Peer-reviewed version available at *Remote Sens.* **2020**, *12*, 281; doi:10.3390/rs12020281

Article

Harmonization of Landsat and Sentinel 2 for Crop Monitoring in Drought Prone Areas: Case Studies of Ninh Thuan (Vietnam) and Bekaa (Lebanon)

Minh D. Nguyen [1,2], Oscar B. Villanueva [2], Duong D. Bui [3]*, Phong T. Nguyen [1] and Lars Ribbe [2]

[1] Vietnam Academy for Water Resources, Ministry of Agricultural and Rural Development, Hanoi, Vietnam
[2] Institute for Technology and Resources Management in the Tropics and Subtropics (ITT), Cologne University of Applied Science, Germany
[3] National Center for Water Resources Planning and Investigation, Ministry of Natural Resources and Environment, Hanoi, Vietnam
* Correspondence: duongdubui@gmail.com(Duong D. Bui)

Abstract: Proper satellite-based crop monitoring applications at the farm-level often require near-daily imagery at medium to high spatial resolution. The synthesizing of ongoing satellite missions by ESA (Sentinel 2) and NASA (Landsat7/8) provides this unprecedented opportunity at a global scale; nonetheless, this is rarely implemented because these procedures are data demanding and computationally intensive. This study developed a complete stream processing in the Google Earth Engine cloud platform to generate harmonized surface reflectance images of Landsat7,8 and Sentinel 2 missions. The harmonized images were generated for two agriculture schemes in Bekaa (Lebanon) and Ninh Thuan (Vietnam) during the period 2018-2019. We evaluated the performance of several pre-processing steps needed for the harmonization including image co-registration, brdf correction, topographic correction, and band adjustment. This study found the miss-registration between Landsat 8 and Sentinel 2 images, varied from 10 meters in Ninh Thuan, Vietnam to 32 meters in Bekaa, Lebanon, and if not treated, posed a great impact on the quality of the harmonized data set. Analysis of a pair overlapped L8-S2 images over the Bekaa region showed that after the harmonization, all band-to-band spatial correlations were greatly improved from (0.57, 0.64, 0.67, 0.75, 0.76, 0.75, 0.79) to (0.87, 0.91, 0.92, 0.94, 0.97, 0.97, 0.96) in bands (blue, green, red, nir,swir1,swir2, ndvi) respectively. Finally, we demonstrated the high potential of the harmonized data set for crop mapping and monitoring. Harmonic (Fourier) analysis was applied to fit the detected unimodal, bimodal and trimodal shapes in the temporal NDVI patterns during one crop year in Ninh Thuan province. Derived phase and amplitude values of the crop cycles were combined with max-NDVI as an R-G-B image. This image highlighted croplands in bright colors (high phase and amplitude) and non-crop areas in grey/dark (low phase/amplitude). Generated harmonized data sets (30m spatial resolution) over the two studied sites along with GEE scripts/app used in the study are provided for public usage and testing.

Keywords: Landsat; Sentinel 2; harmonization; crop monitoring; Google Earth Engine

---

1. Introduction

In recent decades, advances in technology and algorithms have made satellite remote sensing played an ever-increasing role in crop monitoring [1,2]. However, current global satellite missions do not possess enough temporal-spatial resolution to capture crop growth's dynamic and heterogeneity at the farm level. Number of studies pointed out that high temporal resolution products (e.g. MODIS, MERIS) are generally too coarse (from 250 m to a few kilometers) to capture cropland heterogeneity

Peer-reviewed version available at Remote Sens. 2020, 12, 281; doi:10.3390/rs12020281

[3,4]. Meanwhile, medium spatial resolution products (e.g. Landsat 30m spatial resolution) potentially miss the observations at critical growth stages due to long revisit time (16 days)[4–7]. Furthermore, the applicability of optical remote sensing gets extra challenges in tropical regions with frequent cloud cover (e.g. Vietnam). Not enough satellite observations produce composite images with cloud contamination and eventually reduce the quality of crop mapping [8] or influence the study of crop's phenology [9]. [10] reported properly land monitoring applications would require near-daily observations at medium spatial resolution. One of the effort to archive better data resolution is called spatio-temporal image fusion (or sensor fusion). This approach generates fine spatial resolution images while trying to maintain the frequency by synthesizing coarse spatial-high temporal products (e.g. MODIS) with high spatial-low temporal products (e.g. Landsat) [3,4,11–13]. In sensor fusion, coincidence pairs of coarse spatial-high temporal images and high spatial-low temporal images acquired on the same dates (or temporally close) are correlated to find information which is then used to downscale coarse spatial-high temporal images to the resolution of the high spatial resolution images [12]. Three common image fusion techniques can be identified including image-pair-based, spatial unmixing-based and hybrid methods [13,14]. However, multi-sensors fusion is reported involving large uncertainty [13], because of (1) low registration accuracy due to the significant difference in the sensors's resolution [14] and (2) spectral signatures of small objects can be lost in the fused images [14,15].

Recently, there is an increase in the number of medium spatial resolution EO satellites. Since 2013, NASA launched Landsat 8, which is currently operating alongside Landsat 7. The combination of Landsat 7 and Landsat 8 generates three to four observations per month. Since 2015 Sentinel 2 constellation from the European Space Agency is providing global scale imagery within 5-10 days revisit time at 10 to 60 meters resolution. The proven compatibility between Landsat and Sentinel 2 bands producing the opportunity for near-daily global temporal coverage at medium resolution by merging their observations [10,16,17].

Nevertheless, synthesizing (or harmonizing) Landsat7/8 and Sentinel 2 is still an intricate process that requires several data transformation steps [10,18]. A research project initiated by NASA has taken into account Bidirectional Reflectance Distribution Functions (BRDF) correction, sensor misregistration, bands re-scaling, and re-projection, as well as small band adjustment [10]. These steps were applied so that the multi-sensors images can be reasonably stackable for consistent time series analysis. BRDF model was applied to account for the differences in the field of view angles among satellites because after atmospheric correction, this variation is exaggerated [19]. In extreme cases, the differences in the view angle for a ground target can increase up to 20 degree [19]. Additionally, as the consequence of different image registration references, sensor misregistration between Landsat 8 and Sentinel 2 varies geographically and can exceed one Landsat pixel (30 meters) [16,17].

Besides sensor transformation, pre-treatment of Sentinel 2 and Landsat images require some attention too. For example, the same atmospheric correction model should be applied to both sensors to reduce residual errors from using different atmospheric correction (AC) methods [20,21]. On the other hand, unlike predecessor satellites (e.g. Landsat, ASTER, MODIS), Sentinel 2 sensors lack thermal infrared bands, therefore established thermal-based cloud mask algorithms that work well for Landsat (e.g. FMASK) do not guarantee yield similar performance for Sentinel 2. Sentinel 2 cloud detection and optimization are reported as the main issue in the NASA's harmonized product [10].

As a consequence, given the unprecedented opportunity to improve the spatio-temporal resolution of EO imagery at a global scale by harmonizing Landsat and Sentinel 2 images; nonetheless, this is rarely implemented because these procedures are data demanding and computationally intensive.

Meanwhile, emerging cloud computing platforms such as Google Earth Engine (GEE) which has the planetary-scale archives of remote sensing data [22] including Landsat, Sentinel 2, significantly reduce the work for data management and speed up analyzing process [23]. Built-in functions/algorithms within the GEE platform help simplify many pre-processing steps allowing focus on the interpretation of the core algorithms [22,24].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2019 doi:10.20944/preprints201910.0275.v1

Peer-reviewed version available at Remote Sens. 2020, 12, 281; doi:10.3390/rs12020281

The objective of this study is to develop a complete stream processing for the harmonization of Landsat - Sentinel 2 in Google Earth Engine (GEE) to harness the benefit of coherent data structure, built-in functions and computation power in the Google Cloud. In this study, we adapt the BRDF (MODIS-based fixed coefficients c-factor) [25,26] and the topographic correction model (modified Sun-Canopy-Sensor Topographic Correction) [27]. These models were implemented in GEE by [28]. We adjust the Landsat TOA's bands (blue, green, red, nir, swir1 and swir2) using cross-sensor transformation coefficients derived from [29]. We describe several tests to assess and evaluate the performance of each pre-processing/transforming step. Finally, we demonstrate an application of the harmonized dataset to mapping the dynamic of seasonal cropland in Ninh Thuan, Vietnam.

2. Materials and Methods

2.1. Study regions and input data

For the variety of test sites, we chose Ninh Thuan province (area 3366 km2), located in Vietnam (South East Asia) (Figure 1) and an agriculture scheme called Bekaa (area 898 km2) located in Lebanon (Middle East) (Figure 1). For Ninh Thuan province, we processed total 97 TOA satellite images gathered from 18 images of Landsat 7 (PATH 123, ROW 52), 18 images of Landsat 8 (PATH 123, ROW 52) and 61 images of Sentinel 2 (TILE 49PBN and 49PBP). For Bekaa, we processed total 120 TOA images gathered from 34 images of Landsat 7 (PATH 174, ROW 36 and ROW 37), 19 images of Landsat 8 (PATH 174, ROW 37) and 67 images of Sentinel 2 (TILE 36SYC).

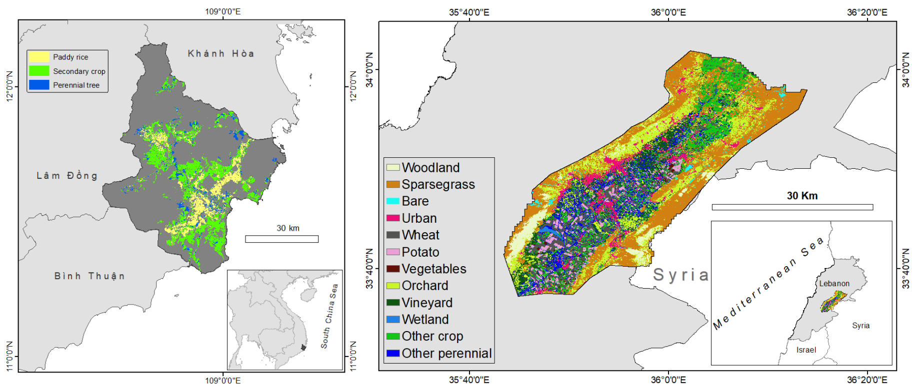

Figure 1. Cropland maps of Ninh Thuan, Vietnam [30] (left) and of Bekaa, Lebanon [31] (right)

2.2. Workflow overview

In general, we design the workflow into three main steps including pre-processing, sensors harmonization and post-processing (Figure 2). In the pre-processing step, we convert the Top of Atmosphere (TOA) images to surface reflectances (SR) (atmospheric correction) filter too cloudy images and mask out high probability cloudy pixels. We applied atmospheric correction via Python API, all the other tasks were Code Editor based. Because the BRDF and topographic correction models require DEM data, they are applied only when the images had been co-registered. Theharmonization step refers to re-projection, rescaling and re-alignment (co-registration) of the Landsat7/8 and Sentinel 2 images. Finally, the post-processing step stacks all the harmonized images into a database of GEE assets. This step also exports harmonized images to Google Drive, making it a shareable geospatial dataset for non-GEE users. The GEE scripts used in the study and links to the generated harmonized datasets that contain surface reflectance images (bands blue, green, red, nir, swir1, swir2, and ndvi at 30 meters) over the two studied sites are provided in the Appendix A.

Peer-reviewed version available at Remote Sens. 2020, 12, 281; doi:10.3390/rs12020281

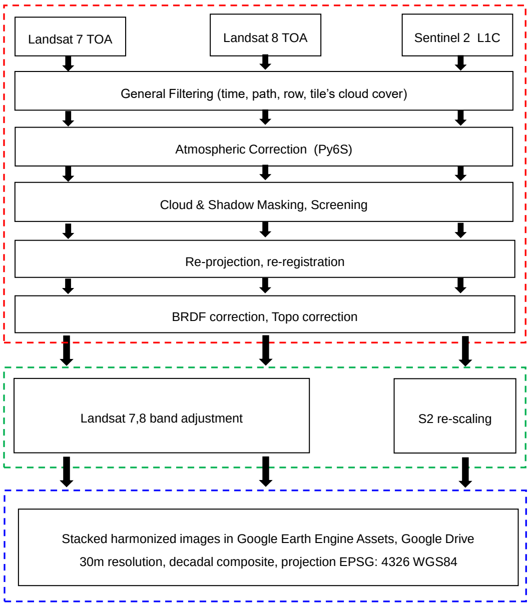

Figure 2. Workflow of the harmonization in GEE

2.3. Atmospheric correction

To reduce residual errors from using different atmospheric correction (AC) methods [20,21], the same AC model called Py6S was applied to all Landsat7/8 and Sentinel 2 TOA images. Py6S is a python interface of 6S radiative transfer model [32] developed by [33] to reduce time and difficulties in setting up numerous input and outputs. Results produced from Py6S will be the same as the results produced from 6S [33]. [34] tested the performance of 6S with the overall relative error was less than 0.8 percent.

This study implemented Py6S in GEE based upon the code shared by [35] (Link Github) which was executed via Python API and Docker container. In the model, the view zenith angle was hardcoded to "0".

2.4. Cloud mask for Landsat images

For Landsat 7/8, cloud and cloud shadow is masked using the BQA band [9] which was generated using the CFMask algorithm. CFMask has been the best overall accuracy among many sates of the art cloud detection algorithms [36].

2.5. Cloud mask for Sentinel 2 images

According to an assessment by [37], the Sentinel 2 L1C product's cloud mask band (QA60), which is generated based on the blue band (B2) and SWIR bands (B11, B12) [38], generally underestimates the presence of clouds. On the contrary, [39] reported that QA60 cloud masks are adjusted to minimize under-detections, which leads, on the other hand, to over-detections. In either case, the performance of the L1C cloud mask is low, especially under critical conditions.

In the GEE environment, [28] applied a cloud scoring algorithm (ee.Algorithms.Landsat.simpleCloudScore) to mask clouds in Landsat 8 and Sentinel 2 images [40]. The algorithm exploits the spectral and thermal properties of cloud that is 'bright and cold but not snow' [41]. However, our eye inspection showed that this Landsat based algorithm did not yield satisfactory results for Sentinel 2 images over Ninh Thuan, Vietnam. This is likely due to the complex atmospheric condition (e.g. high water vapor content) [37] in Ninh Thuan region and lacking a thermal band in Sentinel 2 images.

Peer-reviewed version available at Remote Sens. 2020, 12, 281; doi:10.3390/rs12020281

Inspired by the work of [42] which showed that cloud detection using a machine learning approach can outperform current states of the art threshold-based cloud detection such as Fmask, Sen2Cor or even MAJA which used multi-temporal method for cloud detection [43]. This study combined the QA60's mask with a supervised classification approach in GEE. For every Sentinel 2 scene, we trained the RF classifier using the QA60 band as the base field for stratified random sampling. GEE API simplified this procedure with two built-in algorithms called ee.Classifier.randomForest() and ee.Image.stratifiedSample() [44,45]. Also, we used the Normalized Difference Snow Index (NDSI) to prevent snow from being masked [28]. We used eye visual inspection to check the performance of this procedure, which showed promising results in such a complicated atmospheric condition like Ninh Thuan, Vietnam. Figure A2 demonstrated how the cloud was masked in a cloudy Sentinel 2 scene over Ninh Thuan, Vietnam.

2.6. Cloud shadow detection

Cloud shadow can be predicted using the cloud's shape, height and sun position at the time [46]. However, this method first depends on the cloud identification ability and poses large uncertainty while projecting the cloud's shadow on the earth's surface. This study used Temporal Dark Outlier Mask (TDOM) method which greatly improves the detection of cloud shadow via catching dark pixel anomaly [41]. The TDOM method based on the idea that cloud shadow appears dark and disappears quickly as the cloud moves. The implementation of TDOM in GEE was adapted from [28].

2.7. Co-registration between Landsat and Sentinel 2 images

The miss alignment (or miss registration) between L8 and S2 images varied geographically and can exceed 38 meters [16]. It is mainly due to the residual geolocation errors in the Landsat-8 framework which based upon the Global Land Survey images. In GEE, we used displacement() to measure the displacement between two overlapped S2 and L8 images which were captured at the same time over the studied region. Then displace() function is used to displace or wrap ("rubber-sheet" technique) the L8 image aligned with the S2 image [47]. Because the L8-S2 misalignment is reported stable for a given area and S2 absolute geodetic accuracy is better than L8 [16], this study aligned all Landsat images (same PATH, ROW) using a common base S2 [48]. We also assumed that the misalignment among the same satellite images is neglectable. The co-alignment step described here is purely an image processing technique. It differs from geo-referencing or geo-correcting which involves aligning images to the correct geographic location through ground control points. At the moment, GEE documentation does not explain clearly the underlying of displacement() and displace() algorithms, however, [49] described in great details a similar tool called AROP which is an open-source package designed specifically for registration and orthorectification of Landsat and Landsat-like data.

2.8. Re-projection and Scaling

Because each band can have a different scale and projection [50] therefore band's projection was transformed according to the red band of S2 (WGS84) and band's resolution was rescaled to 30m using 'bicubic' interpolation [51,52].

2.9. BRDF correction

The Bidirectional Reflectance Distribution Functions (BRDF) model is applied to reduce the directional effects due to the differences in solar and view angles between Landsat and Sentinel 2 [10]. The implementation of BRDF correction in GEE was developed by [28] based on results from [25] and [26]. This BRDF is MODIS-based fixed coefficients c-factor, originally developed for Landsat but proven to be working for S2 as well [18,25,26]. The view angle is set to nadir and the illumination is set based on the center latitude of the tile [10].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2019 doi:10.20944/preprints201910.0275.v1

Peer-reviewed version available at Remote Sens. 2020, 12, 281; doi:10.3390/rs12020281

2.10. Topographic correction

Topographic correction accounts for variations in reflectance due to slope, aspect, and elevation. Topographic correction is not always required but can be essential in mountainous or rugged terrain [53,54]. The implementation of topographic correction in GEE was developed by [28]. The method based on the modified Sun-Canopy-Sensor Topographic Correction as described in [27]. The digital elevation model (DEM) used is SRTM V3 product (30m SRTM Plus) which has undergone a void-filling process using open-source data (ASTER GDEM2, GMTED2010, and NED) provided by NASA JPL [55].

2.11. Band adjustment

Although, efforts have been made into the radiometric and geometric calibration of the independently managed Landsat and Sentinel 2 missions so that their bands are compatible [17], small spectral differences in the common bands still exist [10,17,29]. We adjusted the six Landsat bands (blue, green, red, nir, swir1 and swir2) using cross-sensor transformation coefficients (Table 1) derived from [29]'s study. [29] used absolute difference metrics and major axis linear regression analysis over 10,000 image pairs across the conterminous United States to compute these transformation coefficients.

Table 1. Cross-sensor transformation coefficients for Landsat7/8 [29]

	L8		L7	
Bands	Intercept	Slope	Intercept	Slope
Blue	−0.0107	1.0946	−0.0139	1.1060
Green	0.0026	1.0043	0.0041	0.9909
Red	−0.0015	1.0524	−0.0024	1.0568
NIR	0.0033	0.8954	−0.0076	1.0045
SWIR1	0.0065	1.0049	0.0041	1.0361
SWIR2	0.0046	1.0002	0.0086	1.0401

(With: Sentinel2 = Landsat7/8 * Slope + Intercept)

3. Results and Discussions

3.1. Design of the evaluation experiments

Because this study employed several transformation models from other studies, for example, BRDF and topographic correction models from [28]; band adjustment coefficients from [29]; image co-registration from [47]. However, some studies suggested site-specific models may be required for specific areas of study due to inconsistent regression coefficient values obtained across different study areas [18,56,57]. Therefore, we applied several tests to evaluate the effect of each processing/transformation step on two overlapped S2 and L8 images which were captured at the same time over the studied regions. Rectangular areas without cloud, cirrus or saturated pixels were selected for analysis. Tested image IDs, date, time captured are presented in Table 2. Section 3.2 estimated the reduction in sensor mis-registration, section 3.3 calculated per band spatial Pearson's correlation and section 3.4 assessed the temporal correlation of NDVI time series.

3.2. Reducing the sensors mis-registration

Figure 3 showed per pixel offset differences in the tested areas, measured by the magnitude of the vector formed by dX and dY [47], before and after the overlapped pair of L8-S2 images were co-registered using the method described in 2.8. For Bekaa, the offset differences were reduced significantly from [22 - 32] meters to less than 8 meters (mostly less than 2 meters). For Ninh Thuan, Vietnam, the mis-alignment was reduced from 12 meters (maximum) to mostly less than 2

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2019 doi:10.20944/preprints201910.0275.v1

Peer-reviewed version available at Remote Sens. 2020, 12, 281; doi:10.3390/rs12020281

Table 2. Overlapped S2 and L8 images selected for performance evaluation

Products	Acquisition Date	Time	Image ID	Region
L8 _TOA	07/11/2018	08:10:14	LANDSAT/LC08/C01/T1/LC08_174036_20181107	Bekaa
S2 _L1C	07/11/2018	08:30:42	COPERNICUS/S2/20181107T082129_20181107T082732_T36SYC	Bekaa
L8 _TOA	11/08/2017	03:01:30	LANDSAT/LC08/C01/T1_TOA/LC08_123052_20170811	Ninh Thuan
S2 _L1C	11/08/2017	03:23:19	COPERNICUS/S2/20170811T032319_20170811T032319_T49PBN COPERNICUS/S2/20170811T032319_20170811T032319_T49PBP	Ninh Thuan

meters. These results are in agreement with [16] who also found geographically varied mis-alignments between L8-S2. Further analysis in Table 3 showed that the co-registration step contributed the most improvement in band-to-band spatial correlation.

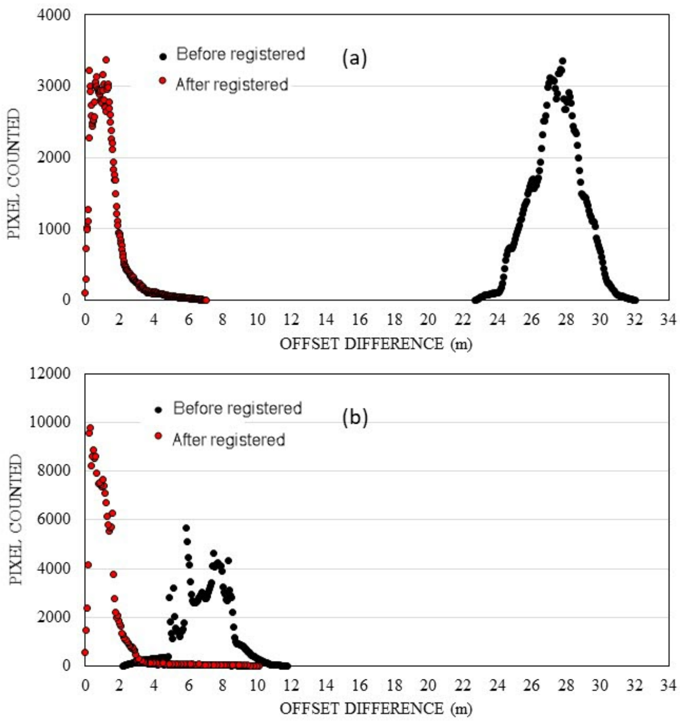

Figure 3. Per-pixel offset differences between L8 and S2 in the tested areas ((a) Bekaa, Lebanon and (b) Ninh Thuan, Vietnam) measured by the magnitude of the vector formed by dX and dY. Black dots are the offset distances between the original images of L8 and S2, red dots are when they were co-registered.

3.3. Band to band spatial correlation

We analyzed the band-to-band correlation over two separated domains, a flat agricultural area (Figure 4a), and a mountainous area (Figure 4b). Each domain has an area of 0.3 km2, without cloud, cirrus or saturated pixels.

Table 3 compared Pearson correlation (r) values of bands (red, nir and ndvi) when each processing/transformation step is applied, in the flat area (Table 3a) and the mountainous area (Table 3b). P0 represents the starting point and P5 represents the last step (band adjustment). For the flat area, correlation values increased significantly from (0.67, 0.75, 0.79) to (0.93, 0.95, 0.96) in bands (red, nir, ndvi) respectively. For the mountainous area, r increased from (0.56, 0.45, 0.63) to (0.77,

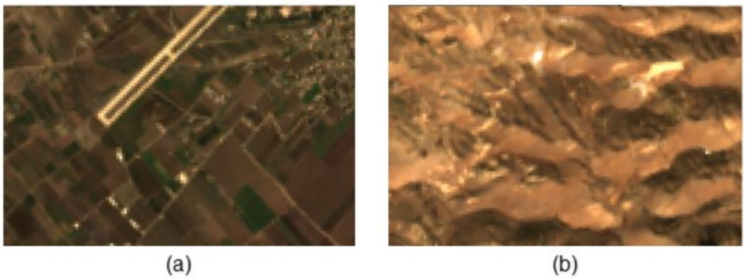

Figure 4. A flat agricultural area (a), and a mountainous area (b) in Bekaa, Lebanon, represented by true-color composites of the Landsat 8, used for L8-S2 band-to-band correlation analysis. Each domain has an area of 0.3 km2, without cloud, cirrus or saturated pixels.

0.72, 0.80) in (red, nir, ndvi) bands. There is a higher correlation occurred in the flat area than in the mountainous area is likely due to the impacts of untreated hill shadow or hill's slope. This result stands in agreement with [53] and [54] who emphasized the importance of properly topographic correction in mountainous or rugged terrain. Table 3 also indicated the co-registration step (P3) contributed the most improvement in band-to-band spatial correlation.

Table 3. L8-S2 cross-comparison of the Pearson correlation (r) values in bands (red, nir and ndvi) when each processing/transformation step is applied, in the flat area (a) and the mountainous area (b)

(a)

	Processing Step	Pearson's r (Red)	Pearson's r (NIR)	Pearson's r (NDVI)
P0	Original (BOA)	0.6665	0.7479	0.7871
P1	Rescale L8 to 10m	0.6917	0.7655	0.808
P2	BRDF correction	0.6917	0.7634	0.8089
P3	Co-registration	0.9268	0.9490	0.9637
P4	Topo correction	0.9270	0.9490	0.9637
P5	Band Adjustment	0.9270	0.9490	0.9641

(b)

	Processing Step	Pearson's r (Red)	Pearson's r (NIR)	Pearson's r (NDVI)
P0	Original (BOA)	0.5647	0.4453	0.6347
P1	Rescale L8 to 10m	0.5773	0.4561	0.6606
P2	BRDF correction	0.5765	0.4525	0.6618
P3	Co-registration	0.7064	0.5860	0.8408
P4	Topo correction	0.7719	0.7193	0.7969
P5	Band Adjustment	0.7719	0.7193	0.7973

For further analysis in the flat domain, Figure A1 presented per-pixel scatter plots of all seven bands (blue, green, red, nir, swir1, swir2, and ndvi), compared (r, bias, and RMSE) before and after the overlapped L8-S2 images were harmonized. These plots showed all bands are in good agreement. Band SWIR1 reached the highest correlation ($r = 0.972$) and band Blue has the lowest ($r = 0.868$).

3.4. Affect of band adjustment to temporal correlation in NDVI time series

Figure 5a and Figure 5b showed NDVI time series at a typical pixel (lat =, long =) before and after the band adjustment was applied. Ones can observe that, before band adjustment, the NDVI values of L8 were systematically lower than that of S2 (Figure 9a), but after the band adjustment, the two datasets matched chronologically (Figure 9b). Figure 9c showed the final harmonized ndvi time series which gathered data from all sensors. There are gaps existed in the time series because cloudy covered images were automatically eliminated in the process.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2019 doi:10.20944/preprints201910.0275.v1

Peer-reviewed version available at *Remote Sens.* **2020**, *12*, 281; doi:10.3390/rs12020281

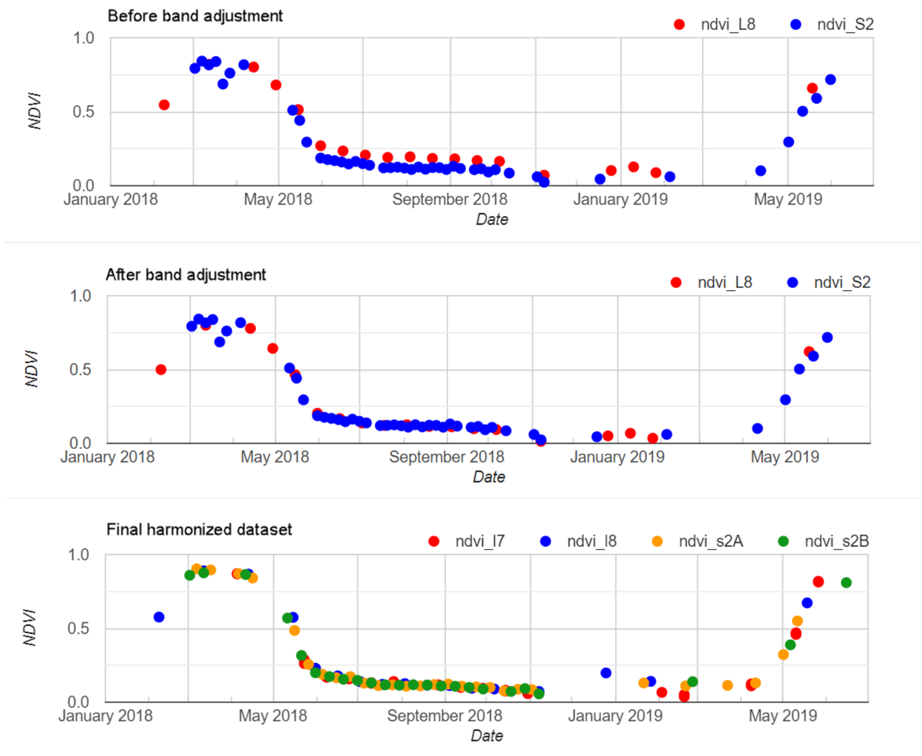

Figure 5. First two plots are the temporal NDVI time series over a typical crop pixel in Bekaa, Lebanon (lat =36.01, long =33.83) before and after band adjustment. The last image showed the harmonized NDVI time series from all sensors. There are gaps in the time series because the cloudy images were eliminated in the processing.

As previously reported in Table 3b, the spatial band-to-band correlation is low in the mountainous region due to the impacts of the hill's slope or remaining of untreated hill shadow. This problem is further visualized in Figure 6 which showed the NDVI time series of a pixel located in a mountainous area (lat = 36.04, long = 33.81). After the processing, Landsat's NDVI values were seen systematically lower than that of Sentinel 2. This result suggested that the topographic correction model can be improved.

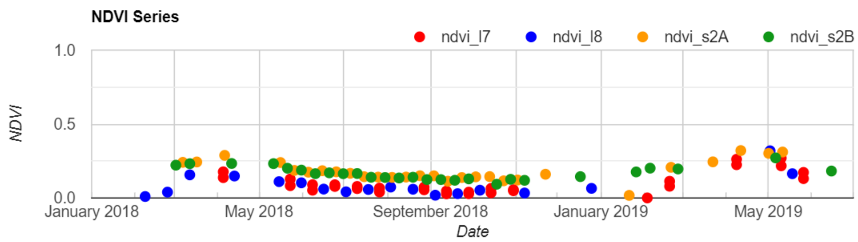

Figure 6. NDVI time series of a pixel located in a mountainous area in Bekaa (lat = 36.04, long = 33.81). After the processing, Landsat's NDVI values were seen systematically lower than that of Sentinel 2

3.5. Assessing the dynamic cropland variation in Ninh Thuan, Vietnam

Ninh Thuan province is the most drought-prone in Vietnam [58]. To cope with water shortage throughout the next dry season (from Jan to Aug), exceeded rainwater during the rainy season (from Aug to Nov) is collected via more than 20 small to medium size reservoirs. Water is irrigated for the next two crop seasons which are winter-spring crop (from Dec to Apr) and summer crop (from May to Aug). Thus, the extended cropland area is largest during the winter-spring season, then reduced

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2019 doi:10.20944/preprints201910.0275.v1

Peer-reviewed version available at Remote Sens. 2020, 12, 281; doi:10.3390/rs12020281

during the summer because of possible water shortage. Meanwhile, the crop during rainy season can be vulnerable to flood [59].

Extended cropland is valuable for the province's Irrigation Management Company (IMC) to calculate water distribution volume and predict water demand for the next season. However, because of seasonal variation, mixed crop rotation and data-scarce, it is difficult for the province to obtain up-to-date and accurate seasonal extended cropland.

As harmonic (or Fourier) analysis has proven useful in characterizing seasonal cycles and variation in land used/land cover types [8,60–63], this study applied harmonic analysis on dense NDVI time series, obtained from the harmonized dataset (L7, L8, and S2), to mapping seasonal cropland in Ninh Thuan during 2018.

Following a methodology described in [8] and implementation of the harmonic model in GEE by [64], we fitted the time series of NDVI data in every pixel. Figure 7 showed NDVI time series and fitted values of regions that have one crop, two crops and three crops per year in Ninh Thuan region. The phase and amplitude values, which were derived from the harmonic models, will be used to express the temporal signature of NDVI.

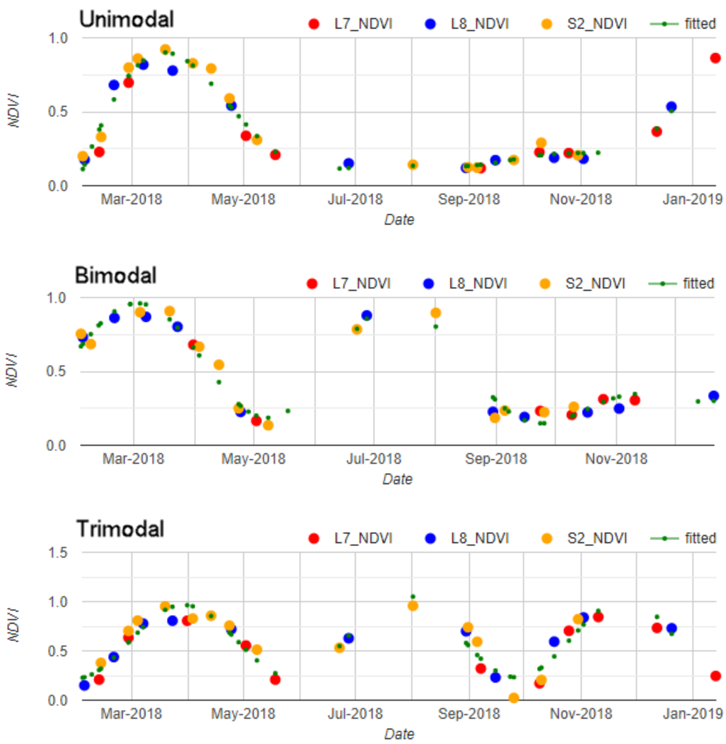

Figure 7. Detected unimodal (a), bimodal (b) and trimodal (c) shapes in the temporal NDVI patterns of different paddy rice areas during 2018. Fitted values (smaller green dots) are used to calculate the phase and amplitude of the cycles.

Since the first harmonic term represents the annual cycle [60], the cropland's variation was identified using a composite image of phase, amplitude (of the first harmonic term) and the max NDVI (Figure 8). Because NDVI at cropland pixels are characterized with high temporal variation, high angle or sharp turn at the peak of crop growth, and high max NDVI values, croplands were highlighted in the Figure 8 as bright colored pixels. Meanwhile, black or gray pixels represent non-cropland. Within the scope of this study, we only interest in the cropland location, although, specific crop types can be further identified using a rule-based approach and ground truth parameters [8].

Peer-reviewed version available at *Remote Sens.* **2020**, *12*, 281; doi:10.3390/rs12020281

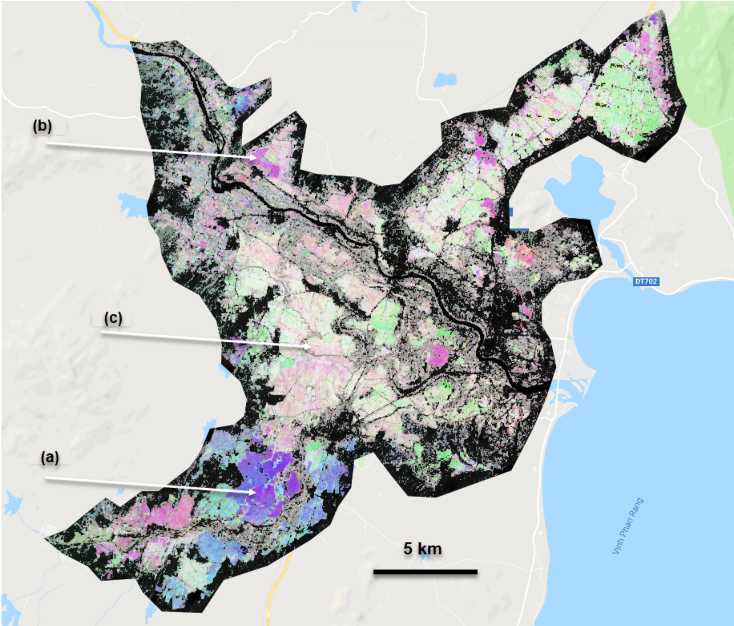

Figure 8. Cropland variation characterized by R-G-B composite image from amplitude, phase (of the first harmonic term) and max-NDVI values. This map highlighted croplands in Ninh Thuan during 2018 as colored pixels (high phase and amplitude) and other types of land as grey/dark pixels (low phase/amplitude).

4. Conclusions

In the presented paper, we demonstrated a complete stream workflow in Google Earth Engine to generate harmonized Landsat – Sentinel 2 images for two agriculture schemes in Bekaa, Lebanon, and Ninh Thuan, Vietnam. We evaluated the performance of several pre-processing steps necessary for the harmonization including image co-registration, brdf correction, topographic correction, and band adjustment. Band adjustment, although, has little impact on L8-S2 spatial correlation, it is valuable for matching temporal spectral time series. The offset difference between L8 and S2 images was as large as 32 meters in the Bekaa region and if not treated, posed a great impact on the quality of the harmonized dataset. Although a topographic correction model was applied, the low performance was observed in mountainous areas.

The merging of multiple sensors improved crop monitoring as it increased temporal resolution and provided more observations during the growing season. Dense observations also omit the need for data smoothing techniques. We demonstrated an application of the harmonized dataset by mapping the extended cropland via harmonic analysis for Ninh Thuan province in 2018.

Author Contributions: M.N. contributed to the study's design, data analysis, programming and writing the manuscript, O.V. and L.R. made contribution to the design, statistical analyses and supervised at all stages of the study, O.V, L.R., D.B., P.N. instructed and discussed the results. All authors read and approved the submitted manuscript

Funding:

Acknowledgments: The author acknowledges support by Vingroup Innovation Foundation (VINIF) in project code VINIF.2019.DA17 and Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number NE/S002847/1

Conflicts of Interest:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2019 doi:10.20944/preprints201910.0275.v1

Peer-reviewed version available at *Remote Sens.* **2020**, *12*, 281; doi:10.3390/rs12020281

Appendix A

Generated harmonized datasets that contain surface reflectance images (bands blue, green, red, nir, swir1, swir2, and ndvi at 30 meters) over the two studied sites are provided for public usage and testing. Data link (Google drive): https://drive.google.com/open?id=1no0MmpL_WA8BWzFRmmUWGPMt-JYMtI-P. GEE app to inspect the NDVI time series and the detected croplands in Ninh Thuan: https://ndminhhus.users.earthengine.app/view/cropninhthuan2019. All GEE scripts used in the study are documented at https://github.com/ndminhhus/geeguide.

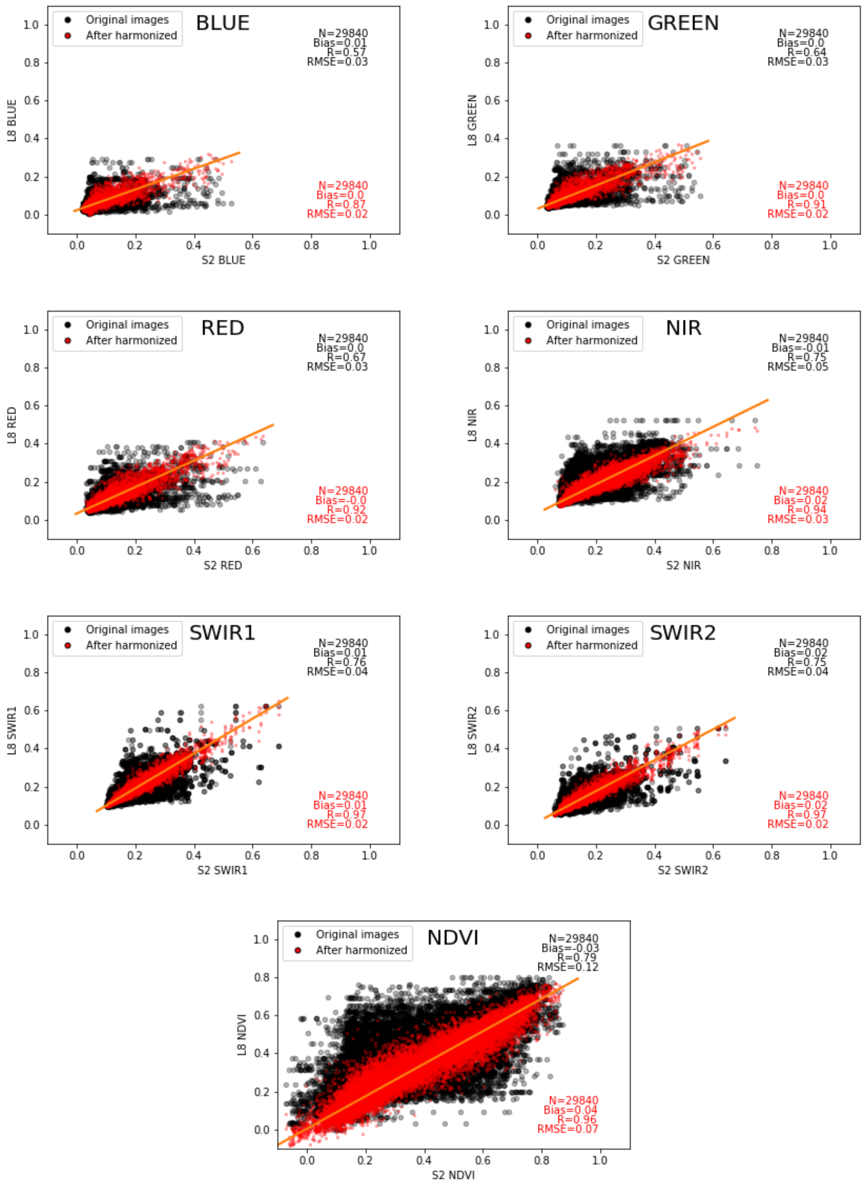

Figure A1. Per-pixel scatters plots of all seven bands (blue, green, red, nir, swir1, swir2, and ndvi) for the flat domain in Bekaa, provided N (total number of the pixels), r, bias, and Root Mean Square Error-RMSE after the overlapped L8-S2 images were harmonized. The straight line represents the linear regression.

Peer-reviewed version available at Remote Sens. 2020, 12, 281; doi:10.3390/rs12020281

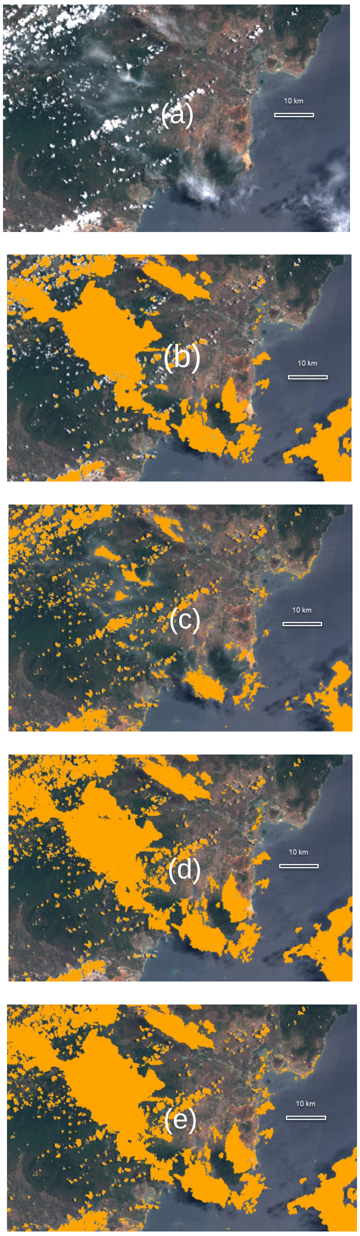

Figure A2. Demonstration of cloud masking steps. (a) Cloudy true color image, (b) Cloud & cirrus masked (yellow) using only QA60 Band, (c) cloud mask using combination of Red and Aerosol Band (B4 & B1), (d) cloud mask using random forest classification, band QA60 was used as training field (e) cloud mask combined all together. This scene was acquired by Sentienl2B on May 12, 2019 over Ninh Thuan, Vietnam (id = COPERNICUS/S2/20190513T030549_20190513T032056_T49PBN).

References

1. Bégué, A.; Arvor, D.; Bellon, B.; Betbeder, J.; De Abelleyra, D.; PD Ferraz, R.; Lebourgeois, V.; Lelong, C.; Simões, M.; R Verón, S. Remote sensing and cropping practices: A review. *Remote Sensing* **2018**, *10*, 99.
2. Cheng, T.; Yang, Z.; Inoue, Y.; Zhu, Y.; Cao, W. Preface: Recent advances in remote sensing for crop growth monitoring, 2016.
3. He, M.; Kimball, J.; Maneta, M.; Maxwell, B.; Moreno, A.; Beguería, S.; Wu, X. Regional crop gross primary productivity and yield estimation using fused landsat-MODIS data. *Remote Sensing* **2018**, *10*, 372.
4. Yang, D.; Su, H.; Zhan, J. MODIS-Landsat Data Fusion for Estimating Vegetation Dynamics - A Case Study for Two Ranches in Southwestern Texas. *Proceedings of 1st International Electronic Conference on Remote Sensing* **2015**.
5. Hansen, M.C.; Loveland, T.R. A review of large area monitoring of land cover change using Landsat data. *Remote sensing of Environment* **2012**, *122*, 66–74.
6. Whitcraft, A.; Becker-Reshef, I.; Justice, C. A framework for defining spatially explicit earth observation requirements for a global agricultural monitoring initiative (GEOGLAM). *Remote Sensing* **2015**, *7*, 1461–1481.
7. Hilker, T.; Wulder, M.A.; Coops, N.C.; Seitz, N.; White, J.C.; Gao, F.; Masek, J.G.; Stenhouse, G. Generation of dense time series synthetic Landsat data through data blending with MODIS using a spatial and temporal adaptive reflectance fusion model. *Remote Sensing of Environment* **2009**, *113*, 1988–1999.
8. Ghazaryan, G.; Dubovyk, O.; Löw, F.; Lavreniuk, M.; Kolotii, A.; Schellberg, J.; Kussul, N. A rule-based approach for crop identification using multi-temporal and multi-sensor phenological metrics. *European Journal of Remote Sensing* **2018**, *51*, 511–524.
9. USGS. US Geological Survey (USGS) Landsat Collection 1 Level-1 Quality Assessment Band. https://www.usgs.gov/land-resources/nli/landsat/landsat-collection-1-level-1-quality-assessment-band?qt-science_support_page_related_con=0#qt-science_support_page_related_con, 2019. Accessed: 2019-09-09.
10. Claverie, M.; Ju, J.; Masek, J.G.; Dungan, J.L.; Vermote, E.F.; Roger, J.C.; Skakun, S.V.; Justice, C. The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote sensing of environment* **2018**, *219*, 145–161.
11. Boschetti, L.; Roy, D.P.; Justice, C.O.; Humber, M.L. MODIS–Landsat fusion for large area 30 m burned area mapping. *Remote Sensing of Environment* **2015**, *161*, 27–42.
12. Ma, J.; Zhang, W.; Marinoni, A.; Gao, L.; Zhang, B. An improved spatial and temporal reflectance unmixing model to synthesize time series of landsat-like images. *Remote Sensing* **2018**, *10*, 1388.
13. Wang, Q.; Zhang, Y.; Onojeghuo, A.O.; Zhu, X.; Atkinson, P.M. Enhancing spatio-temporal fusion of modis and landsat data by incorporating 250 m modis data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2017**, *10*, 4116–4123.
14. Dong, J.; Zhuang, D.; Huang, Y.; Fu, J. Advances in multi-sensor data fusion: Algorithms and applications. *Sensors* **2009**, *9*, 7771–7784.
15. Behnia, P. Comparison between four methods for data fusion of ETM+ multispectral and pan images. *Geo-Spatial Information Science* **2005**, *8*, 98–103.
16. Storey, J.; Roy, D.P.; Masek, J.; Gascon, F.; Dwyer, J.; Choate, M. A note on the temporary misregistration of Landsat-8 Operational Land Imager (OLI) and Sentinel-2 Multi Spectral Instrument (MSI) imagery. *Remote Sensing of Environment* **2016**, *186*, 121–122.
17. Barsi, J.A.; Alhammoud, B.; Czapla-Myers, J.; Gascon, F.; Haque, M.O.; Kaewmanee, M.; Leigh, L.; Markham, B.L. Sentinel-2A MSI and Landsat-8 OLI radiometric cross comparison over desert sites. *European Journal of Remote Sensing* **2018**, *51*, 822–837.
18. Zhang, H.K.; Roy, D.P.; Yan, L.; Li, Z.; Huang, H.; Vermote, E.; Skakun, S.; Roger, J.C. Characterization of Sentinel-2A and Landsat-8 top of atmosphere, surface, and nadir BRDF adjusted reflectance and NDVI differences. *Remote sensing of environment* **2018**, *215*, 482–494.
19. Franch, B.; Vermote, E.; Skakun, S.; Roger, J.C.; Santamaria-Artigas, A.; Villaescusa-Nadal, J.L.; Masek, J. Towards Landsat and Sentinel-2 BRDF normalization and albedo estimation: a case study in the Peruvian Amazon forest. *Frontiers in Earth Science* **2018**, *6*, 185.
20. Ilori, C.O.; Pahlevan, N.; Knudby, A. Analyzing Performances of Different Atmospheric Correction Techniques for Landsat 8: Application for Coastal Remote Sensing. *Remote Sensing* **2019**, *11*, 469.

Peer-reviewed version available at Remote Sens. 2020, 12, 281; doi:10.3390/rs12020281

21. Doxani, G.; Vermote, E.; Roger, J.C.; Gascon, F.; Adriaensen, S.; Frantz, D.; Hagolle, O.; Hollstein, A.; Kirches, G.; Li, F.; others. Atmospheric correction inter-comparison exercise. *Remote Sensing* **2018**, *10*, 352.
22. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* **2017**, *202*, 18–27.
23. Kumar, L.; Mutanga, O. Google Earth Engine applications since inception: Usage, trends, and potential. *Remote Sensing* **2018**, *10*, 1509.
24. Kennedy, R.; Yang, Z.; Gorelick, N.; Braaten, J.; Cavalcante, L.; Cohen, W.; Healey, S. Implementation of the LandTrendr Algorithm on Google Earth Engine. *Remote Sensing* **2018**, *10*, 691.
25. Roy, D.P.; Li, J.; Zhang, H.K.; Yan, L.; Huang, H.; Li, Z. Examination of Sentinel-2A multi-spectral instrument (MSI) reflectance anisotropy and the suitability of a general method to normalize MSI reflectance to nadir BRDF adjusted reflectance. *Remote Sensing of Environment* **2017**, *199*, 25–38.
26. Roy, D.P.; Zhang, H.; Ju, J.; Gomez-Dans, J.L.; Lewis, P.E.; Schaaf, C.; Sun, Q.; Li, J.; Huang, H.; Kovalskyy, V. A general method to normalize Landsat reflectance data to nadir BRDF adjusted reflectance. *Remote Sensing of Environment* **2016**, *176*, 255–271.
27. Soenen, S.A.; Peddle, D.R.; Coburn, C.A. SCS+ C: A modified sun-canopy-sensor topographic correction in forested terrain. *IEEE Transactions on geoscience and remote sensing* **2005**, *43*, 2148–2159.
28. Poortinga, A.; Tenneson, K.; Shapiro, A.; Nquyen, Q.; San Aung, K.; Chishtie, F.; Saah, D. Mapping Plantations in Myanmar by Fusing Landsat-8, Sentinel-2 and Sentinel-1 Data along with Systematic Error Quantification. *Remote Sensing* **2019**, *11*, 831.
29. Chastain, R.; Housman, I.; Goldstein, J.; Finco, M. Empirical cross sensor comparison of Sentinel-2A and 2B MSI, Landsat-8 OLI, and Landsat-7 ETM+ top of atmosphere spectral characteristics over the conterminous United States. *Remote sensing of environment* **2019**, *221*, 274–285.
30. NinhThuan-DONRE. Shapefiles "paddy rice, secondary crop and perennial trees in Ninh Thuan, surveyed 2015". *Project data* **2015**.
31. eLEAF B.V. Summary Methodology of Level 3 Land Cover Mapping. *Project data* **2015**.
32. Vermote, E.F.; Tanré, D.; Deuze, J.L.; Herman, M.; Morcette, J.J. Second simulation of the satellite signal in the solar spectrum, 6S: An overview. *IEEE transactions on geoscience and remote sensing* **1997**, *35*, 675–686.
33. Wilson, R. Py6S: A Python interface to the 6S Radiative Transfer. http://rtwilson.com/academic/Wilson_2012_Py6S_Paper.pdf, 2012. Accessed: 2019-09-09.
34. Kotchenova, S.Y.; Vermote, E.F. Validation of a vector version of the 6S radiative transfer code for atmospheric correction of satellite data. Part II. Homogeneous Lambertian and anisotropic surfaces. *Applied optics* **2007**, *46*, 4455–4464.
35. Sam, M. Atmospheric correction of Sentinel 2 imagery in Google Earth Engine using Py6S. https://github.com/samsammurphy/gee-atmcorr-S2, 2018. Accessed: 2019-09-09.
36. Foga, S.; Scaramuzza, P.L.; Guo, S.; Zhu, Z.; Dilley Jr, R.D.; Beckmann, T.; Schmidt, G.L.; Dwyer, J.L.; Hughes, M.J.; Laue, B. Cloud detection algorithm comparison and validation for operational Landsat data products. *Remote sensing of environment* **2017**, *194*, 379–390.
37. Coluzzi, R.; Imbrenda, V.; Lanfredi, M.; Simoniello, T. A first assessment of the Sentinel-2 Level 1-C cloud mask product to support informed surface analyses. *Remote sensing of environment* **2018**, *217*, 426–443.
38. ESA. Level-1C Cloud Masks - Sentinel-2 MSI Technical Guide - Sentinel Online. https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-2-msi/level-1c/cloud-masks, 2019. Accessed: 2019-09-09.
39. Clerc, S.; Devignot, O.; Pessiot, L. S2 MPC TEAM S2 MPC Data Quality Report - 2015-11-30. https://sentinels.copernicus.eu/documents/247904/685211/Sentinel-2+Data+Quality+Report, 2015. Accessed: 2019-09-09.
40. GEE. Landsat Algorithms in Google Earth Engine API. https://developers.google.com/earth-engine/landsat, 2019. Accessed: 2019-09-09.
41. Housman, I.; Chastain, R.; Finco, M. An Evaluation of Forest Health Insect and Disease Survey Data and Satellite-Based Remote Sensing Forest Change Detection Methods: Case Studies in the United States. *Remote Sensing* **2018**, *10*, 1184.
42. Zupanc, A. Improving Cloud Detection with Machine Learning. https://medium.com/sentinel-hub/improving-cloud-detection-with-machine-learning-c09dc5d7cf13, 2017. Accessed: 2019-09-09.

43. Hagolle, O.; Huc, M.; Pascual, D.V.; Dedieu, G. A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENμS, LANDSAT and SENTINEL-2 images. *Remote Sensing of Environment* **2010**, *114*, 1747–1755.
44. Breiman, L. Random forests. *Machine learning* **2001**, *45*, 5–32.
45. GEE. Supervised Classification in Google Earth Engine API. https://developers.google.com/earth-engine/classification, 2019. Accessed: 2019-09-09.
46. Hollstein, A.; Segl, K.; Guanter, L.; Brell, M.; Enesco, M. Ready-to-use methods for the detection of clouds, cirrus, snow, shadow, water and clear sky pixels in Sentinel-2 MSI images. *Remote Sensing* **2016**, *8*, 666.
47. GEE. Registering Images in Google Earth Engine API. https://developers.google.com/earth-engine/register, 2019. Accessed: 2019-09-09.
48. Claverie, M.; Jeffrey, G.M.; Junchang, J.; Jennifer, L.D. Harmonized Landsat-8 Sentinel-2 (HLS) Product User's Guide. https://hls.gsfc.nasa.gov/wp-content/uploads/2017/08/HLS.v1.3.UserGuide_v2-1.pdf, 2017. Accessed: 2019-09-09.
49. Gao, F.; Masek, J.G.; Wolfe, R.E. Automated registration and orthorectification package for Landsat and Landsat-like data processing. *Journal of Applied Remote Sensing* **2009**, *3*, 033515.
50. GEE. Projections in Google Earth Engine API. https://developers.google.com/earth-engine/projections, 2019. Accessed: 2019-09-09.
51. GEE. Resampling and Reducing Resolution in Google Earth Engine API. https://developers.google.com/earth-engine/resample, 2019. Accessed: 2019-09-09.
52. Keys, R. Cubic convolution interpolation for digital image processing. *IEEE transactions on acoustics, speech, and signal processing* **1981**, *29*, 1153–1160.
53. COLBY, J. TOPOGRAPHIC NORMALIZATION IN RUGGED TERRAIN. *PHOTOGRAMMETRIC ENGINEERING AND REMOTE SENSING* **1991**, *57*, 531–537.
54. Vanonckelen, S.; Lhermitte, S.; Van Rompaey, A. The effect of atmospheric and topographic correction methods on land cover classification accuracy. *International Journal of Applied Earth Observation and Geoinformation* **2013**, *24*, 9–21.
55. GEE. SRTM Digital Elevation Data 30m. https://developers.google.com/earth-engine/datasets/catalog/USGS_SRTMGL1_003, 2019. Accessed: 2019-09-09.
56. Mandanici, E.; Bitelli, G. Preliminary comparison of sentinel-2 and landsat 8 imagery for a combined use. *Remote Sensing* **2016**, *8*, 1014.
57. Flood, N. Comparing Sentinel-2A and Landsat 7 and 8 using surface reflectance over Australia. *Remote Sensing* **2017**, *9*, 659.
58. VAWR. Report "Strengthening the agro-climatic information system to improve the agricultural drought monitoring and early warning system in Vietnam (NEWS), pilot study in the Ninh Thuan province". *Project inception report* **2017**.
59. NinhThuan-MPI. Report "Review and update irrigation planning of the Ninh Thuan province to 2020, vision to 2030 under climate change scenario". *Project report* **2015**.
60. Jakubauskas, M.E.; Legates, D.R.; Kastens, J.H. Crop identification using harmonic analysis of time-series AVHRR NDVI data. *Computers and electronics in agriculture* **2002**, *37*, 127–139.
61. Jakubauskas, M.E.; Legates, D.R.; Kastens, J.H.; others. Harmonic analysis of time-series AVHRR NDVI data. *Photogrammetric engineering and remote sensing* **2001**, *67*, 461–470.
62. Jakubauskas, M.; Legates, D.R. Harmonic analysis of time-series AVHRR NDVI data for characterizing US Great Plains land use/land cover. *International Archives of Photogrammetry and Remote Sensing* **2000**, *33*, 384–389.
63. Shumway, R.H.; Stoffer, D.S. *Time series analysis and its applications: with R examples*; Springer, 2017.
64. Nick, C. Time Series Analysis in Earth Engine. https://goo.gl/lMwd2Y, 2017. Accessed: 2019-09-09.