'Sugar from space': using satellite imagery to predict cane yield and variability

AG27
'SUGAR FROM SPACE': USING SATELLITE IMAGERY TO PREDICT CANE YIELD AND VARIABILITY

By

J.S. MUIR[1], A.J. ROBSON[1], M.M. RAHMAN[1]

[1]*Agricultural Remote Sensing Team, Precision Agriculture Research Group, University of New England, Armidale, NSW*

Jasmine.Muir@une.edu.au

KEYWORDS: Satellite imagery, sugarcane, yield, variability, time series

Abstract

Satellite imagery has been demonstrated to be an effective technology for producing accurate pre-harvest estimates in many agricultural crops. For Australian sugarcane, yield forecasting models have been developed from a single date SPOT satellite image acquired around peak crop growth. However, a failure to acquire a SPOT image at this critical growth stage, from continued cloud cover or from competition for the satellite, can prevent an image being captured and therefore a forecast being made for that season. In order to reduce the reliance on a single image capture and to improve the accuracies of the forecasts themselves, time series yield prediction models have been developed for eight sugarcane growing regions using multiple years of free Landsat satellite images. In addition to the forecasting of average regional yield, an automated computational and programming procedure enabling the derivation of crop vigour variability (GNDVI) maps from the freely available Sentinel 2 satellite imagery was developed. These maps, produced for 15 sugarcane growing regions during the 2017 growing season, identify both variations in crop vigour across regions and within every individual crop. These outputs were made available to collaborating mills within each growing region. This paper presents the accuracies achieved from the time series yield forecasting models versus actual 2017 yields for the respective regions, as well as provides an example of the derived mapping outputs.

Introduction

Satellite imagery has been used to provide accurate pre-harvest yield estimates in many agricultural crops (Bolton and Friedl, 2013). While a number of studies have used satellite imagery to provide a pre-harvest estimate of sugarcane yield, these models have been developed for growing conditions different to those experienced in Australia i.e. Brazil and Reunion Island (Fernandes et al., 2011; Morel et al., 2014). Additionally, whilst these models were demonstrated to be accurate for predicting yield at the local scale, the models were not tested at the regional level. Robson *et al.* (2012) developed a yield forecasting model based on satellite imagery specifically for each Australian sugarcane growing region, to account for region specific climate and growing conditions. The models used the relationship between the historic yield achieved at the individual block level and the corresponding Green Normalised Difference Vegetation Index (GNDVI) extracted from a SPOT 5, 6 or 7 satellite image (acquired around the time of peak GNDVI) to predict yield in tonnes of cane per hectare (TCH). GNDVI has been shown to be less prone to reflectance saturation at higher leaf area index (LAI) than the commonly used Normalised Difference Vegetation Index (NDVI) (Robson *et* *al*., 2012). The yield prediction model initially developed for the Bundaberg region, has now been replicated across many Australian growing regions.

The disadvantage of this method is the difficulty to acquire a satellite image during the window of peak GNDVI (from March to June) when cloud cover persists, or global competition for the satellite is at its highest. In addition, the acquisition of SPOT imagery does incur a cost (approximately $4 000 to $5 000 per sugarcane growing region) (Rahman *et al*., 2017), which can add up if multiple images are required during the growing season. To address these limitations, 'time series' yield models from historic Landsat imagery have been developed for each growing region following the process described by (Rahman and Robson, 2016; Rahman *et al*., 2017). By developing an understanding of historic crop growth trends within each region and matching those trends with annual regional yield, the time series approach has been shown to be more sensitive to seasonal variations in climate and to be less dependent on forecasting yield during the peak growth period. The time series models have also allowed for predictions of regional yield to be made earlier in the growing season i.e. January or February.

In addition to regional forecasting, an image processing methodology, using the Python programming language, to automate the pre-processing of satellite imagery and the derivation of classified crop vigor (GNDVI) and yield maps for all crops within the participating sugar growing regions has been developed. During the 2017 season, Sentinel-2 and SPOT crop variability (GNDVI) maps were delivered to 15 sugarcane growing regions (~90,000 crops), yield maps from SPOT imagery and developed Landsat time series yield prediction models for eight sugarcane growing regions (Tully, Mackay, Burdekin, Herbert, Condong, Harwood, Broadwater) (models will be developed in the future for other regions).

This paper provides an overview of the image processing methodologies and time series yield models developed, as well as validation of the accuracies of yield predictions achieved for the eight growing regions in 2017. The mapping outputs from this research will provide industry with a tool that can accurately identify the spatial and temporal variability in crop performance at both the region and block level. With this information varietal and crop class responses can be clearly observed as well as variations in crop vigour resulting from nutritional constraints, irrigation efficiencies and the incidences of pest and disease outbreaks. Accurate regional forecasts provided to each mill will support pre-harvest planning, forward selling and marketing.

Methods

The research spans 15 sugarcane growing regions, from Mulgrave in northern Queensland to Harwood in New South Wales (NSW) (Figure 1). These growing regions experience very different climatic conditions i.e. rainfall, temperature, humidity and solar radiation. Previous research has shown that a generic relationship between SPOT satellite imagery and TCH across all sugarcane growing regions is not possible due to these climatic differences and therefore individual models for each region are required (Robson *et al*., 2012).

SPOT 6/7 satellite imagery was purchased for each growing region during the peak GNDVI period (assumed to be the same as peak growing period) from March to June in 2017. Early season (i.e. January to March) Sentinel-2 imagery corrected to top of atmosphere reflectance was downloaded from the website https://remotepixel.ca. For eight of the regions (Tully, Mackay, Herbert, Burdekin, Bundaberg, Condong, Harwood and Broadwater) all available Landsat imagery were downloaded via the National Computation Infrastructure Thredds server: http://dapds00.nci.org.au/thredds/catalogs/rs0/catalog.html. Landsat 5, 7 and 8 surface reflectance imagery as well as the corresponding cloud and cloud shadow masks were downloaded, resulting in ~ 100 individual Landsat scenes per sugarcane growing region. The spatial resolutions for the three satellite platforms used in this study are SPOT 6/7 (6 m) Sentinel-2 (10 m) and Landsat (30 m).

Fig. 1— Sugarcane growing regions.

Historic and current mill block data were sourced as GIS spatial layers from each respective mill and saved as ESRI shapefiles. The GNDVI (Eq. 1) was calculated for each SPOT 6/7 and Sentinel-2 satellite image and saved as a new image output. For cloud affected areas in for SPOT 6/7 imagery, individual crop boundaries influenced by cloud and cloud shadow were manually removed before further analysis; whilst for sentinel 2 imagery the F-Mask cloud masking algorithm was adapted to automatically create cloud and shadow masks http://pythonfmask.org. The mill boundary layers were used to extract the GNDVI values from each of the satellite images.

$$GNDVI = \frac{R_{NIR}-R_{Green}}{R_{NIR}+R_{Green}} \quad \text{Eq. 1}$$

For both the Sentinel-2 imagery and SPOT 6/7 imagery the K-Means unsupervised classification method was used to classify the GNDVI into eight classes of low to high GNDVI across each region (i.e. eight classes spread across all blocks) and also to classify each individual block into eight classes (i.e. eight classes within each block). These sugarcane variability maps were then provided to mills and growers in both an electronic format i.e. 2D map and as an image file for display in a GIS system.

Using the GIS block boundaries, the mean GNDVI values from each SPOT image were extracted for each region. Previously developed exponential equations relating yield to GNDVI for each region were applied to predict the regional yield in TCH (Robson *et al*., 2012).

The Landsat time series processing is depicted in Figure 2 and is further described by Rahman *et al*., (2017) and Rahman and Robson (2016).

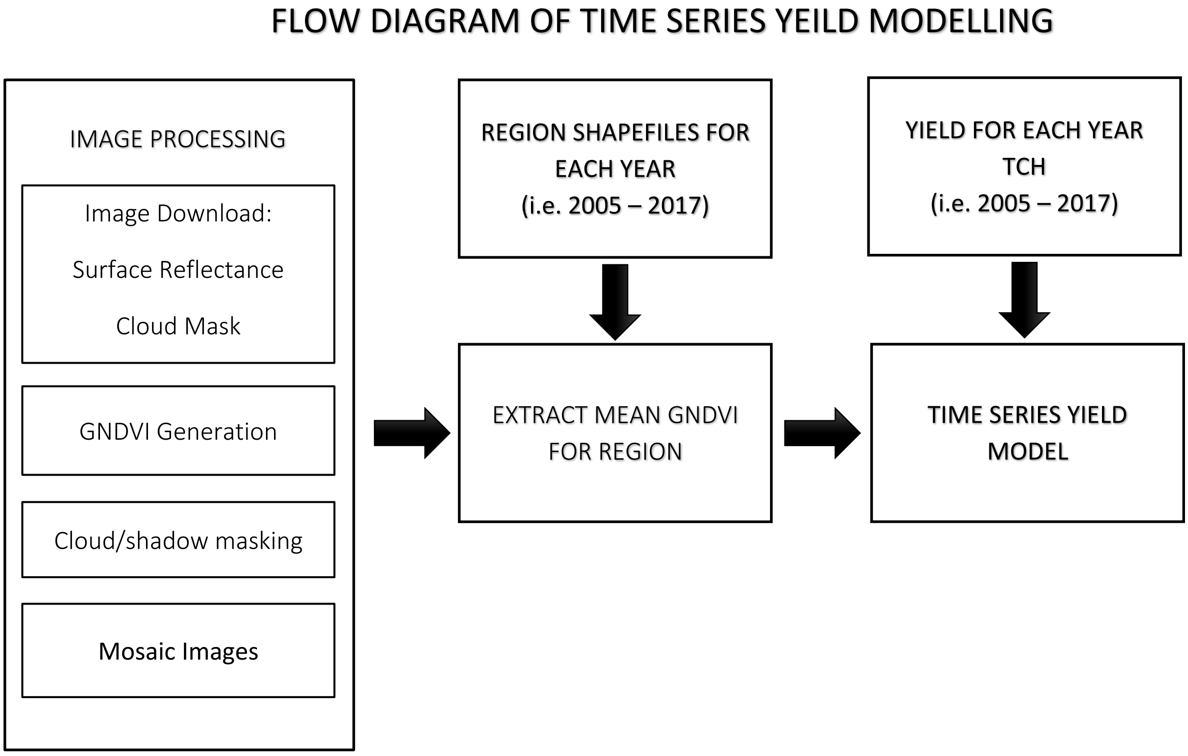


Fig. 2— Landsat time series yield prediction workflow.

GNDVI images were created from each Landsat image. Where two Landsat images were required to cover a sugarcane growing region the GNDVI images were mosaicked together to form one image. Masks were applied to remove cloud and cloud shadow areas from further analysis. GIS block boundaries provided for each growing season were used to extract the mean GNDVI from each masked Landsat GNDVI image acquired during that corresponding season. All extracted GNDVI values were then plotted against the 'season day' (between 1 November to 30 June) and a polynomial fitted to the data for all years. This polynomial form was then "vertically shifted" to match the GNDVI points for each individual year, and the maximum GNDVI derived from the "vertically shifted" polynomial for each year was then regressed against the corresponding annual regional yield. For the prediction of the 2017 seasonal yield, a Landsat image was acquired early in the growing season (i.e. February) and the maximum GNDVI calculated using the polynomial relationship derived from the 'time series' analysis. The maximum GNDVI value was then converted to predict average regional yield by using the linear algorithm produced from the historic relationship between annual maximum GNDVI and average yield.

In order to validate the accuracies of both the single date SPOT yield forecast and that achieved from the Landsat time series, measures of 'average yield to date' were sourced from each respective mill. Note that as harvesting was not yet completed for most sugarcane growing regions, 'average yield to date' takes into account cane that had already been milled (around 90% of the regional harvest), and an estimate of what had not yet been harvested.

Results and Discussion

The development of the single date SPOT imagery yield algorithm is an ongoing process where the relationship between the yield achieved by every individual crop and its corresponding GNDVI value for each growing region is added to previous year's data. Whilst this relationship for some regions such as Bundaberg has been found to be relatively stable across many seasons, even under extreme weather events (Figure 3), other regions are quite variable. Variations in this annual relationship are likely driven by seasonal climate variability, timing of imagery capture and the satellite image 'look angle' at the time of capture. For the latter, thresholds of <15 degrees are specified. However, greater values can occur, such as during the 2016 season where a number of SPOT images had look angles over 30 degrees, resulting in very low GNDVI values and therefore a low value for predicted yield. This stability over time has obvious implications on the accuracies of predictions in current growing seasons and as such was why the alternative time series models were developed.

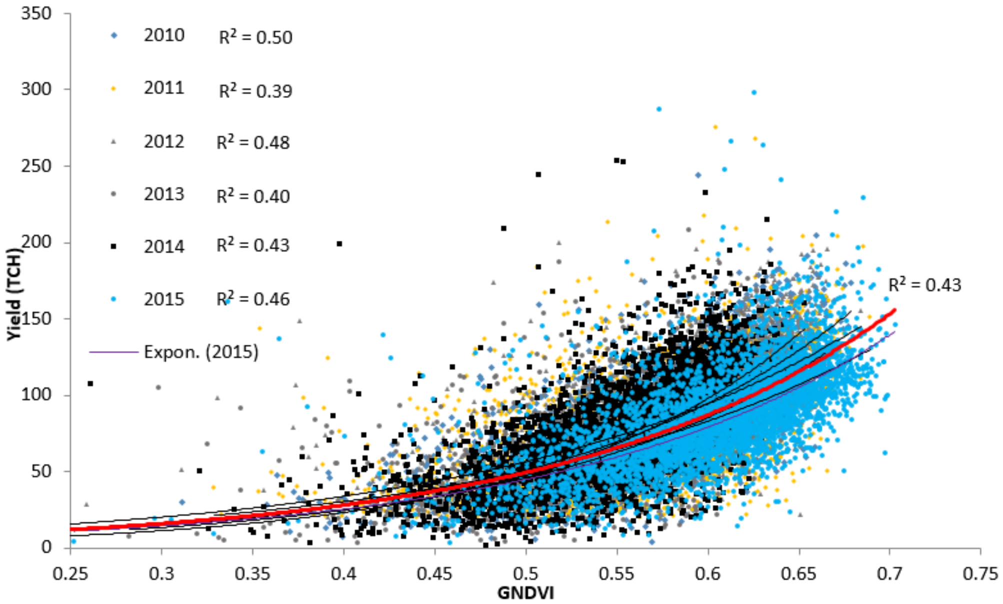


Fig. 3— Spot 5/6/7 GNDVI vs block level TCH for each year and the exponential model fitted to the data with the R^2 values. The thick red line shows the model using data for all years. Example shown is for the Bundaberg region.

Time series models were developed for eight growing regions including Bundaberg, Burdekin, Herbert, Mackay, and Tully as well as two models (1 and 2 year cane) for each of the three NSW growing regions Condong, Broadwater and Harwood. Figure 4a provides an example of a Landsat time series GNDVI model and depicts the annual average growth trend of all crops within the Tully growing region over a 10 year period (2007-2016). Each point represents the average regional crop GNDVI value extracted from a Landsat image, with 70 images used to develop this relationship. The strong quadratic relationship (R^2= 0.8) implies that the underlying growth of crops is similar across years within this region, particularly up until May, when the points start to exhibit more variation around the regression line. The development of this time series model not only serves as a benchmark for future seasons i.e. identify how a current season is progressing in comparison with previous years but allows any major deviations, such as those associated with severe weather events or biosecurity outbreaks to be quickly identified. Figure 4b identifies the strong relationship achieved between the model derived maximum GNDVI for each growing season versus the average yield achieved (R^2 = 0.89). For the forecasting of yield within a current season, the quadratic equation is vertically shifted so the GNDVI value for the current season sits on the quadratic line. The maximum GNDVI can be calculated from the “vertically shifted” quadratic equation presented in Figure 4a, as early as January or February, and then converted to a predicted average regional yield using the linear algorithm presented in Figure 4b. Further discussion and explanation of this model development is given in Rahman and Robson (2016).

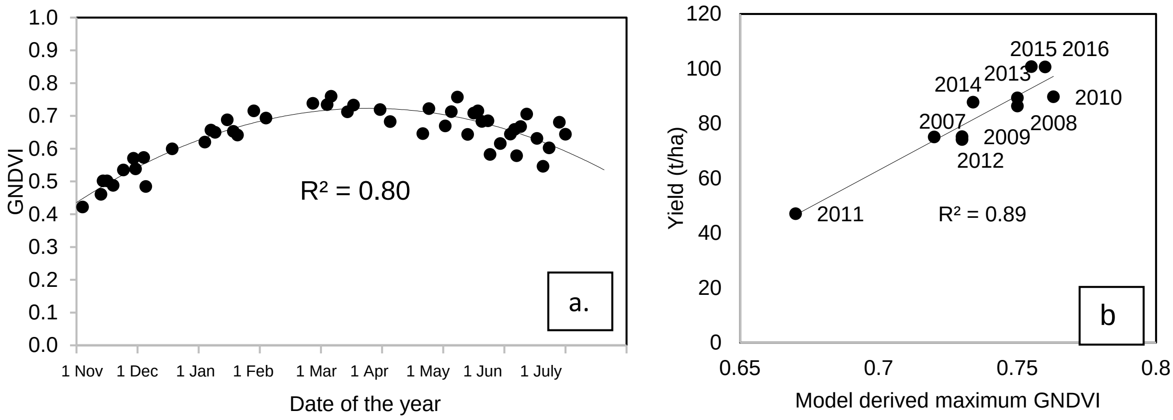


Fig. 4—a. Landsat time series GNDVI using a fitted 2nd order polynomial using all available image dates with less than 50% cloud cover for the Tully region. b. relationship between maximum model derived GNDVI and associated TCH for each year 2005 to 2016.

Figure 5 presents the prediction accuracies of annual regional yield for eight growing regions (2017), using both the single capture SPOT and the Landsat time series methods. It should be noted that the reported mill harvest in TCH is only 'harvest to date', as the 2017 harvest had yet to be completed in many regions. The single date SPOT prediction was not completed for either Tully or Mackay due to insufficient historical imagery available to develop the model. In future years this data will be sourced and the single date algorithm will be created for these regions. In addition, the Herbert single date SPOT prediction was only based on the southern half of the region, due to persistent cloud preventing a full capture of the region during the critical period of peak growth.

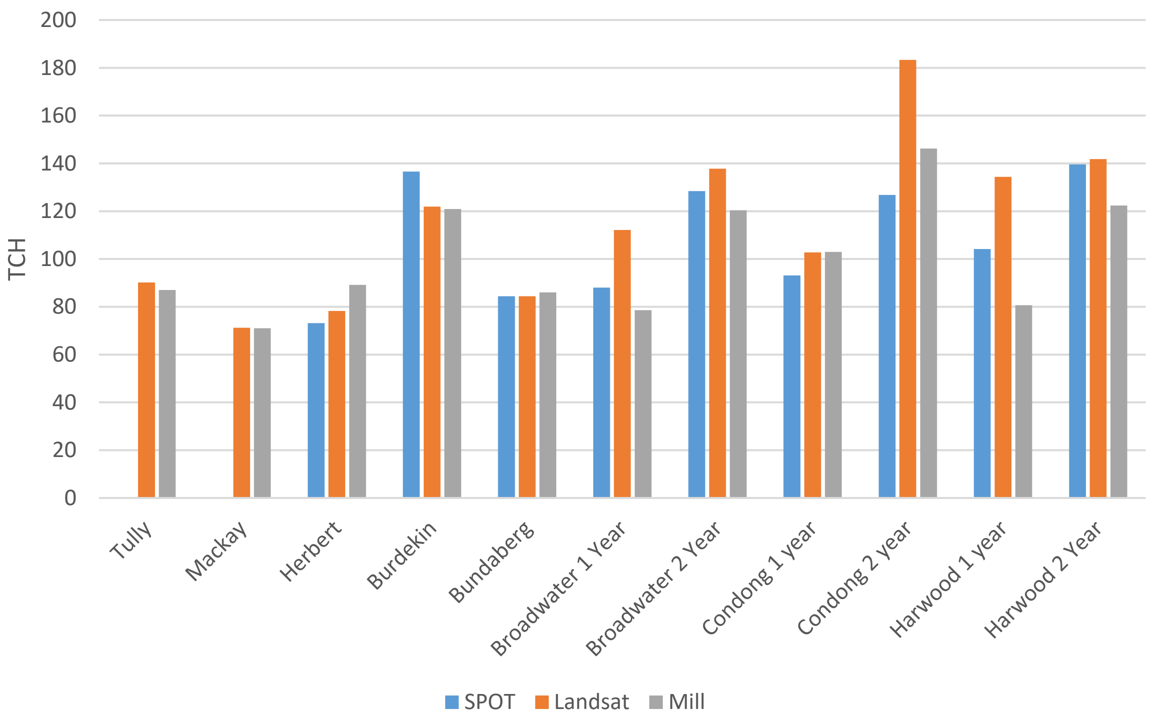
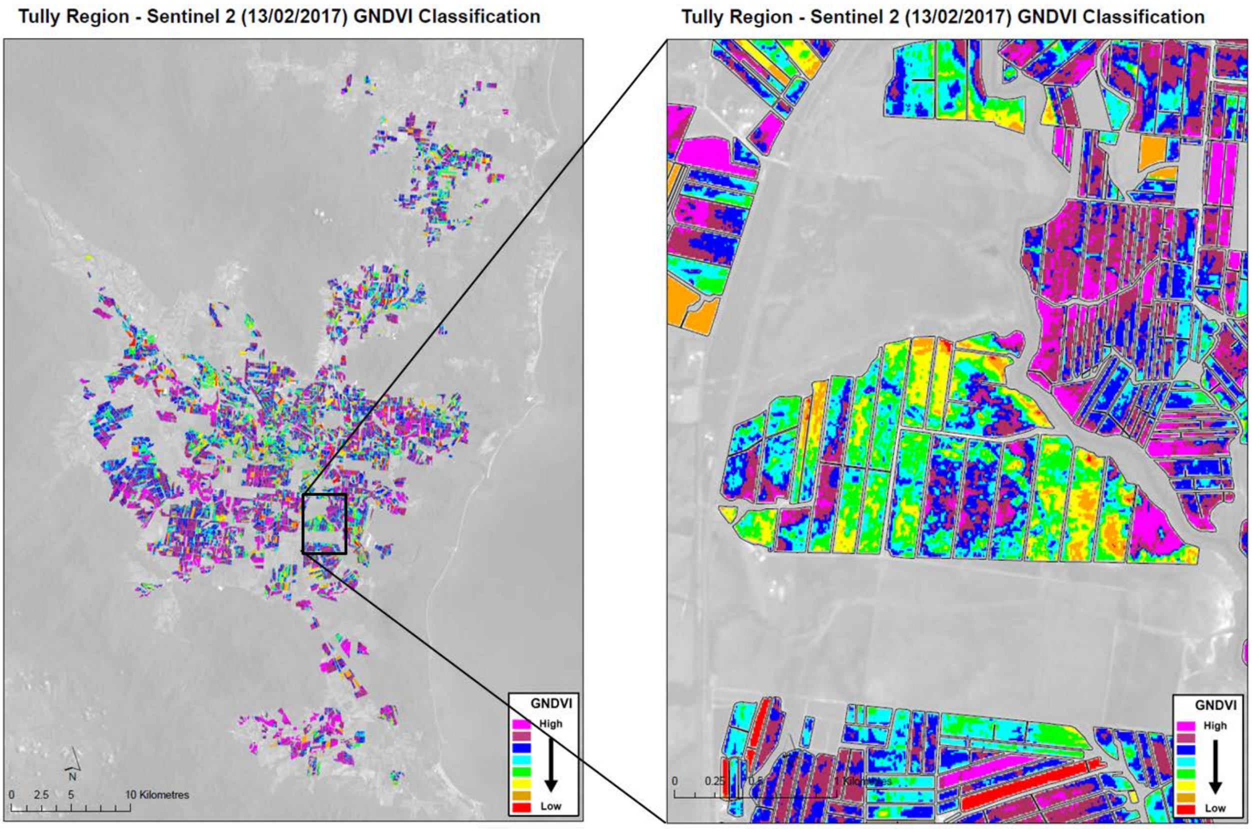


Fig. 5— Comparison of TCH predicted from SPOT 6/7 single date imagery, Landsat time series, and actual harvested TCH to date (data sourced from the mills).

For all regions other than the NSW sugarcane growing regions (Broadwater, Condong and Harwood) the time series models were found to be more accurate in forecasting average regional yield than the single date models derived from SPOT imagery. The lower accuracy of the Landsat time series models for NSW may be due to the complexity of the 1-year and 2-year farming system. As such, further investigation of the NSW datasets will be conducted in an attempt to improve the models.

As well as the forecast of annual yield, a number of mapping products were generated from both SPOT and Sentinel-2 imagery. Examples of K-Means eight class crop variability mapping, ranking GNDVI from low to high, derived from Sentinel-2 satellite imagery at both the regional and block level are shown in Figure 6. Regional level maps (Figure 6 (top)) compare the performance (vigor and yield) of all crops within the growing region prior to harvest. This provides mills with a clear representation of those sub-regions that are performing better than others and can assist with harvest scheduling or identifying the incidence of pest, disease or isolated flooding events. The block level variability mapping (Figure 6 (bottom)) offers greater benefit at the grower level as it clearly shows the variability in crop vigor (derived yield) at the farm and sub block level. With this information growers can better understand the spatial variation that is occurring within their crops and as such can implement targeted agronomy and possibly variable rate management to minimise inputs and maximise production. The image product is supplied to growers or agronomists as a "Geotiff" or "ASCII" file format, common input formats for farm machinery software used in precision agriculture applications, i.e. the block classification image can be loaded onto tractor software and used to specify rates of nitrogen application.



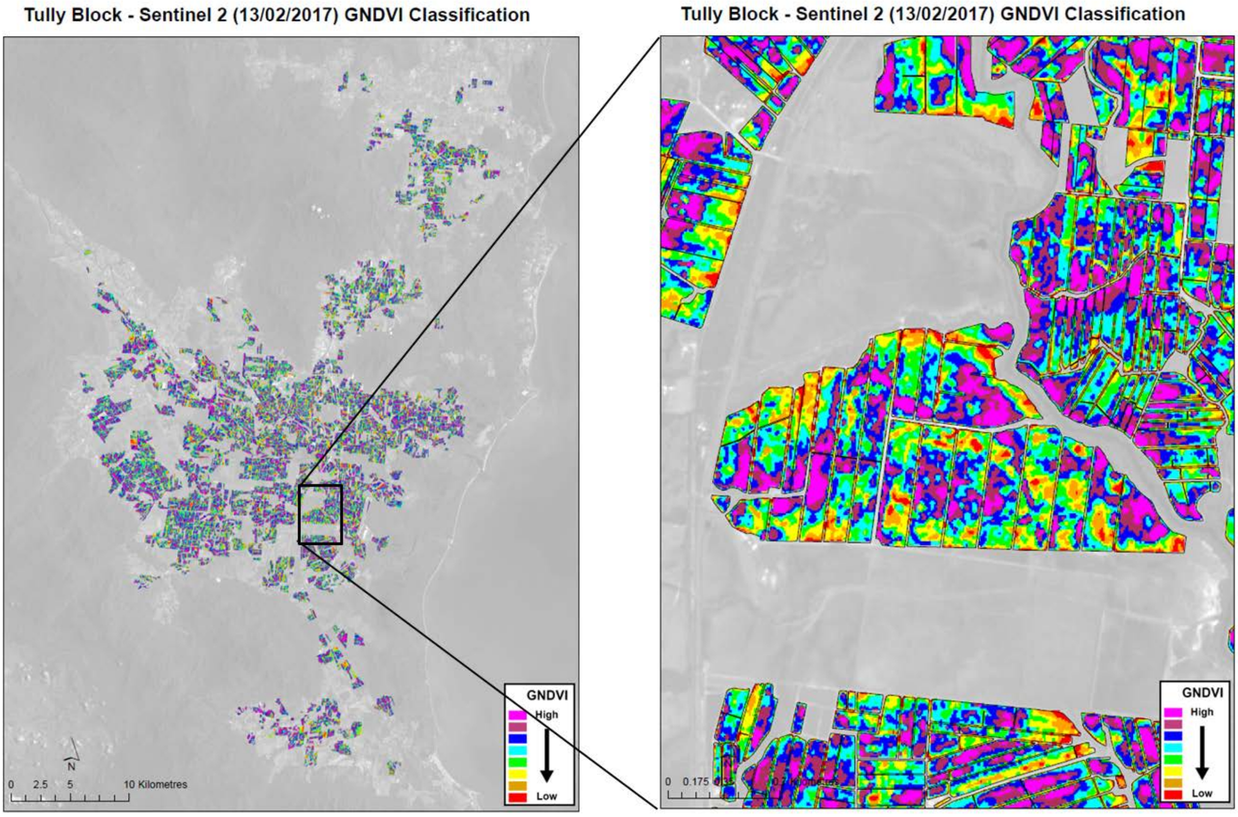


Fig. 6— Example of Sentinel-2 K-Means classification applied across all crops within a growing region (top) and to individual blocks (bottom) for the Tully sugarcane growing region. The entire region is shown (left) for each classification, as well as a zoomed in version to show the difference between the “regional” and “block” classifications.

An important aspect of our research is the dissemination of the derived mapping and imagery products. Currently a commercialisation plan is being developed that will support the automation and delivery of image products to all levels of industry.

Conclusions

These methods offer significant benefit to many levels of the Australian sugar industry. Accurate regional yield forecasts can assist mills with harvest scheduling and forward selling decisions; whilst the mapping outputs provide a strong understanding of the spatial and temporal variation in crop performance at the regional to individual crop level. These maps can be used as a tool for understanding yield variability, as well as for identifying a wide range of abiotic and biotic constraints on production.

Acknowledgements

We wish to acknowledge that this research is fully funded by Sugar Research Australia. We have also received support from Neil Flood through the Joint Remote Sensing Research Program to develop satellite image processing methods.

REFERENCES

Abdel-Rahman EM, Ahmed FB (2008) The application of remote sensing techniques to sugarcane (*Saccharum* spp. hybrid) production: a review of the literature. *International Journal of Remote Sensing* 29, 3753–3767.

Bolton DK, Friedl MA (2013) Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agricultural and Forest Meteorology* 173, 74–84.

Fernandes JL, Rocha JV, Lamparelli RAC (2011) Sugarcane yield estimates using time series analysis of spot vegetation images. *Scientia Agricola*, 68, 139–146.

Morel J, Todoroff P, Bégué A, Bury A, Martiné JF, Petit M (2014) Toward a Satellite-Based System of Sugarcane Yield Estimation and Forecasting in Smallholder Farming Conditions: A Case Study on Reunion Island. *Remote Sensing*, 6, 6620–6635.

Rahman MM, Muir JS, Robson AJ (2017) Simple Multi-Temporal Landsat Algorithms for the Yield Prediction of Sugarcane Crops in Australia, in: *Proceedings of the 7th Asian-Australasian Conference on Precision Agriculture*. Presented at the 7th Asian-Australasian Conference on Precision Agriculture, Hamilton, New Zealand.

Rahman MM, Robson AJ (2016) A Novel Approach for Sugarcane Yield Prediction Using Landsat Time Series Imagery: A Case Study on Bundaberg Region. *Advances in Remote Sensing*, 05, 93–102.

Robson AJ, Abbott C, Lamb D, Bramley R (2012) Developing sugarcane yield prediction algorithms from satellite imagery. *Proceedings of the Australian Society of Sugarcane Technologist*, 34.