Fine scale mapping of fractional tree canopy cover to support river basin management

Sicong Gao[1], Pascal Castellazzi[1], and Tanya Doody[1]

[1]CSIRO

November 9, 2020

Abstract

Management of water, regionally, nationally and globally will continue to be a priority and complex undertaking. In riverine systems, biotic components like flora and fauna, play critical roles in filtering water so it is available for human use and consumption. Preservation of ecosystems and associated ecosystem functions is therefore vital. In highly regulated large river basins, natural ecosystems are often supported through provision of environmental flows. Flow delivery, however, should be underpinned by rigorous monitoring to identify and prioritise biotic water requirements. Broadscale monitoring solutions are thus integral and for woody tree vegetation species, this is can be via measurement of field evapotranspiration, regionally scaled using remote sensing. However, as there is generally a mismatch between field data collection area and remote sensing pixel size, new methods are required to proportion tree evapotranspiration based on tree fractional canopy area per pixel. Within, we present a novel method to derive tree fractional canopy cover (FTCC) at 20 m resolution, in semi-arid and arid floodplain areas. The method employs LiDAR as a canopy area field measurement proxy (10 m resolution). Sentinel-1 and Sentinel-2, radar and multispectral imagery, were used in Random forest analysis, undertaken to develop a predictive FTCC model trained using LiDAR for two regions in the Murray-Darling Basin. A predictor model, combing the results of both regions, was able to explain between 85-91% of FTCC variation when compared to LiDAR FTCC, output in 10% increments. Development of this method underpins the advancement of woody vegetation monitoring to inform environmental flow management in the Murray-Darling Basin. The method and fine scale outputs will also be of value to other catchment management concerns such as altered catchment water yields related to bushfires and as such, has application to water management worldwide.

Sicong Gao[1,2]*, Pascal Castellazzi[2,3], Tanya M Doody[2]

[1] Centre for Applied Water Science, University of Canberra, Canberra, Australia

[2] CSIRO Land and Water, PMB 2, Glen Osmond, SA 5064, Australia

[3] Deep Earth Imaging FSP, Land and Water, CSIRO, Glen Osmond, SA 5064, Australia

* corresponding author: steve.gao@csiro.au

Abstract

Management of water, regionally, nationally and globally will continue to be a priority and complex undertaking. In riverine systems, biotic components like flora and fauna, play critical roles in filtering water so it is available for human use and consumption. Preservation of ecosystems and associated ecosystem functions is therefore vital. In highly regulated large river basins, natural ecosystems are often supported through provision of environmental flows. Flow delivery, however, should be underpinned by rigorous monitoring to identify and prioritise biotic water requirements. Broadscale monitoring solutions are thus integral and for woody tree vegetation species, this is can be via measurement of field evapotranspiration, regionally

scaled using remote sensing. However, as there is generally a mismatch between field data collection area and remote sensing pixel size, new methods are required to proportion tree evapotranspiration based on tree fractional canopy area per pixel. Within, we present a novel method to derive tree fractional canopy cover (FTCC) at 20 m resolution, in semi-arid and arid floodplain areas. The method employs LiDAR as a canopy area field measurement proxy (10 m resolution). Sentinel-1 and Sentinel-2, radar and multispectral imagery, were used in Random forest analysis, undertaken to develop a predictive FTCC model trained using LiDAR for two regions in the Murray-Darling Basin. A predictor model, combing the results of both regions, was able to explain between 85-91% of FTCC variation when compared to LiDAR FTCC, output in 10% increments. Development of this method underpins the advancement of woody vegetation monitoring to inform environmental flow management in the Murray-Darling Basin. The method and fine scale outputs will also be of value to other catchment management concerns such as altered catchment water yields related to bushfires and as such, has application to water management worldwide.

Key words

River Red Gum, radar, Sentinel, Murray-Darling Basin, catchment water management, evapotranspiration, vegetation, SAR

Introduction

River basin management continues to challenge humanity worldwide in finding a balance between anthropogenic hydrological water needs and those of the environment (Hoekstra *et al.* , 2012; Wheeler *et al.* , 2017; Bouckaert *et al.* , 2018). Both surface water and groundwater resources continue to diminish via over abstraction and declining rainfall across many countries, resulting in poor riverine ecological function and environmental condition (Vörösmarty *et al.* , 2010; Rolls and Bond, 2017; Bouckaert *et al.* , 2018). To arrest environmental decline in Australia's food-bowl, the Murray-Darling Basin (MDB; Figure 1), the Murray-Darling Basin Plan was developed to ensure sustainable water use while supporting reliant industries and the environment (MDBA, 2009).

Recognising the MDB will remain a highly managed system, central to the Murray-Darling Basin Plan is water for the environment to support and where possible, restore ecological condition to achieve long-term environmental outcomes in the absence of natural flows (MDBA, 2009). However, prioritising when and where environmental water is required across the MDB is extremely complex and to a large extent, relies on*in-situ* monitoring of ecological assets such as fish, birds and both woody and non-woody vegetation. As the MDB covers ~14% of Australia's land area (or 1,000,000 km^2), reliance on *in-situ* monitoring is not only costly, it impedes the ability to understand large-scale ecological water requirements and ecological responses. Remote sensing methods, however, can provide robust broad scale monitoring options, underpinned by *in-situ* observations.

Over the last decade, it has become evident that field measurement of water loss from floodplain woody vegetation via transpiration and evapotranspiration (ET), provides a way to observe forest or woodland ecological condition and water needs (Doody *et al.* , 2015; Jarchow*et al.* , 2017; Wallace *et al* ., 2019). When tree water requirements are not met, both transpiration and ET decline in response to tree water saving adaptations such as stomatal closure to prevent xylem cavitation and tree death (Baird *et al.* , 2005; Doody*et al.* , 2009, 2015). Severity of decline is related to extent of continued water deficit/drought. Broadscale monitoring of woody vegetation transpiration and ET during drought can highlight tree condition decline trajectories and inform prioritisation of environmental water over both short and longer timeframes. Similarly, monitoring tree response to increasing water availability is equally important in a water management context related to decisions around where and when environmental water should be delivered.

Considerable investigation has demonstrated that remote sensing provides a broadscale means to monitor woody vegetation water loss via ET (Guerschman *et al.* , 2009; Glenn *et al.* , 2011; Mu *et al.* , 2011). In particular, MODIS imagery, providing 8-day composite greenness products (normalised difference vegetation index (NDVI) and enhanced vegetation index (EVI), provides excellent temporal resolution from the year 2000 from which to develop vegetation monitoring solutions (Glenn *et al.* , 2011; Nagler *et al.* , 2016). Evapotranspiration algorithms, such as those developed by Mu (2013), Guerschman (2009) and Nagler (2016), lend themselves to broadscale ecological monitoring. However, the MODIS spatial scale of 250 m can be limiting when it comes to calibrating remote sensing outputs with *in-situ* data collected over smaller scales, as often happens. For example, in the MDB, collection of woody vegetation ET typically occurs using 50 x 50 m plots, with highly heterogenous canopy cover (Doody *et al.* , 2015).

To augment development of temporal fine scale remote sensing monitoring methods which are applicable to ecological vegetation systems worldwide, a method to downscale low spatial resolution ET estimates based on high spatial resolution FTCC maps requires development. Improved downscaling will reduce the risk of over or underestimating vegetation ET during the field calibration process and improve large-scale remote sensing ET estimates. While several fractional vegetation cover products exist in Australia (Guerschman *et al.* , 2015; Guerschman and Hill, 2018; DEE, 2019), the information they impart is not suitable to downscale ET, due to lack of fine scale classification of vegetation cover or discrimination between groundcover vegetation and trees.

The objective of this study, therefore, was to develop a method to determine fractional tree canopy cover (FTCC) at a pixel resolution of 20 m. The intent was to provide a fine scale evaluation of the proportion of tree canopy in each MODIS pixel and field measurement area. Understanding the amount of remote sensing ET related directly to tree ET (from field measurement) allows improved calibration accuracy between the two. This in turn, improves broadscale ET estimates used for monitoring, by accounting for heterogenous canopy cover in each pixel. The innovative method reported, combines radar (Sentinel-1) and multi-spectral (Sentinel-2) imagery for the first time, to estimate FTCC. Imagery based on LiDAR (Light Detection and Ranging) data was employed as a surrogate for field canopy cover and used to train and build a model that can estimate FTCC across large areas of the MDB. Given the paucity of methods to identify and map canopy cover at fine scales, the research presented within is likely to be important to many aspects of environmental management and hydrology, specifically catchment water management and improving our understanding of the underlying hydrological processes related to vegetation presence and absence.

Study area

Two floodplain regions in the southern MDB, Yanga National Park and Barmah National Park (Figure 1), were selected for this investigation based on local knowledge and availability of LiDAR remote sensing for training data. Yanga National Park (34°27'S, 143°48'E) forms a large component of the Lower Murrumbidgee floodplain which lies along a section of the Murrumbidgee River in New South Wales (Doody *et al.* , 2015). Barmah National Park (36°00'S, 144°56'E), is located to the south-east of Yanga in the Barmah-Millewa Forest along the River Murray. Both floodplain systems are composed of the native riparian tree species, *Eucalyptus camaldulensis* (River Red Gum) as well as *E. largiflorens* (Black Box). Notably, Barmah National Park is home to the largest *E. camaldulesis* forest in the world and is thus recognised under the Ramsar Convention (Hale and Butcher, 2011).

Both regions are semi-arid with low annual precipitation with ~425 and 323 mm year^{-1} at Barmah and Yanga respectively (Bureau of Meteorology, 2020), and high evaporative demand (~1600 mm year^{-1}; The Long Paddock, Queensland Government). Austral summers are hot with mean maximum temperature between 31-33^{O}C, while winter is cool with mean maximum temperature of 14-16 ^{O}C (Bureau of Meteorology, 2020).

Method

A 6 x 6 km region of interest (ROI) was selected within each floodplain area to reasonably manage large data volumes. In summary, remote sensing imagery, including airborne LiDAR, Sentinel-1 and Sentinel-2, was obtained to develop a method to derive FTCC. LiDAR data was used to provide a three-dimensional representation of the ROI's and derive high-resolution FTCC data. Representing *in-situ* data, the LiDAR derived FTCC was used to calibrate a Random forest model based on Sentinel-1 and -2. As the Sentinel dataset is open-access, freely accessible and available globally, the technique can be implemented over regional or continental scales if training data (direct *in-situ*FTCC measurement, or as here, a LiDAR surrogate) are available.

LiDAR data

LiDAR (Light Detection And Ranging) remote sensing uses pulsed light waves from an airborne laser to measure distance of Earth objects from an aircraft via reflectance of light (Dubayah and Drake, 2000). The returned wavelengths and time combined, allow three dimensional representations of the reflected surface to be constructed. When LiDAR is collected over natural environments, 3-D reconstruction of canopy structure provides fine resolution field representation of the study location. Airborne LiDAR data for each ROI was obtained from ELVIS (https://elevation.fsdf.org.au/), a spatial data portal. Each tile covered 2 km $\times$ 2 km. Acquisition date was September 2009 and 2015 for Yanga and Barmah, respectively, and was performed with two different LiDAR sensors (Leica ALS50-II -Yanga; Trimble AX60 - Barmah).

The Yanga dataset was collected 0.50 km above the earth surface with a swath width of 1.6 km and swath overlap of 20%. Similarly, the Barmah dataset was measured at a height of 0.85 km with a swath width of 1 km and swath overlap of 30%. Sensors recorded an average point spacing of around 4.0 and 4.4 points per m^2 for Yanga and Barmah, respectively.

Retrieving vegetation height and fractional tree canopy cover from LiDAR data

Tree structural information for both ROIs was retrieved from LiDAR tile data. Each tile includes a dense collection of 'points' based on reflectance time and georeferencing information, such as x and y coordinates, point heights and point return 'types' (related to time each point returns to the sensor and height of the object). FUSION software (http://forsys.cfr.washington.edu/fusion.html) was implemented to partition a digital surface model and digital terrain model from the raw LiDAR data (Boehm *et al.* , 2013). The digital surface model was applied to approximate elevation of each grid cell. The digital terrain model was used to estimate elevation of the ground surface. A canopy height model (Koukoulas and Blackburn, 2005) was created by subtracting the digital terrain model from the digital surface model at 1 m spatial resolution. The canopy height map was then converted from point clouds to pixels. A FTCC product was derived from the canopy height model using all LiDAR points reflected from 2 m above the ground surface (referred to as LiDAR FTCC). As the objective of the study was to map tree canopy cover, smaller shrubs and bushes were excluded (Equation 1). R package ForestTools was applied to identify dominant treetops and tree crown radius from the canopy height model. A moving window was created to scan the canopy height model and tag treetops that depended on the highest point in the window. The 'watershed' method was implemented to outline tree crowns (Beucher and Meyer, 1993). Finally, from the canopy height model, tree number was counted as well as the tree height and crown radius.

$FTCC = \frac{numbers\ of\ pixels\ (height>2m)}{\text{total pixel numbers at given area}}$(Equation 1.)

Sentinel-1 data

Sentinel-1A and 1B satellites carry C-band Synthetic Aperture Radar (SAR) sensors. They are part of the European Space Agency's Copernicus mission, and were launched in 2014 (Sentinel-1A) and 2016 (Sentinel-1B). They are the first globally acquiring SAR sensors, providing dual-polarized (VV and VH) C-band SAR images with a 12-day repeat path frequency. Over land, Interferometric Wide imaging mode is the default automatic imaging mode, with a nominal sensing resolution of 20 (Azimuth) by 5 m (Range).

A Sentinel-1 Ground Range Detected image acquired in May 2016 was obtained from Sentinel Australia Regional Access (SASA; https://copernicus.nci.org.au/). Processing was performed with the Sentinel Application Platform (SNAP) and included updating the orbital metadata, thermal noise removal, border noise removal, calibration, range doppler terrain correction and conversion to decibel (Filipponi, 2019). VV and VH bands were converted to Sigma Nought backscattering coefficients, which includes a compensation for Line-Of-Sight variations in Range.

Sentinel-2 data

The Sentinel-2 satellites consist of two satellites, launched in 2015 (Sentinel-2A) and 2017 (Sentinel-2B), respectively. Each carry multispectral sensors with 13 spectral bands recording visible, near-infrared and short-wave infrared regions of the electro-magnetic wave spectrum. The revisit time of Sentinel-2 is 10 days.

Sentinel-2 Level 1C (L1C) top-of-atmosphere data with less than 10% cloud cover, collected in May 2016, was downloaded from SASA. The original tile (100 km × 100 km) was cropped to the ROIs. Sen2cor was applied to obtain bottom of atmosphere reflectance, converting data from L1C to atmospherically corrected L2A (Main-Knorn *et al.* , 2015). Ten bands were selected and these represent vegetation functional and structural information (Verrelst *et al.* , 2012). Bands include B2-B8, B8a, and B11-B12 from Sentinel-2. All bands where relevant, were resampled to 20 m x 20 m (Table 1) .

Random forest regression analysis

Random forest regression, proposed by Breiman (2001), is an assembling machine learning algorithm that can be applied to high-dimensional spatial dataset analysis. Random forest starts with a random selection of subset data from a training dataset, then creates decision trees for each sample. A 'voting' method is then implemented for the prediction of each decision tree. The most voted prediction is selected as the final result among all individual decision trees (Gislason *et al.* , 2006).

Random forest regression was employed to determine the relationship between LiDAR FTCC and Sentinel-1 and Sentinel-2 bands. Before applying the Random forest regression, Sentinel-1 and 2 bands and the canopy height model were resampled to the same spatial resolution of 20 m. VV and VH bands from Sentinel-1 were resampled to 20 m based on Sentinel-2 image resolution using bilinear interpolation. In order to retrieve FTCC from the canopy height model at Sentinel-2 spatial resolution, a 20 m fishnet grid was created. FTCC was calculated based on equation 1 from the canopy height model for each fishnet grid. With resampling of Sentinel-1, Sentinel-2 and LiDAR FTCC, 733,800 pixels for both Yanga and Barmah ROIs were created for Random forest training and validation.

Three models were created using 'randomForest' (R package) which included single models trained and predicated for both the Yanga and Barmah ROIs (RF_{Yanga} and RF_{Barmah} where RF is Random forest) and a model that combined data from both ROIs (RF_{all}). For each model, the dataset was split into 70% training and 30% validation by random sampling. A ten-fold cross-validation was implemented to keep the best performance of each Random forest model.

Statistical analysis

Root Mean Square Error (RMSE) was applied to analyse the performance of the Random forest predictor model. The RMSE is defined as;

$$RMSE = \frac{\sqrt{\sum_{i=1}^{n} (x_{ret,i} - x_{pre,i})^2}}{N} \text{(Equation 2.)}$$

The $x_{ret,i}$ and $x_{pre,i}$ are LiDAR FTCC and FTCC predicted by the Random forest model, respectively. N is the number of pixels used for prediction. The coefficient of determination was applied to check the relationship between LiDAR FTCC and predicted FTCC for the ROIs. Hence, higher R^2 indicates the regression model fits the LiDAR FTCC, and lower RMSE indicates better predictions of the Random forest models. Data processing, statistical analysis and visualisation were conducted in R scientific computation environment (R core team version 3.6) and associated packages obtained from the comprehensive R archive network (http://cran.r-project.orj).

Results

Assessment of Random forest models

For the Yanga ROI, there was a moderate correlation between LiDAR FTCC and predicted FTCC (Figure 2a; RF_{Yanga}), with the Random forest model underpredicting. The correlation was higher (0.8, p-value < 0.01) at Barmah (Figure 2b; RF_{Barmah}). The model trained using both Barmah and Yanga data (RF_{all}), derived significantly improved results, exhibiting a strong correlation between LiDAR FTCC and predicated FTCC for both ROI's (Figure 2c and d). The RF_{all} predictor model could explain 85% of the FTCC variation for Yanga and 91% for Barmah. The RMSE indicates the accuracy of RF_{all} was higher than RF_{Yanga} and RF_{Barmah}, strongly demonstrating the combined model (RF_{all}) contains more information than RF_{Yanga} and RF_{Barmah} individually.

Prediction of the fractional tree canopy cover for Yanga and Barmah

The percentage of FTCC per 20 m pixel was predicted for both Yanga (Figure 3a) and Barmah (Figure 3d) ROIs. High FTCC (> 65%) was found along the Murrumbidgee River and around the periphery of 'Irrigation Lake' at Yanga, in both the predicted and LiDAR FTCC images (Figure 3a and b). Low FTCC can be observed predominantly in the south of the ROI related to bare land and/or low shrubs such as*Muehlenbeckia florulenta* (Lignum). The patterns of predicted FTCC and LiDAR FTCC generally showed high consistency, although areas indicated within the red boxes suggest errors (Figure 3c). These are likely related to a mismatch between the spatial resolution of the training data (20 m) and LiDAR data (10 m).

Within the Barmah ROI, average FTCC was much higher across the ROI (72%) compared to Yanga (26%) which is not surprising given the ROI is in the heart of a forested area. Similar to Yanga, predicted FTCC (Figure 3d) displayed similar patterns to LiDAR FTCC (Figure 3e) when predicted using the RF_{all} model. Measurement errors were detected in the LiDAR data (red areas; Figure 3e) which explains the missing section of river channel in that figure and it's prominence as an error (yellow pixels) in Figure 3f, which is not an error of prediction.

Relative importance of remote sensing products in the predictive model

One of the most efficient functions of the Random forest model is to reveal the relative importance of imaging techniques and bands (spectral for multi-spectral or polarization for SAR) contributing to the developed predictive models (Fassnacht *et al.* , 2014). Sentinel-2 band B12, a shortwave infrared band with a central wavelength of 2202.4 nm (Vaudour *et al.* , 2019) proved most significant to predict FTCC across the three independent predictor models (Figure 4). Prior studies note the importance of SWIR with respect to leaf water content (Tucker, 1980; Han *et al.* , 2019) and remote sensing. Band B11, another SWIR band, contributed an average of 51% to the final result (Figure 4), while bands B2, 3, and 4 were all vital to the RF_{all}model, with an average importance of 46%. The contribution of Sentinel-1 bands was less important than Sentinel-2 bands. However, of note is the 'cross-polarized' VH band, which contains information on complex volume scattering (a typically dominant radar scattering mechanism in tree canopy covers) which is the most information-rich of the two SAR bands.

SWIR bands are sensitive to variation in leaf area index and leaf water content (Asner and Lobell, 2000; Ghulam *et al.* , 2008). SWIR can account for up to 89% of leaf area index variation based on simulations of a radiative transfer model (Bowyer and Danson, 2004; Wang *et al.* , 2008). In addition, prior studies have noted the importance of leaf water thickness variations which are strongly presented in SWIR bands due to the low absorption of light by water (Asner and Lobell, 2000; Ghulam *et al.* , 2008). Hence, the obvious difference of leaf area index and leaf water content for canopy and soil contributes significantly in Random forest model training.

Considering the lower ranking of the Sentinel-1 bands, the RF_{all} model was trained without Sentinel-1 bands. Correlations were weaker than the previous model at Yanga (R^2 = 0.84, p-value < 0.01) and Barmah (R^2 = 0.86, p-value < 0.01), respectively (data not shown). The results suggest that Sentinel-1 bands play an important role in the Random forest model training to accurately derive FTCC.

Discussion

Understanding vegetation water requirements and losses are important to inform environmental water management and underpin equitable water sharing plans. Given the significant advances in digital technology and high costs of *in-situ* monitoring, new innovative cost-effective methods are vital to monitor large land tracts both in Australia and other regions across the world (Manfreda *et al.* , 2018). Woody vegetation ET can provide a line of evidence to improve monitoring and inform water management, however downscaling of low spatial resolution data is required to provide robust remotely sensed ET estimates. The performance of the RF_{all} predictor model presented within, indicates that a model has been developed that can accurately predict FTCC for both sparse and densely vegetated areas semi-arid and likely arid, floodplain environments.

As mentioned previously, while other fractional vegetation products are available (Guerschman *et al.* , 2015; Guerschman and Hill, 2018; DEE, 2019) the classification and spatial resolution of these did not suit the purpose of improving remotely sensed ET outputs. Guerschman and Hill (2018), for example, provide landscape fractional cover including percent photosynthetic vegetation, non-photosynthetic vegetation and bare soil across 250 m MODIS pixels. In contrast, the model presented here, provides FTCC in 10% increments of canopy cover related only to trees.

Important outcomes of method development

LiDAR imagery collected from the regions of interest (Yanga and Barmah National Parks) proved invaluable to the development of the reported method. LiDAR provides a proxy for field derived canopy cover, against which Sentinel data was trained. As the LiDAR output is composed of 'point clouds' representing 3-D land surface features, it was possible to separate trees over 2 m in height from other surface features, to provide 'field-based' canopy cover. The results, from a remote sensing perspective, are also important to understand critical bands that are required to monitor vegetation and water to inform future satellite development.

Additional method application

While remote sensing methods can be used to derive FTCC such as aerial imagery (Melville *et al.* , 2019), LiDAR (Wasser *et al.* , 2013) and fine resolution satellite imagery like WorldView2 and 3 (Immitzer *et al.* , 2018), acquiring imagery is costly and requires 'tasking' (i.e. imagery it is not collected regularly and needs to be ordered) for specific areas of interest. As a result, national scale imagery is not available and temporal availability is poor. In comparison, developing a method using open-access Sentinel-1 and -2 imagery, provides a mechanism to monitor vegetation cover change from 2015 and into the future at desired intervals such as monthly, seasonally or annually, depending on the application.

The FTCC method, is however, likely to be very valuable to other areas of catchment water management. The significant bushfires across southern Australia over the summer of 2019/2020 are likely to have significant future impacts on water resources and especially changes to water yield in both quality and quantity over the next decade (Brown, 1972; Lee, 2020; Moreno *et al.* , 2020). The FTCC method would enable accurate estimates of tree area, pre and post bushfires, to underpin future hydrological catchment yield forecasting. Current methods are unlikely to be suitable to disentangle woody tree vegetation, which is a dominant water user, from other vegetation sources. This may lead to errors in water yield estimation pre and post fires. Tree reduction also increases streamflow locally, although this is quickly reversed as regeneration occurs, particularly in Australia with bush tolerant native species (Kuczera, 1987; Brookhouse *et al.* , 2013). There is an opportunity to link broadscale FTCC predictions with modelling of water fluxes through Land Surface Models, enabling modelling to understand the effects of fires (or any land cover changes) on hydrologic fluxes (Barlage and Zeng, 2004; Fang *et al.* , 2018). As severe bushfires have also featured in other areas around the world such as the United States and Europe, the method is relevant internationally.

Sources of error

Sentinel data was trained against LiDAR which was collected between 2009-2015. The very high correlation between LiDAR FTCC and predicted FTCC, provides some confidence that although time has passed, substantial changes to vegetation crowns were not apparent in the trained areas. As vegetation might have changed slightly between the training data (LiDAR) and the covariates (multi-spectral and SAR bands), part of the error is actually not attributable to FTCC modelling, i.e. the discrepancy between the LiDAR FTCC and the predicted FTCC represent a maximum error margin. Yanga LiDAR, collected in 2009, occurred before the break in the Millennium Drought from 1997 to 2009 (Leblanc *et al.* , 2012), while Barmah LiDAR was collected after (2015). This might explain the poorer prediction at Yanga using the $\mathrm{RF}_{\mathrm{Yanga}}$model as substantial improvement in tree canopy crowns occurred over the 2010-2012 flood period (Doody *et al.* , 2015) leading to a discrepancy between amount of crown cover pre and post flood. The match at Barmah was likely higher due to closer match between imagery dates (2015 LiDAR and 2016 Sentinel). While the date gap between Yanga imagery is not ideal, it was suitable for this project, however more recent LiDAR imagery is preferred. Additional sources of error could have been introduced to the model from use of two different LiDAR collecting platforms and differences in their acquisition altitudes as well as the spatial mismatch between 20 m training data and 10 m LiDAR data.

Further research

While the initial method shows considerable promise for widespread application and identification of FTCC, to scale the method across the MDB, additional areas will need to be trained to incorporate vegetation in different climate and especially rainfall zones. It is unclear if regions with higher rainfall will fit the $\mathrm{RF}_{\mathrm{all}}$ model, so further investigation is required. The objective moving forward, is to provide a universal model to predict FTCC across the MDB and examine reducing FTCC resolution further to <10 m. Building further upon that, will be investigation of the feasibility of producing FTCC timeseries over the period of Sentinel availability (˜5 years), focusing on seasonal and annual predictions which will be valuable for monitoring of temporal vegetation canopy cover change at a fine resolution. As mentioned in relation to bushfire and water yield research applications, provision of <10 m woody tree canopy cover would substantially improve vegetation water use estimates based on tree area and aid forecasts of how water yield and hydrologic fluxes (ET, recharge and runoff) will change into the future.

Conclusions

The aim of the reported study was to predict woody vegetation FTCC at 20 m resolution for floodplain vegetation and evaluate predictions using LiDAR data. This study has shown that a combining predictor model was able to explain up to 91% of FTCC variation, returning an acceptable RMSE at our study sites. Individual models ($\mathrm{RF}_{\mathrm{Yanga}}$ and $\mathrm{RF}_{\mathrm{Barmah}}$) displayed weaker correlations and larger errors when compared to the combined model. Analysis of sensor band importance suggests SWIR is the most important band which contributes mostly to model training as it is sensitive to variation in leaf area index and leaf water content. Additionally, Sentinel-1 (radar) band contributions cannot be ignored for Random forest model training. Our presented approach will prove useful in expanding knowledge of remote sensing ET related directly to tree ET, improving estimations at a finer spatial resolution. This study will be significant to further our collective understanding of floodplain vegetation response to climatic conditions and catchment water management.

Acknowledgements

We would like to acknowledge funding contributions from the Murray-Darling Basin Authority which have contributed to this research as well as funding contribution by CSIRO's Deep Earth Imaging Future Science Platform.

References

Asner GP, Lobell DB. 2000. A biogeophysical approach for automated SWIR unmixing of soils and vegetation. *Remote Sensing of Environment* **74** (1): 99–112 DOI: 10.1016/S0034-4257(00)00126-7

Baird KJ, Stromberg JC, Maddock T. 2005. Linking riparian dynamics and groundwater: An ecohydrologic approach to modeling groundwater and riparian vegetation. *Environmental Management* **36** (4): 551–564 DOI: 10.1007/s00267-004-0181-z

Barlage M, Zeng X. 2004. The effects of observed fractional vegetation cover on the land surface climatology of the community land model. *Journal of Hydrometeorology* DOI: 10.1175/1525-7541(2004)005<0823:TEOOFV>2.0.CO;2

Beucher S, Meyer F. 1993. The morphological approach to segmentation: the watershed transformation. *Mathematical morphology in image processing* **34** : 433–481

Boehm HDV, Liesenberg V, Limin SH. 2013. Multi-temporal airborne LiDAR-survey and field measurements of tropical peat swamp forest to monitor changes. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **6** (3): 1524–1530 DOI: 10.1109/JSTARS.2013.2258895

Bouckaert F, Wei Y, Hussey K, Pittock J, Ison R. 2018. Improving the role of river basin organisations in sustainable river basin governance by linking social institutional capacity and basin biophysical capacity. *Current Opinion in Environmental Sustainability* **33** : 70–79 DOI: 10.1016/j.cosust.2018.04.015

Bowyer P, Danson FM. 2004. Sensitivity of spectral reflectance to variation in live fuel moisture content at leaf and canopy level. In *Remote Sensing of Environment* Elsevier Inc.; 297–308. DOI: 10.1016/j.rse.2004.05.020

Breiman L. 2001. Random forests. *Machine learning* **45**(1): 5–32

Brookhouse MT, Farquhar GD, Roderick ML. 2013. The impact of bushfires on water yield from south-east Australia's ash forests. *Water Resources Research* **49** (7): 4493–4505 DOI: 10.1002/wrcr.20351

Brown JAH. 1972. Hydrologic effects of a bushfire in a catchment in south-eastern New South Wales. *Journal of Hydrology* **15**(1): 77–96 DOI: 10.1016/0022-1694(72)90077-7

DEE. 2019. Fractional Cover. *Digital Earth Australia* Available at: https://www.ga.gov.au/dea/products/fc [Accessed 4 November 2020]

Doody TM, Colloff MJ, Davies M, Koul V, Benyon RG, Nagler PL. 2015. Quantifying water requirements of riparian river red gum (Eucalyptus camaldulensis) in the Murray-Darling Basin, Australia - implications for the management of environmental flows. *Ecohydrology* **8** (8): 1471–1487 DOI: 10.1002/eco.1598

Doody TM, Holland KL, Benyon RG, Jolly ID. 2009. Effect of groundwater freshening on riparian vegetation water balance. *Hydrological Processes* **23** (24): 3485–3499 DOI: 10.1002/hyp.7460

Dubayah RO, Drake JB. 2000. Lidar Remote Sensing for Forestry. Oxford Academic. DOI: 10.1093/JOF/98.6.44

Fang L, Zhan X, Hain CR, Liu J. 2018. Impact of Using Near Real-Time Green Vegetation Fraction in Noah Land Surface Model of NOAA NCEP on Numerical Weather Predictions. *Advances in Meteorology***2018** DOI: 10.1155/2018/9256396

Fassnacht FE, Hartig F, Latifi H, Berger C, Hernández J, Corvalán P, Koch B. 2014. Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass.*Remote Sensing of Environment* **154** (1): 102–114 DOI: 10.1016/j.rse.2014.07.028

Filipponi F. 2019. Sentinel-1 GRD Preprocessing Workflow.*Proceedings* **18** (1): 11 DOI: 10.3390/ecrs-3-06201

Ghulam A, Li ZL, Qin Q, Yimit H, Wang J. 2008. Estimating crop water stress with ETM+ NIR and SWIR data. *Agricultural and Forest Meteorology* **148** (11): 1679–1695 DOI: 10.1016/j.agrformet.2008.05.020

Gislason PO, Benediktsson JA, Sveinsson JR. 2006. Random forests for land cover classification. In *Pattern Recognition Letters* North-Holland; 294–300. DOI: 10.1016/j.patrec.2005.08.011

Glenn EP, Doody TM, Guerschman JP, Huete AR, King EA, McVicar TR, Van Dijk AIJM, Van Niel TG, Yebra M, Zhang Y. 2011. Actual evapotranspiration estimation by ground and remote sensing methods: the Australian experience. *Hydrological Processes* **25** (26): 4103–4116 DOI: 10.1002/hyp.8391

Guerschman JP, Hill MJ. 2018. Calibration and validation of the Australian fractional cover product for MODIS collection 6. *Remote Sensing Letters* **9** (7): 696–705 DOI: 10.1080/2150704X.2018.1465611

Guerschman JP, Van Dijk AIJM, Mattersdorf G, Beringer J, Hutley LB, Leuning R, Pipunic RC, Sherman BS. 2009. Scaling of potential evapotranspiration with MODIS data reproduces flux observations and catchment water balance observations across Australia. *Journal of Hydrology* **369** (1–2): 107–119 DOI: 10.1016/j.jhydrol.2009.02.013

Guerschman JP, Scarth PF, McVicar TR, Renzullo LJ, Malthus TJ, Stewart JB, Rickards JE, Trevithick R. 2015. Assessing the effects of site heterogeneity and soil properties when unmixing photosynthetic vegetation, non-photosynthetic vegetation and bare soil fractions from Landsat and MODIS data. *Remote Sensing of Environment***161** : 12–26 DOI: 10.1016/j.rse.2015.01.021

Hale J, Butcher R. 2011. Ecological Character Description for the Barmah Forest Ramsar Site

Han D, Liu S, Du Y, Xie X, Fan L, Lei L, Li Z, Yang H, Yang G. 2019. Crop Water Content of Winter Wheat Revealed with Sentinel-1 and Sentinel-2 Imagery. *Sensors* **19** (18): 4013 DOI: 10.3390/s19184013

Hoekstra AY, Mekonnen MM, Chapagain AK, Mathews RE, Richter BD. 2012. Global Monthly Water Scarcity: Blue Water Footprints versus Blue Water Availability. *PLOS ONE* **7** (2): e32688 DOI: 10.1371/JOURNAL.PONE.0032688

Immitzer M, Böck S, Einzmann K, Vuolo F, Pinnel N, Wallner A, Atzberger C. 2018. Fractional cover mapping of spruce and pine at 1 ha resolution combining very high and medium spatial resolution satellite imagery.*Remote Sensing of Environment* DOI: 10.1016/j.rse.2017.09.031

Jarchow CJ, Nagler PL, Glenn EP. 2017. Greenup and evapotranspiration following the Minute 319 pulse flow to Mexico: An analysis using Landsat 8 Normalized Difference Vegetation Index (NDVI) data. *Ecological Engineering* **106** : 776–783 DOI: 10.1016/j.ecoleng.2016.08.007

Koukoulas S, Blackburn GA. 2005. Mapping individual tree location, height and species in broadleaved deciduous forest using airborne LIDAR and multi-spectral remotely sensed data. *International Journal of Remote Sensing* **26** (3): 431–455 DOI: 10.1080/0143116042000298289

Kuczera G. 1987. Prediction of water yield reductions following a bushfire in ash-mixed species eucalypt forest. *Journal of Hydrology* **94** (3–4): 215–236 DOI: 10.1016/0022-1694(87)90054-0

Leblanc M, Tweed S, Van Dijk A, Timbal B. 2012. A review of historic and future hydrological changes in the Murray-Darling Basin. *Global and Planetary Change* DOI: 10.1016/j.gloplacha.2011.10.012

Lee S. 2020. Impact of Wildfire on Annual Water Yield in Large Watersheds

Main-Knorn M, Pflug B, Debaecker V, Louis J. 2015. Calibration And Validation Plan For The L2a Processor And Products Of The Sentinel-2 Mission. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **XL -7/W3** (7W3): 1249–1255 DOI: 10.5194/isprsarchives-XL-7-W3-1249-2015

Manfreda S, McCabe MF, Miller PE, Lucas R, Madrigal VP, Mallinis G, Dor E Ben, Helman D, Estes L, Ciraolo G, et al. 2018. On the use of unmanned aerial systems for environmental monitoring. *Remote Sensing* DOI: 10.3390/rs10040641

MDBA. 2009. *Water Act 2007* . Available at: https://www.legislation.gov.au/Details/C2016C00469

Melville B, Fisher A, Lucieer A. 2019. Ultra-high spatial resolution fractional vegetation cover from unmanned aerial multispectral imagery.*International Journal of Applied Earth Observation and Geoinformation* DOI: 10.1016/j.jag.2019.01.013

Moreno HA, Gourley JJ, Pham TG, Spade DM. 2020. Utility of satellite-derived burn severity to study short- and long-term effects of wildfire on streamflow at the basin scale. *Journal of Hydrology*DOI: 10.1016/j.jhydrol.2019.124244

Mu Q, Zhao M, Running SW. 2011. Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sensing of Environment* **115** (8): 1781–1800 DOI: 10.1016/j.rse.2011.02.019

Mu Q, Zhao M, Running SW. 2013. MODIS Global Terrestrial Evapotranspiration (ET) Product (NASA MOD16A2/A3) Collection 5.*Numerical Terradynamic Simulation Group* Available at: https://www.researchgate.net/publication/315700715 [Accessed 4 November 2020]

Nagler PL, Doody TM, Glenn EP, Jarchow CJ, Barreto-Munoz A, Didan K. 2016. Wide-area estimates of evapotranspiration by red gum (*Eucalyptus camaldulensis*) and associated vegetation in the Murray-Darling River Basin, Australia. *Hydrological Processes*30 (9): 1376–1387 DOI: 10.1002/hyp.10734

Rolls RJ, Bond NR. 2017. Environmental and Ecological Effects of Flow Alteration in Surface Water Ecosystems. In *Water for the Environment: From Policy and Science to Implementation and Management* Elsevier Inc.; 65–82. DOI: 10.1016/B978-0-12-803907-6.00004-8

Tucker CJ. 1980. Remote sensing of leaf water content in the near infrared. *Remote Sensing of Environment* **10** (1): 23–32 DOI: 10.1016/0034-4257(80)90096-6

Vaudour E, Gomez C, Fouad Y, Lagacherie P. 2019. Sentinel-2 image capacities to predict common topsoil properties of temperate and Mediterranean agroecosystems. *Remote Sensing of Environment***223** : 21–33 DOI: 10.1016/j.rse.2019.01.006

Verrelst J, Munoz J, Alonso L, Delegido J, Rivera JP, Camps-Valls G, Moreno J. 2012. Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and -3. *Remote Sensing of Environment* **118** : 127–139 DOI: 10.1016/j.rse.2011.11.002

Vorosmarty CJ, McIntyre PB, Gessner MO, Dudgeon D, Prusevich A, Green P, Glidden S, Bunn SE, Sullivan CA, Liermann CR, et al. 2010. Global threats to human water security and river biodiversity. *Nature***467** (7315): 555–561 DOI: 10.1038/nature09440

Wallace TA; Gehrig SL; Doody TM; Davies MJ; Walsh R; Fulton C; Cullen R; Nolan M. 2019. A-multiple-lines of evidence approach for prioritising environmental watering of wetland and floodplain trees (in press).

Ecohydrology (in press)

Wang L, Qu JJ, Hao X, Zhu Q. 2008. Sensitivity studies of the moisture effects on MODIS SWIR reflectance and vegetation water indices.*International Journal of Remote Sensing* **29** (24): 7065–7075 DOI: 10.1080/01431160802226034

Wasser L, Day R, Chasmer L, Taylor A. 2013. Influence of Vegetation Structure on Lidar-derived Canopy Height and Fractional Cover in Forested Riparian Buffers During Leaf-Off and Leaf-On Conditions.*PLoS ONE* DOI: 10.1371/journal.pone.0054776

Wheeler SA, Loch A, Crase L, Young M, Grafton RQ. 2017. Developing a water market readiness assessment framework. *Journal of Hydrology***552** : 807–820 DOI: 10.1016/j.jhydrol.2017.07.010

Table 1. Remote sensing data used to retrieve fractional tree canopy cover. NIR is near-infrared and SWIR is short-wave infrared.

Sensor	**Bands**	**Spatial resolution (m)**	**Acquisition date**
Sentinel-1	VV VH	20	Yanga 16^{th} May 2016 Barmah 11^{th} May 2016
Sentinel-2	B2 blue	10	Yanga 19^{th} May 2016 Barmah 6^{th} May 2016
	B3 green	10	
	B4 red	10	
	B5 vegetation red edge	20	
	B6 vegetation red edge	20	
	B7 vegetation red edge	20	
	B8 NIR	10	
	B8A NIR	20	
	B11 SWIR	20	
	B12 SWIR	20	
Airborne Lidar	-	0.23	Yanga 1^{st} Sep 2009 Barmah 10^{th} Sep 2015

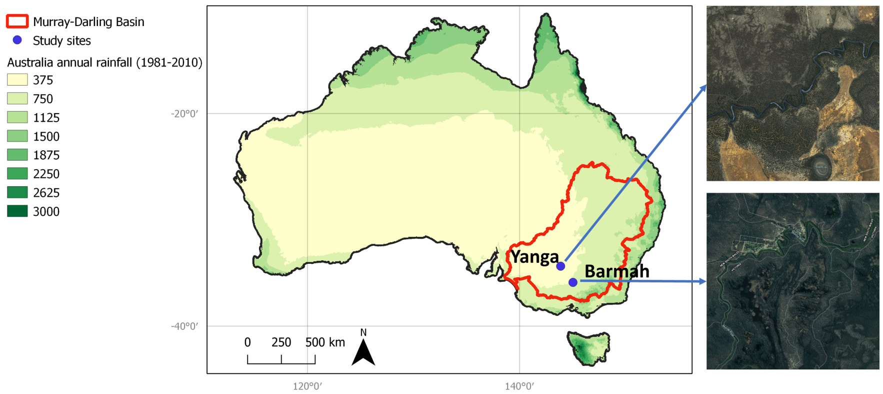
Figures captions

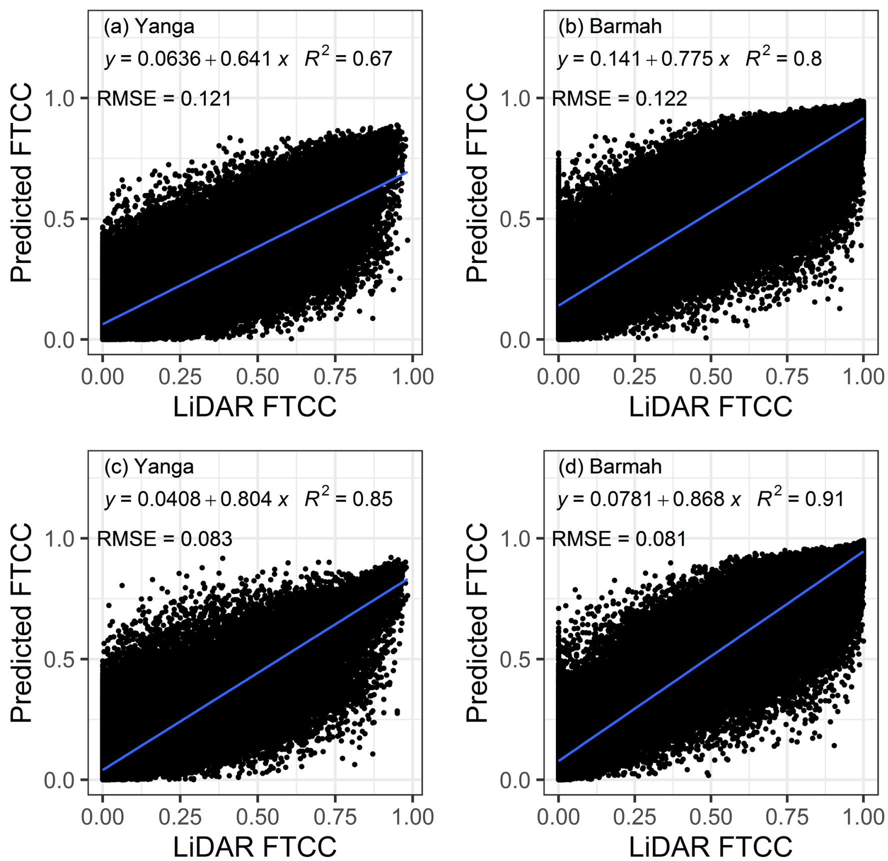
Figure 1. The location and climatic context of Yanga National Park and Barmah National Park (image courtesy of Google Earth on Aug/2018).

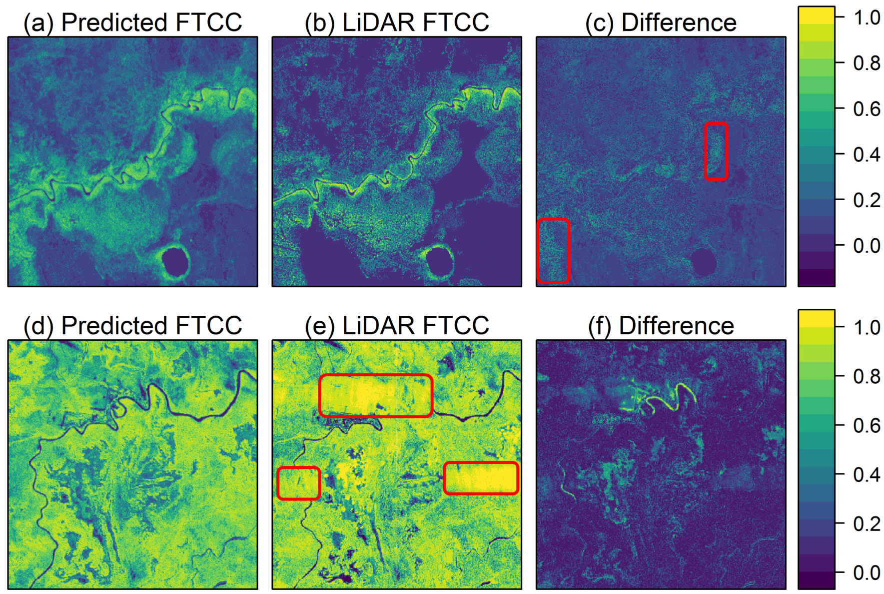
Figure 2. Scatterplots showing LiDAR FTCC against predicted FTCC (a) Yanga and (b) Barmah using individual site-specific models (RFYanga and RFBarmah). The results of predicted FTCC using the model trained using both Yanga and Barmah data (RFall) are shown in (c) for Yanga and (d) for Barmah. The blue line indicates the regression line.

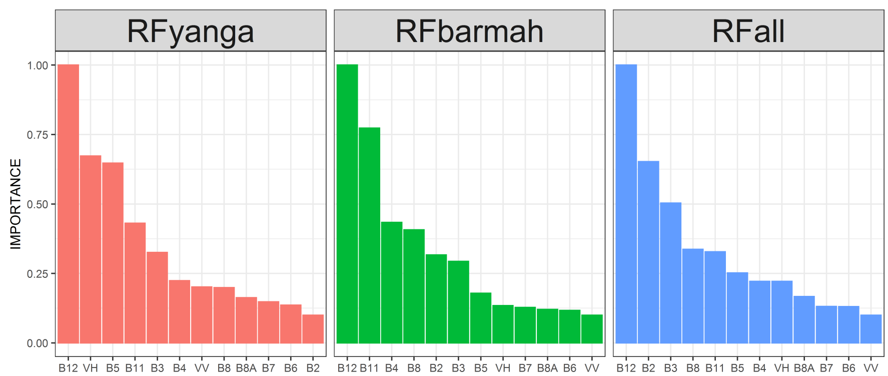
Figure 3. Spatial outputs for the Yanga region of interest, mapping (a) predicted fractional tree canopy cover (FTCC); (b); LiDAR FTCC and (c) the difference between the two images. Similarly, spatial outputs for the Barmah region of interest, mapping (d) predicted fractional tree canopy cover (FTCC); (e); LiDAR FTCC and the (f) difference between the two images.

Figure 4. Bands of importance for Sentienl-1 and Sentinel-2 for FTCC estimation for the RFYanga, RF-Barmah and RFAll Random forest models. The importance score was scaled between 0.1 and 1.


Murray-Darling Basin
Study sites
Australia annual rainfall (1981-2010)
375
750
1125
1500
1875
2250
2625
3000
-20°0′
-40°0′
120°0′
140°0′
Yanga
Barmah
0 250 500 km
N


(a) Yanga
$y = 0.0636 + 0.641\ x$ $R^2 = 0.67$
RMSE = 0.121
(b) Barmah
$y = 0.141 + 0.775\ x$ $R^2 = 0.8$
RMSE = 0.122
(c) Yanga
$y = 0.0408 + 0.804\ x$ $R^2 = 0.85$
RMSE = 0.083
(d) Barmah
$y = 0.0781 + 0.868\ x$ $R^2 = 0.91$
RMSE = 0.081
Predicted FTCC
LiDAR FTCC
0.0
0.5
1.0
0.00
0.25
0.50
0.75
1.00


(a) Predicted FTCC
(b) LiDAR FTCC
(c) Difference
(d) Predicted FTCC
(e) LiDAR FTCC
(f) Difference
1.0
0.8
0.6
0.4
0.2
0.0


RFyanga
RFbarmah
RFall
IMPORTANCE
1.00
0.75
0.50
0.25
0.00
B12 VH B5 B11 B3 B4 VV B8 B8A B7 B6 B2
B12 B11 B4 B8 B2 B3 B5 VH B7 B8A B6 VV
B12 B2 B3 B8 B11 B5 B4 VH B8A B7 B6 VV