradiance arriving at the ground is mainly related to the uncertainty of the WVC ($\delta_{1,4}$), AOT ($\delta_{1,5}$), aerosol types ($\delta_{1,6}$), the MODTRAN 5 model ($\delta_{1,7}$), and solar irradiance ($\delta_{1,8}$). In this study, the uncertainties caused by the WVC and AOT are analyzed using Monte Carlo approaches by assuming an uncertainty of 12% in WVC and 0.01 in AOT [23]. To analyze the uncertainty caused by the mixed aerosol type of the rural and desert aerosols at the Baotou site, two new groups of simulated total irradiance are generated when the desert aerosol model and rural model are used, and their relative difference is calculated as $\delta_{1,6}$ (see Figure 10). In addition, $\delta_{1,7}$ is estimated to be approximately 2% [24]; the Thuillier solar irradiance model used 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 [25,26], and the value of $\delta_{1,8}$ at each wavelength is calculated using an interpolation method. The uncertainty of the MODTRAN 5 model is the dominating factor, approximately 2%, while the uncertainty of the WVC contributes less. Figure 10a shows an example of the uncertainties of total irradiance (400–1000 nm) arriving at the ground caused by WVC, AOT, aerosol types, solar irradiance, and MODTRAN 5 model on 7 April 2018. Figure 10b shows the combined uncertainty results of the estimated total irradiance on the study dates for reading clarity. Table 2 shows the combined uncertainty results of the estimated total irradiance arriving at the ground, corresponding to bands 2, 3, 4, and 8 of Sentinel-2B/MSI, and the corresponding uncertainties are within 3.1%.

Table 2. The combined uncertainty results of total irradiance arriving at the ground in bands 2, 3, 4, and 8 of Sentinel-2B/MSI.

Sample No.	Date	Band 2	Band 3	Band 4	Band 8
1	3/25/2018	2.89%	2.69%	2.48%	2.53%
2	4/7/2018	2.82%	2.63%	2.43%	2.46%
3	5/14/2018	2.74%	2.57%	2.40%	2.52%
4	6/13/2018	2.72%	2.56%	2.37%	2.39%
5	6/26/2018	2.64%	2.49%	2.33%	2.37%
6	7/13/2018	2.65%	2.52%	2.39%	2.47%
7	9/21/2018	2.89%	2.69%	2.52%	2.73%
8	10/1/2018	2.92%	2.73%	2.49%	2.46%
9	10/4/2018	2.96%	2.74%	2.62%	3.06%
10	10/11/2018	2.76%	2.59%	2.41%	2.49%
11	10/24/2018	2.62%	2.47%	2.38%	2.72%
12	10/31/2018	2.64%	2.49%	2.34%	2.48%



(a)



(b)

Figure 10. The combined uncertainty of the total irradiance arriving at the ground covering 400 nm to 1000 nm on the study dates. (a) Uncertainties of total irradiance caused by WVC, AOT, aerosol types, solar irradiance, and the MODTRAN 5 model on 7 April 2018; (b) uncertainties of total irradiance on the study dates.

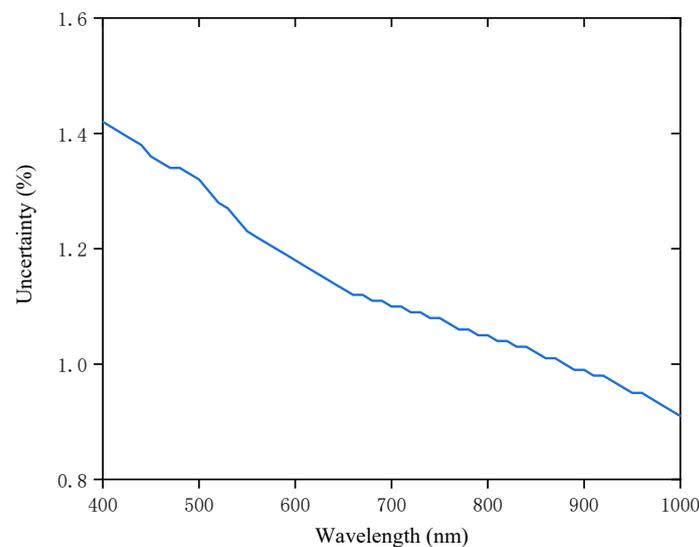
Based on the uncertainties of the measured reflected radiance and the total irradiance arriving at the ground, the combined uncertainties of the retrieved surface reflectance, corresponding to bands 2, 3, 4, and 8 of Sentinel-2B/MSI, are calculated and listed in Table 3. The uncertainties of the retrieved surface reflectance tend to decrease in all bands except for band 8, which would be caused by vapor absorption at approximately 820 nm and 940 nm.

Table 3. The combined uncertainty results of the retrieved surface reflectance in bands 2, 3, 4, and 8 of Sentinel-2B/MSI.

Sample No.	Date	Band 2	Band 3	Band 4	Band 8
1	3/25/2018	3.79%	2.96%	2.77%	2.92%
2	4/7/2018	3.74%	2.91%	2.72%	2.86%
3	5/14/2018	3.68%	2.85%	2.69%	2.90%
4	6/13/2018	3.67%	2.84%	2.66%	2.79%
5	6/26/2018	3.61%	2.78%	2.63%	2.77%
6	7/13/2018	3.62%	2.80%	2.68%	2.86%
7	9/21/2018	3.79%	2.95%	2.80%	3.09%
8	10/1/2018	3.82%	2.99%	2.77%	2.85%
9	10/4/2018	3.85%	3.00%	2.89%	3.39%
10	10/11/2018	3.70%	2.86%	2.70%	2.88%
11	10/24/2018	3.60%	2.76%	2.68%	3.08%
12	10/31/2018	3.61%	2.77%	2.64%	2.87%

2. Uncertainty analysis of surface spatial heterogeneity (denoted as δ_{heter}).

The surface spatial heterogeneity of the desert target is estimated from the atmosphere-corrected airborne hyperspectral images over the Baotou site. This value is determined by the ratio of the standard deviation and the mean value of the target region in the image at the spatial resolution of the automated observation system (0.13 m) (see Figure 11). The uncertainty results of the surface spatial heterogeneity in bands 2, 3, 4, and 8 of Sentinel-2B/MSI are 1.32%, 1.22%, 1.12%, and 1.03%, respectively.

**Figure 11.** The uncertainty of the surface spatial heterogeneity covering 400 nm to 1000 nm for the desert target.

3. Uncertainty results of surface BRDF characteristics (denoted as δ_{brdf})

Because the satellite sensor observes the target at an off-nadir angle and the target is a non-Lambert body, uncertainty would result from the VZA differences between the sensor and automated observation system. In this study, to analyze this term of uncertainty, the Roujean model, as expressed by Equation (2), is constructed with the aid of BRDF in situ measurements of the target. On the basis of this Roujean model [14], the desert target BRDF in the viewing geometries of the Sentinel-2B/MSI sensor and the corresponding BRDF at nadir could be simulated, and their relative difference covering 400–1000 nm is calculated as the uncertainty of the surface BRDF characteristics (see Figure 12).

Consequently, the uncertainty of the surface BRDF characteristics for desert target in bands 2, 3, 4, and 8 of Sentinel-2B/MSI are 2.10%, 1.95%, 1.89%, and 1.92%, respectively.

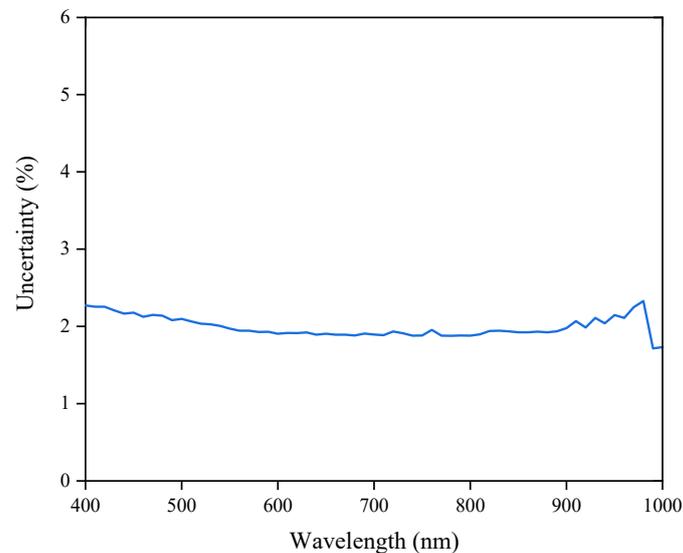


Figure 12. The uncertainty of the surface BRDF characteristics covering 400 nm to 1000 nm for the desert calibration target.

4. Uncertainty results of $\delta_{TOA,ref}$

According to the error transfer theory, the combined uncertainty of the target surface reflectance ($\delta_{combin,ref}$) is calculated from the above estimated uncertainties of the retrieved surface reflectance ($\delta_{BOA,ref}$), heterogeneity (δ_{heter}), and BRDF (δ_{brdf}). Then, based on the Monte Carlo simulation approach, the TOA reflectance uncertainty ($\delta_{TOA,ref}$) caused by $\delta_{combin,ref}$ is derived. Figure 13 shows the values of $\delta_{TOA,ref}$ for the desert target in 400–1000 nm on the study dates, and the corresponding values in bands 2, 3, 4, and 8 of Sentinel-2B/MSI are also given in Table 4. Note that $\delta_{TOA,ref}$ tends to increase with the increase in wavelength and is within 4% for bands 2, 3, 4, and 8 of Sentinel-2B/MSI.

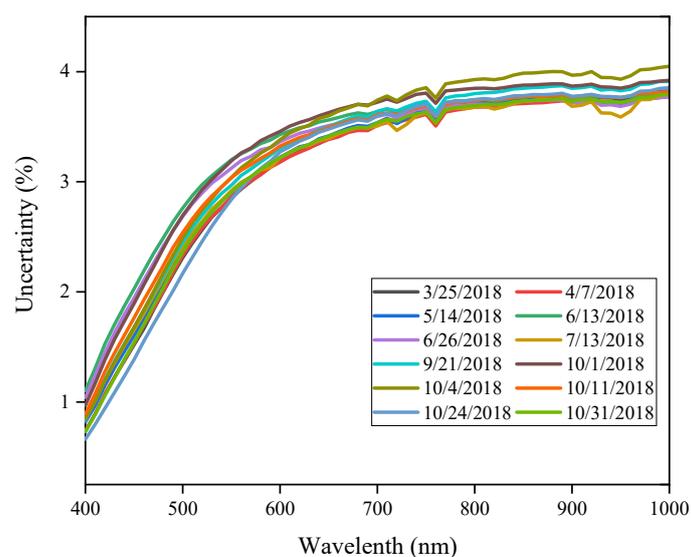


Figure 13. The uncertainty results of the predicted top-of-atmosphere (TOA) reflectance associated with the combined uncertainty of the desert target surface reflectance covering 400 nm to 1000 nm on the study dates.

Table 4. The uncertainty results of the predicted TOA reflectance caused by surface characteristics in bands 2, 3, 4, and 8 of Sentinel-2B/MSI.

Sample No.	Date	Band 2	Band 3	Band 4	Band 8
1	3/25/2018	2.17%	2.92%	3.44%	3.73%
2	4/7/2018	2.19%	2.93%	3.43%	3.69%
3	5/14/2018	2.25%	2.97%	3.47%	3.74%
4	6/13/2018	2.64%	3.24%	3.60%	3.76%
5	6/26/2018	2.56%	3.18%	3.54%	3.72%
6	7/13/2018	2.28%	2.98%	3.45%	3.70%
7	9/21/2018	2.32%	3.04%	3.55%	3.83%
8	10/1/2018	2.56%	3.24%	3.67%	3.86%
9	10/4/2018	2.35%	3.11%	3.65%	3.95%
10	10/11/2018	2.41%	3.09%	3.54%	3.76%
11	10/24/2018	2.04%	2.93%	3.52%	3.77%
12	10/31/2018	2.22%	2.97%	3.46%	3.71%

4.2.2. TOA Reflectance Uncertainty Caused by Atmospheric Characteristics (Denoted as δ_{atm})

Similar to the uncertainty analysis of the aforementioned simulated total irradiance, the uncertainty of the predicted TOA reflectance caused by atmospheric parameters is mainly related to three factors, namely, WVC, AOT, and aerosol type. By assuming that the uncertainties of the measured WVC and AOT are 12% and 0.01, the Monte Carlo approach is used to analyze their resulting uncertainties in TOA reflectance. In addition, the TOA reflectance uncertainty caused by aerosol type is also calculated when the desert aerosol model and rural model are used. The combined uncertainty of the simulated TOA reflectance caused by the uncertainties of WVC, AOT, and aerosol type (δ_{atm}) is shown in Figure 14, and the corresponding uncertainty in bands 2, 3, 4, and 8 of Sentinel-2B/MSI are given in Table 5. It is worth noting that δ_{atm} for each sample is within 0.71%.

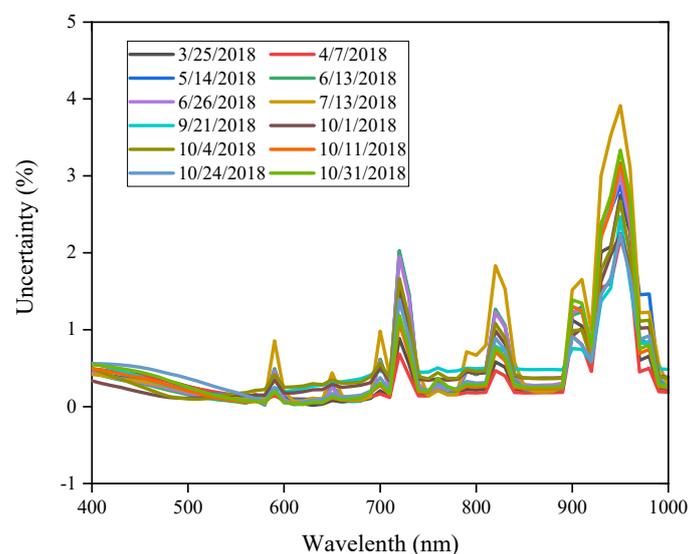
**Figure 14.** Combined uncertainty of the predicted TOA reflectance caused by atmospheric parameters on the study dates.

Table 5. Combined uncertainty of the predicted TOA reflectance caused by atmospheric parameters in bands 2, 3, 4, and 8 of Sentinel-2B/MSI.

Sample No.	Date	Band 2	Band 3	Band 4	Band 8
1	3/25/2018	0.29%	0.12%	0.07%	0.30%
2	4/7/2018	0.29%	0.14%	0.10%	0.24%
3	5/14/2018	0.24%	0.12%	0.14%	0.50%
4	6/13/2018	0.18%	0.06%	0.11%	0.50%
5	6/26/2018	0.20%	0.07%	0.12%	0.52%
6	7/13/2018	0.23%	0.11%	0.17%	0.71%
7	9/21/2018	0.24%	0.14%	0.33%	0.54%
8	10/1/2018	0.12%	0.15%	0.27%	0.50%
9	10/4/2018	0.13%	0.18%	0.32%	0.53%
10	10/11/2018	0.24%	0.07%	0.10%	0.34%
11	10/24/2018	0.39%	0.14%	0.14%	0.39%
12	10/31/2018	0.29%	0.08%	0.10%	0.37%

4.2.3. Overall Uncertainty Results of TOA Reflectance (Denoted as $\delta_{overall}$)

In addition to $\delta_{TOA,ref}$ and δ_{atm} , the uncertainties from the MODTRAN 5 model and solar irradiance should also be considered when calculating the combined overall uncertainty of the predicted TOA reflectance according to the error transfer theory ($\delta_{overall}$). The uncertainties caused by the MODTRAN 5 model and solar irradiance are the same as $\delta_{1,7}$ and $\delta_{1,8}$, respectively, because of their independent transfer characteristics. Figure 15 shows the $\delta_{overall}$ in the spectral range of 400–1000 nm, and the corresponding values in bands 2, 3, 4, and 8 of Sentinel-2B/MSI are shown in Table 6. The $\delta_{overall}$ in band 8 is larger than that in bands 2, 3, and 4, and the $\delta_{overall}$ for all bands is within 4.6%.

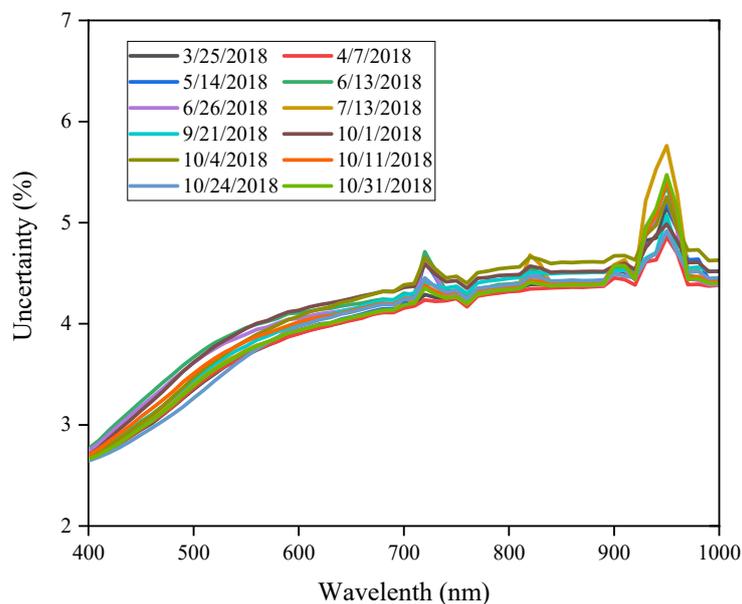
**Figure 15.** Overall uncertainty results of the TOA radiance covering 400 nm to 1000 nm on the study dates.

Table 6. Overall uncertainty of the TOA reflectance in bands 2, 3, 4, and 8 of Sentinel-2B/MSI.

Sample No.	Date	Band 2	Band 3	Band 4	Band 8
1	3/25/2018	3.28%	3.73%	4.09%	4.38%
2	4/7/2018	3.29%	3.74%	4.07%	4.34%
3	5/14/2018	3.33%	3.77%	4.11%	4.41%
4	6/13/2018	3.60%	3.99%	4.22%	4.43%
5	6/26/2018	3.54%	3.93%	4.17%	4.40%
6	7/13/2018	3.35%	3.78%	4.10%	4.43%
7	9/21/2018	3.37%	3.83%	4.19%	4.48%
8	10/1/2018	3.54%	3.99%	4.29%	4.51%
9	10/4/2018	3.39%	3.88%	4.28%	4.60%
10	10/11/2018	3.43%	3.86%	4.17%	4.41%
11	10/24/2018	3.21%	3.74%	4.16%	4.42%
12	10/31/2018	3.31%	3.77%	4.10%	4.37%

4.3. KCRV and Degrees of Equivalence Based on Multiple Calibration Samples

With the aid of the overall uncertainty results of the TOA reflectance and the calibration accuracy of Sentinel-2B/MSI (approximately 5%), the uncertainty of the relative differences between the simulation and observation for each sample in bands 2, 3, 4, and 8 of Sentinel-2B/MSI can be calculated (see Table 7). Subsequently, according to the calibration results for 12 samples and their corresponding uncertainties, the weight coefficient for each sample, KCRV, and degrees of equivalence are determined using the aforementioned method. In this process, the cutoff uncertainty agreed upon by CCPR is given as 6.10%, 6.32%, 6.52%, and 6.64% for bands 2, 3, 4, and 8 of Sentinel-2B/MSI, respectively. After adjusting by the cutoff uncertainty, the calibration uncertainty for each sample (see Table 8) is used to calculate the weight coefficient according to Equation (7). The weight coefficients for 12 calibration samples in the four bands of Sentinel-2B/MSI are given in Table 9. The comparison of Table 6 with Table 9 reveals that the weight coefficient has a negative correlation with the adjusted validation uncertainties; the greater the adjusted uncertainties are, the smaller the weight coefficients are; the correlation coefficients are approximately -1 in the four bands of Sentinel-2B/MSI.

Subsequently, according to Equations (8) and (9), the weighted average value of the relative differences (y), namely, the KCRV and the corresponding uncertainties ($u(y)$) in each band, are determined and shown in Table 10. The weighted average values of the relative differences are 3.75%, 5.11%, 6.09%, and 5.03% for the four bands, and the uncertainties are 1.84%, 1.87%, 1.90%, and 1.93%, respectively.

Table 7. The uncertainty of the relative difference between the simulation and observation in bands 2, 3, 4, and 8 of Sentinel-2B/MSI.

Sample No.	Date	Band 2	Band 3	Band 4	Band 8
1	3/25/2018	6.58%	6.57%	6.58%	6.66%
2	4/7/2018	6.57%	6.55%	6.57%	6.63%
3	5/14/2018	5.99%	6.25%	6.47%	6.64%
4	6/13/2018	6.64%	6.61%	6.61%	6.68%
5	6/26/2018	6.63%	6.60%	6.59%	6.64%
6	7/13/2018	6.60%	6.58%	6.58%	6.64%
7	9/21/2018	6.01%	6.27%	6.52%	6.70%
8	10/1/2018	6.69%	6.67%	6.68%	6.73%
9	10/4/2018	6.02%	6.31%	6.57%	6.78%
10	10/11/2018	6.04%	6.30%	6.51%	6.65%
11	10/24/2018	6.63%	6.61%	6.61%	6.68%
12	10/31/2018	5.97%	6.24%	6.46%	6.63%

Table 8. The adjusted uncertainties for 12 calibration samples in bands 2, 3, 4, and 8 of Sentinel-2B/MSI.

Sample No.	Date	Band 2	Band 3	Band 4	Band 8
1	3/25/2018	6.58%	6.57%	6.58%	6.66%
2	4/7/2018	6.57%	6.55%	6.57%	6.64%
3	5/14/2018	6.10%	6.32%	6.52%	6.64%
4	6/13/2018	6.64%	6.61%	6.61%	6.68%
5	6/26/2018	6.63%	6.60%	6.59%	6.64%
6	7/13/2018	6.60%	6.58%	6.58%	6.64%
7	9/21/2018	6.10%	6.32%	6.52%	6.70%
8	10/1/2018	6.69%	6.67%	6.68%	6.73%
9	10/4/2018	6.10%	6.32%	6.57%	6.78%
10	10/11/2018	6.10%	6.32%	6.52%	6.65%
11	10/24/2018	6.63%	6.61%	6.61%	6.68%
12	10/31/2018	6.10%	6.32%	6.52%	6.64%

Table 9. The weight coefficients for 12 calibration samples in bands 2, 3, 4, and 8 of Sentinel-2B/MSI.

Sample No.	Date	Band 2	Band 3	Band 4	Band 8
1	3/25/2018	0.0785	0.0810	0.0831	0.0837
2	4/7/2018	0.0788	0.0814	0.0835	0.0841
3	5/14/2018	0.0913	0.0876	0.0847	0.0842
4	6/13/2018	0.0772	0.0800	0.0823	0.0832
5	6/26/2018	0.0774	0.0803	0.0828	0.0841
6	7/13/2018	0.0779	0.0807	0.0832	0.0841
7	9/21/2018	0.0913	0.0876	0.0848	0.0827
8	10/1/2018	0.0760	0.0786	0.0807	0.0818
9	10/4/2018	0.0913	0.0876	0.0833	0.0808
10	10/11/2018	0.0913	0.0876	0.0847	0.0839
11	10/24/2018	0.0774	0.0801	0.0823	0.0832
12	10/31/2018	0.0913	0.0876	0.0847	0.0841

Table 10. The KCRV and its uncertainty in bands 2, 3, 4, and 8 of Sentinel-2B/MSI.

	Blue Band	Green Band	Red Band	NIR Band
y	3.75%	5.11%	6.09%	5.03%
$U(y)$	1.84%	1.87%	1.90%	1.93%

Furthermore, in this study, it is assumed that the 12 calibration samples are independent from each other. A chi-squared test is applied to check the independence of the results obtained. The observed chi-squared values χ_{obs}^2 are 2.89, 4.98, 7.20, and 4.66, and there are 11 degrees of freedom. According to the chi-squared test condition, this chi-squared test is successful, which means that the 12 calibration samples are actually independent from each other.

Finally, the degrees of equivalence for the 12 calibration samples and the corresponding uncertainties are given in Tables 11 and 12, respectively. Note that samples 2–4, 7, 8, 11, and 12 are much closer to the KCRV value than the other samples, and the degrees of equivalence of these eight samples are lower than 0.05, especially for sample 3.

Table 11. The degrees of equivalence for 12 calibration samples in bands 2, 3, 4, and 8 of Sentinel-2B/MSI.

Sample No.	Date	Band 2	Band 3	Band 4	Band 8
1	3/25/2018	0.0200	0.0420	0.0548	0.0273
2	4/7/2018	0.0324	0.0297	0.0085	0.0410
3	5/14/2018	0.0148	0.0299	0.0303	0.0230
4	6/13/2018	0.0123	0.0399	0.0603	0.0487
5	6/26/2018	0.0218	0.0247	0.0567	0.0146
6	7/13/2018	0.0251	0.0407	0.0925	0.1013
7	9/21/2018	0.0004	0.0384	0.0375	0.0187
8	10/1/2018	0.0206	0.0027	0.0046	0.0412
9	10/4/2018	0.0711	0.0394	0.0795	0.0348
10	10/11/2018	0.0366	0.0899	0.0459	0.0061
11	10/24/2018	0.0210	0.0142	0.0037	0.0270
12	10/31/2018	0.0333	0.0414	0.0426	0.0297

Table 12. The uncertainties of the degrees of equivalence for the 12 calibration samples in bands 2, 3, 4, and 8 of Sentinel-2B/MSI.

Sample No.	Date	Band 2	Band 3	Band 4	Band 8
1	3/25/2018	6.32%	6.30%	6.30%	6.37%
2	4/7/2018	6.30%	6.28%	6.29%	6.35%
3	5/14/2018	5.81%	6.04%	6.24%	6.35%
4	6/13/2018	6.38%	6.34%	6.34%	6.39%
5	6/26/2018	6.36%	6.33%	6.31%	6.36%
6	7/13/2018	6.34%	6.31%	6.30%	6.36%
7	9/21/2018	5.81%	6.04%	6.23%	6.41%
8	10/1/2018	6.43%	6.40%	6.40%	6.45%
9	10/4/2018	5.81%	6.04%	6.29%	6.50%
10	10/11/2018	5.81%	6.04%	6.24%	6.37%
11	10/24/2018	6.36%	6.34%	6.34%	6.39%
12	10/31/2018	5.81%	6.04%	6.24%	6.35%

5. Discussion

In this study, the field calibration is used to evaluate the radiometric performance of Sentinel-2B/MSI, and first the difference between the simulated radiance and the observed value for each sample is calculated. However, since the field calibration is affected by the coupled effects of surface and atmospheric characteristics, the relative differences acquired under different surface and atmospheric conditions differ from each other and have different uncertainties, making it difficult to compare the consistency among multiple samples. In view of this, a new method for estimating a weighted average value from multiple calibration results is proposed according to the CCPR Key Comparison K1-a, and this value is accepted as the KCRV 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. In this method, the weight coefficient for each sample is determined according to its corresponding uncertainty; the derived KCRV has very low uncertainty and becomes close to the “true” value when the number of calibration samples increases. This value could also be used as a reference value for conducting a consistency comparison of the calibration results from multiple samples. The relative difference for each sample is compared with the KCRV, and the degrees of equivalence are calculated. The lower degrees of equivalence are, the more accurate the sample is. This method is preliminarily applied to Sentinel-2B/MSI using the desert target at the Baotou site, as their calibration results are different and independent from each other since they have obviously different atmosphere characteristics of the study date.

Due to the limitation on the matchup of satellite observations and in situ measures, only 12 calibration samples are used in this study. These samples are concentrated in the warm months of March

to October since the Baotou site has very cold and snowy winters, and the automatic measurement system does not work very well at low air temperatures. In the future, many more calibration samples over different surface types and synchronized ground measurements will be required, since the KCRV sensitivity depends on the number of calibration samples, and the uncertainty of the KCRV will decrease. This result indicates that the KCRV will be closer to the “true” value when more samples participate in the field calibration. To analyze the influence of the number of calibration samples on the uncertainty of the KCRV, we randomly select different numbers of calibration samples to determine the uncertainty of the KCRV, and the results are shown in Figure 16. The uncertainty of the derived KCRV decreases dramatically with the increase in sample number, which means that the derived KCRV becomes much closer to the “true” value as the sample number increases.

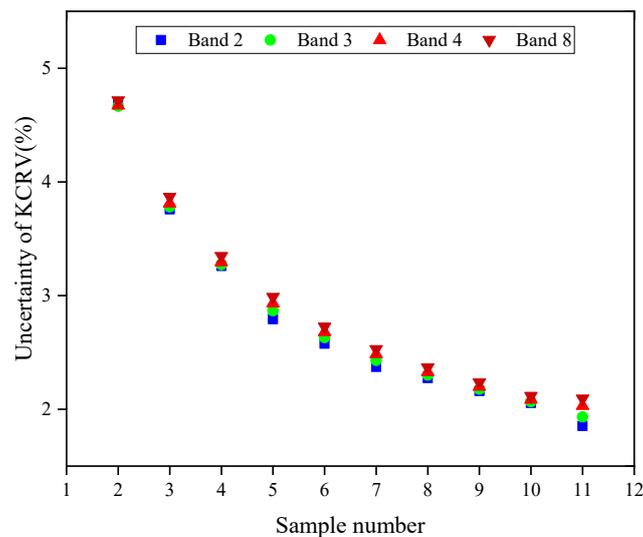


Figure 16. The uncertainty of the KCRV derived with different calibration samples.

Moreover, a case study is performed using these datasets at the Baotou site to test the proposed method for determining the KCRV from multiple calibration results, and this method could also be applied to other instrumented calibration sites in the framework of CEOS, especially RadCalNet, since the surface and atmospheric characteristics around the automatic measurement systems are included in RadCalNet sites, such as the La Crau site, the Railroad Valley site, and the Gobabeb site, and the uncertainty associated with the BOA and TOA reflectance is provided on the RadCalNet portal for each site. This situation makes it very feasible and effective to provide the input parameters for determining the KCRV by combining all the calibration results from RadCalNet sites. In the future, we will intend to apply this method to Sentinel-2B/MSI using the four demonstrated sites of RadCalNet so that the proposed method will be tested and can be improved. This test is already in our work schedule and will be carried out very soon.

6. Conclusions

Evaluating the radiometric performance of in-orbit sensors using field calibration during their long-term operational phase via ground targets is an essential quality assurance step that needs to be performed regularly. Field calibration using multiple samples under different surface and atmospheric conditions can be an effective way to reduce the uncertainty of extrapolating parameter estimates from a single-sample calibration. Due to the different surface and atmospheric characteristics, significant differences in the calibration results and the associated uncertainties for those multiple samples are generated. In view of this, a reference value, which approaches the “true” value as much as possible, should be derived and used as a bridge for comparing the consistency of those calibration results.

This paper proposes a novel weighted average method for evaluating the sensors' radiometric performances using multiple samples under different atmospheric and surface conditions, and a reference value close to the corresponding "true" value, namely, the KCRV, is calculated for comparing each calibration result. In this method, the weight coefficient assigned to each sample is calculated according to the calibration uncertainty of each sample, and a single KCRV is calculated as a weighted mean value when the calibration samples are independent from each other. The KCRV would be a state-of-art reference value for performing the consistency comparison among various calibration samples, and the degrees of equivalence for each sample are calculated. The lower degrees of equivalence are, the more accurate the sample is.

To test the method, it is also applied to the field calibration of Sentinel-2B/MSI using the desert target at the Baotou site, China. A single KCRV with low uncertainty for the 12 calibration samples is derived for bands 2, 3, 4, and 8 of Sentinel-2B/MSI. The KCRVs of the relative differences between the simulated and observed TOA reflectances are 3.75%, 5.11%, 6.09%, and 5.03% for the four bands, and the uncertainties are 1.84%, 1.87%, 1.90%, and 1.93%, respectively. Subsequently, the calibration results of the 12 samples are compared with derived KCRV, and samples 2–4, 7, 8, 11, and 12 are closer to the KCRV value than are the other samples.

Moreover, to thoroughly validate this method, RadCalNet sites will also be used to further increase samples so that the KCRV that is derived by combining all the calibration results from RadCalNet sites can approach the "true" value as much as possible.

Author Contributions: C.G. and C.L. conceived and designed the experiments; Y.L. and J.L. performed the experiments and uncertainty analysis; L.M., Z.W., and Y.Z. contributed the calibration; Q.H. contributed the Sentinel-2B/MSI satellite data processing; S.Q., E.Z., Y.Q., and N.W. contributed to the acquisition of in situ data; C.G. wrote the paper. All authors have read and agreed to the published version of the manuscript.

Funding: The work has been supported by the National Key R&D Program of China under Grants 2018YFB0504800, 2018YFB0504804, by the National Natural Science Foundation of China under Grants 41671370, 41601398, and 6170011903, by the Bureau of International Co-operation Chinese Academy of Sciences under Grant 181811KYSB20160040, and by the Priority Research Program of Chinese Academy of Sciences under Grant XDA13030402.

Acknowledgments: The authors would like to thank the European Space Agency for providing Sentinel-2B/MSI satellite data. Thanks are also given to the anonymous reviewers.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Chander, G.; Markham, B. Revised landsat-5 TM radiometric calibration procedures and postcalibration dynamic ranges. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 2674–2677. [[CrossRef](#)]
2. Li, Z.L.; Tang, B.H.; Wu, H.; Ren, H.; Yan, G.; Wan, Z.; Trigo, I.F.; Sobrino, J.A. Satellite-derived land surface temperature: Current status and perspectives. *Remote Sens. Environ.* **2013**, *131*, 14–37. [[CrossRef](#)]
3. Duan, S.B.; Li, Z.L.; Tang, B.H.; Wu, H.; Tang, R.L. Generation of a time-consistent land surface temperature product from MODIS data. *Remote Sens. Environ.* **2014**, *140*, 339–349. [[CrossRef](#)]
4. Duan, S.B.; Li, Z.L.; Li, H.; Göttsche, F.M.; Wu, H.; Zhao, W.; Leng, P.; Zhang, X.; Coll, C. Validation of Collection 6 MODIS land surface temperature product using in situ measurements. *Remote Sens. Environ.* **2019**, *225*, 16–29. [[CrossRef](#)]
5. Wu, A.; Angal, A.; Xiong, J.; Cao, C. Using CEOS reference standard test sites to track the calibration stability of NOAA-19 AVHRR reflective solar channels. In Proceedings of the SPIE, Toulouse, France, 22–23 September 2010.
6. Czaplá-Myers, J.; Mccorkel, J.; Anderson, N.; Biggar, S. Earth-observing satellite intercomparison using the radiometric calibration test site at Railroad Valley. *J. Appl. Remote Sens.* **2017**, *12*, 012004. [[CrossRef](#)]
7. White, D.R. On the analysis of measurement comparisons. *Metrologia* **2004**, *41*, 122–131. [[CrossRef](#)]
8. Khlevnoy, B. Final report on CCPR-S1: Spectral radiance 220nm to 2500nm, Metrologia. *Metrologia* **2008**, *45*, 02001. [[CrossRef](#)]

9. Emma, W.; Cox, M.; Nigel, F.; Peter, H. Final report on CIPM: Key comparison K1-a spectral irradiance 250 nm to 2500 nm. *Metrologia* **2006**, *43*, 02003.
10. Bouvet, M.; Thome, K.; Berthelot, B.; Bialek, A.; Czaplá-Myers, J.; Fox, N.P.; Goryl, P.; Henry, P.; Ma, L.; Marcq, S.; et al. RadCalNet: A radiometric calibration network for Earth observing imagers operating in the visible to short wave infrared spectral range. *Remote Sens.* **2019**, *11*, 2401. [[CrossRef](#)]
11. Main-Knorn, M.; Pflug, B.; Louis, J.; Debaecker, V.; Müller-Wilm, U.; Gascon, F. Sen2Cor for Sentinel-2. In Proceedings of the SPIE, San Diego, CA, USA, 6–10 August 2017.
12. Liu, Y.K.; Li, C.R.; Ma, L.L.; Wang, N.; Qian, Y.G.; Tang, L.L. An automatic reflectance-based approach to vicarious radiometric calibrate the Landsat8 operational land imager. In Proceedings of the IGARSS, Fort Worth, TX, USA, 23–28 July 2017; pp. 4699–4702.
13. Liu, Y.K.; Ma, L.L.; Wang, N.; Qian, Y.G.; Shi, Q.; Li, C.R.; Tang, L.L. Vicarious radiometric calibration/validation of Landsat-8 operational land imager using a ground reflected radiance-based approach with Baotou site in China. *J. Appl. Remote Sens.* **2017**, *11*, 044004. [[CrossRef](#)]
14. Roujean, J. A bidirectional reflectance model of the Earth's surface. *J. Geophys. Res.* **1992**, *97*, 455–468. [[CrossRef](#)]
15. Guide to the Expression of Uncertainty in Measurement (GUM)—Supplement 1: Numerical Methods for the Propagation of Distributions. 2004. Available online: http://www.ufrgs.br/medterm/areas/area-i/ISO_GUM_sup1.pdf (accessed on 25 September 2018).
16. Cox, M.; Harris, P. An outline of supplement 1 to the guide to the expression of uncertainty in measurement on numerical methods for the propagation of distributions. *Meas. Tech.* **2005**, *48*, 336–345. [[CrossRef](#)]
17. Couto, P.R.G.; Damasceno, J.C.; de Oliveira, S.P. Monte Carlo simulations applied to uncertainty in measurement. In *Theory and Applications of Monte Carlo Simulations*; Chan, V.W.K., Ed.; IntechOpen: London, UK, 2013. [[CrossRef](#)]
18. Liu, Y.K.; Ma, L.L.; Wang, N.; Zhao, Y.G.; Gao, C.X.; Li, C.R.; Tang, L.L. Uncertainty analysis of vicarious radiometric calibration of optical sensor using a Monte Carlo statistical approach. In Proceedings of the SPIE, San Diego, CA, USA, 11–15 August 2019.
19. Boris, K.; Victor, S.; Bernard, R.; Charles, G.; Howard, Y.; Arnold, G.; Dieter, T.; Juergen, H. CCPR-S1 Supplementary comparison for spectral radiance in the range of 220nm to 2500nm. *Metrologia* **2009**, *46*, S174–S180.
20. Cox, M.G. The evaluation of key comparison data. *Metrologia* **2002**, *39*, 589–595. [[CrossRef](#)]
21. Kokoska, S.; Nevison, C. Statistical tables and formulae. In *Critical Values for The Chi-Square Distribution*; Springer: New York, NY, USA, 1989.
22. Li, L.; Dai, C.H.; Wu, Z.F.; Wang, Y.F. Laboratory calibration and temperature research of spectral radiometric instruments. *Spectrosc. Spect. Anal.* **2019**, *39*, 1965–1969.
23. Holben, B.N.; Eck, T.F.; Slutsker, I.; Tanre, D.; Buis, J.P.; Setzer, A.; Vermote, E.; Reaga, J.A.; Reagan, Y.J.; Kaufman, Y.J.; et al. AERONET—A federated instrument network and data archive for aerosol characterization. *Remote Sens. Environ.* **1998**, *66*, 1–16. [[CrossRef](#)]
24. Berk, A.; Anderson, G.; Acharya, P.; Shettle, E. *MODTRAN5 5.2.1 User's Manual*; Spectral Sciences Inc.: Burlington, MA, USA, 2011.
25. Thuillier, G.; Hers, M.; Simon, P.C.; Labs, D.; Mandel, H.; Gillotay, D.; Foujols, T. The solar spectral irradiance from 200 to 2400 nm as measured by the SOLSPEC spectrometer from the ATLAS and EURECA missions. *Sol. Phys.* **2003**, *214*, 1–22. [[CrossRef](#)]
26. Thuillier, G.; Hers, M.; Simon, P.C.; Labs, D.; Mandel, H.; Gillotay, D.; Foujols, T. The visible solar spectral irradiance from 350 to 850 nm as measured by the SOLSPEC spectrometer during the Atlas 1 mission. *Sol. Phys.* **1998**, *177*, 41–61. [[CrossRef](#)]

