

# Random forest classification: A case study of dryland crop cover mapping in the Victorian Mallee using Sentinel-2A, Sentinel-3, and MODIS imagery

**S. Sabaghy<sup>a</sup>, K.J. Sheffield<sup>a</sup>, and S.A. Robson<sup>b</sup>**

<sup>a</sup> *Agriculture Victoria Research, Department of Jobs, Precincts and Regions, AgriBio, 5 Ring Road, Bundoora, 3083, Victoria, Australia*

<sup>b</sup> *Agriculture Victoria Research, Department of Jobs, Precincts and Regions, 110 Natimuk Rd, Horsham, 3400, Victoria, Australia*

*Email: [sabah.sabaghy@ecodev.vic.gov.au](mailto:sabah.sabaghy@ecodev.vic.gov.au)*

**Abstract:** Land cover characterises the land surface, which is fundamental information for wind and water erosion management. Of particular interest are different crop types since different crop residues provide different levels of protection against wind or water erosion and have different risk characteristics. This study examines the use of random forest algorithm for accurate crop classification across the Victorian Mallee to support wind erosion risk monitoring activities. The random forest is one of the most popular and efficient machine learning algorithms for land cover classification.

The Mallee is an economic region located in north-west Victoria and covers a total area of approximately 40,000 km<sup>2</sup>. It is characterized by flat lands and low-lying areas. The dominant light sandy top-soils, low annual rainfall and characteristic strong winds makes the Mallee one of the most vulnerable areas to wind erosion in Australia. Wind erosion assessment by the National Landcare Program (NLP) identified that 97% of the Mallee Management Unit's dryland cropping areas and 81% of its grazing areas are highly or moderately impacted by wind erosion. This represents the third and second highest figures respectively when compared to all other Australian Natural Resource Management (NRM) regions.

This study initially used a single date MODIS image - a freely available satellite imagery - for land cover mapping. However, as the MODIS satellite platform approaches mission end, the study also investigated alternative multispectral images including Sentinel-2A and Sentinel-3 synergy data for land cover mapping with the aim to compare different platforms and identify the most promising satellite(s) data for crop cover type classification.

The use of random forest algorithm for crop type classification indicated satisfactory overall accuracy (greater than 90%) and kappa coefficient (greater than 0.80). The MODIS-based land cover map performed the best, with an overall accuracy of 97.9% and a kappa coefficient of 0.96. A kappa value close to 1 assures a strong agreement between classified image and reference data. The second-best classification accuracy was achieved for the Sentinel-3 synergy-based land cover map which reached an overall accuracy of 94.2% and kappa value of 0.88. These results show the potential of Sentinel-3 synergy imagery in deriving representative spatial variability of land cover.

In this study, the random forest was trained to assign a discrete land cover to each pixel. Future work could expand this research is to develop and test random forest model/s that can classify mixed pixels. This paper focuses on imagery acquired during 2019, future work will apply these techniques to imagery in other years to assess the robustness of this approach. Assessment of these techniques across different crop growing regions will further ascertain the usefulness of random forest classification of land cover in different climatic zones and growing regions.

**Keywords:** *Crop cover mapping, image classification, random forest, machine learning, agriculture*

## 1. INTRODUCTION

The Mallee is an economic region located in north-west Victoria and covers a total area of approximately 40,000 km<sup>2</sup>. It is characterized by flat lands and low-lying areas. The dominant light sandy top-soils, low annual rainfall and characteristic strong winds makes the Mallee one of the most vulnerable areas to wind erosion in Australia. Wind erosion assessment by the National Landcare Program (NLP) identified that 97% of the Mallee Management Unit's dryland cropping areas and 81% of its grazing areas are highly or moderately impacted by wind erosion. This represents the third and second highest figures respectively when compared to all other Australian Natural Resource Management (NRM) regions.

Wind erosion risk and cover condition is currently monitored across the Mallee region through the Mallee Wind Erosion and Land Management monitoring project being undertaken by Agriculture Victoria. Land cover data are fundamental information that can be used to inform assessments of not only land cover practices (and their associated risks in terms of wind and water erosion), but potential erosion mitigation factors such as vegetation cover. Land cover characterises the land surface (e.g. cereal crop, pasture grass, native woody vegetation) and is particularly useful as different crop residues provide different levels of protection from erosion and have different risk characteristics compared with other land covers such as pastures and horticulture. To support wind erosion risk monitoring activities, this study establishes methods to produce an annual land cover map across the Mallee region, capturing key broadacre crop types (cereal, legume, canola) as well as pasture and bare ground.

The land cover mapping is based on classification of freely available optical data including Moderate Resolution Imaging Spectroradiometer (MODIS) Terra, Sentinel-2A and Sentinel-3 synergy using the random forest algorithm. This study initially used a single date MODIS image for land cover mapping. However, as the MODIS satellite platforms approach mission end, the study was expanded to use alternative multispectral images for land cover mapping, aiming to compare different platforms and identify the most promising satellite(s) data for crop cover types classification in the event of a MODIS failure. The random forest, one of the most popular and efficient machine learning algorithms for land cover classification (Song et al., 2017; Jin et al., 2018), was used to classify land cover across the Mallee Catchment Management Authority (CMA). The dataset is delivered as a crop type spatial dataset and map, which will inform wind erosion risk assessments. The spring transect (Coombes et al., 2019) undertaken as part of the monitoring framework captures land cover, cover management, and the presence of livestock. The spring transect land cover categories follow those published in the Victoria Land Information System (VLUIS) and is used as calibration and validation data to produce a land cover map for the Mallee.

## 2. DATA

### 2.1. REMOTE SENSING DATA

A single date MODIS MOD13Q1 image and a single date Sentinel-2A image captured on 01 October 2019 during the crop growing season, as well as a Sentinel-3 synergy Level-2 V10 which is a 10-day synthesis surface reflectance product composited for 14 October 2019 were used in this study. These observation dates are the closest date to the transect survey completion time. Single date imagery delivers more accurate land cover map in semi-arid regions (such as Mallee) than the multitemporal satellite images (Langley et al. 2001). The Sentinel-3 synergy product is corrected to ensure the continuity of the SPOT vegetation product (Lamquin et al., 2020; [Sentinels Copernicus website](https://sentinels.copernicus.eu/)). MODIS data was downloaded from <https://search.earthdata.nasa.gov/search>. Access to the Sentinel-2A and Sentinel-3 synergy data was provided through The Copernicus Open Access Hub (<https://scihub.copernicus.eu/>). These multi-spectral images were resampled from their original spatial resolution to a common 50 m grid using nearest neighbour resampling method. Details of the imagery bands including band name, spatial resolution and their corresponding central wavelength (if applicable) are summarized in Table 1.

### 2.2. CALIBRATION AND VALIDATION DATA

The spring transect data was used as calibration and validation data. This data which was collected during roadside surveys conducted across the Mallee in September 2019 and recorded observed land cover at pre-determined points. 1143 paddocks were surveyed during this spring transect survey using an application called 'Collector for ArcGIS'. This data collection method is based on a pre-defined data entry form. An example of this form is shown in Figure 1. The spring transect was completed between the 19th and 25th of September 2019 to capture crop type before harvest activities started.

The land cover information collected during the spring transect survey were used to train and evaluate the forest model. This transect data was processed to include data within a footprint corresponding to a

homogenous (in terms of ground cover) MODIS pixel, which was collected during the survey, to eliminate the confusion around crop type classification. The field data was then resampled to a 50 m x 50 m area, corresponding to the resampled satellite imagery pixel footprints. Heterogenous pixels at this scale were also removed to mitigate confusion between crop types.

**Table 1.** List of spectral bands used for land cover type classification.

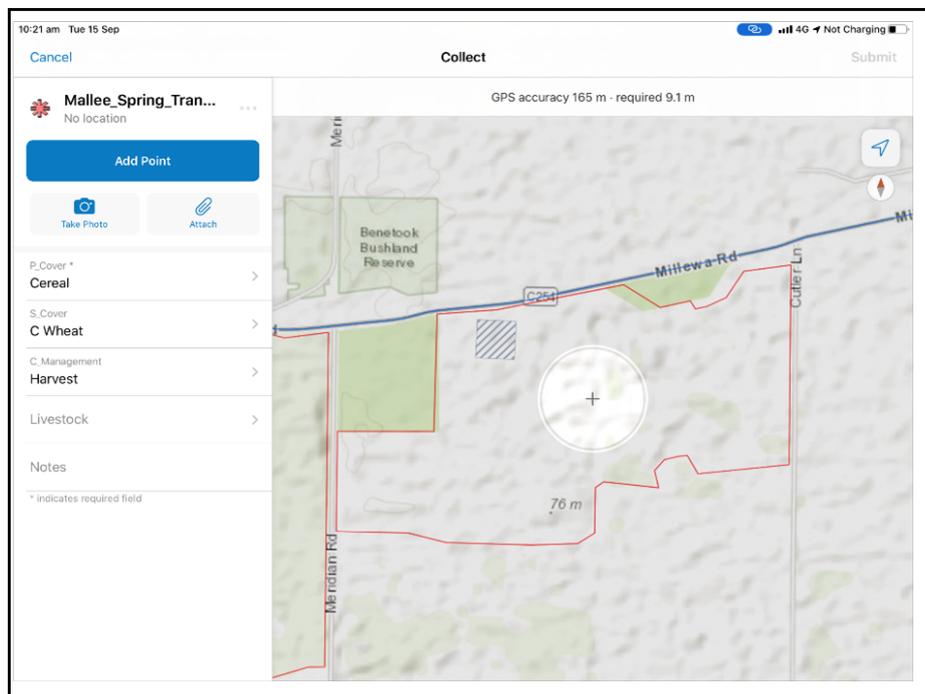
MODIS			Sentinel-2A			Sentinel-3 synergy		
Bands	Spatial resolution	Central wavelength (nm)	Bands	Spatial resolution	Central wavelength (nm)	Bands	Spatial resolution	Central wavelength (nm)
NDVI	250 m	NA	NDVI <sup>1</sup>	10 m	NA	NDVI	1 km	NA
EVI	250 m	NA	EVI <sup>2</sup>	10 m	NA	Blue	1 km	450
Blue	250 m	469	B2(Blue)	10 m	492.4	Red	1 km	645
Red	250 m	645	B3(Green)	10 m	559.8	NIR	1 km	835
NIR	250 m	858	B4(Red)	10 m	664.6	MIR	1 km	1665
MIR	250 m	2130	B5* <sup>3</sup>	20 m	704.1			
			B6*	20 m	740.5			
			B7*	20 m	782.8			
			B8(NIR)	10 m	832.8			
			B8A <sup>4</sup>	20 m	864.7			
			B11(SWIR)	20 m	1613.7			
			B12(SWIR)	20 m	2202.4			

<sup>1</sup> NDVI = (NIR - Red) / (NIR + Red) = (B8 - B4) / (B8 + B4)

<sup>2</sup> Sentinel-2A EVI = 2.5 \* (B8 - B4) / ((B8 + 6.0 \* B4 - 7.5 \* B2) + 1.0)

<sup>3</sup> B5\*, B6\*, and B7\* bands are vegetation red edge.

<sup>4</sup> B8A band is narrow NIR band.



**Figure 1.** Screenshot sample of the predefined entry form in Collector for ArcGIS app.

### 3. METHOD

#### 3.1. RANDOM FOREST CLASSIFICATION

The random forest or random decision forest classifier was used in this study to classify land cover type. It is one of the most popular and efficient machine learning algorithms for land cover classification (Song et al., 2017; Jin et al., 2018). This is a tree-based ensemble machine learning method that constructs an ensemble of

classification and regression trees at training time and accounts for correlation and interactions among features to output the classes. Python code was developed to optimize the random forest model to generate the Mallee land cover data. The optimization of random forest model resulted in increased accuracy percentages in the product and crop cover types. The random trees classification procedure includes: i) developing a training model to split data into explanatory branches, and ii) performing the random trees classification on a per-pixel basis, based on the input training feature file. A schematic of the approach to classify the land cover types (i.e. cereal, legume, oilseed, bare and pasture) over the Mallee region is shown in Figure 2.

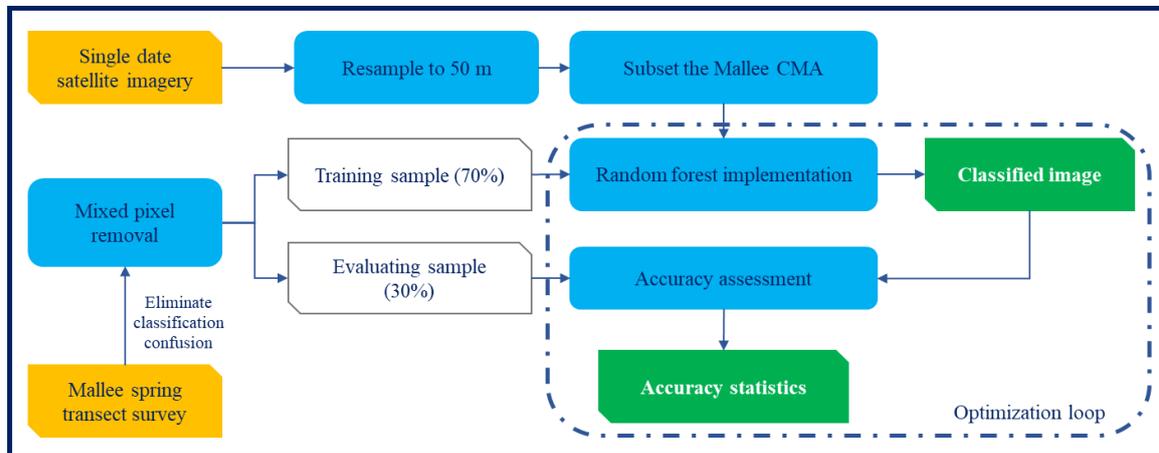


Figure 2. A schematic of dryland cropping land cover classification procedure.

### 3.2. EVALUATION

The evaluation process involved calculation of land classification accuracy metrics (i.e. user accuracy, producer accuracy, Kappa coefficient, and F-score) based on a comparison of the classified map and Mallee transect data as ground truth data. The remaining 30% of land cover data was used for this accuracy assessment. The accuracy assessment is a measure of how well the classified land cover map agrees with what was recorded during the spring transect surveys.

The user accuracy clarifies how well a classified pixel represents that category on the ground. The user accuracy is calculated by taking the ratio of the total number of correct classifications for a category and the total number of pixels that were classified in that category. The producer accuracy represents how well reference pixels of the ground cover type are classified. To calculate the producer accuracy, the number of correctly classified pixels in each category is divided by the number of reference pixels of that category. The Kappa coefficient measures the agreement between classification and ground truth data, compared with a random assignment of class to a pixel. A kappa value of 1 represents complete agreement between classified image and ground data, while a value of 0 represents no agreement. The F-score or F-measure, which is commonly used for machine learning techniques, is a measure of the accuracy of the model. It combines the producer and user accuracy of the model and is defined as the harmonic mean of the user and producer accuracy of the model. The F-score value varies between 0 and 1. An F-score value of 1 indicates perfect producer and user accuracy, while the lowest possible value of 0 is obtained, if either the producer or user accuracy is zero.

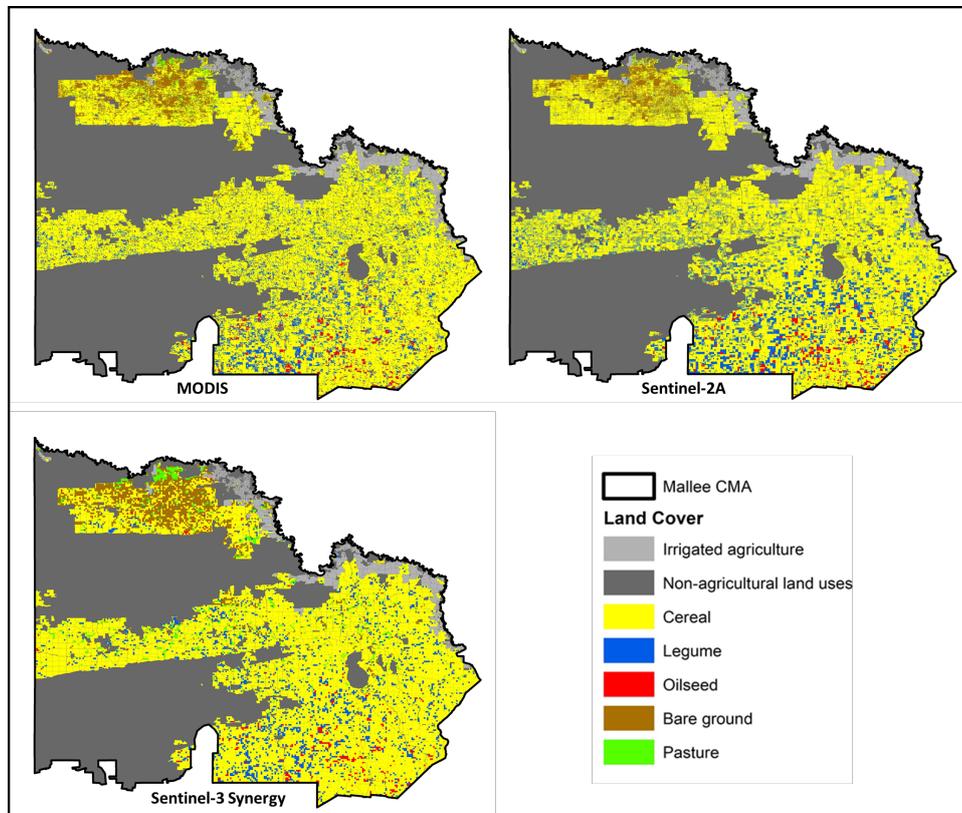
## 4. RESULTS

The Mallee regional land cover maps generated from the MODIS, Sentinel-2A and Sentinel-3 synergy imagery are shown in Figure 3. This map excludes public land, built-up and non-cropped regions as well as irrigated areas based on information obtained from the VLUIIS 2010/2011 (DJPR 2015) and 2016/2017 (DJPR 2018) information, DELWP land cover “dryland cropping” classification data (DELWP 2020) and the ABARES catchment scale land use of Australia 2018 land use data (ABARES 2019). The accuracy assessment of these land cover maps is summarized in Table 2. The land cover classification based on the MODIS imagery represented the highest level of overall accuracy (97.9%) supported by a high kappa value of 0.96. A kappa value close to 1 shows a strong agreement between classified image and ground truth data.

The land cover map produced using the Sentinel-3 synergy imagery represented the second most accurate crop type classification with overall accuracy of 94.2% and kappa value of 0.88. Details in Figure 4 indicate that classes in all crop cover maps were well classified with low levels of misclassification between classes. However, among different crop types, legumes and cereals were slightly confused.

## 5. DISCUSSION AND CONCLUSION

The Mallee regional land cover map was created based on a random forest classification of satellite imagery that are available to the public free of charge. The Mallee regional land cover data is intended for use at a regional or landscape scales. The classification process was developed to assess the capability of various multispectral images including MODIS, Sentinel-2A and Sentinel-3 synergy to discriminate crop types.



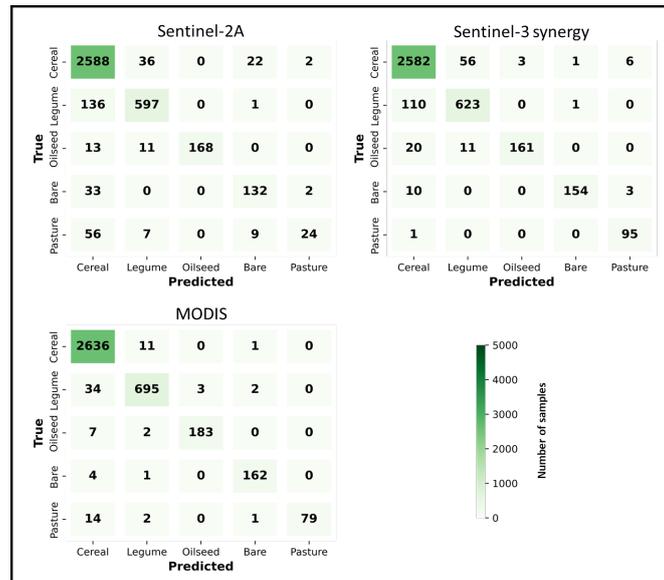
**Figure 3.** Cropping and pasture land cover across the Mallee CMA area acquired in October 2019. These land cover maps are created using MOD13Q1, Sentinel-2A and Sentinel-3 synergy Level-2 V10 images resampled to 50 m spatial resolution.

**Table 2.** Overall accuracy of the crop cover maps along with the summary of accuracy for each class of crop types in each 2019 crop cover maps.

Satellite imagery	Overall Accuracy	Kappa Coefficient	Crop types	Producer Accuracy	User Accuracy	F-score
Sentinel-2A	91.5%	0.81	Cereal	0.92	0.98	0.95
			Legume	0.92	0.81	0.86
			Oilseed	1.00	0.88	0.93
			Bare	0.80	0.79	0.80
			Pasture	0.86	0.25	0.39
Sentinel-3	94.2%	0.88	Cereal	0.95	0.98	0.96
			Legume	0.90	0.85	0.88
			Oilseed	0.98	0.84	0.90
			Bare	0.99	0.92	0.95
			Pasture	0.91	0.99	0.95
MODIS	97.9%	0.96	Cereal	0.98	1.00	0.99
			Legume	0.98	0.95	0.96
			Oilseed	0.98	0.95	0.97
			Bare	0.98	0.97	0.97
			Pasture	1.00	0.82	0.90

This classification process is not without errors, which should be taken into account when using the data. The classification accuracy is influenced by many factors including the amount of ground data used (in total and for each class), the original resolution of the satellite imagery, the characteristics of spectral bands of each multispectral image, the vegetation growth patterns during the season, climatic conditions, and soil reflectance. These factors, either individually or in combination, can have a greater or lesser impact on classification accuracy dependent on the year. The unprecedented dry seasonal conditions experienced in the

Mallee during 2019, and subsequent impact on crop and pasture growth, would have had an impact on the classification accuracy observed. Therefore, the accuracy of the data (both overall and by class) should be considered when interpreting this data.



**Figure 4.** Confusion matrix for the 2019 Mallee regional land cover classifications.

In this study, the random forest was trained to assign a discrete land cover to each pixel. Development of a random forest model which can classify mixed pixels should be investigated as an expansion to this research. The main focus of interest would be to classify mixed pixels in remotely sensed images applying the sub-pixel classification techniques in accordance with other studies that address the problem of sub-pixel classification. For instance, Kumar et al. (2017) used the subpixel algorithms to disintegrate Landsat data spectrum into its constituent spectra and obtained abundances of Substrate (S), Vegetation (V) and Dark object (D) classes. These S-V-D abundance maps were subsequently classified using a random forest classifier. Sanpayao et al. (2017) also applied the random forest into the subpixel classification by classifying the portion of each pixel using the new voting scheme of decision trees based on the maximum a posteriori probability estimate.

The increased availability of remotely sensed earth observations allows the accelerated land cover mapping across the globe. Representative training datasets are required as the key component in mapping accurate land cover through supervised classification techniques (Viana et al. 2019) such as random forest. Restrictive access to such field data imposes a limit on the application and development of such supervised classification techniques. An alternative option when little to no field data exists could be the application of unsupervised classification methods, creating training samples using the existing land cover products, migrating prior training data, and using expert knowledge through visual interpretation of other products (Chaves et al. 2020).

The use of random forest algorithm for crop type classification indicated satisfactory overall accuracy (greater than 90%) and kappa coefficient (greater than 0.80). Note that the procedure of mixed pixel removal from the training data helped with increasing the accuracy of crop cover maps and reducing the classification confusion between crop types. Classifying the MODIS imagery based on the processed transect data which excluded heterogeneous pixels resulted in overall accuracy of 97.9% which was slightly higher than the accuracy of classified map that was created using both homogenous and heterogeneous field data pixels (97.6%). However, increased accuracy due to mixed pixel removal was more significant for the Sentinel-based land cover maps compared to that of MODIS-based one; Sentinel-2A and Sentinel-3 synergy-based land cover maps experienced an overall accuracy increase of 1.5% (from 90% to 91.5%) and 1.3% (from 92.9% to 94.2%), respectively.

The objective of this study was to evaluate the performance of various multispectral images (namely MODIS, Sentinel-2A, and Sentinel-3 synergy) for land cover mapping. The results showed that the MODIS-based land cover map best captured the spatial variability of crop types reaching the overall accuracy of 97.9% and the user accuracy of at least 82% for different crops. This overall accuracy is higher than the one reported for the MODIS-based land cover map (92%) in Song et al. (2021) which classified corn and soybean crop cover types. The outperformance of MODIS-based land cover map is unlike the results published in Song et al.

Sabaghy et al., Random forest classification: A case study of crop cover mapping in the Victorian Mallee using Sentinel-2A, Sentinel-3, and MODIS data

(2021). Based on this study, MODIS-based land cover map had the lowest performance compared to that of Sentinel-2, Sentinel-1, and Landsat based land cover maps.

The outstanding performance of MODIS-based crop cover map was followed by that of Sentinel-3 synergy-based land cover map which reached an overall accuracy of 94.2% and kappa value of 0.88. As the MODIS mission is approaching the end of its lifetime, the Sentinel-3 synergy platform appears to be a promising alternative to the MODIS imagery for classification of land cover types.

Classification of Sentinel-2A imagery for crop type mapping revealed the lowest level of overall accuracy compared to the other classified satellite data. An overall accuracy of 91.5% was achieved for the Sentinel-2A based land cover map. This level of accuracy is comparable to the accuracy reported for the classified Sentinel-2 in Thanh et al. (2018), but still lower than the overall accuracy of 96.6% reported in Song et al. (2021) for Sentinel-2 based crop cover map. This paper focuses on imagery acquired during 2019, future work will apply these techniques to imagery in other years and regions to assess the robustness and broader applicability of these results.

## AKNOWLEDGEMENT

This project is supported by Mallee Catchment Management Authority (MCMA), through funding from the Australian Government's National Landcare Program. The authors would like to acknowledge Rachel Coombes, Martin Hamilton, and Darryl Pearl (Agriculture Victoria) for undertaking the Mallee spring transect survey.

## REFERENCES

- ABARES, 2019. Catchment scale land use of Australia - Update December 2018. Australian Bureau of Agricultural and Resource Economics and Sciences. Canberra, CC BY 4.0, <https://data.gov.au/data/dataset/8d5d0a09-d100-407b-b326-6e775025feec>.
- Chaves, M., Picoli, M., Sanches, I., 2020. Recent Applications of Landsat 8/OLI and Sentinel-2/MSI for Land Use and Land Cover Mapping: A Systematic Review. *Remote Sensing*, 12(18), 3062.
- Coombes, R., Sheffield, K., Robson, S., 2019. Monitoring Wind Erosion and Land Management in the Victorian Mallee: Spring Transect Report 2019. Agriculture Victoria (Department of Jobs, Precincts and Regions), Victoria.
- DELWP, 2020. Victorian Land Cover Time Series. Department of Environment, Land, Water and Planning, <https://www.environment.vic.gov.au/biodiversity/Victorias-Land-Cover-Time-Series>
- DJPR, 2018. Victorian Land Use Information System 2016-17. Department of Jobs, Precincts and Regions, <https://discover.data.vic.gov.au/dataset/victorian-land-use-information-system-2016-2017>.
- DJPR, 2015. Victorian Land Use Information System 2010/11. Department of Jobs Precincts and Regions, <https://discover.data.vic.gov.au/dataset/victorian-land-use-information-system-2010-2011>.
- Jin, y., Liu, X., Chen, Y., Liang, X., 2018. Land-cover mapping using Random Forest classification and incorporating NDVI time-series and texture: a case study of central Shandong, *International Journal of Remote Sensing*, 39:23, 8703-8723.
- Kumar, U., Ganguly, S., Nemani, R.R., Raja, K.S., Milesi, C., Sinha, R., Michaelis, A., Votava, P., Hashimoto, H., Li, S., Wang, W., Kalia, S., Gayaka, S., 2017. Exploring Subpixel Learning Algorithms for Estimating Global Land Cover Fractions from Satellite Data Using High Performance Computing, *Remote Sensing*, 9, 1105.
- Lamquin N, Clerc S, Bourg L, Donlon C., 2020. OLCI A/B Tandem Phase Analysis, Part 1: Level 1 Homogenisation and Harmonisation, *Remote Sensing*, 12(11):1804.
- Langley, S., Cheshire, H., Humes, K., 2001. Comparison of single date and multitemporal satellite image classifications in a semi-arid grassland, *Journal of Arid Environments*, 49, 401-411.
- Sanpayao, M., Kasetkasem, T., Rakwatin, P., Isshiki, T., 2017. A Subpixel Classification Algorithm Based on a Random Forest Approach, 8th International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES), 7-9 May 2017, Chonburi, Thailand.
- Song, I., Huang, W., Hansen, M. C., Potapov, P., 2021. An evaluation of Landsat, Sentinel-2, Sentinel-1 and MODIS data for crop type mapping, *Science of Remote Sensing*, 3(100018), ISSN 2666-0172.
- Song, Q., Hu, Q., Zhou, Q., Hovis, C., Xiang, M., Tang, H., Wu, W., 2017. In-Season Crop Mapping with GF-1/WFV Data by Combining Object-Based Image Analysis and Random Forest, *Remote Sensing*, 9(11), 1184.
- Thanh N. P., Kappas, M., 2018. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery, *Sensors* .18(1):18.
- Viana, C.M.; Girão, I.; Rocha, J., 2019. Long-Term Satellite Image Time-Series for Land Use/Land Cover Change Detection Using Refined Open Source Data in a Rural Region, *Remote Sensing*, 11, 1104.