EXTENDED ABSTRACT

Estimating grapevine crop coefficients at high resolution using open-source satellite

Hankun Luo[1], Diego Roy Guevara Torres[1], Vinay Pagay[1]
*Corresponding author: vinay.pagay@adelaide.edu.au

[1] School of Agriculture, Food and Wine, The University of Adelaide, Adelaide, SA, Australia 5064

Keywords: irrigation scheduling, crop coefficient, satellite, spectral unmixing

INTRODUCTION

Climate change results in increasing water stress due to co-effects of rising evapotranspiration (ET) and decreased precipitation over the past 65 years (Spinoni et al. 2019). Though mild water deficits can improve fruit and wine quality, severe shortage of water to grapevines negatively influences both grapevine productivity and fruit quality (van Leeuwen et al. 2024). To address these adverse drought effects on grapevines, one direct and effective solution is the application of supplemental irrigation with appropriate schedules (Schlank et al. 2024). For winegrapes, ET-based irrigation scheduling has shown to result in higher yields, bunch numbers, and crop water use efficiency, compared to other methodologies like grower's experience and soil-based measures (Schlank et al. 2024).

The ET method measures evaporation from soil and transpiration from vegetation. Crop evapotranspiration (ETc) can be computed according to derived formula (Allen et al. 1998):

$$ETc = ET0 \times kc$$

ET0 is the 'Reference ET', which represents the rate of evapotranspiration from a reference surface, often a 10 cm tall grass, and can be calculated via the Penman-Monteith model (Allen et al. 1998), with data commonly obtained from automatic weather stations. A key parameter in the ETc computation is the crop coefficient (kc), a dimensionless parameter related to the canopy size and leaf area, and is highly variable during the growing season due to varying canopy structure, fraction of ground covered by vegetation, training system, and pruning level, amongst others (Pereira et al. 2020).

Compared to other time-consuming, costly, cumbersome and often single-point or small-scale kc measurement methods like lysimeter or flux tower, remote sensing (RS) has high temporal flexibility and broader spatial representativeness by capturing structural, spectral, thermal, or microwave information from land covers (Gautam and Pagay 2020, Gautam et al. 2021). Amongst the mainstream RS platforms including Manned Aircraft System (MAS), Unmanned Aerial Vehicle (UAV), satellites provide low-cost, long-term time-series data with broad spatial coverage, which benefits regional predictive analysis (Gautam and Pagay 2020, Govi et al. 2024). Airborne platforms, by contrast, are more challenging to operate and collect data at such a large scale. Additionally, the similarity between kc and satellite-derived vegetation indices has driven the use of low-cost remote sensing technologies to estimate kc across various spatial and temporal scales (Gautam et al. 2021).

The current limitation of using satellite data is the low resolution of the dataset; typical pixel sizes or ground sampling distances (GSD) are greater than the area occupied by a single grapevine canopy. Furthermore, vineyard inter-rows are often covered by non-vine vegetation such as cover crops or weeds (Gautam and Pagay 2020), which result in a 'mixed' satellite pixel that contain a combination of grapevines, bare soil, weeds, and/or cover crops, all of which add "noise" to the vine spectral signal (Govi et al. 2024). Separation of land cover types to obtain specific VI values can help to improve the accuracy of RS kc predictions (Quintano et al. 2012). To the best of our knowledge, there are no reports on unmixing data of low- to medium-spatial-resolution satellite data for grapevine kc computation. This study aims to fill this gap.

RESEARCH OBJECTIVES

This study aims to explore the use of remote sensing (RS) methods which utilizes free medium-spatial-resolution satellite data (Sentinel-2 imagery) to improve kc estimation accuracy, enabling more affordable precision irrigation. Our work addresses challenges associated with relatively small-sized agricultural lands such as vineyards, where their row and plant spacing, being smaller than the satellite's GSD, lead to spectral index "mixing." Leveraging spectral unmixing techniques and machine learning, this study discusses how to solve significant discrepancies between the actual values for grapevines and the spectral indices derived from satellite mixed pixels, and highlights how to offer a cost-effective solution for irrigation-scheduling and precision irrigation in vineyards and other irrigated crops.

MATERIAL AND METHODS

Vineyard site

The ROI for this study is situated in Alex 88, Coonawarra, Limestone Coast zone, South Australia. It is a Cabernet Sauvignon vineyard of the Wynns Coonawarra Estate (37°17'08.500" S, 140°49'37.900" E). We had one experimental block (approx. 1 ha), in which we had 32 sampling points for ground data collection and RS analysis.

Satellite data acquisition and pre-processing

The satellite multi-spectral imagery is sourced from the Harmonized Sentinel-2 MSI: Multispectral Instrument, Level-2A (Stl-2), provided by the EU/ESA/Copernicus program. We accessed the Stl-2 dataset ("COPERNICUS/S2_SR_HARMONIZED") via Google Earth Engine (GEE). The GEE offers an interactive code editor, where we developed scripts to perform band math and time-series analysis. To eliminate the interference of clouds and shadows, we first filtered and cloud-masked the dataset by s2_cloudless approach, incorporating the cloud probability dataset to improve cloud detection accuracy. To correct the potential image 'shift' error due to position and orientation (pose) during satellite image collection, we applied 'buffering' to add a margin of tolerance around sampling points and avoid data omission due to shift errors. Based on Sentinel-2's resolution, we manually created circles centred on each sampling point with a radius of 10 meters. The median value of the pixels covered by the circle was then calculated. For the vegetation index, we selected Normalized difference vegetation index (NDVI). It is reported that NDVI has a strong relationship with kc (Gautam and Pagay 2020) and the bands (red and near infrared) used to calculate NDVI have highest spatial resolution in Stl-2 compared to other bands. Sampling times were based on the phenological phases reported for the Coonawarra region in Longbottom et al. (2022). We collected spectral data within the ROI from December to February for the 2019/20 and 2020/21 seasons, covering from flowering to véraison, and pre-harvest, which were downloaded as raster files.

UAV data acquisition

All high-resolution images were captured by a UAV (DJI Matrice 600 Pro, Da-Jiang Innovations Science and Technology Co., Ltd., Shenzhen, China). The images were collected over three vintages—2018/19, 2019/20, and 2020/21—and align with the same months as the satellite data. For details on UAV flight and photography, refer to Gautam et al. (2021).

Ground-truthing data

Ground measurements of kc were estimated using the Paso Panel. The methods and formulas used can be found in Gautam et al. (2021).

Spectral unmixing

To separate the grapevine NDVI from mixed pixels, we performed linear spectral unmixing. First, using high-resolution imagery in ArcGIS Pro (version 3.1.0), we applied a random tree classification (maximum number of trees = 50, maximum tree depth = 30, maximum number of samples per class = 1000). The classified result was then processed with a "fishnet" segmentation and 'zonal statistics' (counted sum of pixels in every grid) to calculate the proportion of grapevine area within each fishnet cell for each sampling point. Raster data of NDVI were also incorporated to ArcGIS for unmixing calculations according to the following formula (Quintano et al. 2012):

$$NDVImixed = fgrape \times NDVIgrape + fsoil \times NDVIsoil + fcc \times NDVIcc$$

Where f = fraction of area, cc = cover crops. Due to the lack of in-situ NDVI measurements for cover crops and soil, we found "reference points" for these two land cover types. We ensured that the reference soil matched the type in Alex East (B4) and selected nearby grassland as the cover crop reference. Using Google Earth Engine (GEE), we downloaded the time-series NDVI data for these reference points over five years (2019-2024) as the representative NDVI for soil and cover crops for different growth stages in the linear unmixing process.

Modelling and validation

For the two vintages, a total of 192 unmixing NDVI data points were collected. These data were randomly split into training and testing sets at a 7:3 ratio for model validation. Using R Studio (2024 09.0 for Windows), we first separated the data by year, then combined them to analyse linear correlations between variables with a correlogram. Subsequently, we applied a simple linear model reported by Trout and Johnson (2007), as well as a Random Forest Regression (Gautam et al. 2021), to train and test all data. Model performance was evaluated using R squared (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE); for linear regression, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) is added as effectiveness indicators.

RESULTS

There were notable differences in the correlations between variables in 2020 and 2021 vintages. In the 2020 correlogram (Figure 1a), higher unmixed NDVI generally corresponded to higher kc values ($r = 0.347$, $p < 0.01$), whilst in the next vintage (Figure 1b), this positive correlation was weakened significantly in both strength and statistical significance ($r = 0.133$, $p > 0.05$). Mixed NDVI displayed a similar trend with kc in 2020, though the statistical parameters were stronger and more significant ($r = 0.617$, $p < 0.001$) than that of unmixed values. However, during the 2021 growing season, mixed NDVI exhibited a negative correlation with kc, with low statistical significance ($p > 0.05$). The seasonal relationship between our computed canopy area (c_area) and kc also exhibited differences. In the comprehensive chart (Figure 1c) which evaluates the correlation relationships of each variable (kc, NDVI, and unmixed NDVI) across both vintages, crop coefficients had trends more closely aligned with the 2020 result.

The results indicate a high level of heterogeneity in the vineyard, as evidenced by the UAV RGB imagery. We observed varying levels of weed cover vigour around different sampling points across months. Additionally, factors such as vine shadows, the mosaic of bare soil and grass cover, and discarded grape clusters from thinning all interfered with spectral unmixing. Furthermore, at the same sampling point, differences in leaf color intensity, canopy density, and other vine characteristics were evident across years. These variabilities also support our later findings in modelling. As Table 1 indicates, due to its relative robustness, mixed NDVI demonstrates higher accuracy (R2 = 0.224) in predicting kc than unmixed NDVI (R2 = 0.049) within the linear model; but the overall performance of the linear model remains suboptimal (R2 of all groups are low than 0.3), which reveal the sensitivity of linear model to high heterogeneity in vineyards.

On the other hand, more complex machine learning models, like Random Forest (RF), demonstrated strong adaptability to variability (Table 1). Gonzalo-Martín and colleagues (2017) pointed out that RF offers several key advantages over other well-known classifiers. Its non-parametric nature allows flexibility across various datasets, and it often achieves higher classification accuracy than traditional classifiers when suitable attributes are selected for training pattern characterization. Additionally, RF can identify feature importance, enhancing interpretability, and it demonstrates robustness against imbalanced class distributions.

In the RF model (Table 1), unmixed NDVI showed a much stronger ability (R2 = 0.771) to predict kc compared to mixed NDVI (R2 = 0.621), indicating that spectral unmixing contributed to improving the model's performance. Additionally, the addition of canopy area as a model parameter significantly improved the performance of the kc model based on mixed NDVI (R2 = 0.819), bringing it closer to that of the unmixed NDVI model with the same leaf area inclusion (R2 = 0.846).

CONCLUSION

We found that spectral unmixing and machine learning approaches improved kc predictions based on satellite remote sensing of vineyards. For a vineyard with high temporal and spatial heterogeneity, linear models may be insufficient for accurate generalization; in contrast, more complex machine learning models, such as Random Forest, can better adapt to the variability within the vineyard. Additionally, remote sensing (RS) data combined with grapevine canopy structural information (e.g. canopy area) resulted in improved model performance, which suggests that simple field measurements of the canopy may offer a cost-effective approach for RS irrigation scheduling and precision irrigation in vineyards.

ACKNOWLEDGEMENTS

The authors are grateful to Dr Sami Rifai (University of Adelaide) for providing valuable advice on the use of Google Earth Engine. We also thank Dr Rochelle Schlank and Dr Deepak Gautam for providing the high-resolution images and ground-truthing data.

REFERENCES

Allen, R.G., Pereira, L.S., Raes, D. and Smith, M. (1998) Crop evapotranspiration: Guidelines for computing crop water requirements. (FAO: Rome).

Gautam, D., Ostendorf, B., and Pagay, V. (2021) Estimation of grapevine crop coefficient using a multispectral camera on an unmanned aerial vehicle. Remote Sensing 13, 2639.

Gautam, D. and Pagay, V. (2020) A review of current and potential applications of remote sensing to study the water status of horticultural crops. Agronomy 10, 140.

Govi, D., Pappalardo, S. E., De Marchi, M., and Meggio, F. (2024) From space to field: combining Satellite, UAV and agronomic data in an open-source methodology for the validation of NDVI maps in precision viticulture. Remote Sensing 16, 735.

Pereira, L.S., Paredes, P., Melton, F., Johnson, L., Wang, T. López-Urrea, R., Cancela, J.J. and Allen, R.G. (2020) Prediction of crop coefficients from fraction of ground cover and height. Background and validation using ground and remote sensing data. Agricultural Water Management 241, 106197.

Quintano, C., Fernández-Manso, A., Shimabukuro, Y. E. and Pereira, G. (2012) Spectral unmixing. International Journal of Remote Sensing 33, 5307-5340.

Schlank, R., Kidman, C.M., Gautam, D., Jeffery, D.W., and Pagay, V. (2024) Data-driven irrigation scheduling increases the crop water use efficiency of Cabernet Sauvignon grapevines. Irrigation Science 42, 29-44.

Spinoni, J., Barbosa, P., de Jager, A., McCormick, N., Naumann, G., Vogt, J.V., Magni, D., Masante, D. and Mazzeschi, M. (2019) A new global database of meteorological drought events from 1951 to 2016. Journal of Hydrology: Regional Studies 22, 100593.
Trout, T.J. and Johnson, L.F. (2007) Estimating crop water use from remotely sensed NDVI, crop models, and reference ET. (Proceedings of the USCID Fourth International Conference on Irrigation and Drainage: Sacramento, California) pp. 275-285.
van Leeuwen, C., Sgubin, G., Bois, B., Ollat, N., Swingedouw, D., Zito, S. and Gambetta, G.A. (2024) Climate change impacts and adaptations of wine production. Nature Reviews Earth & Environment 5, 258-275.

TABLE AND FIGURE

Table 1. Comparison in model performance of kc prediction.

Model variables	Model types	R^2	RMSE	MSE	AIC	BIC
Mixed NDVI	Linear regression	0.224	0.09	0.008	-250	-242
	Random Forest	0.621	0.066	0.004	n/a	n/a
Mixed NDVI + canopy area	Linear regression	0.229	0.09	0.008	-249	-238
	Random Forest	0.819	0.05	0.0025	n/a	n/a
Unmixed NDVI	Linear regression	0.067	0.099	0.009	-224	-213
	Random Forest	0.771	0.059	0.003	n/a	n/a
Unmixed NDVI +canopy area	Linear regression	0.049	0.1	0.01	-224	-215
	Random Forest	0.846	0.053	0.0028	n/a	n/a
Canopy area	Linear regression	0.038	0.101	0.01	-222	-214
	Random Forest	0.736	0.0036	0.06	n/a	n/a

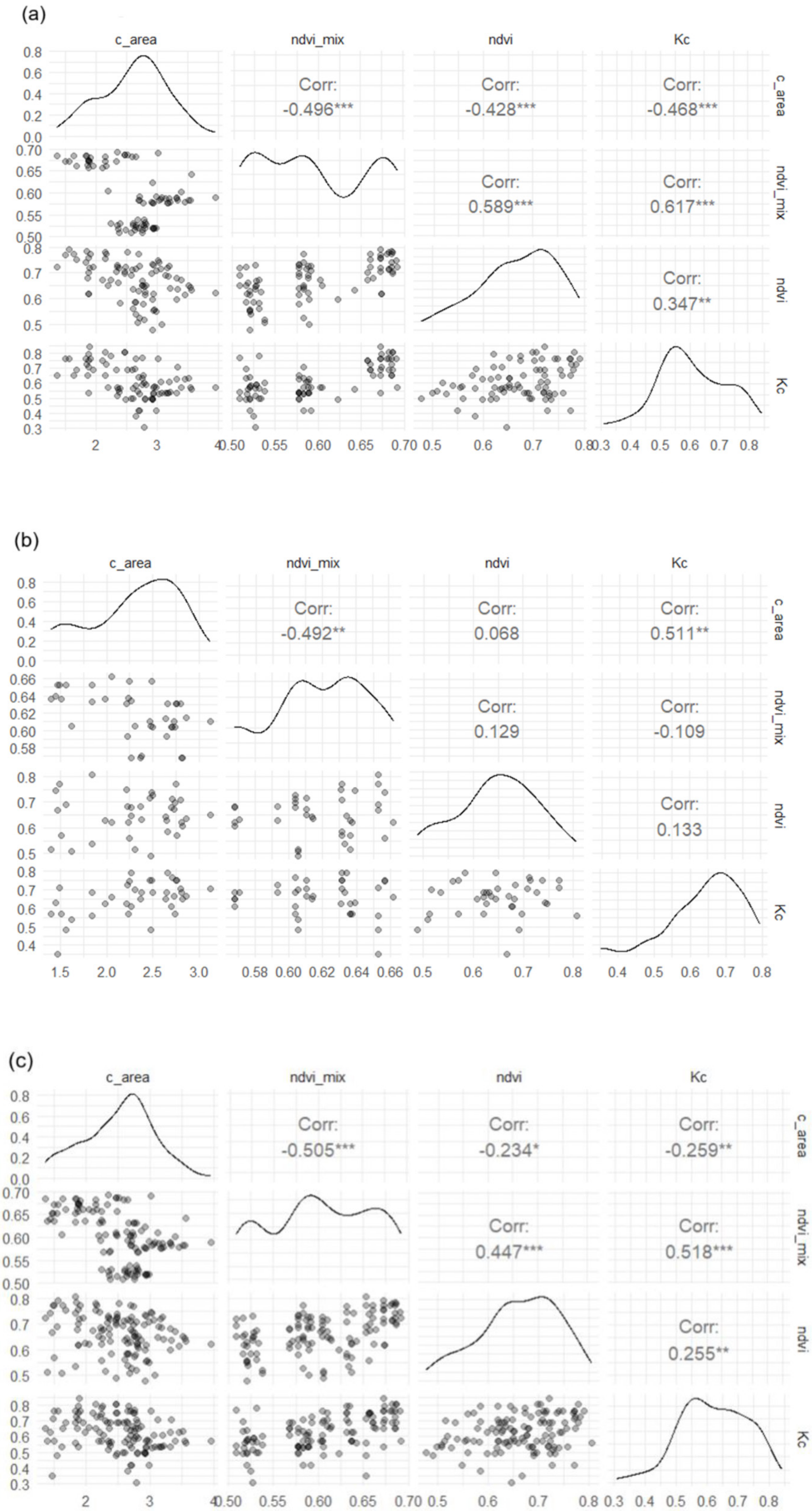

Figure 1. Correlograms of variables in (a) 2020, (b) 2021, (c) total. In this figure, c_area = canopy area, ndvi_mix = NDVI of mixed pixel, ndvi = unmixed NDVI of grapevines, Kc = crop coefficient; the number under each variable indicates Pearson value (r), and * denotes statistical significance (p-value), where *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.