Data-driven modelling for nowcasting of soil water for dryland cropping in Australia

Niranjan Wimalathunge and T.F.A. Bishop

Sydney Institute of Agriculture, School of Life & Environmental Science, The University of Sydney, Central Ave, Eveleigh, Sydney, NSW 2015, Email: niranjan.manikkuacharige@sydney.edu.au

Abstract
The term 'soil water nowcasting' signifies the insight, planning and risk management that can be undertaken once the knowledge of the current status of soil water storage plants can access is established. Understanding infiltration and water flow through the soil layers for building soil water models is key to nowcasting. The approach here uses a data-driven model, which combines an unsaturated flow multi-bucket model with a machine-learning algorithm to improve the model estimates using local information. This method can utilise commonly available geospatial datasets as model drivers. The soil bucket size is represented by the Soil Landscape Grid of Australia combined with PTFs to estimate the bucket size enabling the output to have a 90 m spatial and up to 1 m vertical resolution. Soil water maps were produced with good prediction quality (concordance = 0.75 ; RMSE= 0.05 cm^3 cm^{-3}) for the root zone.

Keywords
Data-cube, multi-bucket water balance model, unsaturated flow, machine learning

Introduction

There are no perfect data streams for all soil water applications because they vary spatially, temporally, and quality. Empirical models can combine these data streams accounting for different spatial, temporal and spatial-temporal resolutions. Datasets can be gathered using 'bottom-up,' *i.e.,* proximal sensors and 'top-down,' *i.e.,* space-borne sensors approaches and other geospatial datasets such as Soil Landscape Grid of Australia (SLGA) (Wimalathunge and Bishop, 2019). However, the challenge here is to fuse available data sources with observations to estimate more meaningful soil water content in space-time.
With the advances in the soil water measuring sensor technologies, relatively low cost of the sensor units, wireless data transferability and near-real-time availability of the datasets through APIs, there are now new avenues to get high spatial and temporal coverage of soil water variability. Soil monitoring networks are used within farm-scale, regional-scale and country-scale. The observations retrieved from soil monitoring networks can combine with data sources gathered from 'bottom-up' and 'top-down' approaches and other soil water model outputs to develop empirical models using modern machine learning techniques (Lokers et al., 2016).
With the ever-increasing availability of geospatial data sources, managing and re-arranging such datasets plays an essential role in empirical model development and prediction framework development. These geospatial datasets and soil water streams allow the creation of spatial-temporal data called "data-cube". The data cube is the basic unit in the modelling framework (Nativi et al., 2017), which applies machine learning algorithms, *e.g.,* random forest, support vector machines, deep neural networks. In this work, a data cube describing the study area's soil water characteristics is designed, then estimated soil water, and presented in spatial maps for topsoil, subsoil, and root-zone at key times of the agricultural systems.

Methods

Study Area
The study focused on the Kyeamba sites and their observations over 18 yrs (2001-2018) from the OzNet Hydrological monitoring network situated in agricultural landscapes in south-eastern Australia. All the soil moisture probes in the network are calibrated. Kyeamba Creek is a small catchment that covers an area of 600 km^2. The land use is predominantly grazing for sheep and cattle. The

topography is gentle slopes, and there are some undisturbed areas with native vegetation on the steeper slopes. The mean annual precipitation of Kyeamba is ~600 mm.

Map Production Area

The map production area (35 km^2) covers four soil moisture sites (K3, K4, K5 and K7), two creeks (Kyeamba and Teatree) and two grazing farms (Ireland Angus and Kyeamba Station Woolshed).

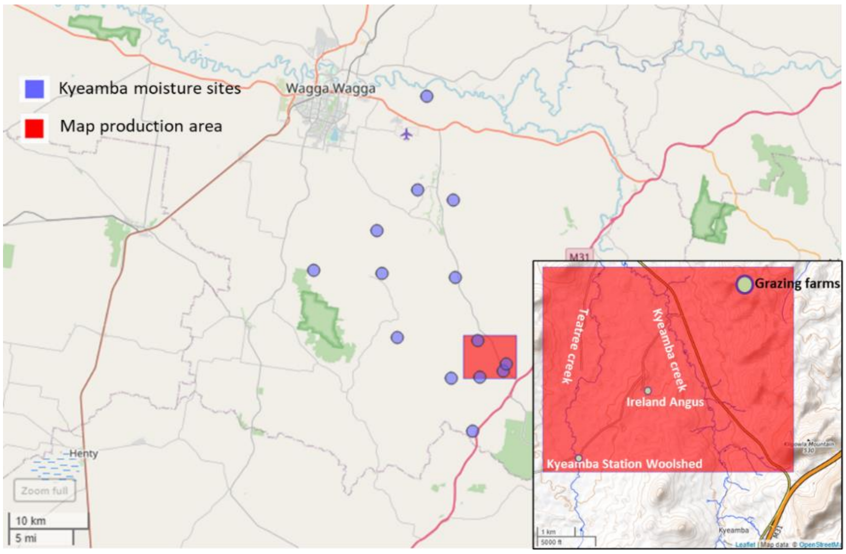

Figure 1. Kyeamba soil moisture probe network and map production area overlaying a Google map image (Google, 2021).

Water balance model

We used the water balance (WB) model developed by Wimalathunge & Bishop (2019), which add space and time daily soil moisture estimates to the data cube. This WB is a multi-layer, process-based model that better represents the vertical soil moisture variation. It is also an unsaturated model where water infiltrates through layers freely and continuously according to the soil properties. SLGA soil depth intervals are the WB model's layer thickness. The corresponding clay, sand and bulk density values were used to calculate the saturated volumetric moisture content (θs) using a pedotransfer function (PTF), which is developed by Padarian et al. (2014). The soil is assumed to be uniform within each horizontal layer, and the water flows vertically through the soil layers. The infiltration continues for all layers, and excess soil water beyond the 60–100 cm layer is assumed to be deep drainage and lost to the system. It also is assumed that runoff only occurs when both Layer 0–5 cm and Layer 5–15 cm are saturated. The modelling depth is 1 m, and the predictions spatial resolution is 90 m, as determined by the SLGA data. The model is run on each SLGA raster cell with the corresponding value for the Scientific Information for Land Owners (SILO) rainfall and Modis Global Evapotranspiration Project (MOD16) evapotranspiration (ET). ET is assumed to be an equal contribution of evaporation and transpiration; however, no ET is lost in the process.

Data-cube and incorporation of observations into the model

Incomplete representation of physical processes can lead to errors in process-based models. For example, variable rooting depths, which has a major impact on evapotranspiration, is absent from the WB model as we assume a uniform depth of 1m. Therefore, we integrate the WB model with a machine learning algorithm, Random Forest (RF), which can include the local information on soil properties and topography to increase the relevance of predictions. As shown in Table 1, covariates that vary in the spatial, temporal, and spatial-temporal domain are organised in a data cube describing the study area's soil moisture. Discounted ET, precipitation and EVI were added as additional covariates, which use a weighting function to account for past or antecedent conditions since the soil moisture at any time depends on the current and prior conditions of these covariates (Wang et al., 2011; Lessels and Bishop, 2013).

Table 1 data-cube for soil water prediction.

	Covariate	Source	Resolution
Spatial	Slope, aspect and solar radiation	DEM from Geoscience Australia	90 m raster
	Soil order	ASRIS	90 m raster
	clay % (0-30,30-100 cm)	SLGA	90 m raster
Spatial & temporal	Evapotranspiration and discounted ET	MODIS	500 m, 8 day
	precipitation and discounted precipitation	BOM	5 km, daily
	EVI and discounted EVI	MODIS	500 m, 16 day
	WB soil moisture	Wimalathunge & Bishop (2019)	90 m, daily
Temporal	Month (1-12)	-	Monthly

Evaluation of model performance
The overall performance of the hybrid model was assessed with the observed Kyeamba moisture probes at the depth interval :(i) topsoil (0-30 cm); subsoil (30-100 cm), and; root-zone (0-100 cm). Leave-one-out-site cross-validation (LOOSCV) was used to check the quality of the model predictions for all sites. LOOSCV is the generalisability of the model, which involves training a Random Forest model for all sites excluding one site and then predicting at that site. This is repeated sequentially for all sites, so for each site, we have completely independent predictions.
Two statistics were calculated to evaluate model performance: (i) Lin's concordance correlation coefficient (LCCC); and (ii) the root-mean-square error (RMSE).

Mapping soil water
Soil water maps at 90 m resolution are created using RF for topsoil, subsoil and root-zone by averaging (weighted) soil water at each layer. These maps give a snapshot of soil moisture at key times in agricultural systems, assisting with management decisions. The mapping area is described in Figure 1.

Results and Discussion

Assessment of Random Forest model predictions
The estimate soil water at three depth interval of topsoil, subsoil and root-zone (Figure 2) with the prediction quality: topsoil (Concordance = 0.69, Accuracy = 0.05 $cm^3\ cm^{-3}$); subsoil (Concordance = 0.72, Accuracy = 0.04 $cm^3\ cm^{-3}$); and root-zone (Concordance = 0.75, Accuracy = 0.05 $cm^3\ cm^{-3}$). The subsoil is more predictable than the topsoil, with better accuracy. Subsoil moisture does not vary as much as topsoil moisture due to less infiltration of rainfall, greater clay content and low root biomass (Figure 3).

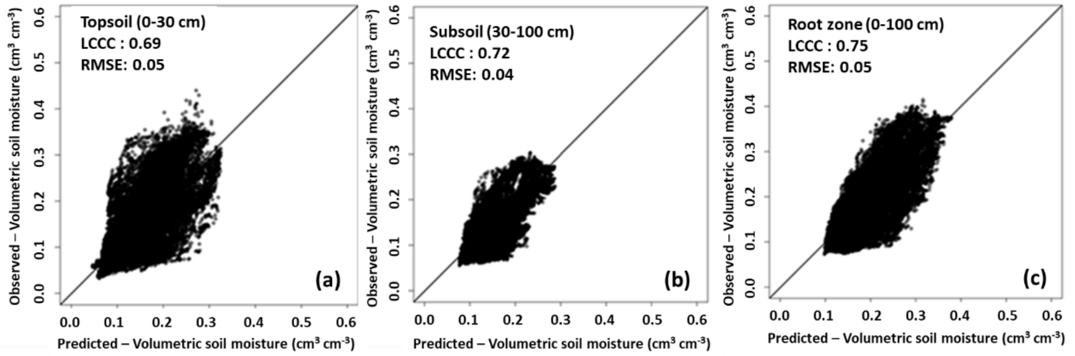

Figure 2. Observed vs predicted soil moisture for Kyeamba sites: (a) Topsoil; (b) Subsoil; (c) Root zone

Spatial maps (Figure 3) were created to map soil moisture at specific times of the year in relation to a winter cropping season—near to sowing, mid-season and after the harvest. April maps show low moisture conditions before sowing, whereas the August map shows high soil water conditions due to the winter rain. December maps show low moisture content due to the use of moisture by crops and dry summer conditions.

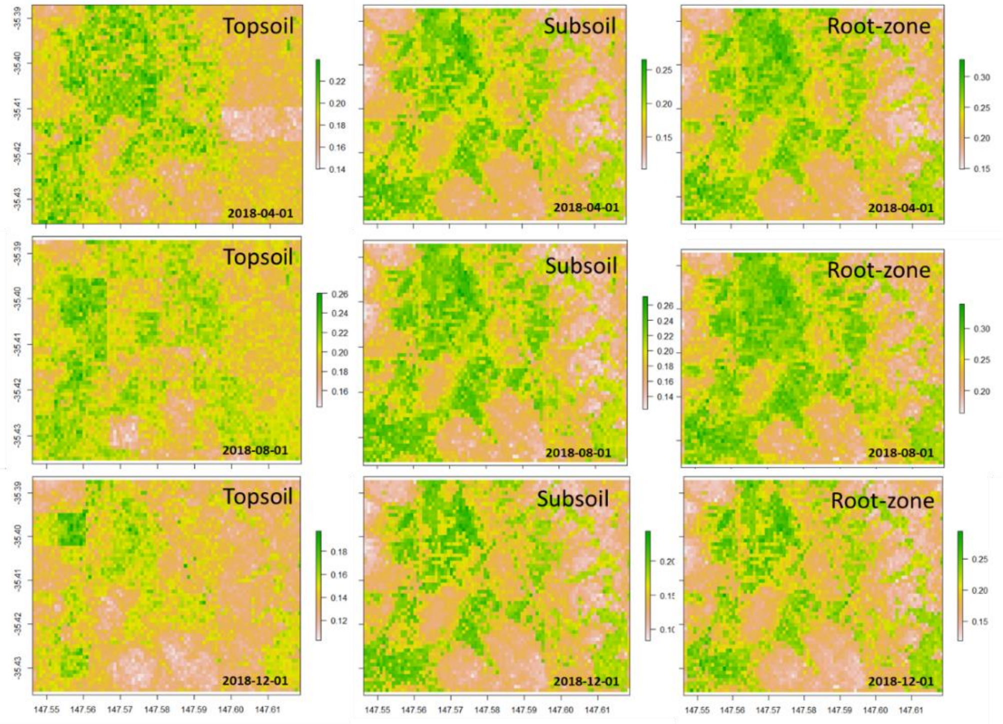

Figure 3. Soil moisture (cm^3 cm^{-3}) maps: topsoil, subsoil and root-zone for 1 April, 1 August and 1 December on a 90 m grid

Conclusion

Soil moisture nowcasting is a primary factor in agricultural productivity. Growers traditionally rely on winter rainfall, but rainfall patterns can change with significant rainfall outside the regular season. This work addresses the problems in nowcasting soil water while meeting the grower's soil water requirements. The results are presented at ~90 m resolution for different soil profile depths at key points in an agricultural landscape. Importantly, the water balance model can be developed at a continental scale. Incorporating local soil moisture measurements (vertical), *i.e.,* OzNet soil moisture probes, makes the model estimates more sensible. Further, the model estimates can be improved spatially in future using satellite soil moisture products, *e.g.,* SMAP and Sentinel.

References

Lessels, J. and Bishop, T. (2013). Estimating water quality using linear mixed models with stream discharge and turbidity. Journal of Hydrology, 498, pp.13-22.

Padarian et al. (2014). Predicting and mapping the soil available water capacity of Australian Wheatbelt. Geoderma Regional, 2-3, pp.110-118.

Wang et al. (2011). Load estimation with uncertainties from opportunistic sampling data – A semiparametric approach. Journal of Hydrology, 396(1-2), pp.148-157.

Wimalathunge, N. and Bishop, T., 2019. A space-time observation system for soil water in agricultural landscapes. Geoderma, 344, pp.1-13.

Lokers, R., Knapen, R., Janssen, S., van Randen, Y. and Jansen, J., 2016. Analysis of Big Data technologies for use in agro-environmental science. Environmental Modelling & Software, 84, pp.494-504.

Nativi, S., Mazzetti, P. and Craglia, M., 2017. A view-based model of data-cube to support big earth data systems interoperability. Big Earth Data, 1(1-2), pp.75-99.