Band Selection in Sentinel-2 Satellite for Agriculture Applications

Tianxiang Zhang, Jinya Su, Cunjia Liu, Wen-Hua Chen

Department of Aeronautical and Automotive Engineering, Loughborough University

Loughborough, LE11 3TU, U.K

E-mail: T.Zhang@lboro.ac.uk, J.Su2@lboro.ac.uk, C.Liu5@lboro.ac.uk, W.Chen@lboro.ac.uk

Hui Liu and Guohai Liu

School of Electrical and Information Engineering, Jiangsu University

Zhenjiang, Jiangsu, China

E-mail: amity@ujs.edu.cn, ghliu@ujs.edu.cn

Abstract— **Various indices are used for assessing vegetation and soil properties in satellite remote sensing applications. Some indices, such as NDVI and NDWI, are defined based on the sensitivity and significance of specific bands. Nowadays, remote sensing capability with a good number of bands and high spatial resolution is available. Instead of classification based on indices, this paper explores direct classification using selected bands. Recently launched Sentinel-2A is adopted as a case study. Three methods are compared, where the first approach utilizes traditional indices and the latter two approaches adopt specific bands (Red, NIR, and SWIR) and full bands of on-board sensors, respectively. It is shown that a better classification performance can be achieved by directly using the three selected bands compared with the one using indices, while the use of all 13 bands can further improve the performance. Therefore, it is recommended the new approach can be applied for Sentinel-2A image analysis and other wide applications.**

Keywords- Sentinel-2A; remote sensing; agriculture; supervised classification; machine learning

I. INTRODUCTION

Over the past few decades, satellite remote sensing has been playing a crucial role in forest monitoring, disaster management and agricultural applications [1-3]. Various satellites possess different retrieval characteristics due to their customized sensors. Remote sensing images may be produced by optical sensors with a good number of spectral bands and require tailored analysis depending on specific applications. So far, a widely used approach in satellite remote sensing is to define various indices to facilitate the classification of diverse land covers or plants under various stresses (e.g. Normalized Difference Vegetation Index (NDVI) and Normalized Difference Water Index (NDWI)). NDVI proposed by Rouse et al [4] has the ability to classify land covers in remote sensing area, as well as vegetation vitality. This index is defined by the reflectance of red (RED) band and near infrared (NIR) band since they sense very different depths through vegetation canopies. RED channel locates in the strong chlorophyll absorption region while NIR channel has high vegetation canopy reflectance in this area [5]. Thus, this index can be applied to classify land covers. NDWI was proposed by Gao [6] to assess water status by the combination of NIR and short wave infrared (SWIR)

channel because both are located in the high reflectance plateau of vegetation canopies and sense similar depth in vegetation canopies. Absorption by vegetation liquid water near NIR is negligible, and weak liquid absorption near SWIR is present. Therefore, canopy scattering enhances the water performance. In the past, lots of research has been conducted to link these two indices with other indices of interest (e.g. Vegetation Water Content (VWC), or Equivalent Water Thickness (EWT)) to generate a classification map of interest [7-10].

For remote sensing applications, band information is of paramount importance in the phase of satellite data analysis and interpretation. The launch of Sentinel-2A is a key part of Global Monitoring for Environment and Security (GMES) program supported by the European Space Agency (ESA) and European Commission (EC) ensuring a better data continuity than other relevant satellites, such as SPOT and Landsat satellite series, due to its high spatial resolution and short revisit time. To obtain more retrieval information, its Multispectral Instrument (MSI) is an important component on this satellite as shown in Fig. 1. The MSI holds an anastigmatic telescope with three mirrors with a pupil diameter of about 150mm minimizing thermos-elastic distortions, and the optical design has been optimized to achieve state-of-the-art imaging quality across its 290km field of view [11-12]. MSI also features 13 spectral bands ranging from visible, NIR to SWIR at different resolutions. This configuration is selected as the best compromise between user requirements and mission performance. Four bands at 10m resolution meet the basic requirements for land classification; six bands at 20m resolution provide additional information on vegetation detecting. The remaining three bands at 60m contribute to atmospheric and geophysical parameters [12]. The launch of Sentinel-2B in March, 2017 shorten the revisit time into 5 days, which means Sentinel-2 series have the shortest revisit time among mainstream freely available satellites until now.

In the aforementioned literature, most of the research is focused on the NDVI or NDWI calculation and their usefulness in land cover classification, water content evaluation, etc., by exploiting the specific spectral bands of satellites. Although the NDVI and NDWI have been widely used due to their simplicity along with clear

physical meanings, there still exist several limitations. For example, in land cover classifications [5], NDVI usually saturates when vegetation coverage become dense (i.e., Leaf Area Index (LAI), i.e., the one-sided green leaf area per unit ground surface area, reaches around 3) and no longer sensitive to vegetation changes [5]. Although compared with NDVI, NDWI saturates at a later stage, it also results in limited performance [6]. Besides, it is generally not an easy task to determine an appropriate threshold for index-based classification approaches. To avoid the problems in index-based classification approaches and further explore the potential of all the latest available capabilities of new satellites, the benefit of using selected or even all spectral bands of Sentinel-2A will be investigated using machine leaning techniques in land cover classifications.
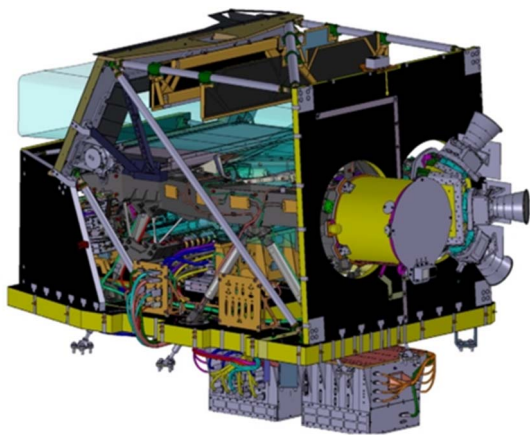

Fig. 1 Multi Spectral Imager View on Sentinel-2A

On the other hand, little has been done in the literature to classify Sentinel-2A images by using machine learning methods and explore the benefits of the availability of more spectral bands of this satellite in classification. Consequently, this paper will compare different classification approaches according to indices and different bands. Three approaches are studied and compared where in the first approach, NDVI and NDWI are treated as the features, in the second approach the three bands are directly adopted including Red, Near Infrared (NIR) and Short Wave Infrared (SWIR), and in the third approach all 13 bands available on Sentinel-2A satellite are used. It is discovered that in comparison with the index-based approach, a better classification performance can be achieved by directly adopting the three selected bands. Moreover, the use of all 13 bands can further improve the classification performance due to increased bands and consequently information.

The rest of this paper is organized as follows. Data source acquisition is discussed in Section II. The methodology is described in Section III. NDVI/NDWI index-based classification and the spectral bands based classification algorithms are conducted and compared in Section IV. Finally, conclusions are drawn in Section V along with future work.

II. DATA SOURCES

Sentinal-2A satellite: Landsat8 and Sentinel-2A are the most advanced satellites with freely available data for long-term high-frequency remote sensing applications. The former one was launched in 2013 with Operational Land Imager (OLI) sensor offering high quality multispectral images at 15m, 30m, 100m and with a 16 days revisit time [13-15]. The latter one consists of Sentinel-2A and Sentinel-2B equipped with Multispectral Instruments (MSI) capable of acquiring 13 bands information at different spatial resolutions (10m, 20m and 60 m). The band information for Landsat8 and Sentinel-2A are shown in Fig. 2. Compared with Landsat 8, Sentinel-2A is more popular due to its fine properties including increased number of bands, shorter revisit time, and higher spatial resolution. Sentinel-2A provides more details in NIR band range and SWIR band range, which is helpful for agriculture, forest monitoring, and natural disaster management applications. A drawback of Sentinel-2A compared with Landsat 8 is without thermal infrared bands. Number of band and spatial and temporal resolution determine the quality of spectral image; consequently, Sentinel-2A satellite is more suitable for remote sensing applications [14].

All Sentinel-2A satellite images could be freely downloaded from Sentinel Hub, which was developed by ESA (https://scihub.copernicus.eu/). Besides, freely available satellite information analysis software SNAP is also provided, which in comparison with QGIS and ENVI, is specially customized for Sentinel series. This software could read all the information Sentinel series provide and export any data to other relative analysis software in next steps.

Site selection: In supervised learning, ground truth data is of vital importance such that different approaches can be evaluated and compared. To study and compare the performance of different land cover classification algorithms, an area that LUCAS (Loughborough University Centre for Autonomous Systems) performs flight tests regularly is chosen as an example site in this paper. The remote sensing data for the site of interest can be selected on the aforementioned website and downloaded. The basic information of this chosen field (see, Fig. 3) including location, spectral bands, pixel information, cloud cover percentage and date are summarized in TABLE I.

TABLE I: DETAILS OF THE SITE FOR CASE STUDY

Location	Bands	Pixel	Cloud Cover (%)	Date
52.761N, 1.287W; 52.749S, 1.25E	13	250*123	1.1171	05/01/2017

Sentinel-2 Bands	Central Wavelength (μm)	Resolution (m)
Band1 - Coastal Aerosol	0.443	60
Band2 - Blue	0.490	10
Band3 - Green	0.560	10
Band4 - Red	0.665	10
Band5 - Near Infrared	0.705	20
Band6 - Near Infrared	0.740	20
Band7 - Near Infrared	0.783	20
Band8 - Near Infrared	0.842	10
Band 8A - Near Infrared	0.865	20
Band 9 - Water Vapour	0.945	60
Band 10 - Shortwave Infrared (Cirrus)	1.375	60
Band 11 - Shortwave Infrared	1.610	20
Band12 - Shortwave Infrared	2.190	20

(a)

Landsat 8 Bands	Central Wavelength (μm)	Resolution (m)
Band1 - Blue	0.443	30
Band2 - Blue	0.483	30
Band3 - Green	0.563	30
Band4 - Red	0.665	30
Band5 - Near Infrared	0.865	30
Band6 - Shortwave Infrared	1.610	30
Band7 - Shortwave Infrared	2.200	30
Band8 -Panchromatic	0.590	15
Band 9 -Cirrus	1.405	30
Band 10 - Thermal Infrared	10.9	100
Band 11 - Thermal Infrared	12	100

(b)

Fig. 2 Bands for Sentinel-2A (a) and Landsat 8 (b)

To achieve the best results for NDVI and NDWI calculation, RED band is chosen as Band 4, NIR band and SWIR band are selected as Band 8a and Band 11 respectively in Sentinel-2A.

Fig. 3 Site selected for the case study

III. METHODOLOGY

Overall procedure: The whole process of land cover classification using satellite remote sensing images can be divided into pre-processing and data analysis stage, as shown in Fig. 4. SNAP software is to pre-process the data downloaded from Sentinel Hub and calculate the related

indices. Some specific areas could be labelled on the original data, such that the NDVI and NDWI data can be exported in excel format. The data analysis stage is performed by MATLAB using different classification algorithms.

Resampling, Atmospheric Correction and Subset selection are necessary in pre-processing satellite images. Resampling ensures that images of each band have the same resolution and number of pixels. Subset selection allows re-choosing specific areas of interests. Atmospheric Correction algorithms are based on the Atmospheric/Topographic Correction for Satellite Imagery by Richter [16]. This method performs atmospheric correction according to libRadtran radiative transfer model that is run to generate a large look-up table accounting for various atmospheric conditions, solar geometries and ground elevations [17]. This simplified model runs much faster than a full model to invert the radiative transfer equation and to calculate bottom-of-atmosphere reflectance. Therefore, all gaseous and aerosol properties of the atmosphere are both derived by the algorithm and aerosol optical thickness or water vapor content are derived from the images respectively. SNAP software offers a plugin to make atmospheric corrections called Sen2Cor [18]. Atmospheric correction is an integrated part in the process of Sentinel-2A satellite image processing. Fig.5 provides the RGB map of sentinel-2A data for the selected site after atmospheric corrections.
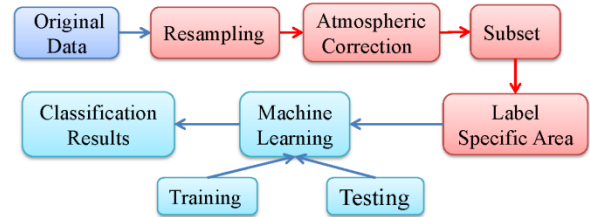
Fig. 4 The flowchart for land covers classification using Sentinel-2A image data

(b)

Fig. 5 RGB image from Sential-2A for the chosen site after atmospheric correction

NDVI and NDWI: Index-based classification requires NDVI and NDWI as features; consequently, they were first calculated in SNAP. As mentioned in Section II, Band 4, Band 8A and Band 11 are chosen as RED, NIR and SWIR band, respectively. According to the formulae in Eq. (1) and (2), NDVI and NDWI can be calculated and their maps are shown in Fig. 6 and Fig. 7 respectively. It can be seen that NDVI ranges from -0.26 to 0.748 and

NDWI ranges from -0.169 to 0.576, where various colors represent different values with corresponding "colormap".

$$NDVI = (Band\ 8A - Band\ 4) / (Band\ 8A + Band\ 4) \quad (1)$$

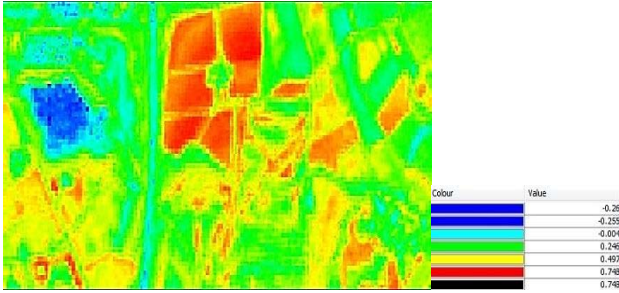$$NDWI = (Band\ 8A - Band\ 11) / (Band\ 8A + Band\ 11) \quad (2)$$
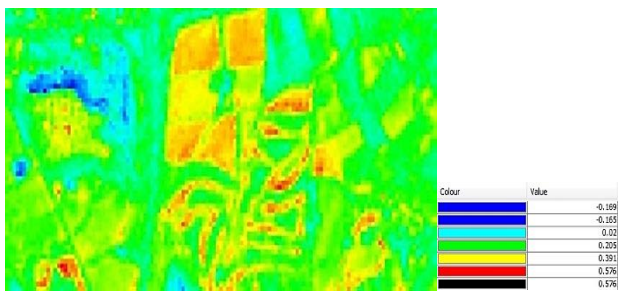

Fig. 6 NDVI map calculated in SNAP

Fig. 7 NDWI map calculated in SNAP

Ground truth: Labelling specific areas is a key part in SNAP level. Due to its compatibility with Sentinel-2A, the procedure is easy to realize than other satellite software. By using the ground truth obtained from Google map, UAV photography we collected and our onsite experience, four classes of land covers (i.e. crop (No. 1), tree (No. 2), water (No. 3), and road (No. 4)) are labelled as shown in Fig. 8. The average reflectance over bands for different classes is shown in Figure. 9. The reflectance differences at different bands lay the foundation for machine learning based classification. The labelled areas on these images could be exported to an excel file, including location details, band details and NDVI, NDWI values.

Machine learning: The land cover classification problem can be solved by using supervised learning algorithms. Supervised classification builds the implicit relationship between feature vector and target variable (i.e. class label) by learning from limited labelled training data [19]. With the trained classification model, prediction can be made on new feature data such that its class label can be determined. To avoid overfitting the labelled data is usually divided into training set and testing set. Different classification algorithms have been developed in the literature including decision trees, discriminant analysis, Support Vector Machines (SVMs), nearest neighbor, neural network, just to name a few. In this paper, SVMs is chosen as the classifier since in comparison with other approaches SVMs generalizes well even with small training samples. Particularly, Quadratic SVMs with 50% holdout validation is chosen based on

our previous experience, where its implementation is conducted in MATLAB using classifier learner with built-in functions.

Classification algorithms: At first, classification by using one index (i.e., NDVI or NDWI) is performed. However, it is discovered that one index results in very poor result (the result is omitted due to lack of space), this is mainly due to the fact that one feature is not enough for the land cover classification problem with four different classes in this study. Consequently, only the classification methods with relatively satisfying performance are presented in this paper. According to the different features selected for classification, three different classification methods are studied and compared in this paper, where the first approach treats NDVI and NDWI as feature vector, the second one directly adopts the three specific relevant bands including Red, Near Infrared (NIR) and Short Wave Infrared (SWIR) as features, and in the third approach all 13 bands available on Sentinel-2A satellite are used.
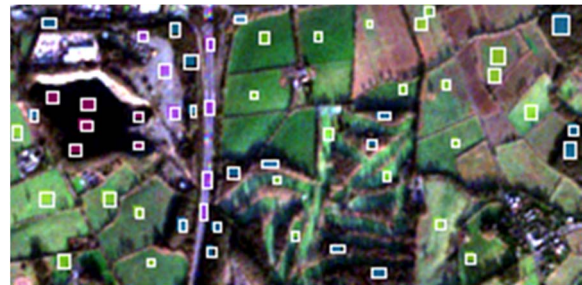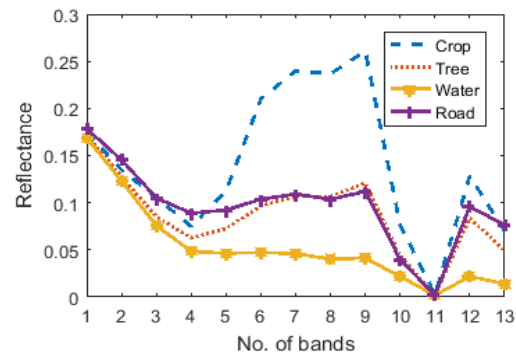

Fig. 8 Four different labelled classes

Fig. 9 Average reflectance for different classes

IV. CLASSIFICAITON COMPARISIONS

In this section, the algorithms discussed in Section. III are implemented, particularly the performance of three different classification methods are evaluated and compared by using confusion matrix (see, Figs. 10-12). On the confusion matrix plot, the rows correspond to the predicted class (i.e. Output class), and the columns show the truth class (i.e. Target class). More explanations on confusion matrix will be given where necessary.

A. NDVI, NDWI based classification

As discussed in Classification algorithms of Section. III, in the index-based classification approach, rather than use one index due to poor performance, both NDVI and

NDWI are treated as features. The confusion matrix for this approach is given in Fig. 10. In this figure, the diagonal cells in green show the number and percentage of correct classification. For example, 219 samples are correctly classified as crop corresponding to 39.0% of all 561 samples. The off-diagonal cells show where the mistakes come from. For example, in first row, 6 of tree samples are incorrectly classified as crop corresponding to 1.1% of all 561 samples. The rightmost column shows the accuracy for each predicted class, while the bottom row shows the accuracy for each true class. For example, out of 225 crop prediction, 97.3% are correct and 2.7% are wrong; out of 222 crop samples, 98.6% are correctly predicted as crop and 1.4% are wrongly predicted as tree. The cell at the right bottom shows the overall accuracy, which for this algorithm is 92.7%.

The result shows that classification based on empirical or semi-empirical approach has a relatively high accuracy. This is mainly due to the fact that NDVI can effectively reflect vegetation status and NDWI is valid for water content evaluation. Both of them can partly capture the main characteristics of the land covers of interest. It should also be noted that the main misclassification is that road is misclassified as tree. This is mainly because there is little chlorophyll in tree in winter and consequently poses challenges in distinguishing between tree and road.

Fig. 10 NDVI/NDWI-based classification result

B. RED, NIR, SWIR based classification

Instead of using empirical or semi-empirical indices in satellite remote sensing as in Section IV.A, the specific relevant bands including RED, NIR and SWIR are directly adopted as features for supervised classification in this part. The classification results are shown in Fig. 11, where the overall accuracy is 94.1%. Different from index-based classification where specific mathematical operations are performed on the three bands, machine learning algorithm can automatically build the relationship between the three

bands and class label by learning from labelled training samples. Comparing the performance of these two approaches, one can discover that classification by directly using RED, NIR, SWIR band is more effective than NDVI, NDWI based classification. Therefore, classifying Sentinel-2A spectral images by using selected bands along with machine learning techniques is an effective approach.

Fig. 11 RED, NIR, SWIR based classification result

Fig. 12 Full band information classification

C. Full bands information classification

It should be noted that there are 13 spectral bands on Sentinel-2A satellite, which provides a great amount of information for remote sensing applications. It would be of interest to verify whether full band information can further improve the performance or not. To this end, all 13 bands are further treated as features for classification in the third approach, where the classification results are

shown in Fig. 12. It can be seen from the figure that the overall accuracy increases to 98.8%. from 94.1%. Moreover, the misclassification rates between road class and tree class also reduce obviously. This demonstrates that incorporating more related band information can further improve the classification performance.

By comparing these different classification algorithms, it can be discovered that supervised learning based classification using more spectral bands outperforms empirical or semi-empirical indices based approaches. Employing more related bands on Sentinel-2A satellite can improve the classification performance. Overall, machine learning based classification by using the spectral bands of Sentinel-2A satellite is a promising solution for agriculture remote sensing applications.

V. Conclusions

This paper develops a novel approach to analyze satellite remote sensing images, particularly Sentinel-2A satellite images using machine learning techniques. Three classification methods are studied and compared, namely index-based classification (NDVI, NDWI), specific relevant bands (RED, NIR, SWIR) based classification and all available bands based classification. By using a case study of land cover classification with four classes, it is shown that the method employing all available bands of Sentinel-2A satellite result in the best performance while the use of only three highly relevant bands also yields quite promising results. Overall the classification methods directly using specific relevant bands using supervised learning outperform the classic index based classification methods. Some limits of the index based classification could be removed by the direct use of spectral bands through the learned regression function between vegetation water content or soil moisture and certain bands of Sentinel-2A. The proposed algorithm can be applied to forest vegetation monitoring, vegetation physiological status detecting and irrigation decisions [20, 21].

References

[1] C. J. Tucker, "Remote sensing of leaf water content in the near infrared," *Remote Sens. Environ.*, vol. 10, no. 1, pp. 23–32, 1980.

[2] J. Peñuelas, I. Filella, C. Biel, L. Serrano, and R. Savé, "The reflectance at the 950–970 nm region as an indicator of plant water status," *International Journal of Remote Sensing*, vol. 14, no. 10, pp. 1887–1905, 1993.

[3] M. J. Hill, "Vegetation index suites as indicators of vegetation state in grassland and savanna: An analysis with simulated SENTINEL 2 data for a North American transect," *Remote Sens. Environ.*, vol. 137, pp. 94-111, 2013.

[4] W. Rouse, R. H. Haas, and D. W. Deering, " Monitoring vegetation systems in the Great Plains with ERTS", Texas A&M Univ, TX, United States, 1974, pp. 309-317.

[5] Y. Gao et al., "Optical Sensing of Vegetation Water Content : A Synthesis Study," *IEEE J. Sel. Top. Appl. EARTH Obs. Remote Sens.*, vol. 8, no. 4, pp. 1–9, 2015.

[6] B. C. Gao, "NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space," *Remote Sens. Environ.*, vol 58, no. 3, pp. 257-266, 1996.

[7] T. J. Jackson et al., "Vegetation water content mapping using Landsat data derived normalized difference water index for corn and soybeans," *Remote Sensing of Environment.*, vol. 92, no. 4, pp. 475-482, 2004.

[8] D. Chen, J. Huang, and T. J. Jackson, "Vegetation water content estimation for corn and soybeans using spectral indices derived from MODIS near- and short-wave infrared bands," *Remote Sens. Environ.*, vol. 98, no. 2, pp. 225-236, 2005.

[9] D. A. Roberts, R. O. Green, and J. B. Adams, "Temporal and spatial patterns in vegetation and atmospheric properties from AVIRIS," *Remote Sens. Environ.*, vol. 62, no. 3, pp. 223–240, 1997.

[10] M. Tugrul Yilmaz, E. Raymond Hunt Jr, and T. J. Jackson, "Remote sensing of vegetation water content from equivalent water thickness using satellite imagery," *Remote Sens. Environ..,*vol. 112, no. 5, pp. 2514-2522, 2008.

[11] E. S. A. Communications, Sentinel-2 ESA's optical high-resolution mission for GMES operational services. 2012.

[12] M. Drusch et al., "Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services," *Remote Sens. Environ.*, vol. 120, pp. 25-36, 2012.

[13] M. Pesaresi, C. Corbane, A. Julea, A. J. Florczyk, V. Syrris, and P. Soille, "Assessment of the added-value of sentinel-2 for detecting built-up areas," *Remote Sens.*, vol. 8, no. 4, pp. 299, 2016.

[14] E. Mandanici and G. Bitelli, "Preliminary Comparison of Sentinel-2 and Landsat 8 Imagery for a Combined Use," *Remote Sens.*, vol. 8, no. 12, pp. 1014, 2016.

[15] H. van der Werff and F. van der Meer, "Sentinel-2A MSI and Landsat 8 OLI provide data continuity for geological remote sensing," *Remote Sens.*, vol. 8, no. 11, pp. 883 2016.

[16] R. Richter, "Atmospheric / Topographic Correction for Satellite Imagery (ATCOR - 2/3 User Guide)," ATCOR-2/3 User Guid. Version 6.3, pp. 1–71, 2007.

[17] C. Emde et al., "The libRadtran software package for radiative transfer calculations (version 2.0.1)," *Geosci. Model Dev.*, vol. 9, no. 5, pp. 1647–1672, 2016.

[18] B. Pflug, J. Bieniarz, V. Debaecker, J. Louis, and U. Müller-Wilm, "Some Experience Using SEN2COR."

[19] S. J. Russell, P. Norvig, J. F. Canny, J. M. Malik, D. D. Edwards, and S. J. S. Jonathan, "A modern approach." *Artificial Intelligence.* Prentice-Hall, Egnlewood Cliffs 25 (1995): 27.

[20] Y. Yi, D. Yang, D. Chen, and J. Huang, "Retrieving crop physiological parameters and assessing water deficiency using MODIS data during the winter wheat growing period," *Can. J. Remote Sens.*, vol. 33, no. 3, pp. 189-202, 2007.

[21] C. Slave, "Analysis of Agricultural Areas Using Satellite Images," *Int. J. Acad. Res. Environ. Geogr.*, vol. 1, no. 2, pp. 26–32, 2014.