Radiometric Normalization of Multi-Temporal Landsat and Sentinel-2 Images Using a Reference MODIS Product through Spatiotemporal Filtering

Wenxia Gan, *member*, *IEEE*, Hessah Albanwan, and Rongjun Qin, Senior *Member*, *IEEE*

***Abstract*—Radiometric normalization is an essential preprocessing step for almost all remote sensing applications such as change detection, image mosaic and 3D reconstruction. This paper proposes a novel radiometric normalizing method based on spatiotemporal filtering using a reference MODIS product. This differs from traditional RRN (Relative radiometric normalization) methods in two-folds: first, the number of reference images is more than one which introduces more complexities than RRN with a single reference image; second, the resolution of MODIS product is significantly lower thus requiring the algorithms to accommodate scale differences. To address, our approach extends the traditional spatiotemporal filtering method with per image bias that representing both internal (e.g. sensor characteristics) and external (e.g. atmosphere and topography) against the reference data. In addition, we use the Kullback-Leibler divergence metric to statistically determine the resemblance degree between the temporal images for weighting. We applied our proposed method to normalize Landsat OLI, ETM+, and Sentinel MSI using MODIS NBAR product, covering two study areas of 30 × 15 km^2 and 32 × 52 km^2 respectively , and we show a notable radiometric consistency over both temporal and spatial dimension after the processing through three comparative experiments with state-of-the-art methods: 1) 3-7% improvement in the contexts of transfer learning which favors only images with consistent radiometric properties; 2) mosaic results using our processed images shows no apparent seamlines as compared with images processed by other methods;**

***Index Terms*—Inter-sensor normalization, Radiometric consistency, Wavelet Transform, Landsat 8 OLI, Landsat 7 ETM+, Sentinel MSI, Kullback-Leibler divergence, Bias Term, Transfer learning Classification**

I. Introduction

RADIOMETRIC normalization is an essential preprocessing step for many remote sensing applications such as change detection, image mosaic and 3D reconstruction, etc.[1][2], [3]. In general, there are two main categories of radiometric normalization: 1) absolute radiometric normalization (ARN, interchangeable with Absolution Radiometric Correction for multitemporal images), and 2) Relative Radiometric Normalization (RRN) based on respectively whether or not the absolute/global reflectance measures are needed as the desired output[4][5]. ARN methods requires information such as sensor responses, radiometric calibration coefficients, viewing angles, sun angles, atmospheric conditions, topography data and in-situ data[6],[7],[8], which is oftentimes unavailable. On the contrary, RRN methods do not require prior information such as weather or aerosol depths; it corrects the images using a single reference image, and requires the reference image to be noise-free and spectrally well-balanced [6], [9], [10]. Recent studies have suggested that integrating ARN and RRN for radiometric calibration can effectively achieve absolute and consistent normalization [2][11]. On one hand, ARN methods are able to correct specific types of noises of individual images such as atmospheric noises, viewing angle-induced bias, topography-induced bias, while being considered as the most rigorous solution for radiometric correction, it does suffer from modeling errors introduced by the fact that a single or multiple corrections models are not able to comprehensively cover the varying images under other unknown sensory or environmental conditions. As a result, the ARN processed images remain to be temporally inconsistent. On the other hand, RRN methods with the aim to homogenize spectral responses across temporal images do not demand for measures and thus can only support limited applications, e.g. regional change detection and qualitative spatiotemporal analyses, thus are much less demanding in terms of the needed in-situ data. Integrating both ARN and RRN for processing remote sensing raw images are obviously advantageous to achieve products that can be used for a wider scope of applications.

However, a simple sequential application of both methods is

(Corresponding author: Rongjun Qin, qin.324@osu.edu)

Wenxia Gan is with the School of Civil Engineering and Architecture, Wuhan Institute of Technology, China. The work was performed while she was a visiting scholar in the Department of Civil, Environment and Geodetic Engineering, the Ohio State University (e-mail: ganwenxiagw@gmail.com).

Hessah Albanwan., is with the Department of Civil, Environment and Geodetic Engineering, The Ohio state University, Columbus, USA (e-mail: lbanwan.1@buckeyemail.osu.edu).

Rongjun Qin is with the Department of Civil, Environment and Geodetic Engineering and the Department of Electrical and Computer Engineering, and Translational Data Analytics Institute, The Ohio state University, Columbus, USA (email: qin.324@osu.edu).

potentially problematic: first, the effort of acquiring in-situ data is still needed for model-inversion based ARN; second, RRN methods are often developed to accommodate scenarios in which one reference image is used, while only selecting one out of multiple ARN-corrected reference images can potentially produce propagated errors. Meanwhile, there are RRN studies perform the inter-sensor RRN by developing a global fixed linear model based on a large number of archived synchronous images from the multi-sensors to transform the TOA reflectance or surface reflectance, or night-time light data of one sensor to another[12], [13][14], which obviously can be problematic to accommodate local variations. Therefore, normalizing the images using pre-processed reflectance products can be potentially more viable, as at least the sensor specific variations have been pre-accommodated when these data are converted to a surface reflectance product following a physical-based correction procedure. Since obtaining corresponding in-situ auxiliary data (i.e. weather, aerosol depth etc.) at a high resolution for ARN is potentially impractical, we consider to utilize low-resolution and standard reflectance product such as Nadir BRDF-adjusted reflectance (NBAR) product from Moderate Resolution Imaging Spectroradiometer (MODIS) as the reference images, which brings clear advantages and challenges: the advantage is that such data are often for global coverage and has very high temporal resolution (i.e. on a daily basis) and with good radiometrical consistency and continuity over both the space and time dimensions, while the challenges are obviously the resolution being a factor of 20 less than high-resolution images such as Landsat and Sentinel-2 and thus made it hard to perform accurate inter-sensor radiometric normalization[15]. In this paper, we address these challenges by proposing a novel spatiotemporal filtering model that extends a traditional spatiotemporal filtering method [16] in two ways: 1) we have incorporated a per image bias which accommodate corrections from multiple reference images and 2) use the Kullback-Leibler divergence(KL divergence) metric to statistically determine the resemblance degree between the temporal images for optimal weight determination. This model inherits the non-parametrization nature of the spatiotemporal filtering method to accommodate local variations, hence this new model may be able to yield images cross different sensors with accurate reflectance to the level of well-calibrated MODIS dataset, as well as temporally consistent results for various remote sensing image processing and applications such as change detection, classification and mosaics. Our contribution of this work is mainly two-fold: 1) we have proposed novel method that extend the existing spatiotemporal filtering method by incorporating a per-image bias term to accommodate systematic radiometric corrections using multiple low-resolution reference images; 2) we have experimentally demonstrated that the proposed method achieves a leveraged relative consistency and global consistency over state-of-the-art methods, through both spectral analysis, transfer learning, and global mosaicking applications.

The rest of the paper is organized as follows: Section 2 briefly introduces relevant works related to radiometric normalization; section 3 describes the general methodology of the proposed work; section 4 presents the experimental results and performs quantitative analysis and evaluation through typical remote sensing applications and section 5 concludes this paper by discussing its pros and cons.

II. Related Works and Consideration

A. *Related works*

As mentioned in Section I (Introduction), both ARN and RRN are two major categories of methods in radiometric normalization. In-situ measurements are crucial for absolute algorithms [6], [17]. RRN algorithms (i.e. RRN) tend to homogenize images to a reference image to remove all the spectral variations unrelated to the land surface change using mathematical models. It is considered to be a solution for applications that does not require absolute reflectance measures, whereas the need for absolute reflectance is still essential for global scale applications. Studies have shown that exploiting the spatial and temporal information can provide accurate radiometric information of targeted objects in the scene[16]. [16] developed a 3D spatiotemporal filtering algorithm to utilize the temporal images to enhance radiometric properties and consistency among images. Their method eliminates the necessity to have a reference image to normalize the subject images, and overall, they showed improvements in the temporal consistency of data and noise and artifacts reduction by 15% in their experiments[16]. Their method models the consistency by weighting the measure of spectral differences in the temporal direction with respect to the reference image, which selectively take the temporal images to homogenize the spectral information in the image stack, this mathematically modeled solution however, minimize the discrepancy wherever possible without incorporating the absolute measures of the radiance. As a result, improperly weighted parameters may simply homogenize all spectrums saturating all possible seasonality and phenological differences between the images [2], [6], [18][19].

Another line of work considers RRN cross different sensory data, often known as inter-sensor calibration, which focuses on calibrating sensors by correcting the pixel values using the radiometric calibration coefficient based on some reference sensory data[20], [21]; however, it is still regarded as a relative correction as it does not address the uncertainties from the ambient (systematic) differences between sensors. The sensor-to-sensor bias can be somehow addressed by determined bias through prediction-based models[20], [22], which however require at least a few of the overlapped images to be captured as the same time, which generates limitations of such methods, and in addition, the bias modeling can be subject to modeling errors, as the sensor-to-sensor bias can be subject to many sources such as the sensor responses, time dependent and location dependent factors, which may be difficult to predict using a few overlapping inter-sensor images captured at the same time.

It was shown that integrating both absolute and relative approaches can provide better results in terms of improving radiometric consistency and minimizing distortions [2],

[23][19], [24][25][11]. Most approaches apply atmospheric correction as the first step for ARN to get reliable Pseudo-invariant features (PIFs). PIFs being features that present reliable holders of spectrums that often do not change (e.g. concrete surfaces), and this is followed by a feature-based RRN [2], [24], [26][11]. These methods vary with the ARN and RRN methods: for instance, [4] suggested using an atmospherically corrected reference image to normalize every band in the target images individually using PIFs. Similarly, [27] improved PIFs selection for the RRN by first conducting an absolute correction using dark object subtraction (DOS); [11] proposed a method called "mixed radiometric normalization (MRN)", where they firstly used fast line-of-sight atmospheric analysis of spectral hypercubes (FLAASH) to convert from DN to surface reflectance, and then used iteratively reweighted multivariate alteration detection (IR-MAD) to radiometrically normalize and scale the temporal images.

B. Considerations in our proposed approach

Satellite data such as MODIS recording data with global coverage on an almost daily basis have been very helpful to provide comprehensive spectral information about the Earth, as well as their imaging conditions (e.g. atmosphere, slope, viewing angles) [21]. Because it has a very frequent revisiting time, and there exist the best available in-situ data with equivalently or lower resolution, the MODIS coarse resolution reflectance product (MODIS Nadir BRDF-adjusted reflectance (NBAR) product) have trustworthy stability, accuracy and temporal consistency in radiance, which can potentially serve as a reliable reference image for radiometric normalization for other satellite imageries such as Landsat[28]. However, a direct application of existing RRN or inter-sensor calibration methods may produce artifacts due to the large resolution differences, as both traditional down-scaling and transformation-based methods (e.g. wavelet transform [29]) may generate artifacts such as aliasing and blocking artifacts[30][31]. The sensor-level biases may be subject to two major sources: 1) sensor specific biases such as static spectrum response and 2) time-varying inter-sensor biases that may be depending on specific scenes, objects and in-situ variables that are not measurable. Therefore, the algorithms taking low-resolution ARN products (i.e. MODIS Nadir BRDF-adjusted reflectance (NBAR) product) for normalizing high resolution data (e.g. Landsat and Sentinel), must consider components addressing the large resolution differences, as well as components considering leveraging both relatively radiometric consistency between temporally neighboring images and overall consistency with respect to the low-resolution ARN product.

III. METHODOLOGY

Our proposed work is described in Fig. 1, in which we specified the possible data to be normalized (but not limited to). Our method starts with preparing and preprocessing both low resolution ARN product (MODIS NBAR) and the to-be-normalized data (e.g. Landsat 7 Enhanced Thematic Mapper plus(ETM+), Landsat 8 Operational Land Imager(OLI), and Sentinel-2 Multispectral Instrument(MLI)), components shown in white boxes, in which we used the discrete wavelet analysis (DWT) for data down-scaling. The light-grey boxes indicate works associated with our proposed spatiotemporal filtering method and dark-grey boxes implies evaluation and validation of our work.

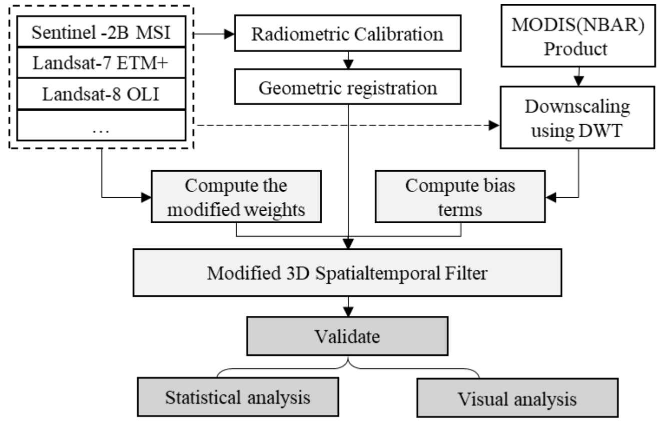


Fig. 1. Flow chart of the proposed method.

A. *Data pre-processing*

The data preprocessing involves several steps:

- *Geo-registration and initial radiometric calibration*: Since the images are collected from varying sources, we perform the registration for geometric alignment. The data from MODIS, Sentinel, Landsat ETM+ are geo-registered using the Landsat-8/OLI data of which the Level 1 product accuracy is geometrically corrected with terrain correction and a global sample of ground control points (L1T)[32]; All the high - resolution images are geometrically co-registered using the Image Registration workflow tool of ENVI 5.3 software [33] and with the residuals reported as 14m, and the MODIS data are co-registered based on the image-based control points using ArcMap 10.6 software [34]. Meanwhile, the Sentinel-2B MSI images are resampled to 30m as same as the Landsat-8/OLI image. When TOA reflectance is not available, the initial radiometric calibration is performed to convert the DN for all satellites to TOA reflectance using the calibration coefficients provided by the satellite sensors.

- *2D-Discrete Wavelet transformation (DWT)*: The 2-D Discrete Wavelet Transformation-based decomposition and reconstruction is utilized to downscale the coarse resolution MODIS product to 30m to be same as the Landsat OLI data, where the fine-resolution data is used to provide the spatial details. The procedure is composed of the following steps and performed using MATLAB program: first, we perform multi-level decomposition, where we use the fine-resolution image to obtain the approximation [low–low (LL)] and detail [high–low (HL), low–high (LH), and high–high (HH)] components. The decomposition level is set as $\log_2(ratio)$, determined by the spatial resolution ratio of the fine resolution data and the coarse resolution data, which is a factor of 4 in our work as the spatial resolution ratio is approximately 16 for 30-meter resolution data, either Landsat or resampled Sentinel image, since the majority of MODIS's bands have 500 m spatial resolution (4 bands out of the 6), thus, we choose the 500 m as a reference to downscale all bands, i.e. coarse-resolution/fine-resolution=500 m / 30 m≈16. The next step is to replace the component LL by the coarse resolution product, and finally, perform the Wavelet reconstruction using inverse discrete wavelet transform (IDWT) to obtain the downscaled fine-resolution of the MODIS product. For more details on DWT refer to [29].

B. *The proposed algorithm*

Based on our prior developed work[16], we propose a modified version of the 3D spatiotemporal filter, in which our goal is to enhance the temporal consistency at the same time maintaining a relatively accurate reflectance values with respect to the reference MODIS product. As mentioned in Section 2.2, to account for different types of errors (e.g. sensor-specific and time-varying errors), we propose to add a per-image bias term modeling such non-parametric variations, and to account for the large resolution differences between the high resolution image (e.g. Landsat and Sentinel) and low resolution MODIS product, we propose to incorporate KL divergence[35] as a statistical similarity measure for effective weighting in the spatiotemporal filtering.

1) *Non-parametric and per-image bias*

The generic form for correcting the uncertainty in satellite data can be expressed as in (1):

$$I_F = I_p - \varepsilon + \Delta \tag{1}$$

Where, I_F is the corrected image, I_p is the input image, ε is the random noise, and Δ (an image grid with the same dimension of I_p) is the bias correcting term to cover the systematic error (as discussed in Section 2.1.). We at a first step eliminate random noises using the traditional 3D spatiotemporal filtering following the method in [14] as follows:

$$I_p - \varepsilon = \frac{1}{w_q} \sum\nolimits_{q \in S} w_q \, I_q \tag{2}$$

Where w_q is the aggregated weight composed of the spatial, spectral, and temporal weights over the image space S, and I_q is the temporal images processed for each pixel and each band individually. More details on the filter and on the enhanced version of this filter are described in the following section (Section 3.3.B).

Since the input images vary in their spectral and spatial distributions from the reference MODIS product, we use a bias term to correct and match their spectral values to the reference. The bias term can be decomposed into two sources (in (3)); the first bias term Δ_1 is used to model the per-pixel inter-sensor bias(in (4)), and the second bias term Δ_2 (a grid with a constant value) (in (5)) models a per-image bias to leverage the potential mis-match of spatial resolution between the reference image and the image to be corrected, for example, the MODIS and Landsat image have a resolution difference of a factor of 20, and a mere Δ_1 correction might potentially saturate the high resolution Landsat image grid with low-resolution MODIS image grid. The weight λ leveraging both can be empirically determined based on the type of images to be processed.

$$\Delta = (1-\lambda)\,\Delta_1 + \lambda\,\Delta_2 \tag{3}$$

Specifically, Δ_1 is a residual grid that measures the spectral difference between the downscaled reference product (after applying 2D-DWT) and the filtered image $(I_p - \varepsilon)$:

$$\Delta_1 = D_p - (I_p - \varepsilon) \tag{4}$$

Δ_2 is fixed throughout the computation for each high-resolution image and their corresponding low resolution MODIS product (determined as the image captured from the closest date), and is measured by taking the spectral difference between the mean value of the coarse resolution product C_p and the mean value of the filtered image $\left(I_p - \varepsilon\right)$ as follows:

$$\Delta_2 = Mean\left(C_p\right) - Mean\left(I_p - \varepsilon\right) \tag{5}$$

Substituting (3) to (1), we obtain:

$$I_F = I_p - \varepsilon + (1-\lambda)\,\Delta_1 + \lambda\,\Delta_2 \tag{6}$$

By further substituting (2-5) to (6) we obtain:

$$I_F = (1-\lambda)\frac{1}{w_p}\sum\nolimits_{q\in S} w_q\, I_q + \lambda D_p + Mean\left(C_p\right) - Mean\left(\frac{1}{w_p}\sum\nolimits_{q\in S} w_q\, I_q\right) \tag{7}$$

2) The enhanced spatiotemporal filtering algorithm

The 3D spatiotemporal filtering method proposed by [16] is a typical RRN method; it utilizes the temporal images to radiometrically calibrate the images and eliminate the noise and random errors (refer to (2)). The aggregated weight w_q in (2) in the filter is described in the subsequent equation:

$$w_q = w_{Spatial} \times w_{Band} \times w_{Temporal} \tag{8}$$

The spatial weight can be further characterized using two terms the spatial distance $w_{Spatial\_Distance}$ and the spectral value $w_{spectral\_value}$ between every pixel and its neighboring pixels (see (9)):

$$w_{Spatial} = w_{Spatial\_Distance} \times w_{spectral\_value} \tag{9}$$

Using 2D Gaussian kernel function G (•), we can measure the weight in the spatial and spectral dimension to reduce the spatial inconsistency while preserving its spatial detail:

$$w_{Spatial\_Dist} = exp\left(-\frac{(x_p-x_q)^2+(y_p-y_q)^2}{\sigma_{S,D}}\right) \tag{10}$$

$$w_{Spatial\_value} = exp\left(-\frac{\left(I_p(x_p,y_p) - I_q(x_q,y_q)\right)^2}{\sigma_{S,V}}\right) \tag{11}$$

Where (x_p, y_p) are the coordinates of the pixel and (x_q, y_q) are the corresponding neighboring pixels in spatial or temporal dimension, $\sigma_{S,D}$ and $\sigma_{S,V}$ are the spatial and spectral value bandwidths to determine the extent of filtering. Since this filter operates on every band individually, a Delta Dirac weighting function is used to assure operation only on similar bands:

$$w_{Band} = \begin{cases} 1, if\ band(p) = band(q) \\ 0, if\ band(p) \neq band(q) \end{cases} \tag{12}$$

As to the temporal weight, the original filter only computes the weight through measuring the differences between the spectral differences of the centric pixel of the filter, this however is not robust enough to noises caused by potentially the large resolution differences of the original data.

$$w_{Temporal\_V} = exp\left(-\frac{\left((I_{p,t_0}(x_p,y_p) - I_{p,t_i}(x_p,y_p)\right)^2}{\sigma_{T,V}}\right) \tag{13}$$

Where, I_{p,t_0} is the spectral value of pixel (x_p, y_p) in the current image, I_{p,t_i} is the temporal neighboring image, and $\sigma_{T,V}$ refers to the bandwidth of this component in the spatiotemporal filter. To robustify such a measure, we incorporate the patch-based KL divergence and the reflectance value difference are computed to determine the new temporal weight.

$$w_T = w_{Temporal\_V} \times w_{Temporal\_KL} \tag{14}$$

The KL divergence is a mathematical and statistical measure of how one probability distribution is different from another reference probability distribution, and the weight $w_{Temporal\_K\_L}$ is calculated based on a patch centered around every pixel to measure the similarity among different temporal images

$$w_{Temporal\_K\_L} = exp\,(-\frac{KL(I_{p,t_0}, I_{p,t_i})^2}{\sigma_{T,KL}}) \tag{15}$$

IV. Experimental results and analysis

We perform our experiment on two study areas as described in Section IV.A (also See Fig. 2), and validate our results both qualitatively and quantitatively: to demonstrate the global radiometric consistency, we consider a mosaic experiment that stitches data of two regions separately processed by our proposed method to evaluate their seamlines. The quantitative evaluation considers a before-and-after comparative study using our and other existing methods, and the evaluations include 1) sampling analysis of the spectral and temporal consistency; 2) we presume that the well-normalized data will perform better in classification and transfer learning tests, we therefore evaluate the results by analyzing classification and transfer learning practices on data processed by our and other comparable normalization methods including the method of [16] and the IR-MAD [36] based RRN method. Our choice of parameters is empirical for the window size $w = 5$, $\sigma_{S,D} = 3$, $\sigma_{S,V} = 30$, and we inherit the conclusion that $\sigma_{T,V} = 0.2$ is optimal as stated in [16], these parameters remain throughout the experiments. For the modified temporal weight in which we use a patch-based KL divergence weight, the patch size is set to 30 × 30 pixels allowing us to monitor the change in the spectral values around each pixel. The $\sigma_{T,KL}$ is also set as 0.2 through our experiments. The bias terms weights λ in (3) is empirically set as 0.3.

A. *Data description*

Our experimental dataset is consisting of two study areas, the study area-I covers an area of 30 × 15 km^2 in Illinois in U.S. (41.25–41.39 N, 88.37–88.74 W), the study area-II covers an area of 35 × 52 km^2 in Missouri in U.S. (38.22–38.70 N, 90.71–91.13 W), shown in Fig. 2, which both includes a variety of land covers (i.e. water surfaces, forests, impervious surface, Cropland), and the forest in study area-I are mainly evergreens forest, and corn is the primary components of the cropland. We perform our experiments on two regions of the study area individually, respectively outlined in the red and light green boxes. Our dataset includes multi-temporal satellite images from Landsat-7 ETM+, Landsat-8 OLI, and Sentinel-2 MSI[37], [38][39], and the associated MODIS nadir BRDF-adjusted reflectance (NBAR) product of the same days. Table II summarizes detailed information of these sensors. The test datasets are randomly chosen (in our experiment they are from 2016), and their temporal footage covers the growth season thus we are able to evaluate our method on its adaptability to seasonal and phenological changes, and details of data are listed in Table I. To ensure fair comparison with other methods, our method performs normalization only on common bands among different satellite sensors. Table III summarizes the six bands for normalization in our experiments, including Blue, Green, Red, NIR, SWIR-1, and SWIR-2, which we refer these bands in the rest of the paper as B1 to B6, respectively.

TABLE III.
THE OVERLAPPING BANDS USED IN THE EXPERIMENT

	Blue	GREEN	RED	NIR	SWIR-1	SWIR-2
MODIS -NBAR	3	4	1	2	6	7
Landsat-7 ETM+	1	2	3	4	5	7
Landsat-8 OLI	2	3	4	5	6	7
Sentinel-2B MSI	2	3	4	8	11	12

B. *The qualitative analysis*

The normalization results of study-area-I-region-2 and study-area-II (see Section A.) from different methods are shown in Fig. 3 and note unit of the images are in radiance and the images use the same scaling. Comparing the last row (of both Fig. 3(a) and Fig. 3(b)) to the other rows, it can be seen that our proposed method matches much better to the reference images, while the method of (Albanwan and Qin, 2018) and IR-MAD(the Time 5, Landsat-8 OLI image acquired in 20160912 and Time 3, Landsat-8 OLI image acquired in 20160506 are respectively used as reference for study-area-I-region-2 and study-area-II) shows a good consistency in the temporal direction (from left to right), while apparently it averages through all the images and thus lacks fidelity in radiance. Specifically in Fig.3(a), the MODIS image of date 07/26 appears to be an artefact that shows high level of NIR component. Our proposed method preserves such an NIR content with high level of spatial details, while the method of [14] tends to average its magnitude that differs significantly from the MODIS reference image, and the IR-MAD method not able to not able to preserve the reflectance change pattern of either the original image or the MODIS reference data.

We have also performed a mosaic experiment of the two regions in study-area-I to show the radiometric compliance of the two regions after being processed separately. Fig. 4. shows the mosaic output for the raw data, coarse resolution MODIS product, the 3D spatiotemporal filter [14], and our proposed method. These images are simply stitched using the mosaic tool of ENVI 5.3 software without any post-processing such as feathering[33], [40]. There is an apparent seam line in the raw data where it is presented as the TOA reflectance for the two regions (see Fig. 4). Meanwhile, the patches of the MODIS reflectance product in the seamed area looks smoother than the original raw images, this provides a good basis for normalizing high resolution images from different scenes. The results of our proposed methods show a clear advantage to leverage the global radiometric consistency even though images are from different regions and are processed separately.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3069155, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

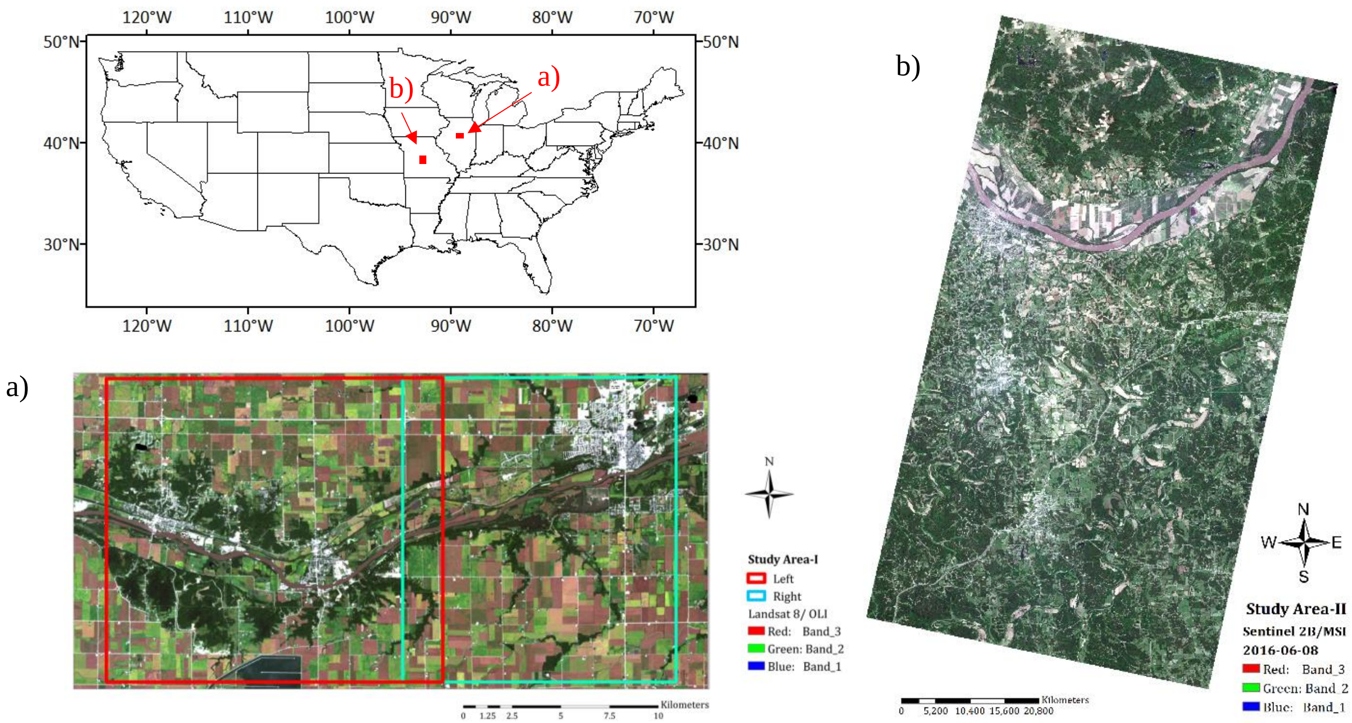


Fig. 2. The study area. a) The study area-I, of which the left side is region 1 and the right side is region 2. b) The study area-II.

TABLE I.
TEMPORAL INFORMATION OF THE DATASET OF STUDY AREA-II

Study-area-I-Region 1 (left image) (500 × 500 pixels)			Study area-I-Region 2 (right image) (500 × 500 pixels)			Study Area-II (1190 × 1740 pixels)		
ID	**Sensor**	**Date**	**ID**	**Sensor**	**Date**	**ID**	**Sensor**	**Date**
			Time:1	Landsat-7 ETM+	20160413	**Time:1**	Landsat-7 ETM+	20160404
			Time:2	Landsat-7 ETM+	20160515	**Time:2**	Landsat-8 OLI	20160412
Time:1	Landsat 8 OLI	20160523	**Time:3**	Landsat-8 OLI	20160523	**Time:3**	Landsat-7 ETM+	20160506
			Time:4	Landsat-8 OLI	20160726	**Time:4**	Sentinel-2B MSI	20160608
Time:2	Landsat 8 OLI	20160912	**Time:5**	Landsat-8 OLI	20160912	**Time:5**	Landsat-8 OLI	20160615
			Time:6	Landsat-7 ETM+	20160920	**Time:6**	Landsat-7 ETM+	20160911
Time:3	Sentinel-2B MLI	20161013	**Time:7**	Sentinel-2B MSI	20161013	**Time:7**	Landsat-8 OLI	20160919
Time:4	Landsat 8 OLI	20161014	**Time:8**	Landsat-8 OLI	20161014	**Time:8**	Landsat-7 ETM+	20160927
			Time:9	Landsat-7 ETM+	20161107	**Time:9**	Landsat-8 OLI	20161021
Time:5	Landsat 8 OLI	20161115	**Time:10**	Landsat-8 OLI	20161115			

TABLE II.
RELEVANT SENSOR AND PRODUCT INFORMATION

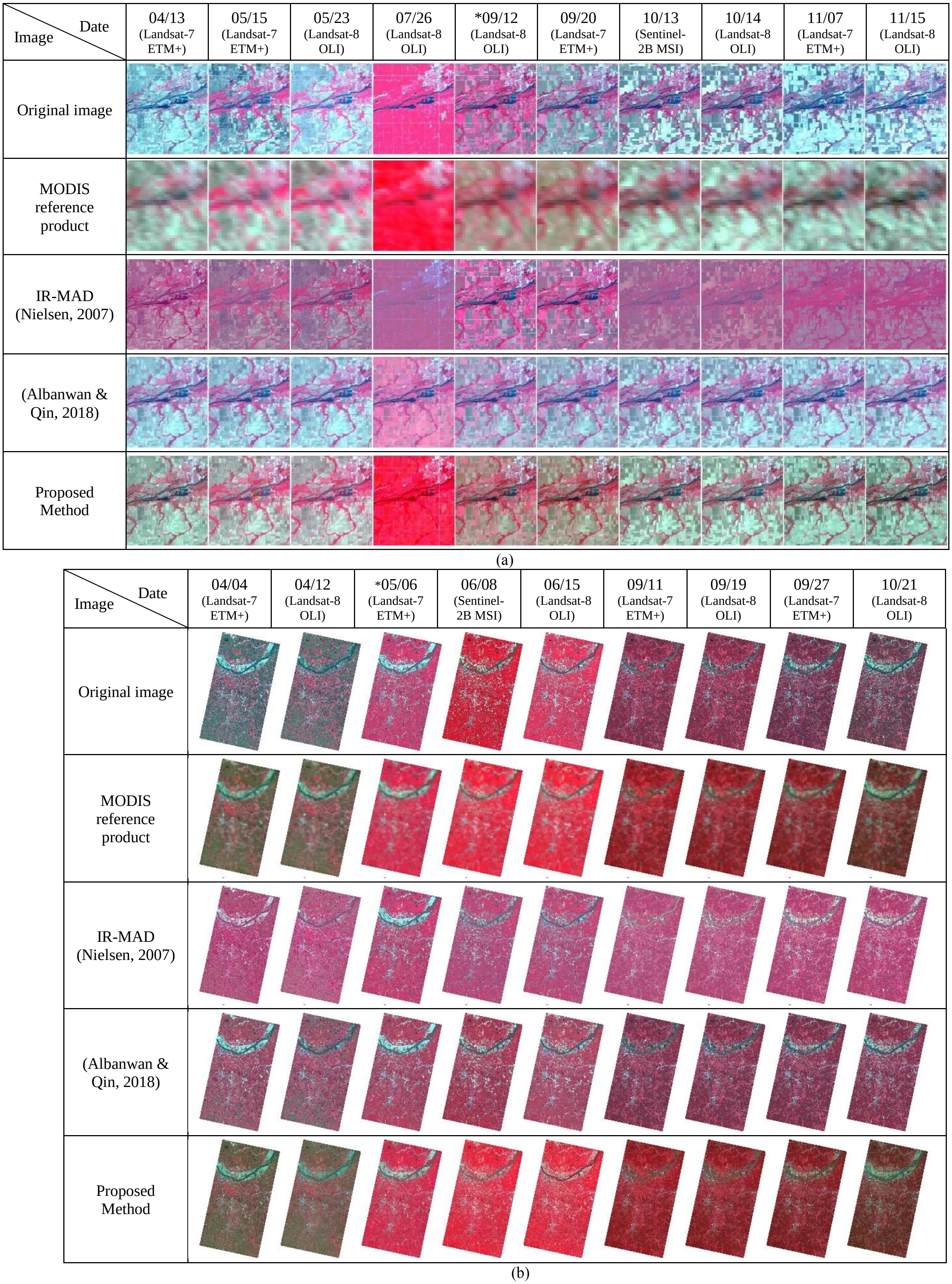
	Landsat TM/ETM+/OLI	**Sentinel-2/MLI**	**MODIS (NBAR) Product**
Spatial resolution	30 (m)	10 - 60 (m) [Bands 1, 9, 10 60(m), Bands 5, 6,7, 8A, 11, 12 20(m), and Bands 2, 3, 4, 8 10 (m)]	500 (m)
Spectral resolution	Landsat 7 ETM+ = 6 (bands) Landsat 8 OLI = 8 (bands)	13 (bands)	7 (bands)
Temporal resolution	16 (days)	10 (days) with one satellite and 5 (days) with 2 satellites	MCD43A4 NBAR product is generated daily via 16 (days) Terra and Aqua MODIS satellite

C. The quantitative and experimental analysis

Our analysis involves the consistency check in the spatial, spectral, and temporal domain. It also includes an accuracy assessment using transfer learning classification of the radiometrically normalized images to evaluate the overall enhancement of the consistency

1) The temporal quality analysis

Fig.5 shows the temporal trend between the reference MODIS product and corresponding original and filtered images



*, the reference image for IR-MAD

Fig. 3. A visual comparison between the original high-resolution data, downscaled MODIS product, the 3D spatiotemporal filter, and our proposed method. Note the image reflectance uses the same scaling so their visuals reflect the absolute reflectance. (RGB: NIR, Red, Green).(a) for the dataset in study-area-I-region 2.(b)for the dataset in study-area-II.

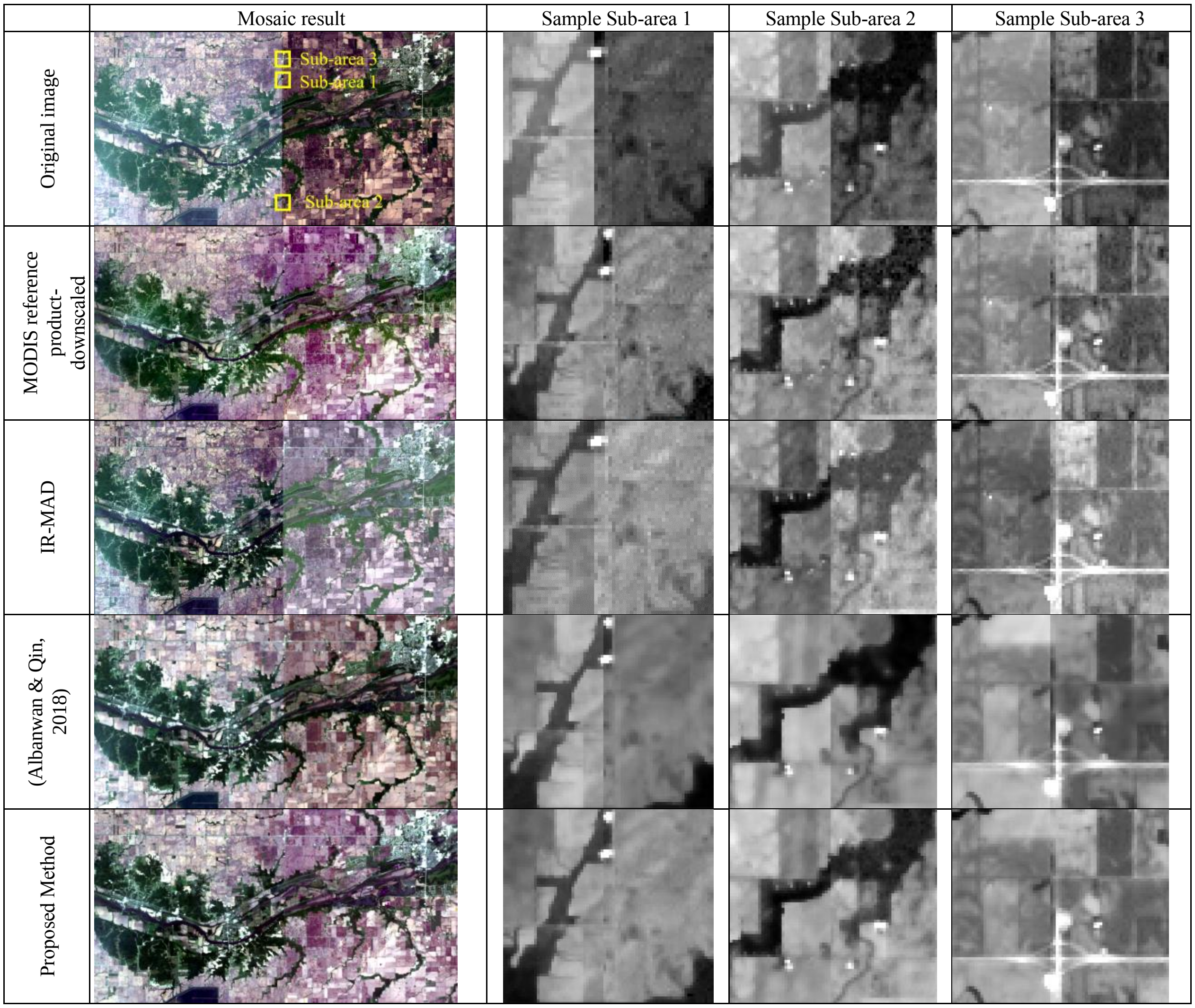
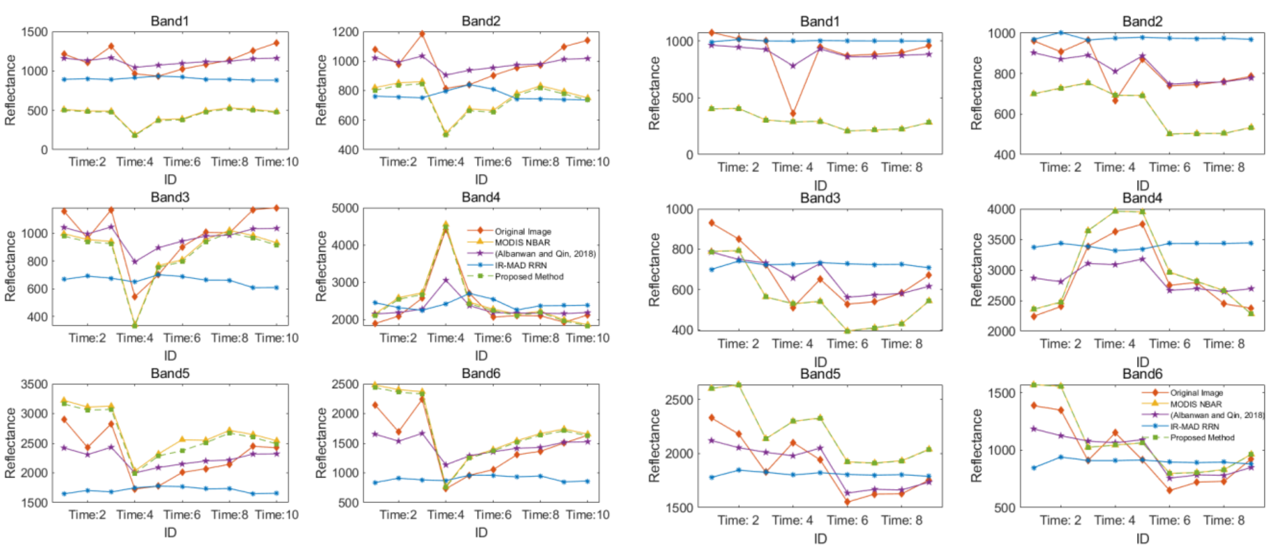


Fig. 4. A visual radiometric consistency comparison using a Mosaic before and after the normalization of region 1 in Time 1 and region 2 in Time 2. Note the image reflectance uses the same scaling so their visuals reflect the absolute reflectance. (RGB: Band 1, 2, 3(Landsat-7 ETM+)/Band 2, 3, 4(Landsat 8/OLI)).



(a) (b)

Fig. 5. The temporal trend comparison between the original image, MODIS's reference data, (Albanwan & Qin, 2018), IR-MAD, and our proposed method for study-area-I-region 2. All figures share the same legend as indicated on the right figures of the second row. Date of the images refer to Table I.(a) for the dataset in study-area-I-region 2.(b)for the dataset in study-area-II.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3069855, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

of study-area-I-region-2 and study-area-II, where the mean reflectance is computed for each band for all times. It can be seen that our proposed method is visually more consistent with the MODIS and at the same time it leverages well the relative temporal consistencies similar to relative methods, while which for example, the method of (Albanwan and Qin, 2018) and IR-MAD show that the results tend to over achieve temporal stability while present a large disparity to absolute radiance measures (as compared to the MODIS product).

We can notice that our proposed method coincides well with the MODIS product (see the blue and green lines marked in triangles and squares in Fig. 5), where they follow the same trend along the time; the slight shifts between the two lines represent systematic errors. The method of (Albanwan & Qin, 2018) and IR-MAD shows that the resulting data in the temporal direction tend to be flat as compared to the MODIS NBAR product (indicated in purple and blue like marked with stars and asterisks accordingly), which basically achieves good temporal consistency with correctly derived absolute values (presumed as the MODIS NBAR values). Spectral quality analysis

For the spectral analysis, we plot the histograms of the original image, MODIS's reference product, (Albanwan & Qin, 2018) filtered image, IR-MAD, and our proposed method for study-area-I-region-2 and a sample of bands (i.e. NIR, Red, and Blue) (see Fig. 6). We note that the band distributions in the original image vary greatly in the reflectance values and ranges from the MODIS reference product with an obvious shift between their means (see first and second rows in Fig. 6). Image normalized using IR-MAD moves slight but with a similar distribution as the original images. The normalized image using (Albanwan & Qin, 2018) method, changes the histograms and their ranges to some extent, might move closer to the average of the corresponding neighboring images of the subject image due to its temporal averaging-effect introduced in the previous subsection. In our proposed method (Fig. 6), where we can see that the centers (i.e. means) and ranges of the bands distributions of our method matches well with the MODIS's reference product, this might be contributed by the correct inter-band spectral relations provided by the well-radiometric-corrected MODIS product.

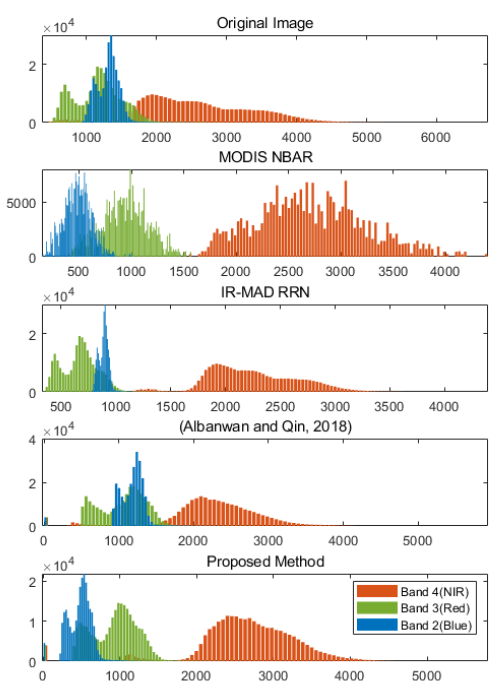


Fig. 6. The spectral analysis using the histograms for a sample of bands: NIR, Red, and Blue comparing the original image, MODIS's reference data, (Albanwan & Qin, 2018), IR-MAD, and our proposed method for region 2. All the figures share the same legend.

2) Visual quality of the normalized data

Visual details are shown in Fig. 7. We note that the noise is reduced in both filtering methods. The spatial details are well preserved in our proposed algorithm, and due to the statistical KL divergence measure we introduced, the results show lower noise than the images normalized by IR-MAD and less blurring effect than the images normalized by (Albanwan & Qin, 2018) (see Fig. 7.). The adaptive filtering with their edge-preserving capability can effectively keep large changes among temporal data. Our approach adopts this concept, in which we keep the local variance of the data by only compensating the radiometric difference using bias terms, at the same time, with the capability of denoising local non-linear distortions using the adaptive filter. The temporal bandwidth of the adaptive filtering can be critical, which we derived from our previous work which has shown a good leverage between the ability to denoise while preserving areas with significant changes temporally [14].

Table IV shows the zero-mean PSNR (Peak Signal-to-Noise Ratio) of the normalized images by (Albanwan & Qin, 2018), IR-MAD, and the proposed method, taking the original image as reference. For the ten images used in our experiment, we find that the proposed method keeps the best averaged quality over the other two methods, which shows that although our proposed method involves the use of low-resolution MODIS data, the normalized images still preserve the spatial details of the data. The IR-MAD method, although yield very high PSNR in some of the images, because it considers per-pixel transformation between image pairs, it performs poorly for images whose reference images are drastically different (e.g. Time 7 vs. Time 5 for IR-MAD).

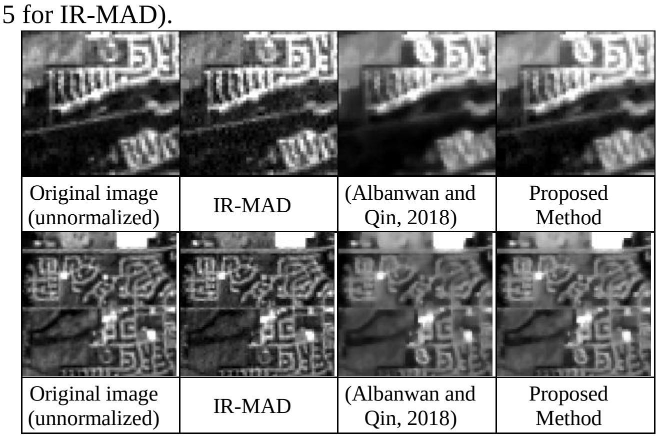


Fig. 7. Two patches from the dataset in study-area-I-region-2 showing the spatial detail.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.202?.30?69?5, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

TABLE IV.
THE ZERO-MEAN PSNR OF THE NORMALIZED IMAGES. (USING THE ORIGINAL IMAGE AS REFERENCE)

Image ID	IR-MAD (Nielsen, 2007)	(Albanwan & Qin, 2018)	Proposed method
Time1	**60.31**	56.04	58.93
Time2	49.07	52.73	**53.45**
Time3	54.73	55.40	**57.94**
Time4	**77.08**	63.35	69.02
Time5	N.A. Reference	64.13	67.04
Time6	**82.94**	59.86	60.47
Time7	42.37	64.96	**66.11**
Time8	61.14	61.53	**64.71**
Time9	**70.98**	61.09	64.98
Time10	57.73	58.03	**62.81**
Average	55.74	59.71	**62.55**

3) *The classification experiments*

We perform a supervised classification using support vector machine (SVM)[41] through a transfer learning practice to evaluate the radiometric consistency between the images before and after the normalization. The classification is performed by training one image and apply to the other image. In general, the classifier trained by one image will yield good classification accuracy on other image if they are radiometric consistent, i.e. the spectral distribution of that image is similar to the image from which the classifier is trained. Although state-of-the-art deep learning methods can be a good option for classification, here we take SVM for two reasons: 1) It has been a standard method land-cover classification of high resolution (Landsat and Sentinel) data in practice, and requires much fewer training samples as compared to deep learning models[42], [43]; 2) It accepts manually-crafted features and is more suitable for controlled experiment. To evaluate the consistency of radiance among different images, we used the radiance values as the only features for classification, this ablates unnecessary roles that more advanced feature plays in classification. Note in this transfer learning experiment we do perform any additional domain adaptation algorithm, rather we train classifiers from one dataset and apply that to other datasets to understand how consistency these datasets are. We assume for well normalized radiometric images, a classifier trained from one image could readily predict reasonable results on other images. Therefore, we use the training information in one reference image to train and test the classification accuracy to the rest of the dataset. In our experiment, we consider four land-cover classes: Forest, Impervious surfaces, Cropland, and Water, based on the land-cover classification system developed by [44]. Fig. 8. shows a sample for the transfer learning classification results for image of time 5 from region 1 using the training data from a reference image of time 4. Classification map of the original image shows many misclassifications and noise, for instance, we can see that at the forest region is mostly classified into cropland. Meanwhile, the water surface in IR-MAD is misclassified into Forest. With the algorithm proposed by (Albanwan & Qin, 2018), we can see better classification results, where forest and water surface are better identified, however, the impervious surfaces are incorrectly detected in many locations. Our proposed method, on the other hand, shows even better

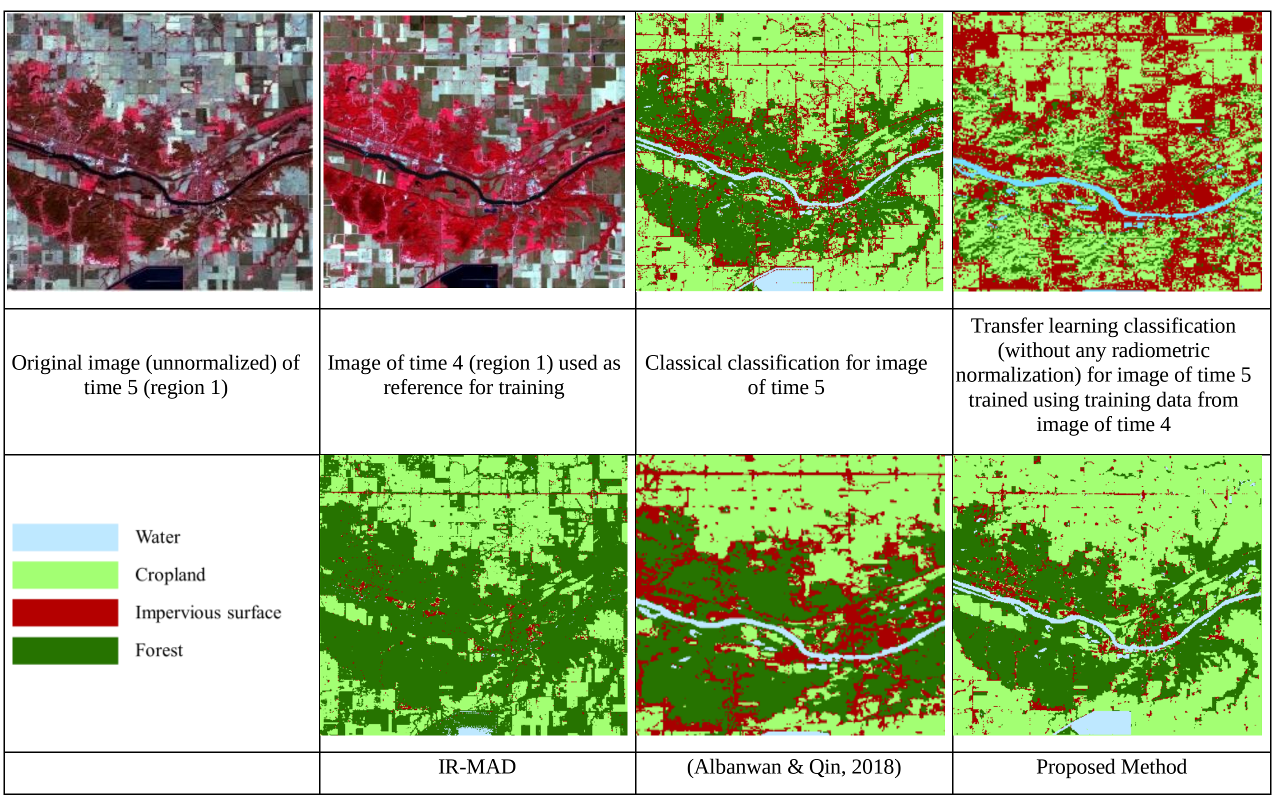


Fig. 8. The transfer learning classification results from region 1 using image of time 5 and reference training data of time 4.

1

TABLE V.
THE TFC RESULTS EVALUATED USING THE OVERALL ACCURACY (OA) AND KAPPA COEFFICIENT (KC) FOR REGION 1

Image ID	Transfer Learning SVM using training data from Time 4							
	Unnormalized		**IR-MAD**		**(Albanwan & Qin, 2018)**		**Proposed method**	
	KC	**OA**	**KC**	**OA**	**KC**	**OA**	**KC**	**OA**
Time: 1	0.6425	75.30%	0.7640	83.27%	**0.9664**	**97.61%**	0.8661	90.44%
Time: 2	0.6189	71.51%	0.7328	80.68%	**0.7703**	**83.47%**	0.7366	80.88%
Time: 3	0.9025	93.03%	0.8582	89.84%	**0.9665**	**97.61%**	0.9331	95.22%
Time: 4	-	-	-	-	-	-	-	-
Time: 5	0.6246	73.71%	0.5628	68.53%	**0.9693**	**97.81%**	0.9161	94.02%
Average	0.6971	78.39%	0.7295	80.58%	**0.9181**	**94.13%**	0.8630	90.14%
Average increase in accuracy			0.0323	2.19%	0.2210	15.74%	0.1659	11.75%

TABLE VI.
THE TFC RESULTS EVALUATED USING THE OVERALL ACCURACY (OA) AND KAPPA COEFFICIENT (KC) FOR REGION 2

Image ID	Transfer Learning SVM using training data from Time 8							
	Unnormalized		**IR-MAD**		**(Albanwan & Qin, 2018)**		**Proposed method**	
	KC	**OA**	**KC**	**OA**	**KC**	**OA**	**KC**	**OA**
Time: 1	0.586	74.78%	0.4528	58.55%	**0.9503**	**96.72%**	0.7962	87.05%
Time: 2	0.9292	95.34%	0.8266	87.91%	**0.9661**	**97.75%**	0.9262	95.16%
Time: 3	0.8315	89.12%	0.8309	88.26%	**0.9661**	**97.75%**	0.9235	94.99%
Time: 4	0.3456	49.57%	0.1593	36.79%	**0.7186**	**80.14%**	0.2451	43.01%
Time: 5	0.5517	67.36%	0.5806	68.91%	**0.9355**	**95.68%**	0.774	84.46%
Time: 6	0.8252	88.08%	0.6927	77.72%	**0.9558**	**97.06%**	0.8795	92.06%
Time: 7	0.9507	96.72%	0.9505	96.72%	**0.9767**	**98.45%**	0.9534	96.89%
Time: 8	-	-	-	-	-	-	-	-
Time: 9	0.7801	85.84%	0.557	69.95%	**0.9714**	**98.10%**	0.9397	96.03%
Time: 10	0.593	72.88%	0.5404	68.57%	**0.9688**	**97.93%**	0.9107	94.13%
Average	0.73878	71.97%	0.65882	65.34%	0.94093	85.96%	0.83457	78.38%
Average increase in accuracy			-0.07996	-6.63%	0.20215	13.99%	0.09579	6.41%

TABLE VII.
THE TFC RESULTS EVALUATED USING THE OVERALL ACCURACY (OA) AND KAPPA COEFFICIENT (KC) AND USING A REFERENCE TRAINING DATA FROM ONE REGION AND APPLY IT TO IMAGE FROM ANOTHER REGION

Training data	Test image	**K-L Divergency**	**Date Interval (Days)**	**Original image**		**IR-MAD**		**(Albanwan & Qin, 2018)**		**Proposed method**	
				KC	**OA**	**KC**	**OA**	**KC**	**OA**	**KC**	**OA**
Time 1 Region 1 (05152016 - ETM+)	Time 1 Region 2 (05232016-OLI)	0.066	8	0.7226	80.07%	0.319	48.35%	0.5708	67.76%	**0.8175**	**87.52%**
Time 2 Region 1 (09122016-OLI)	Time 2 Region 2 (09202016 - ETM+)	0.209	8	0.6631	76.26%	0.6159	72.44%	0.7802	84.92%	**0.8738**	**91.68%**
Time 3 Region 1 (10132016-MSI)	Time 9 Region 2 (11072016-ETM+)	0.117	25	0.7233	77.47%	0.5125	70.71%	0.7409	75.39%	**0.8784**	**78.13%**
Time 4 Region 1 (10142016-OLI)	Time 9 Region 2 (11072016-ETM+)	0.041	24	0.828	88.39%	0.3995	64.99%	0.684	76.95%	**0.8839**	**92.20%**
Time 5 Region 1 (11152016-OLI)	Time 9 Region 2 (11072016-ETM+)	0.088	8	**0.8465**	**89.60%**	0.1178	26.69%	0.6404	73.48%	0.8143	87.18%
Average				0.7567	82.36%	0.39294	56.64%	0.68326	75.70%	**0.8536**	**87.34%**
Average increase in accuracy						-0.36376	-25.72%	-0.07344	-6.66%	**0.0969**	**4.98%**

1

classification outcome, where misclassifications of impervious surfaces are significantly reduced.

Two experiments with the transfer learning classification (TFC) are carried out over the two regions. For the first experiment (visual results shown in Fig. 8.), we train a classifier for one temporal image and apply it to the rest of temporal images, and the goal is to evaluate the degree of radiometric consistency between the normalized temporal images. In the second experiment, we train a classifier from one temporal of one region and apply the classifier to the normalized images of another region, and the goal is to evaluate the level of global radiometry consistency over the space. We indicate the accuracy of TFC using two measure the overall accuracy (OA) and Kappa coefficient (KC) as accuracy indicators.

TFC Experiment I: Table V and VI shows the statistical results of TFC for study-area-I region 1 and 2. The "original image" column refers to the TFC on images with no radiometric normalization, meaning a direct application using the classier trained from time 4 image (e.g. for Table V) to the rest of the images, and statistics under other columns indicate the same operation but on images processed with the respective radiometric normalization methods. The average OA enhancement is about ~ 6-11% in our proposed method (see last rows in Table V and VI). The accuracy enhancement in (Albanwan & Qin, 2018) is notably higher (~14-15%), and this is expected because the goal of the (Albanwan & Qin, 2018) method aims homogenize the spectrums and for areas that has relatively smaller changes, it presents superior results. Our proposed normalization algorithm although has relatively lower improvement, it does leverage the RRN and ARN and thus variations resulted from the effort for keeping the spectrum similar to the MODIS products lead to the relatively less improvement, this can be noted in Table VI, in which the classifier trained on time-8 has a very poor result for time 4 image (taken on July 26th) using our method. This is because that this time-4 image is significantly different from the others in the original MODIS data (shown in Fig. 3.), while the method of (Albanwan & Qin, 2018) correct this image with no constrain, which happen to result in higher accuracy.

TFC Experiment II: To test the global consistency of the normalized results, we perform the transfer learning classification experiment by means of training a classifier on one Landsat/Sentinel-2 image and applied that to an image of a different region, under this comparative condition that these images are processed by different normalization methods. Specifically, for each image from region 1, an SVM classifier will be trained, and then applied to an image from region 2. Our hypothesis is that if the normalized images preserve consistent radiances, a classification model trained from one image in any location, will ideally yield fairly consistent and accurate classification results. In practice, we expect to observe that radiometric consistencies will outperform unnormalized images or images that are only normalized by RRN methods. In this experiment, we take the region-1 images and find the most unlike corresponding region-2 images (selected as the region-2 image with largest KL divergences to this region-1 image). Table VII shows the experimental results, in which training data column indicates the region-1 images used to train the classifier and the testing data column indicates the selected region-2 images that the classifier will be applied to evidence the effectiveness of our approach, we optimized the parameters & hyperparameters of both competing normalization approaches (IR-MAD and the method of. Albanwan & Qin (2018)) to achieve the best results on this dataset. Among these comparing approaches shown in the table (Table VII), we note that our method achieves an average improvement of ~ 5% (see last rows in Table VII) in overall accuracy. We also note that classification accuracy after IR-MAD and (Albanwan & Qin, 2018) normalization decrease: they show a drop in the overall accuracy of ~ 25.72% and 6.66%, respectively. This is reasonable since IR-MAD method is only able to normalize pairs of images at a time, which introduces even larger gaps in a multitemporal image case, and the method of (Albanwan & Qin, 2018) is a RRN method and only accounts for the spectral homogeneity in each region, meaning that the normalization is performed completely in these two regions with no absolute metrics. Our proposed method in this case, outperform in most of the cases, except for the last image, in which the unnormalized data is consistent already. images normalized by a multitemporal image case, and the method of (Albanwan & Qin, 2018) is a RRN method and only accounts for the spectral homogeneity in each region, meaning that the normalization is performed completely in these two regions with no absolute metrics. Our proposed method in this case, outperform in most of the cases, except for the last image, in which the unnormalized data is consistent already

V. Conclusion

In this work, we propose a radiometric normalization method for high resolution images that does not rely on in-situ data, rather on a well radiometric-corrected low-resolution and globally available reference product (such as MODIS's NBAR product). Our method takes advantage of the non-parametric and adaptive characteristics of the spatiotemporal filter and further extends it by an image-to-image bias term to accommodate the per-image and per-pixel differences.

In our proposed method, the bias minimization is carried out adaptively such that on one hand, the resulting normalized images are as consistent as possible temporally, and on the other hand, their absolute radiometric values are as close as possible to the reference low-resolution product. We demonstrate that the proposed method are able to produce images with notably consistent radiometric properties in different aspects: first, the visual analysis of the mosaic shows a good radiometric consistency and the seamlines between two images are barely visible in comparison to the mosaic results of the unnormalized data and of normalized data processed by other normalization methods. Secondly, the normalized images of our method, although involve the use of low-resolution MODIS data, preserves spatial details well and yields sharper and cleaner images than those generated from other methods. The statistical analysis of the transfer learning classification (TFC) experiments shows that the proposed method can consistently improve the accuracy of the classification by 6-11% in the case of TFC on multi-temporal images of the same region. Although one RRN method that is used for validation slightly outperform our method, we show that our proposed method are able to achieve global transferability in TFC experiment II, in which we train classifiers from normalized dataset of one region and applied to another region, and improved the accuracy by ca. 5%, while other RRN methods contrarily reduced the transferability of classifiers over different regions.

Acknowledgments

The authors would like to thank the United states Geological Survey (USGS) department and the European Space Agency (ESA) for the data used in this study and for making the MODIS and Sentinel Data available for public remote sensing community use.

References

[1] G. Kamberova and R. Bajcsy, “Sensor Errors and Uncertainties in Stereo Reconstruction,” *Empir. Eval. Tech. Comput. Vis.*, 1998.

[2] L. Paolini, F. Grings, J. A. Sobrino, J. C. Jiménez Muñoz, and H. Karszenbaum, “Radiometric correction effects in Landsat multi-date/multi-sensor change detection studies,” *Int. J. Remote Sens.*, vol. 27, no. 4, pp. 685–704, 2007, doi: 10.1080/01431160500183057.

[3] S. Vanonckelen, S. Lhermitte, and A. Van Rompaey, “The effect of atmospheric and topographic correction methods on land cover classification accuracy,” *Int. J. Appl. Earth Obs. Geoinformation*, vol. 24, pp. 9–21, 2013, doi: 10.1016/j.jag.2013.02.003.

[4] X. Chen, L. Vierling, and D. Deering, “A simple and effective radiometric correction method to improve landscape change detection across sensors and across time,” *Remote Sens. Environ.*, vol. 98, no. 1, pp. 63–79, 2005, doi: 10.1016/j.rse.2005.05.021.

[5] H. Ghanbari, S. Homayouni, P. Ghamisi, and A. Safari, “Radiometric Normalization of Multitemporal and Multisensor Remote Sensing Images Based on a Gaussian Mixture Model and Error Ellipse,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 11, no. 11, pp. 4526–4533, 2018, doi: 10.1109/JSTARS.2018.2871373.

[6] N. E. Young, R. S. Anderson, S. M. Chignell, A. G. Vorster, R. Lawrence, and P. H. Evangelista, “A survival guide to Landsat preprocessing,” *Ecology*, vol. 98, no. 4, pp. 920–932, 2017, doi: 10.1002/ecy.1730.

[7] V. E. G. Millán, G. A. S. Azofeifa, G. C. Malvárez, G. Moré, X. Pons, and M. Yamanaka-Ocampo, “Effects of topography on the radiometry of CHRIS/PROBA images of successional stages within tropical dry forests,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 6, no. 3, pp. 1584–1595, 2013, doi: 10.1109/JSTARS.2013.2259471.

[8] X. Pons, L. Pesquer, J. Cristóbal, and O. González-Guerrero, “Automatic and improved radiometric correction of Landsat imagery using reference values from MODIS surface reflectance images,” *Int. J. Appl. Earth Obs. Geoinformation*, vol. 33, pp. 243–254, 2014, doi: 10.1016/j.jag.2014.06.002.

[9] Q. Xu, Z. Hou, and T. Tokola, “Relative radiometric correction of multi-temporal ALOS AVNIR-2 data for the estimation of forest attributes,” *ISPRS J. Photogramm. Remote Sens.*, vol. 68, pp. 69–78, 2012, doi: 10.1016/j.isprsjprs.2011.12.008.

[10] D. Yuan and C. D. Elvidge, “Comparison of relative radiometric normalization techniques,” *ISPRS J. Photogramm. Remote Sens.*, 1996, doi: 10.1016/0924-2716(96)00018-4.

[11] Y. Zhang, L. Yu, M. Sun, and X. Zhu, “A mixed radiometric normalization method for mosaicking of high-resolution satellite imagery,” *IEEE Trans. Geosci. Remote Sens.*, 2017, doi: 10.1109/TGRS.2017.2657582.

[12] M. Claverie *et al.*, “The Harmonized Landsat and Sentinel-2 surface reflectance data set,” *Remote Sens. Environ.*, vol. 219, pp. 145–161, 2018, doi: 10.1016/j.rse.2018.09.002.

[13] D. P. Roy *et al.*, “Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity,” *Remote Sens. Environ.*, vol. 185, pp. 57–70, Nov. 2016, doi: 10.1016/j.rse.2015.12.024.

[14] Q. Zheng, Q. Weng, and K. Wang, “Developing a new cross-sensor calibration model for DMSP-OLS and Suomi-NPP VIIRS night-light imageries,” *ISPRS J. Photogramm. Remote Sens.*, vol. 153, pp. 36–47, 2019, doi: 10.1016/j.isprsjprs.2019.04.019.

[15] F. W. Acerbi-Junior, J. G. P. W. Clevers, and M. E. Schaepman, “The assessment of multi-sensor image fusion using wavelet transforms for mapping the Brazilian Savanna,” *Int. J. Appl. Earth Obs. Geoinformation*, vol. 8, no. 4, pp. 278–288, 2006, doi: 10.1016/j.jag.2006.01.001.

[16] H. Albanwan and R. Qin, “A novel spectrum enhancement technique for multi-temporal, multi-spectral data using spatial-temporal filtering,” *ISPRS J. Photogramm. Remote Sens.*, vol. 142, pp. 51–63, Aug. 2018, doi: 10.1016/j.isprsjprs.2018.05.020.

[17] G. Chander *et al.*, “Landsat-5 TM reflective-band absolute radiometric calibration,” *IEEE Trans. Geosci. Remote Sens.*, 2004, doi: 10.1109/TGRS.2004.836388.

[18] J. Barsi *et al.*, “Landsat-8 on-orbit and Landsat-9 pre-launch sensor radiometric characterization,” vol. 1078104, no. October 2018, p. 3, 2018, doi: 10.1117/12.2324715.

[19] D. K. Seo and Y. D. Eo, “Multilayer perceptron-based phenological and radiometric normalization for high-resolution satellite imagery,” *Appl. Sci. Switz.*, 2019, doi: 10.3390/app9214543.

[20] K. Thome, “In-flight intersensor radiometric calibration using vicarious approaches,” 2004.

[21] D. Naughton, “Absolute radiometric calibration of the RapidEye multispectral imager using the reflectance-based vicarious calibration method,” *J. Appl. Remote Sens.*, 2011, doi: 10.1117/1.3613950.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.202[illegible], IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

[22] J. Gorroño *et al.*, "A radiometric uncertainty tool for the sentinel 2 mission," *Remote Sens.*, vol. 9, no. 2, pp. 1–25, 2017, doi: 10.3390/rs9020178.

[23] V. Avitabile, A. Baccini, M. A. Friedl, and C. Schmullius, "Capabilities and limitations of Landsat and land cover data for aboveground woody biomass estimation of Uganda," *Remote Sens. Environ.*, 2012, doi: 10.1016/j.rse.2011.10.012.

[24] T. A. Schroeder, W. B. Cohen, C. Song, M. J. Canty, and Z. Yang, "Radiometric correction of multi-temporal Landsat data for characterization of early successional forest patterns in western Oregon," *Remote Sens. Environ.*, 2006, doi: 10.1016/j.rse.2006.03.008.

[25] D. Helder *et al.*, "Absolute radiometric calibration of landsat using a pseudo invariant calibration site," *IEEE Trans. Geosci. Remote Sens.*, 2013, doi: 10.1109/TGRS.2013.2243738.

[26] A. F. Militino, T. Goicoa, and M. D. Ugarte, "Estimating the percentage of food expenditure in small areas using bias-corrected P-spline based estimators," *Comput. Stat. Data Anal.*, vol. 56, no. 10, pp. 2934–2948, 2012, doi: 10.1016/j.csda.2012.01.009.

[27] Pat S. Chavez Jr., "An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data," *Remote Sens. Environ.*, vol. 24, pp. 459–479, 1988, doi: https://doi.org/10.1016/0034-4257(88)90019-3.

[28] M. Feng, C. Huang, S. Channan, E. F. Vermote, J. G. Masek, and J. R. Townshend, "Quality assessment of Landsat surface reflectance products using MODIS data," *Comput. Geosci.*, 2012, doi: 10.1016/j.cageo.2011.04.011.

[29] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, 1989, doi: 10.1109/34.192463.

[30] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Processing Magazine*. 2005, doi: 10.1109/MSP.2005.1550194.

[31] A. Suwendi and J. P. Allebach, "Nearest-neighbor and bilinear resampling factor estimation to detect blockiness or blurriness of an image," 2006, doi: 10.1117/12.647924.

[32] J. Storey, M. Choate, and K. Lee, "Landsat 8 Operational Land Imager On-Orbit Geometric Calibration and Performance," *Remote Sens.*, vol. 6, no. 11, pp. 11127–11152, 2014, doi: 10.3390/rs61111127.

[33] X. Jin, "ENVI automated image registration solutions. Harris Geospatial Systems whitepaper," 2017. http://www.harrisgeospatial.com/Portals/0/pdfs/ENVI_Image_Registration_Whitepaper.pdf (accessed Mar. 06, 2020).

[34] "Georeferencing toolbar tools—Help | ArcGIS for Desktop." https://desktop.arcgis.com/en/arcmap/10.3/manage-data/raster-and-images/georeferencing-toolbar-tools.htm (accessed Dec. 15, 2020).

[35] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *Ann. Math. Stat.*, 1951, doi: 10.1214/aoms/1177729694.

[36] A. A. Nielsen, "The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data," *IEEE Trans. Image Process.*, 2007, doi: 10.1109/TIP.2006.888195.

[37] J. R. Irons, J. L. Dwyer, and J. A. Barsi, "The next Landsat satellite: The Landsat Data Continuity Mission," *Remote Sens. Environ.*, 2012, doi: 10.1016/j.rse.2011.08.026.

[38] B. Markham *et al.*, "Landsat-8 Operational Land Imager Radiometric Calibration and Stability," *Remote Sens.*, vol. 6, no. 12, pp. 12275–12308, 2014, doi: 10.3390/rs61212275.

[39] European Space Agency, "Sentinel-2 delivers first images," 2015. http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Sentinel-2/Sentinel-2_delivers_first_images (accessed May 06, 2020).

[40] H.-Y. Shum and R. Szeliski, "Construction of Panoramic Image Mosaics with Global and Local Alignment," 2001.

[41] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Transactions on Neural Networks*. 1999, doi: 10.1109/72.788640.

[42] P. Liu, K. K. R. Choo, and L. Wang, "SVM or deep learning? A comparative study on remote sensing image classification," *Soft Comput*, vol. 21, pp. 7053–7065, 2017, doi: https://doi.org/10.1007/s00500-016-2247-2.

[43] Y. Shao and R. S. Lunetta, "Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points," *ISPRS J. Photogramm. Remote Sens.*, vol. 70, pp. 78–87, 2012, doi: https://pubs.acs.org/doi/10.1021/ci0341161.

[44] P. Gong, J. Wang, L. Yu, Y. Zhao, and Y. Zhao, "Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data," *Int. J. Remote Sens.*, vol. 34, no. 7, pp. 2607–2654, 2013, doi: https://doi.org/10.1080/01431161.2012.748992.

Wenxia Gan(M'2010) was born in Hubei, China in 1987. She received the B.S., the M.S. and Ph.D. degrees in Photogrammetry and Remote Sensing from Wuhan University, China, in 2009, 2011, 2015 respectively. She was a visiting scholar at The Ohio State University, Department of Civil, Environmental and Geodetic Engineering from 2018 to 2020. She is currently an assistant professor of the School of Civil Engineering and Architecture in Wuhan Institute of Technology, Wuhan, China. Her research interests include radiometric normalization of Remote Sensing data, application of multiple and hyperspectral Remote Sensing technique on ecology environment. Dr. Gan serves as a reviewer for several international journals in the field of photogrammetry and remote sensing.

Hessah Albanwan received the B.S. degree in civil engineering from Kuwait University, Kuwait, 2014, and the M.S. in civil, geodetic, and environmental engineering from Ohio State University, USA, 2017. She is currently a Ph.D. candidate and is pursuing her Ph.D. degree in Ohio State University. Her research interest includes remote sensing and photogrammetry.

Rongjun Qin (M'2012–SM'2019) received the B.S. degree in computational mathematics from Wuhan University, Wuhan, China, in 2009, and the M.S. and Ph.D. degree in photogrammetry and remote sensing from Wuhan University and ETH, Zurich in 2011 and 2015 respectively. He is currently a faculty member of the Department of Civil, Environmental and Geodetic Engineering, Department of Electrical and Computer Engineering, and the Translational Data Analytics Institute, at The Ohio State University, Columbus. His research interests include photogrammetric 3D reconstruction, remote sensing image classification, UAV images processing, image dense matching and change detection. His research seeks for computational solutions to various geometric and interpretation problems in an urban context using imaging sensors such as aerial/UAV imagery, LiDAR, and satellite multispectral/hyperspectral images. Prof. Qin is the author of RSP (RPC stereo processor) and MSP (multi-stereo processor) used for reconstructing 3D information from 2D images with high quality. Prof. Qin is an associate editor for the Photogrammetric Engineering and Remote Sensing journal. He is also chairing the working group "Satellite Constellation for Remote Sensing" of International Society for Photogrammetry and Remote Sensing Commission. His awards include the first prize of Mathematical Modeling Contest and several other prominent scholarship awards. Prof. Qin serves as a reviewer for more than 15 international journals in the field photogrammetry and remote sensing.