Operational Coregistration of the Sentinel-2A/B Image Archive Using Multitemporal Landsat Spectral Averages

Philippe Rufin, David Frantz, Lin Yan, and Patrick Hostert

***Abstract*— Geometric misalignment between Landsat and Sentinel-2 data sets as well as multitemporal inconsistency of Sentinel-2A and -2B data sets currently complicate multitemporal analyses. Operational coregistration of Sentinel-2A and -2B imagery is thus required. We present a modification of the established Landsat Sentinel Registration (LSReg) algorithm. The modifications enabled LSReg to be included in an operational preprocessing workflow to automatically coregister large volumes of Sentinel-2 imagery with Landsat base images that represent multiannual monthly spectral average values. The modified LSReg was tested for the complete Sentinel-2 archive covering Crete, Greece, which is a particularly challenging region due to steep topographic gradients and high shares of water in Sentinel-2 tiles. A coregistration success rate of 87.5% of all images was obtained with a mean coregistration precision of 4.4 m. The mean shifts of 14.0 m in the x-direction and 13.4 m in the y-direction before coregistration were found, with maxima exceeding four pixels. Time series noise in locations with land cover transitions ($n = 585$) was effectively reduced by 43% using the presented approach. The multitemporal geometric consistency of the Sentinel-2 data set was substantially improved, thus enabling time series analyses within the Sentinel-2 data record, as well as integrated Landsat and Sentinel-2A and -2B data sets. The modified algorithm is implemented in the Framework for Operational Radiometric Correction for Environmental monitoring (FORCE) version 3.0 (https://github.com/davidfrantz/force).**

***Index Terms*— Coregistration, geometric accuracy, Landsat, multisensor, multitemporal, Sentinel-2, time series.**

I. Introduction

TIME series analysis of remotely sensed data enables characterization of land cover, land use, and long-term changes therein. Geometric consistency within single-sensor image time series and between time series obtained from multiple sensors is a vital prerequisite for such analyses [1].

Recently, the integration of medium resolution optical sensors, such as Landsat 8 Operational Land Imager (OLI) and Sentinel-2 (S2) Multispectral Instrument, gained traction, posing high demands to the geometric preprocessing of individual data records and consistency between data sets [2].

The Collection 1 Tier 1 Level 1 Landsat record systematically provides high multitemporal geometric accuracies (<7 m at worst) with absolute geometric accuracy of <13 m, that is, less than half a pixel [3]. The S2 Level 1C data products currently have a 12-m multitemporal accuracy, which means that coregistration errors within a pure S2 time series can already exceed a full pixel in the 10-m visible (VIS) and near-infrared (NIR) bands [4]. The expected geometric error between Landsat and S2 data currently amounts for up to 38 m, which further underlies geographic variation due to varying quality of the Global Land Survey 2000 ground control [5]. These geometric inaccuracies superimpose challenges for single-sensor and particularly multisensor time series analyses. Lacking image-level metadata on geometric accuracy in S2 data adds to the uncertainty on the user side.

Improvements on the absolute geometric accuracy and multitemporal coregistration of both S2 sensors are expected upon the release of the global Geometric Reference Image (GRI) by the European Space Agency (ESA), which is assumed to reduce the multitemporal error to less than 0.3 pixels at 95% [6]. Unfortunately, the release of the GRI has been repeatedly delayed, reprocessing of past S2 data is currently not planned, and the expected global geo-registration accuracy needs to be confirmed in an operational setting. Time series analysis in world regions with inconsistencies in image geometry thus requires operational means to remove pixel and subpixel inconsistencies between S2 and Landsat data, and within S2 data before the release of the GRI [7].

Numerous approaches for automated tie point detection and coregistration of Landsat and S2 data sets were presented recently, and most of them reporting substantial inter- and intrasensor geometric mismatches [2], [7]–[10]. Current automated image coregistration techniques frequently rely on area-based correlation or Fourier-based matching techniques for automated tie point detection [10], [11]. Subsequently, affine or polynomial translation functions, or Random Forest

Manuscript received November 13, 2019; revised February 10, 2020; accepted March 17, 2020. *(Corresponding author: Philippe Rufin.)*

Philippe Rufin and Patrick Hostert are with the Earth Observation Laboratory, Geography Department, Humboldt-Universität zu Berlin, 10099 Berlin, Germany, and also with the Integrative Research Institute on Transformations of Human-Environment Systems, Humboldt-Universität zu Berlin, 10099 Berlin, Germany (e-mail: philippe.rufin@hu-berlin.de; patrick.hostert@hu-berlin.de).

David Frantz is with the Earth Observation Laboratory, Geography Department, Humboldt-Universität zu Berlin, 10099 Berlin, Germany (e-mail: david.frantz@hu-berlin.de).

Lin Yan is with the Center for Global Change and Earth Observations, Michigan State University, East Lansing, MI 48823 USA (e-mail: yanlin@msu.edu).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2020.2982245

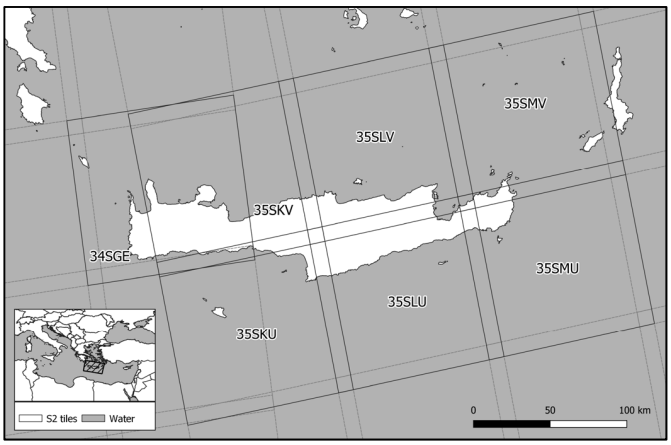

Fig. 1. Study area and Sentinel-2 tiles.

regressions [10] are determined as translation functions based on the respective set of tie points.

The Harmonized Landsat Sentinel Product [2], for instance, was created using a modification of the Automated Registration and Orthorectification Package (AROP) [12], which used a two-layer hierarchical approach with tie point detection and cross correlation matching on each layer with different spatial resolutions. The Landsat Sentinel Registration (LSReg) algorithm [9] constructs a four-layer hierarchical structure, using a feature-based initial tie point detection on the top layer (lowest resolution) followed by area-based least squares matching [13] on every hierarchical layer, and locations matched on all layers are identified as tie points. The hierarchical structure of the LSReg algorithm makes it computationally efficient, and the least squares matching provides higher subpixel geometric accuracy than the cross correlation matching [14]. In addition, the least squares matching in LSReg uses the spectral-angle-mapper similarity measure, which makes it more suitable to account for intersensor differences between spectral bands, more robust to reflectance brightness variations, and enables improved multitemporal consistency when compared with classic correlation-based least squares matching approaches [9]. Thus, we chose to adapt LSReg and specifically tuned the algorithm for operational and fully automated geometric coregistration of S2A/B time series to selected Landsat base images. Emphasis was particularly put on the selection of suitable base images, which can be complicated due to seasonal land surface changes and cloud cover, as well as parameter tuning for finding enough tie points in target S2 images that do not contain much valid data.

II. Study Area and Challenges

The island of Crete in Greece was chosen for developing and testing the approach (Fig. 1). Crete is a particularly challenging region for geometric correction for two reasons. First, a high proportion of open ocean in the individual S2 tiles reduces the land area available for tie point detection. Furthermore, the presence of waves on open water regularly results in white-caps with high reflectance and contrast that tend to be detected as initial tie points. These effects either reduce the number of tie points or trigger the occurrence of pseudo tie points in LSReg, which in combination can lead to distortion during geometric correction. Second, Crete has a strong topographic gradient, with steeply rising terrain reaching elevations of about 2500 m.

III. Data and Methods

A. LSReg 2.0

We used version (2.0) of the LSReg algorithm [9]. Different from the original algorithm, version 2.0 performs an additional step of dense point matching on the bottom hierarchical layer (highest resolution) to provide more tie points for a better fit of the transformation functions between coregistered images. LSReg requires only few inputs. First, an image needs to be designated as the base for coregistration. Second, a target image to be coregistered with the base image is specified. Third, the type of transformation needs to be selected. The tie point matchings are conducted using the NIR bands that provide high contrast across land cover types and reduced sensitivity to atmospheric effects [9], [10]. The LSReg 2.0 algorithm does not undertake cloud masking prior to image coregistration. Further details on the original algorithm can be found in [9].

B. Modifications

We implemented several refinements to adjust the LSReg algorithm for operational coregistration of S2A/B with Landsat-8 images. First, we increased the spatial resolution of the depth-first matching pyramid layers from 10, 30, 60, and 120 m to 10, 20, 40, and 80 m. This will effectively lead to more potential tie point locations in images with large shares of water and in images where only a small part of an S2 data take is intersecting the tile. Moreover, the sampling step for dense matching (highest resolution) was modified to depend on the number of valid land pixels (no cloud, water, or no data), instead of considering the whole image dimension. This change results in a potentially higher abundance and density of tie points in case the land share in an image is low, or if the image only includes a small part of the S2 data take.

Second, LSReg 2.0 does not have hard coregistration failure criteria, that is, the code does not abort with an error message, but only issues a warning that the coregistration might have failed. In this case, manual inspection is recommended. However, for operational implementation into a full processing chain, failure criteria are necessary that stop the execution of the code automatically. Thus, we identified a number of criteria for early termination of the complete processing chain, that is, if there are less than 12 matched tie points and if the predicted image shift is larger than 6 pixels, that is, 60 m.

Third, we increased the threshold for water masking from 5% to 10% NIR reflectance to avoid false detection of tie points on waves and white-caps, which frequently happened with LSReg 2.0 over the open ocean and resulted in a large number of tie points with arbitrary shift vectors.

Fourth, as suggested by Yan *et al.* [9], we considered only affine transformation, as its performance was shown to be comparable to polynomial transformations in earlier

experiments [9], [11], and as it was found to be more robust, especially if the tie points were not distributed across the complete image extent, for example, due to clouds or higher water shares at either side of the image. Furthermore, the affine transformation has been demonstrated to perform well on Landsat and Sentinel 2 images that were terrain-corrected [7], [9].

Fifth, we performed the coregistration subsequent to cloud masking to increase its computational efficiency in an operational setting.

The modified LSReg is completely integrated into the Framework for Operational Radiometric Correction for Environmental monitoring (FORCE) Level 2 Processing System [15], [16], and thus coregistration is now a feasible option in its fully automatic preprocessing chain. The module is implemented between cloud masking [17] and radiometric correction [15], and thus the coregistration benefits from excluding clouds and cloud shadows. In addition, FORCE uses an integrated radiometric correction that corrects for both atmospheric and topographic effects. The latter correction especially benefits from the prior improvement of the geometric accuracy as it substantially improves the alignment between S2 and the Digital Elevation Model to perform this correction and thus reduces correction artifacts, for example, around crests.

C. Compilation of Base Images

We aimed at matching the geometry of the Landsat Collection 1 Tier 1 data record, due to its superior geometric consistency, until the release of the S2 GRI. Furthermore, this increases the consistency in retrospective time series analyses in cases where historic Landsat archives are combined with contemporary S2 images.

The crucial step of selecting suitable base images offers several options. First, a base image without much change relative to all target images can be selected for coregistering multiple target images. While this method facilitates the base image selection, it requires manual intervention and does not account for seasonal variations in reflectance, which might lead to an insufficient number of matched tie points [12]. Second, a selection of individual base images with acquisition dates proximate to the acquisition of the target images may be considered [10]. While this mitigates seasonality-related challenges, the labor-intensive selection of suitable images and the possibility of no cloud-free base image being available for a specific year are the drawbacks. Third, chain correction could be applied where a single base image is defined to correct one target image, which then serves as a base image for the temporally neighboring image in the time series. This method should be used cautiously, as errors are likely to propagate, which may cause systematic shifts in coregistered image time series.

We therefore present an alternative approach that uses monthly Landsat spectral average metrics as base images. To achieve near gap-free coverage, we accumulated all Landsat OLI acquisitions for the five-year period from 2015 to 2019 and calculated monthly mean NIR reflectance images for January through December. The seasonality of land surfaces is thereby mitigated, providing a near gap-free base image for each month.

D. Target Images: Sentinel-2A/B Image Time Series

We aimed at coregistering all available S2A and -2B L1C images covering Crete across seven tiles (Fig. 1). We downloaded images only with a cloud cover below 70% as indicated by the metadata catalog, resulting in a total of 1739 images in the time period between July 2015 and end of December 2018. The L1C images were processed to Level 2 Analysis Ready Data using the FORCE Level 2 Processing System with the incorporated coregistration module as outlined above. The cloud detection module additionally identified 23 images with a cloud cover larger than 90%, which serves as termination criterion for the cloud masking [15].

E. Evaluation of Coregistration Performance

We evaluated the performance of the modified LSReg by calculating the rate of successfully coregistered images, the number of tie-points used for coregistration, model root mean square errors (RMSEs), absolute image shifts, and the noise in original and coregistered normalized difference vegetation index (NDVI) time series. For the latter, we collected 585 pixel locations at the borders of land cover transitions, dispersed across Crete. For each coordinate, we derived noise across the respective NDVI time series using three successive measurements y_i, y_{i+1}, and y_{i+2} acquired at dates day_i, day_{i+1}, and day_{i+2}. We quantified the differences between the center NDVI and the linear interpolation between the two outer measurements as follows [18]:

$$\text{Noise} = \sqrt{\frac{\sum_{i=1}^{n-2}\left(y_{i+1} - \frac{y_{i+2}-y_i}{\text{day}_{i+2}-\text{day}_i}(\text{day}_{i+1}-\text{day}_i) - y_i\right)^2}{N-2}}. \quad (1)$$

IV. Results and Discussion

We applied the coregistration to 1716 S2A/B images using the modified LSReg algorithm, yielding 1501 coregistered images, that is, a success rate of 87.5%. The mean RMSE of the coregistration was 0.44 pixels at 10-m resolution (Fig. 2 top left) and the number of automatically identified tie points for the coregistered images ranged between 1219 and 106 359 with a mean of 16 290 (Fig. 2, top right).

The coregistration failed for 215 images. An inspection of these images' characteristics revealed that those had low data coverage in the S2 tiles, high shares of water, high cloud cover, or the combinations thereof. This reduced the area of cloud-free land observations available for tie point matching and thus too few tie points for coregistration (Fig. 2, bottom left).

The mean image shifts between base and target images before coregistration were 14.0 m (standard deviation: 6.9 m) and 13.4 m (standard deviation: 11.3 m) in the x- and y-direction, respectively. The maximum image shifts were 46.2 in the x-direction and 59.6 m in the y-direction and were confirmed by examining the associated image pairs.

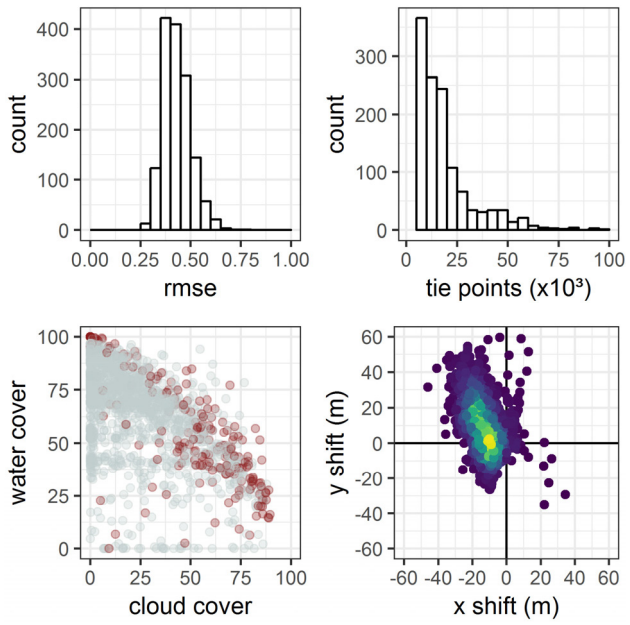

Fig. 2. (Top left) Model RMSE, (Top right) number of tie points detected per image, (Bottom left) scatterplot of percentage water cover and cloud cover for failed (red) and successfully coregistered images (gray), and (Bottom right) density plot of absolute image shifts performed during coregistration.

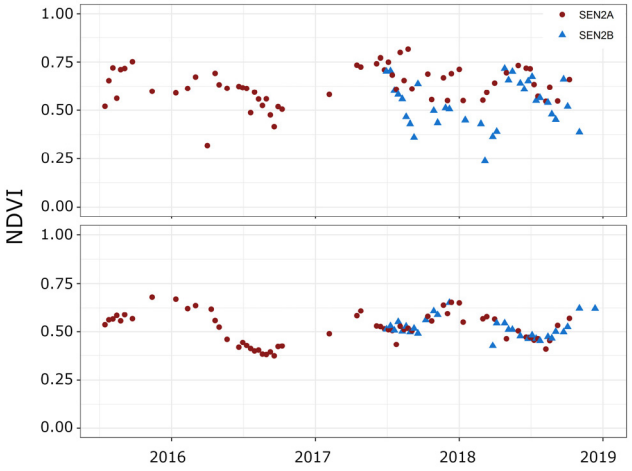

Fig. 3. NDVI time series of Sentinel-2A (red) and -2B (blue) data before (top) and after (bottom) coregistration procedure. Pixel location alternating between open and dense tree canopies at latitude/longitude: 35°34.57737′/24°8.40115′.

A general tendency for North-West shifts was apparent (Fig. 2, bottom right).

The determined geometric shifts in the S2 time series ranged up to six pixels and thus exceeded previously observed shifts [9], [10]. These inconsistencies in the original time series caused spectral variability due to alternating land cover types, partly exceeding the seasonal variability, which hamper analyses of dense image time series, for example, for capturing land surface phenology (Fig. 3). Coregistration drastically improved the consistency of the time series (Fig. 4), with average time series noise being reduced by 42.9%, from 0.086 (standard deviation = 0.029) in the unregistered to 0.049 (standard deviation = 0.018). In 21 locations, slight increases

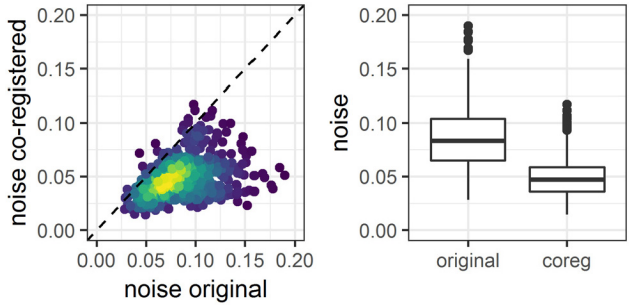

Fig. 4. (Left) Scatterplot of noise in Sentinel-2 NDVI time series before (x-axis) and after (y-axis) coregistration. (Right) Boxplot comparing distribution of noise in original against coregistered time series. Time series noise was quantified for 585 manually selected pixels located at boundaries of different land cover types throughout Crete.

in noise were apparent (mean < 0.008), which relate mostly to the fact that the time series were dynamic but clear observation was relatively sparse.

The presented approach is highly automated and thus suitable for large-area applications. It operates without manual selection of suitable base images through the use of Landsat-based multiyear spectral averages. While mitigating challenges related to seasonal reflectance variation, the procedure does not account for interannual variation in reflectance due to land cover change that might occur in the 5-year period used to generate the mean NIR data. However, similar to the occurrences of clouds, the occurrences of land cover changes only reduce matched tie points at locations of the occurrences (recall that the original LSReg algorithm does not undertake cloud masking). Given the fact that dense point matching was conducted in LSReg 2.0 and we obtained a mean of 16 290 tie points per coregistered image pair, the issue of land cover changes was not found to be a problem. Nevertheless, the aggregation period (here 5 years) for the generation of the Landsat-based base images can be adapted flexibly. In deforestation frontiers of tropical evergreen forests, for instance, low seasonal variability coincides with high rates of land cover change. In such cases, the overall time frame should be narrowed down.

The upcoming Landsat Collection 2 will be geometrically adjusted to the S2 GRI [19], and the forthcoming S2 acquisitions will use the GRI for geo-correction. Consequently, a high geometric consistency of integrated Landsat and S2 time series can be expected for GRI-corrected data. However, coregistration will be needed for using post-Level 1 S2 data until the entire S2A and 2B archive has been reprocessed and the geometric quality targets have been confirmed globally in an operational setting.

V. Conclusion

We presented a modification of LSReg 2.0, which enables the operational geometric coregistration of S2-A/B images based on multitemporal Landsat spectral–temporal metrics. The approach allows for automated subpixel coregistration under challenging conditions, overcoming issues of low data or land coverage in satellite products, topographic gradients, and seasonality. The described modifications of the

LSReg algorithm and the generation of Landsat base images are implemented in the free and open source software FORCE version 3.0 (https://github.com/davidfrantz/force).

ACKNOWLEDGMENT

This research contributes to the Landsat Science Team.

REFERENCES

[1] J. Dong, G. Metternicht, P. Hostert, R. Fensholt, and R. R. Chowdhury, "Remote sensing and geospatial technologies in support of a normative land system science: Status and prospects," *Current Opinion Environ. Sustainability*, vol. 38, pp. 44–52, Jun. 2019, doi: 10.1016/j.cosust.2019.05.003.

[2] M. Claverie *et al.*, "The harmonized landsat and Sentinel-2 surface reflectance data set," *Remote Sens. Environ.*, vol. 219, pp. 145–161, Dec. 2018, doi: 10.1016/j.rse.2018.09.002.

[3] 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, Nov. 2014, doi: 10.3390/rs61111127.

[4] F. Gascon *et al.*, "Copernicus Sentinel-2A calibration and products validation status," *Remote Sens.*, vol. 9, no. 6, p. 584, Jun. 2017, doi: 10.3390/rs9060584.

[5] J. Storey, D. P. Roy, J. Masek, F. Gascon, J. Dwyer, and M. Choate, "A note on the temporary misregistration of Landsat-8 operational land imager (OLI) and Sentinel-2 multi spectral instrument (MSI) imagery," *Remote Sens. Environ.*, vol. 186, pp. 121–122, Dec. 2016, doi: 10.1016/j.rse.2016.08.025.

[6] S. Clerc, "Sentinel-2 L1C data quality report," Data Qual. ESA, Paris, France, Tech. Rep. 44, 2019. Accessed: Oct. 31, 2019. [Online]. Available: https://sentinel.esa.int/documents/247904/685211/Sentinel-2_L1C_Data_Quality_Report

[7] L. Yan, D. P. Roy, Z. Li, H. K. Zhang, and H. Huang, "Sentinel-2A multi-temporal misregistration characterization and an orbit-based sub-pixel registration methodology," *Remote Sens. Environ.*, vol. 215, pp. 495–506, Sep. 2018, doi: 10.1016/j.rse.2018.04.021.

[8] A. Stumpf, D. Michéa, and J.-P. Malet, "Improved co-registration of Sentinel-2 and Landsat-8 imagery for Earth surface motion measurements," *Remote Sens.*, vol. 10, no. 2, p. 160, Jan. 2018, doi: 10.3390/rs10020160.

[9] L. Yan, D. Roy, H. Zhang, J. Li, and H. Huang, "An automated approach for sub-pixel registration of Landsat-8 operational land imager (OLI) and Sentinel-2 multi spectral instrument (MSI) imagery," *Remote Sens.*, vol. 8, no. 6, p. 520, Jun. 2016, doi: 10.3390/rs8060520.

[10] S. Skakun, J.-C. Roger, E. F. Vermote, J. G. Masek, and C. O. Justice, "Automatic sub-pixel co-registration of Landsat-8 operational land imager and sentinel-2A multi-spectral instrument images using phase correlation and machine learning based mapping," *Int. J. Digit. Earth*, vol. 10, no. 12, pp. 1253–1269, Dec. 2017, doi: 10.1080/17538947.2017.1304586.

[11] D. Scheffler, A. Hollstein, H. Diedrich, K. Segl, and P. Hostert, "AROSICS: An automated and robust open-source image co-registration software for multi-sensor satellite data," *Remote Sens.*, vol. 9, no. 7, p. 676, Jul. 2017, doi: 10.3390/rs9070676.

[12] J. Masek, "Automated registration and orthorectification package for landsat and landsat-like data processing," *J. Appl. Remote Sens.*, vol. 3, no. 1, Mar. 2009, Art. no. 033515, doi: 10.1117/1.3104620.

[13] A. Gruen, "Adaptive least squares correlation: A powerful image matching technique," *South Afr. J. Photogramm., Remote Sens. Cartogr.*, vol. 14, no. 3, pp. 175–187, 1985.

[14] A. Gruen, "Development and status of image matching in photogrammetry," *Photogramm. Rec.*, vol. 27, no. 137, pp. 36–57, Mar. 2012, doi: 10.1111/j.1477-9730.2011.00671.x.

[15] D. Frantz, A. Roder, M. Stellmes, and J. Hill, "An operational radiometric landsat preprocessing framework for large-area time series applications," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 3928–3943, Jul. 2016, doi: 10.1109/TGRS.2016.2530856.

[16] D. Frantz, "FORCE—Landsat + Sentinel-2 analysis ready data and beyond," *Remote Sens.*, vol. 11, no. 9, p. 1124, 2019, doi: 10.3390/rs11091124.

[17] D. Frantz, E. Haß, A. Uhl, J. Stoffels, and J. Hill, "Improvement of the Fmask algorithm for Sentinel-2 images: Separating clouds from bright surfaces based on parallax effects," *Remote Sens. Environ.*, vol. 215, pp. 471–481, Sep. 2018, doi: 10.1016/j.rse.2018.04.046.

[18] E. Vermote, C. O. Justice, and F.-M. Breon, "Towards a generalized approach for correction of the BRDF effect in MODIS directional reflectances," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 898–908, Mar. 2009, doi: 10.1109/TGRS.2008.2005977.

[19] J. C. Storey, R. Rengarajan, and M. J. Choate, "Bundle adjustment using space-based triangulation method for improving the landsat global ground reference," *Remote Sens.*, vol. 11, no. 14, p. 1640, Jul. 2019, doi: 10.3390/rs11141640.