Mapping cotton fields using phenology-based metrics derived from a time series of Landsat imagery

Dhahi Al-Shammari[1], Thomas F.A. Bishop[1], Ignacio Fuentes[1], Patrick Filippi[1]

[1] Sydney Institute of Agriculture, School of Life & Environmental Science, The University of Sydney, Central Ave, Eveleigh, Sydney, New South Wales, 2015

Abstract
A phenology-based crop type classification was carried out to map cotton in New South Wales and Queensland, Australia. The workflow was implemented in Google Earth Engine (GEE) platform as it is time efficient and does not require processing in multiple platforms to complete the classification steps. A time-series of images were generated from Landsat 8 Surface Reflectance (L8SR) data. The Normalised Difference Vegetation Index (NDVI)time series was calculated from satellite specifically Landsat imagery and a Harmonic Model (HM) was fitted to it to produce the Harmonised NDVI (H-NDVI). Phase and amplitude images were generated to visualise active cotton in the targeted fields. These images were used as predictor variables with H-NDVI and other raw bands in the Random Forest (RF) classification model. The results of RF proved that both phase and amplitude increased the accuracy of the classification. Moreover, cotton classification accuracy increased as the season progressed.

Keywords
Harmonic model, phase, amplitude, classification, time series

Introduction
Crop type classification is widely studied, and, in most cases, it is performed retrospectively after the growing season (Azar et al., 2016). The conventional way of collecting land-use information is mostly achieved by surveying farmers or conducting land surveys. This is an accurate way of mapping crop type, but it is relatively expensive and time-consuming (Zhong, 2012). Furthermore, within-season classification is required for better management for both farmers and policymakers (Bozbek, et al., 2006). Researchers have found that remote sensing (RS) information such as hyperspectral and multispectral wavelengths offer great potential for collecting temporal and spatial data to support decisions on crop monitoring and management (Chen et al., 2018). Crop type maps from remote sensing are cost/time-effective for large areas of interest. RS-derived products such as NDVI imagery is usually used to detect seasonal changes of vegetation cover over a specific period. Tracking crop cycle is one of the methods to identify crop type, with unique a pattern for each crop type (Guyon and Bréda, 2016). Phenology is one of the traits used to monitor periodic plant life-cycle events and how climate influences these events, and thus it can be used for crop type identification (Zhong et al., 2012). Crops in a particular environment have specific phenological stages at defined time intervals in the year (Rodriguez-Galiano et al., 2015). Therefore, there is a high potential for specific crop discrimination based on its unique characteristics depending on the stage of the growing season. This study looks at the analysis of phenology-based metrics to identify crop type using Random Forest Classification (RF) in Google Earth Engine (GEE).

Methods
Study area and reference data
This study covered the major cotton production areas which stretch from the Macintyre River on the Queensland-NSW border to the Murrumbidgee River in New South Wales. Reference data were provided as field boundaries for the 2013 season which refer to cotton fields for each growing season CottonMap (2014). All field boundaries were further checked, and some were excluded as they did not have any cotton crop for 2013-2014 summer season.

Data preparation within Google Earth Engine platform
The Google Earth Engine (GEE) platform is a powerful tool that allows the acquisition, processing and analysis of data from many remote sensing platforms such as Landsat (Xiong et al., 2017). For example, Shelestov et al. (2017) found that GEE is very useful and enabled them to classify large areas efficiently due to the ability of GEE to handle big data analysis (Shelestov et al., 2017). GEE was used to carry out all steps in data analysis, from acquiring surface reflectance data to an assessment and verification stage. All clouds and cloud shadows were masked using the FMASK algorithm, which identifies and removes pixels which contain clouds and cloud shadows of all images in the collection (Dong et al., 2016). We then prepared three

data sets for the classification. The first dataset consisted of five Landsat images starting from October to the first of January. The second data set consisted of seven Landsat images starting from October until the first of February. The third data set consisted of 15 Landsat images starting from October until the end of May. The idea of conducting three classifications is to examine the potential of within-season mapping of crop type, in this case cotton.

Construction and extraction of the Harmonic Normalised Difference Vegetation Index (NDVI) Time Series, phase and amplitude

A Harmonic model is one of the methods for the construction and analysis of phenological features, thereby identifying different crops and vegetation species depending on their phase (which represents the angle, which defines the offset between the start of the growing season and the peak of the NDVI) and amplitude (which shows the amount of NDVI change over time) (Yan et al., 2015). To build this model, a mathematical transformation of the NDVI is performed to decompose it into unique phase and amplitude (Chen et al., 2018). In our study, a Fourier transformation procedure was applied to decompose NDVI into phase and amplitude components to extract the unique seasonal patterns of cotton crop cycle (fig.1). The HM was performed using equation (1).

$$\mathbf{NDVIt} = \boldsymbol{\beta 0} + \boldsymbol{\beta 1 t} + \boldsymbol{\beta 2 \cos(2\pi\omega t)} + \boldsymbol{\beta 3 \sin(2\pi\omega t)} \quad \mathbf{(1)}$$

where $NDVI_t$ is the value of NDVI at each time period (t), $\beta 1t$ is the linear trend, and $\beta 2\cos(2\pi\omega t) + \beta 3\sin(2\pi\omega t)$ represent the cyclical period of 32 weeks (growing season) for the cotton crop. Many applications ignore the importance of phase information. Phase indicates the time of the year in which the greenness peak occurs which for cotton occurs in March in New South wales region (Jakubauskas et al., 2002). Therefore, we assigned the phase band to be included within our extracted time-series bands. Moreover, amplitude band which is an indicator of the change in NDVI throughout time was used in the time series classification (Dubovyk et al., 2016).

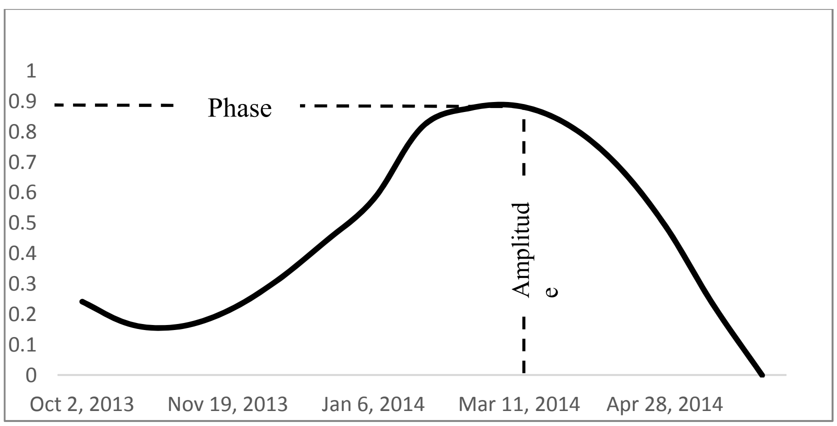

Figure 1. An example of crop cycle (cotton) which shows distinct phase and amplitude for cotton.

Google Earth Engine Random Forest (GEE-RF) classification

The training dataset was further processed prior to the classification to visualise active crops within each field, based on pixel greenness. This was done visually to ensure that all training and validation samples represent cotton fields. Images of phase and amplitude were used to help visualise active crop in cotton fields for that particular season. Phase and amplitude images helped to detect the fields that were completely bare soil which if included in the analysis will decrease the accuracy of the classification.

The next step was to build the model, and this was based on raw bands [bands 1-7, 10-11, Harmonised-NDVI (HNDVI), amplitude and phase]. A pixel-based supervised classification was performed using a Random Forest (RF) classification model. Samples were randomly created within each field to be used in the analysis. These samples were split randomly in GEE to 80% training samples and 20% validation samples. Each sample unit consists of a field boundary which includes values of raw bands, H-NDVI, amplitude and phase through the time.

Results

Accuracy assessment

The accuracy assessment was performed using overall accuracy and kappa coefficient. The overall accuracy is the total number of classified samples divided by the total number of samples. Kappa coefficient is used to evaluate the accuracy of image classification. It ranges from 0, when there is no agreement between training and validation sets, and 1, when there is strong agreement between training and validation sets. Figure.2 (map 1) shows the final cotton map for the study area, but we only show cotton fields (a) near Narrabri, and (b) in the Border rivers regions due to the huge coverage of study area. According to the RF classification, overall accuracy and kappa coefficient increased when using H-NDVI, phase, amplitude and other bands to the classification. Table 1 shows the results of classification achieved from validation datasets. It is clear from table 1 that amplitude and phase helped to increase the accuracy of predictions. For example, the overall accuracies of early classification model improved from 83%, 87% and 88% when using H-NDVI, phase and amplitude, respectively, to 97%, 97%, and 99% when using all features together to classify cotton. Similarly, overall accuracies increased with using combination of features together to classify cotton in mid and late season. Moreover, Kappa increased significantly when using combination of features for the three models. We can see that amplitude was the most important variable in the selected features, followed then by phase, H-NDVI and other parameters. Our study showed that the Fourier Transformations helped to achieve the highest accuracy.

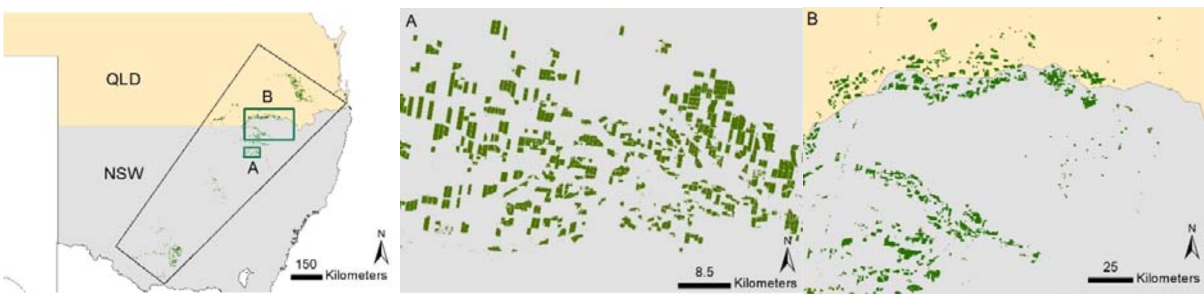

Figure 2. Cotton map 1) locations of classified cotton. a) cotton fields in Narrabri. b) cotton fields in Gwydir and Macintyre region.

Table 1. Lists of overall accuracies achieved of Random Forest (RF) classification. Accuracies achieved using different combinations of mean values of bands (B1-B7, B10 and B11, phase, amplitude and H-NDVI).

Date	**Accuracy assessment**	Raw Bands + H-NDVI + Amplitude + Phase	Raw Bands + H-NDVI + Amplitude	Raw Bands + H-NDVI	H-NDVI	Amplitude	Phase
01-10-13/01-01-14	Accuracy	0.99	0.97	0.97	0.83	0.88	0.87
	Kappa	0.97	0.94	0.91	0.70	0.51	0.69
01-10-13/01-02-14	Accuracy	0.96	0.94	0.98	0.91	0.90	0.90
	Kappa	0.92	0.87	0.91	0.84	0.76	0.78
01-10-13/01-06-14	Accuracy	0.97	0.95	0.98	0.91	0.92	0.88
	Kappa	0.94	0.90	0.97	0.83	0.63	0.72

Discussion

Temporal-spectral information; specifically, amplitude and phase, were the key features which helped to map cotton. Amplitude and phase were the most important spectral features in the classification for the three selected time periods (Fig. 3). In all three classification models, amplitude, phase and H-NDVI were the most important variables. Figure 3 shows variable importance from early, mid and late-season models. The unique amplitude and phase (the angle of which the peak occurs) helped to increase the accuracy of the classification. Generally, accuracies for the time series classification were high when mean H-NDVI and

other raw bands were used but including phase and amplitude gave the classes a unique curve that could be used to discriminate between cotton and other summer crops in the region, e.g. sorghum.
Classification of cotton crops using a RF model yielded high accuracy using a time series of remotely-sensed imagery features for three periods in the growing season. It is important to consider the quality of training samples used in the classification to ensure that masked pixels do not affect the results of classification.

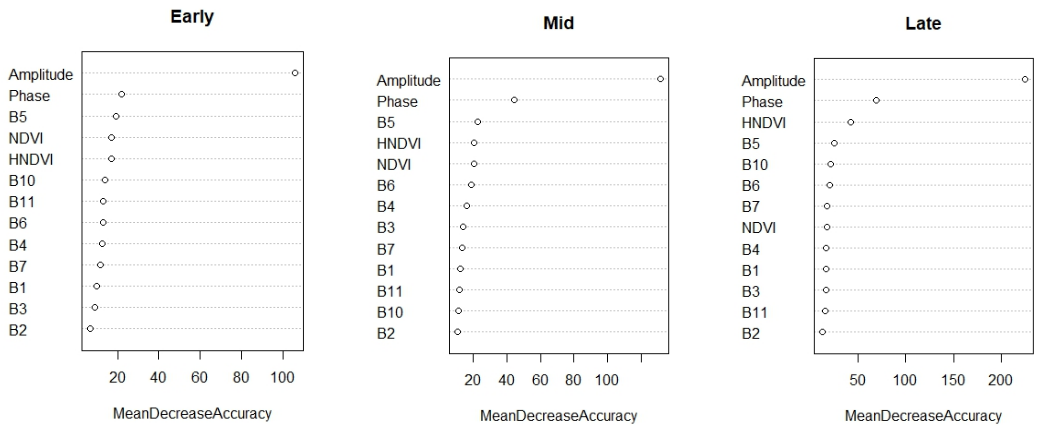

Figure 3. variable importance of input features. (left) Early season classification. (Middle) Mid-season classification. (Right) Late season classification.

Conclusions

This study investigated the potential of mapping cotton fields in NSW and QLD for a single season using phenology-based indices derived from a times series imagery using GEE. The results obtained from this study showed that the addition of Fourier Transformations to the traditional NDVI increased the accuracy of the classification. This approach can be extended by including more crop types in the model and then testing its ability to discriminate between a wide range of crops. Furthermore, different vegetation indices can be included in this model to test their sensitivity to discriminate different classes.

References

Chen J, Chen J, Liu H, et al. (2018) Detection of Cropland Change Using Multi-Harmonic Based Phenological Trajectory Similarity. *Remote Sensing* 10: 1020.

Dong J, Xiao X, Menarguez MA, et al. (2016) Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine. *Remote sensing of environment* 185: 142-154.

Dubovyk O, Landmann T, Dietz A, et al. (2016) Quantifying the impacts of environmental factors on vegetation dynamics over climatic and management gradients of central Asia. *Remote Sensing* 8: 600.

Guyon D and Bréda N. (2016) Applications of Multispectral Optical Satellite Imaging in Forestry. *Land Surface Remote Sensing in Agriculture and Forest.* Elsevier, 249-329.

Jakubauskas ME, Legates DR and Kastens JH. (2002) Crop identification using harmonic analysis of time-series AVHRR NDVI data. *Computers and Electronics in Agriculture* 37: 127-139.

Rodriguez-Galiano VF, Dash J and Atkinson PM. (2015) Characterising the land surface phenology of Europe using decadal MERIS data. *Remote Sensing* 7: 9390-9409.

Shelestov A, Lavreniuk M, Kussul N, et al. (2017) Large scale crop classification using Google earth engine platform. *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS).* IEEE, 3696-3699.

Xiong J, Thenkabail PS, Gumma MK, et al. (2017) Automated cropland mapping of continental Africa using Google Earth Engine cloud computing. *ISPRS Journal of Photogrammetry and Remote Sensing* 126: 225-244.

Yan E, Wang G, Lin H, et al. (2015) Phenology-based classification of vegetation cover types in Northeast China using MODIS NDVI and EVI time series. *International Journal of Remote Sensing* 36: 489-512.

Zhong L, Gong P and Biging GS. (2012) Phenology-based crop classification algorithm and its implications on agricultural water use assessments in California's Central Valley. *Photogrammetric Engineering & Remote Sensing* 78: 799-813.