ASSESSMENT OF BURNED AREA MAPPING METHODS FOR SMOKE COVERED SENTINEL-2 DATA

Alexandru-Cosmin Grivei[1,2], Corina Văduva[2], and Mihai Datcu[2,3]
[1] CECTI, Military Technical Academy "Ferdinand I", Bucharest, Romania
[2] CEOSpaceTech, University Politehnica of Bucharest, Bucharest, Romania
[3] DLR, German Aerospace Center, Oberpfaffenhofen, Germany
Contact author e-mail: alex.grivei@ymail.com

***Abstract*—Wildfires become more frequent in the context of global warming and severe drought in several parts of the globe. Earth observation data can be used to provide information in such cases, but sometimes, when using optical satellite imagery, the evaluation of the effects produced by ongoing large scale forest fires, can be impeded by smoke. It can reduce the accuracy of the information required by disaster management authorities when allocating resources. To improve both the usability of optical remote sensing data and the quality of the obtained information we compare multiple feature extraction, classification, and visual enhancement methods and algorithms for land cover mapping of smoke covered Sentinel-2 data. The demonstration is performed for the 2019 forest fires in Australia.**

Keywords—smoke; wildfire; earth observation; land cover; mapping

I. INTRODUCTION

Worldwide, each year, thousands of wildfires cause great loss of natural resources and human lives. They also have a great impact on the economy of the affected countries [1]. In late 2019 to early 2020, Australia suffered from some of the largest bushfires in history [2], with more than 18 million ha of vegetation and forests destroyed, over one billion animals killed, 34 human losses, and with huge economic costs.

Disaster Management Authorities require all the help they can find when fighting against large scale disasters. Earth observation (EO) data can be used to assess the scale and the propagation trend of ongoing events such as wildfires. In the case of optical remote sensing data, not all acquisitions can be used, as some contain visual impediments such as clouds of smoke resulted from the burned vegetation and forests.

Most of the wildfire monitoring platforms rely on NASA's MODIS (Moderate Resolution Imaging Spectroradiometer) data [3] for their services [4], [5]. An impediment for the MODIS data is the relatively low spatial resolution of 250 m which can make it hard to detect and map smaller, wildfires.

The resolution of the burned area maps, produced with the help of MODIS data, can be improved through different methods, such as the subpixel mapping (SPM) [6] or the edge error (EE) based method [7]. Finer maps can also be achieved through the usage of higher resolution data, such as Landsat-8. Common feature extraction methods used in burned area semantic mapping rely on indices such as modified normalized burn ratio (MNBR) which are applied on both pre and post-fire images to produce change detection maps [8].

Even better spatial resolution maps can be obtainable with the help of Sentinel-2 data. Mapping burned area land cover relies on the existence of high-quality pre and post data and uses common spectral indices for data representation. A machine learning algorithm, such as Support Vector Machine (SVM) [9] is used to classify the represented data to produce burned area maps [10]. Sentinel-2 acquisitions can also be used to created severity semantic maps of burned areas, while also relying on spectral indices such normalized difference vegetation index (NDVI), normalized burn ratio (NBR), and burned area index for sentinel (BAIS) [11].

Most of the available methods for burned area mapping rely on high-quality EO data and seldom tackle the challenge posed by processing smoke clouded acquisitions. Also, the majority require both a pre and post-fire image for the method to work. In this paper, we compare several data representation methods, feature extraction, and supervised classification algorithms to determine the best combination for the determination of land cover semantic class hidden underneath smoke clouds. For our assessments, we used a full size 10980 x 10980 pixels' Sentinel-2 scene over West Sydney, Australia. The processing was done using our novel multispectral data analysis DAS-Tool software plugin [12], for the Sentinel Application Platform (SNAP).

The dataset is presented in Section 2, together with the three different data representation methods which we compare in our analysis. The feature extraction and classification algorithms used in our tests are highlighted in Section 3. Here, we also assess the obtained results concerning the available ground truth. Conclusions and further improvement approaches are given in Section 5.

II. DATA REPRESENTATION METHODS

For our analysis, we used a Sentinel-2 L2A image over West Sydney, obtained on 31 Dec 2019 during the great bushfires from late 2019, early 2020. The acquisition can be observed in Fig. 1, and contains the following land cover classes: low

978-1-7281-5611-8/20/$31.00 ©2020 IEEE

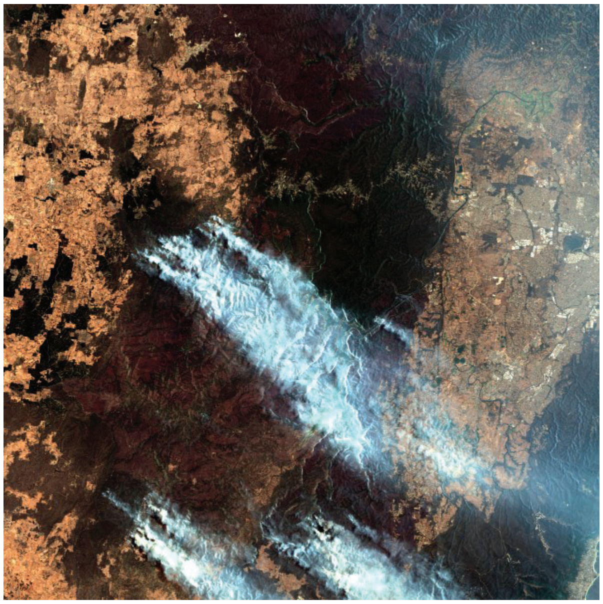

Fig. 1. Sentinel-2 image, West Sydney, Australia, 31 Dec 2019, in RGB representation (bands 4, 3, 2).

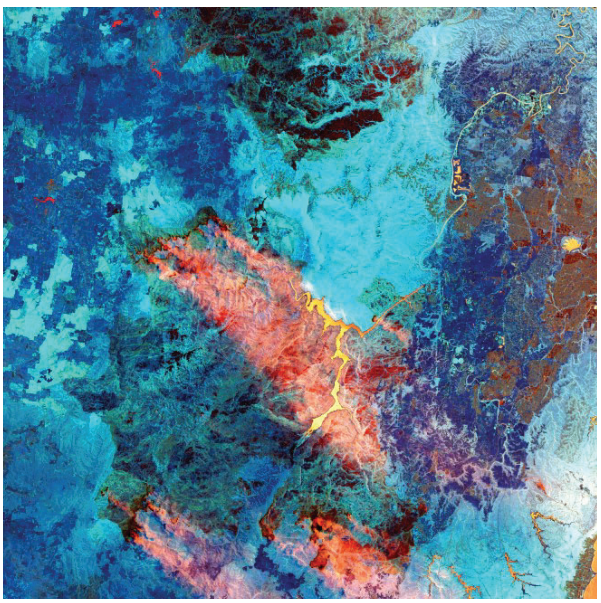

Fig. 2. Sentinel-2 image, West Sydney, Australia, 31 Dec 2019. Pseudo-color representation from MNDWI, NBR, NDBI indices.

density urban fabric, high-density urban fabric (mostly in the right third of the image), forest (central area, and bottom right), water bodies (bottom right corner and central area), grassland (mostly in the top left), agriculture (small patches top right, above the urban area), burned areas (central to low central and central top) and smoke (mainly in the center of the image).

Aiming for a way to highlight the land features under the smoke clouds, we propose three data representations (described in sections *A*, *B*, *C* below): raw data limited to the spectral bands having the highest spatial resolution, specific spectral indexes

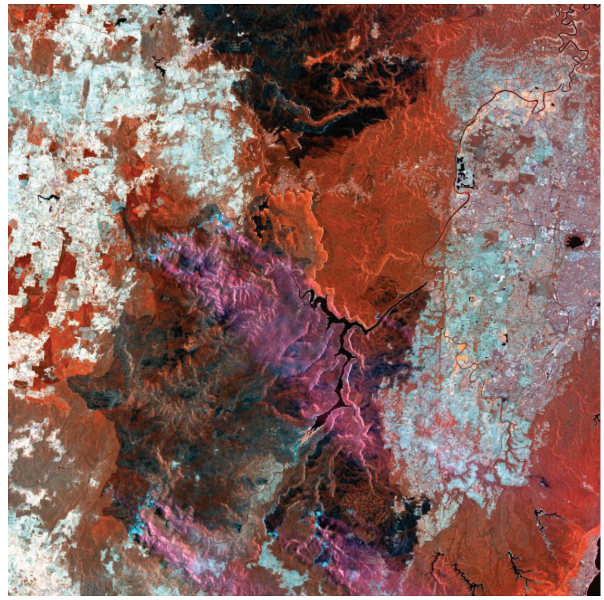

Fig. 3. Sentinel-2 image, West Sydney, Australia, 31 Dec 2019. Pseudo-color representation for FA bands.

representation and dimensionality reduction. In the case of the last two data representation cases, all the bands of the Sentinel-2 L2A product were resampled at the 10m spatial resolution.

A. Spectral bands selection

For our first data representation method, we exploit the spatial resolution and retain only the four 10m resolution bands available in the Sentinel-2 L2A data. These are bands 2, 3, 4, and 8 and will be further referenced as RGB & NIR (red, green, blue, and near-infrared). No other processing was done on these bands. An RGB representation of the data is given in Fig. 1.

B. Spectral indices

For our second representation, we created a product from the Modified Normalized Difference Water Index (MNDWI) (1), Normalized Burn Ratio (NBR) (2), Normalized Difference Built-up Index (NDBI) (3) and Normalized Difference Vegetation Index (NDVI) (4). We chose these four spectral indices as they can emphasize four of the main semantic categories of land cover found in the analyzed scene (water bodies and water vapor, burned areas, built-up areas, and vegetation). A weighted combination of these spectral indexes will help us separate between more specific land cover classes as analyzed in our tests (low density urban, high density urban, water, agriculture, forest, grassland, burned area, smoke).

$$\mathrm{MNDWI} = (\mathrm{Green} - \mathrm{SWIR}) / (\mathrm{Green} + \mathrm{SWIR}) \tag{1}$$

$$\mathrm{NBR} = (\mathrm{NIR}\text{-}\mathrm{SWIR}) / (\mathrm{NIR}+\mathrm{SWIR}) \tag{2}$$

$$\mathrm{NDBI} = (\mathrm{SWIR} - \mathrm{NIR}) / (\mathrm{SWIR} + \mathrm{NIR}) \tag{3}$$

$$\mathrm{NDVI} = (\mathrm{NIR} - \mathrm{Red}) / (\mathrm{NIR} + \mathrm{Red}) \tag{4}$$

The green, red, NIR, and SWIR spectral bands, used in the computation of the spectral indices given in (1) – (4), correspond to Sentinel-2 bands: 2, 3, 8 and 11, respectively. The pseudo-color representation of the resulted product is given in Fig. 2.

The RGB bands are represented by the MNDWI, NBR, and NDBI spectral indices values.

C. Factor analysis

For our third and last data representation method, we used the factor analysis method [13] which maximizes the intra-class correlation through the minimization of inter-class correlation. For this method, we used all of the available 12 bands of the Sentinel-2 L2A product and produced a 3 band pseudo-color product represented in Fig. 3. This method produces the best visual improvement out of the three methods.

III. Benchmarked Algorithms and Results

For each of the data representation methods, we run a series of tests in which we employed tree different patch-based feature extraction algorithms and two supervised classification methods. All of them are part of the DAS-Tool plugin and the tests were made using the SNAP desktop application using a workstation with an Intel i9, 5GHz, 16 thread, processor, and 64GB of RAM.

The tested feature extraction algorithms are: mean and dispersion (MeDiB), 32 bin spectral histogram, and 4 scales and 4 orientations Gabor features as they have shown to generate the best results among the available ones in the plugin [12]. The entire scene of 10980x10980 pixels was divided in 25x25 pixels size patches, resulting in a number of 192,721 samples. For each of these samples, the aforementioned algorithms were applied to each of the component bands (4 bands for the spectral band and spectral indices representations, and 3 for the factor analysis representation).

The two supervised classification algorithms used in the benchmark are Support-Vector Machines (SVM) and k-nearest neighbors (k-NN). We trained the following 8 semantic classes (Fig. 4): low density urban, high density urban, water, agriculture, forest, grassland, burned area, and smoke. The same training samples were used for all the data representation methods.

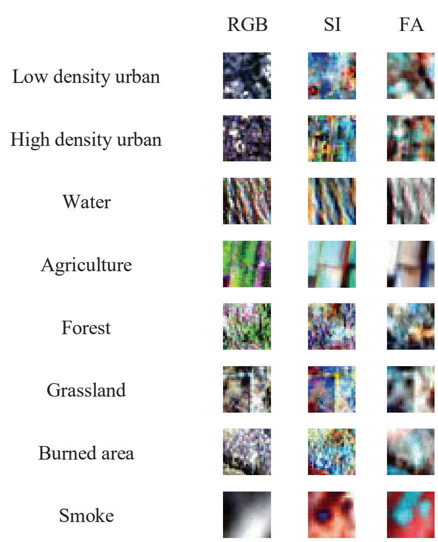

Fig. 4. Examples of training samples and representations in RGB, SI, FA

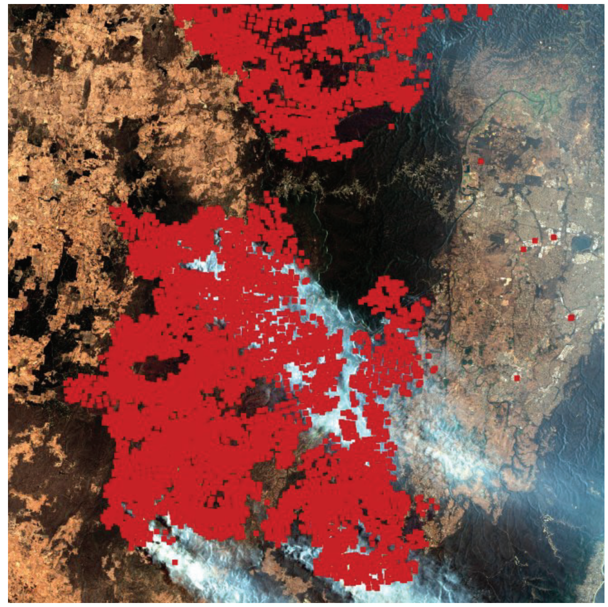

Fig. 5. Burned areas (red) at the end of Dec 2019, West of Sydney, Australia.

To assess our results, we used the data provided by NASA's Fire Information for Resource Management System [14] as ground truth. The data in Fig. 5 represents the burned area map at the end of Dec 2019 for the region of interest.

The performances for all the combinations of methods used in our benchmark are given in Fig. 6. The best-performing one (FA, MeDiB, SVM), manages to obtain an F1-score of 0.8732, while also uncovers land cover classes from 95.32% of the smoke covered data, as presented in Fig. 7.

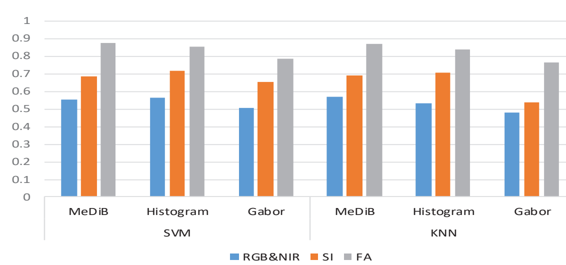

Fig. 6. F1 score for the burned area land cover classification in regard to data representation, feature extraction and classification algorithms.

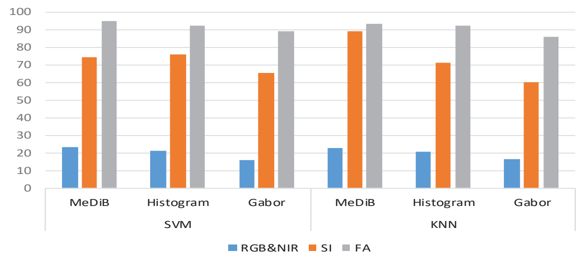

Fig. 7. Percentage of smoke covered data which was classified as the underling land cover class.

The overall best data representation method is FA as it both manages to help the feature extraction and classification algorithms uncover 91.61% (on average) of the smoke obstructed data, and, on average, yields an F1-score of 0.8302. The best combination of feature extraction and classifying algorithms are spectral histogram and SVM. They manage to obtain an average F1-score of 0.7115 for all of the data representation methods.

The best results are provided by the combination of methods: FA (data representation), MeDiB (feature extraction), SVM (classifier). The generated burned area map, represented by the red color, is given in Fig. 8.

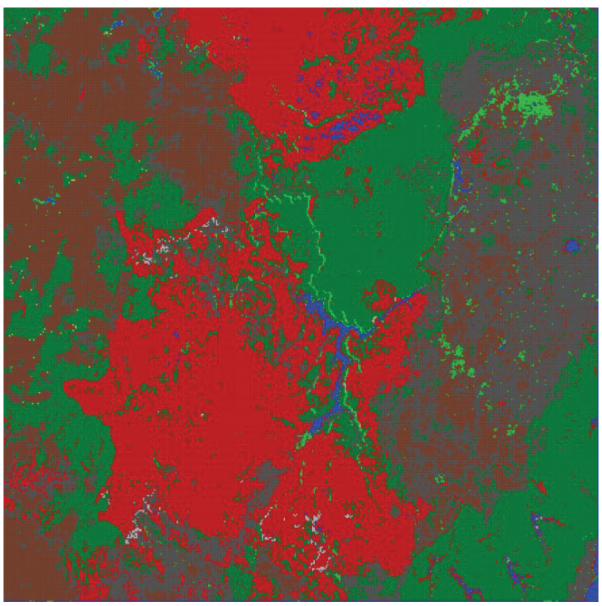

Fig. 8. Patch based (250 x 250 m) classification for SVM using the MeDiB descriptor applied on the FA bands. Legend: low density urban (yellow), high density urban (grey), water (blue), agriculture (light green), forest (dark green), grassland (brown), burned area (red), smoke (light grey)

IV. Conclusion

Data representation plays an important role in data analysis. The benchmark results in this paper show that the dimensionality reduction used to represent Sentinel 2 data can outperform raw data and spectral indexes in a burned area mapping scenarios when there is smoke covering the area of interest. Preliminary analysis of the state of the art methods and algorithms and the experiments in this paper lead us to the combination of FA for data representation, MeDiB for feature extraction and SVM for feature classification for optimum results when mapping burned areas in smoke covered Sentinel-2 data. With a reduced computational burden, the proposed approach provides fast results for forest fires analysis.

Disaster management authorities require information as soon as data is available, even if it sometimes comes covered in smoke clouds. This paper has identified a viable solution for this scenario, which can be used to improve the utilization of remote sensing optical data.

Further work needs to be done to improve the performance and to try to find solutions in the case of dense smoke cloud formations.

Acknowledgment

This work has been performed within the frame of the “Multispectral Data Analysis Toolbox for SNAP – ESA’s SentiNel Application Platform” project, funded by ESA, and it will be made available, through ESA’s webpage as open source.

References

[1] S. Doerr and C. Santin, "Global trends in wildfire and its impacts: Perceptions versus realities in a changing world," *Philosophical Transactions of The Royal Society B Biological Sciences,* vol. 371, no. 1696, 2016.

[2] Wikipedia, "List of major bushfires in Australia," 15 04 2020. [Online]. Available: https://en.wikipedia.org/wiki/List_of_major_bushfires_in_Australia.

[3] "MODIS (or Moderate Resolution Imaging Spectroradiometer)," NASA, [Online]. Available: https://modis.gsfc.nasa.gov/. [Accessed 15 04 2020].

[4] "Global Forest Watch Fires," [Online]. Available: https://fires.globalforestwatch.org/home/. [Accessed 15 04 2020].

[5] "Global Wildfire Information System (GWIS)," European Commission - Copernicus, [Online]. Available: https://gwis.jrc.ec.europa.eu/static/gwis_current_situation/public/index.html. [Accessed 15 04 2020].

[6] F. Ling, Y. Du, Y. Zhang, X. Li, and F. Xiao, "Burned-Area Mapping at the Subpixel Scale With MODIS Images," *IEEE GEOSCIENCE AND REMOTE SENSING LETTERS,* vol. 12, no. 9, pp. 1963-1967, 2015.

[7] M. Humber, L. Boschetti, and L. Giglio, "Assessing the Shape Accuracy of Coarse Resolution Burned Area Identifications," *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING,* 2019.

[8] S. Liu, Y. Zheng, M. Dalponte, X. Tong and Q. Du, "FEATURE-LEVEL FUSION OF LANDSAT-8 OLI-SWIR AND TIR IMAGES FOR FINE BURNED AREA CHANGE DETECTION," *IGARSS 2019,* vol. 9967, p. 9964, 2019.

[9] C. Corinna and V. Vapnik, "Support-vector networks," *Machine Learning,* pp. 273-297, 1995.

[10] N. Georgopoulos, D. Stavrakoudis, and I. Gitas, "OBJECT-BASED BURNED AREA MAPPING USING SENTINEL-2 IMAGERY AND SUPERVISED LEARNING GUIDED BY EMPIRICAL RULES," in *IGARSS 2019*, Yokohama, 2019.

[11] R. Lasaponara, A. Proto, G. Cardettini, V. Varela, and M. Danese, "On the Mapping of Burned Areas and Burn Severity Using Self Organizing Map and Sentinel-2 Data," *IEEE GEOSCIENCE AND REMOTE SENSING LETTERS,* 2019.

[12] A. Grivei, I. Neagoe, F. Georgescu, A. Griparis, C. Vaduva, Z. Bartalis and M. Datcu, "Multispectral data analysis for semantic assessment A